Who Speaks 'Kid'? How Experience with Children Does (and Does Not) Shape the Intelligibility of Child Speech

Madeleine E. Yu[1,2], Angela Cooper[2], and Elizabeth K. Johnson[1,2]

[1]Department of Psychology, University of Toronto

[2]Department of Psychology, University of Toronto Mississauga

Author Note

Madeleine E. Yu https://orcid.org/0000-0002-3252-3485

The authors have no conflict of interest to declare. We thank Thomas St. Pierre, Tiana Simovic, Priscilla Fung, Lisa Hotson, and the Child Language and Speech Studies Lab for their support and assistance. This work was supported by grants awarded to E.K.J. from SSHRC, NSERC, and the CRC program. A preliminary report on this work was presented at the 43rd Annual Meeting of the Cognitive Science Society. All data and analysis code are publicly available at the Open Science Framework and can be accessed at https://osf.io/9fmdr/.

Correspondence regarding this article should be addressed to Madeleine E. Yu, Dept. of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada. Email: madeleine.yu@mail.utoronto.ca

Word count: 7277

Abstract

Child speech deviates from adult speech in predictable ways. Are listeners who routinely interact with children implicitly aware of these systematic deviations, and thereby better at understanding children? Or do idiosyncratic differences in how children pronounce words overwhelm these systematic deviations? In Experiment 1, we use a speech-in-noise transcription task to test who 'speaks kid' among four listener groups: undergraduates (N=48), mothers of young children (N=48), early childhood educators (N=48), and speech-language pathologists (N=48). All listeners transcribed speech by typically developing children and adults. In Experiment 2, we use a similar task to test an additional group of mothers (N=50) on how intelligible they found their own child versus another child. Contrary to previous claims, we find no evidence for an experience-based general child speech intelligibility advantage. However, we do find that mothers understand their own child best. We also observe a general task advantage by speech-language pathologists. Our findings demonstrate that routine (and even extensive) exposure to children may not make all children more intelligible, but that it may instead make particular children one has experience with more intelligible.

Keywords: speech perception; speech production; accent adaptation; child speech intelligibility; talker familiarity benefit

Public Significance Statement

The goal of the study is to determine who (if anyone) ‘speaks kid’. In support of existing evidence, we find that experience with a child makes it easier to understand words produced by that child specifically; however, this frequent experience does not predict a general child speech processing advantage. We did, however, find that speech-language pathologists had a general advantage in understanding words spoken by both adults and children. Our findings clarify existing claims in the literature and provide key insights into the nature of human speech processing in general.

Who speaks 'kid'? How experience with children does (and does not) shape the intelligibility of child speech

Child speech is often characterized as featuring systematic deviations from adult speech. For example, clusters are often simplified, easier-to-produce sounds are substituted for more difficult sounds, and stressed syllables can be reduplicated (e.g., Fikkert, 2010; Vihman, 1993). As such, a word such as *spaghetti* can be produced as *getti*, *shark* as *sock*, and *baby* as *baba*. This can make young children's speech difficult for some adults to understand. But past work has suggested that certain types of listeners — for instance, mothers – are particularly adept at understanding their own child's speech (e.g., Baudonck et al., 2009; Flipsen Jr, 1995; Goehl & Martin, 1987; Weist & Kruppe, 1977; Weist & Stebbins, 1972). Why is this the case? Do primary caregivers, such as mothers and other adults who routinely interact with children, possess a generalized abstract understanding of how children's utterances typically deviate from adult utterances, akin to how familiarity with a specific accent can facilitate adaption to newly encountered talkers who speak in that particular accent (e.g., Bradlow & Bent, 2008)? Or are adults just better at understanding the speech idiosyncrasies of the particular child (or children) that they interact with routinely, in much the same way that listeners find speech by the adults they interact with regularly more intelligible (e.g., Souza et al., 2013)? Here, we probe this question by testing different types of listeners, who have varied experience listening to children, on their ability to understand words produced by familiar and unfamiliar child talkers.

Past research suggests that mothers of young children are uniquely advantaged in understanding child speech, outperforming other adults who do not routinely interact with young children (Baudonck et al., 2009; Flipsen Jr, 1995; Goehl & Martin, 1987; Weist & Kruppe, 1977; Weist & Stebbins, 1972). On the surface, this may seem unsurprising — of course primary caregivers who are accustomed to interacting with children are more aware of how children typically speak (e.g., Fraser & Roberts, 1975; Moerk, 1974; Snow, 1972) — but upon closer inspection, this observation is actually quite interesting. For instance, although children's productions share similarities in how they deviate from adult target forms, each child has their own particular idiosyncrasies (i.e., one child may substitute *r*'s with *w*, while another may simply delete all *r*'s). Indeed, child speech has been shown to feature greater acoustic-phonetic inter- and intra-talker variability compared to adult speech (e.g., Lee et al., 1999; MacDonald et al., 2012), and child speech is known to also change over the course of development. So, possession of a general child speech processing advantage would necessitate an appreciation for a wide variety of typical child speech patterns, and the ability to draw generalizations despite ample individual variation in child talkers. But note that in the adult accent processing literature, speech patterns that are particularly variable or distant from a listener's own variety may be difficult to adapt to (e.g., Cooper et al., 2022). If the same logic applies to child speech, it may be the case that only listeners with extensive experience with the ways in which child speech deviates from adult speech show a general child speech intelligibility advantage.

While weighing the evidence for whether primary caregivers might possess a generalized advantage in the processing of child speech, it is useful to re-examine the existing literature on the matter. Upon careful consideration, it is readily apparent that research on a maternal (or caregiver) advantage in child speech processing has typically confounded two factors: familiarity with the speech of a particular child versus familiarity with children's speech more generally. For instance, past studies have presented mothers with the speech of their own child (Baudonck et al., 2009; Goehl & Martin, 1987; Weist & Stebbins, 1972), and have compared their performances to that of other adults who were tested with the same speech. These studies claimed that mothers' better performance in this context is evidence of a maternal advantage in understanding children; however, this design also leaves open the possibility that mothers find the speech more intelligible because of their familiarity with the specific speech patterns of their own child, and not with children generally. And in those studies that have not confounded talker familiarity with talker age, other methodological issues lead us to question their interpretation. For example, the findings of Weist & Kruppe (1977) were based on the observations of only twelve mothers hearing ten different multiword utterances produced by children during free-play. Not only was the sample size in the study small, but their use of this type of naturally produced speech as stimuli involves the risk of providing listeners with contextual cues that may have aided performance. And while Flipsen Jr (1995) used a controlled set of single word utterances to test mothers, their stimuli were produced by just four speech-delayed children, whose productions were likely more idiosyncratic than that of their non-delayed peers — thereby not addressing the question of whether a generalized advantage for the processing of typically developing children's speech exists. In sum, much of the evidence used to support the argument that primary caregivers such as mothers possess a general child speech processing advantage could be explained by methodological choices, and/or a well-known phenomenon heavily studied in the adult speech processing literature — that is, the talker familiarity benefit.

The talker familiarity benefit refers to the observation that spoken words are easier to process when produced by a familiar talker. That familiar talker can be someone in the listener's social circle, introduced to them in the lab, or their own voice (talker familiarity benefit; Eger & Reinisch, 2019; Holmes et al., 2018; Johnsrude et al., 2013; Levi et al., 2011; Mullennix & Pisoni, 1990; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Souza et al., 2013; Yonan & Sommers, 2000). By nature, speech by familiar talkers is heard frequently and within many different contexts. This enables listeners to become attuned to the particular speaking idiosyncrasies of that talker and to better understand their speech as a result. If familiarity plays a role in child speech intelligibility, it is possible that listeners would only demonstrate an advantage in understanding child speech over others when speech is produced by children that they are familiar with — an “own child speech advantage”.

Although existing evidence for mothers’ performance with child speech can be explained within the framework of the talker familiarity benefit (i.e., the own child speech advantage), it is certainly conceivable that a general child speech advantage exists alongside the own child speech advantage. Indeed, in initial encounters with talkers of an unfamiliar accent, listeners typically experience greater perceptual difficulty (e.g., Dupoux & Green, 1997; Munro & Derwing, 1995). Yet, intelligibility tends to improve the more experience a listener has with a particular type of speech variety, leading to more efficient recognition of words produced in that accent (e.g., Clarke & Garrett, 2004; Dahan & Mead, 2010; Norris et al., 2003; Maye et al., 2008; Weber et al., 2011). It has been argued that listeners adapt to unfamiliar accents at the pre-lexical level, in a potentially lexically-guided fashion (see Norris et al., 2003). In this view, listeners use knowledge of existing word forms to work out the specific mapping between speech in their own native accent and the speech they are adapting to, allowing for more efficient and precise processing. Given that child speech contains systematic deviations from adult forms, one could consider it to be somewhat analogous to accented speech (see Cooper et al., 2020 for discussion), and indeed, the acoustic systematicity within child speech has been leveraged to boost speech intelligibility in Automatic Speech Recognition (ASR) settings (Shivakumar & Georgiou, 2020). It would follow then that understanding child speech would be easier for adults who routinely interact with children. In fact, mothers’ performances were often compared to that of other adults who likely interacted with children to a lesser extent (e.g., undergraduates; Weist & Kruppe, 1977; Weist & Stebbins, 1972) as well. This may indicate that the ability to understand children's productions is similarly related to one's experience, validating prior claims of a "general child speech advantage".

Of course, it could also be the case that a general child speech advantage exists, but not in all adults who routinely interact with children. For example, in the accent adaptation literature, exposure to multiple accented talkers, rather than just one, can facilitate more successful generalization across novel talkers of the same accent (e.g., Bradlow & Bent, 2008; Kraljic & Samuel, 2007; Reinisch & Holt, 2014; Potter & Saffran, 2017). This increased variability in the form of multiple talkers is argued to draw listeners' attention to relevant dimensions in unfamiliar accented speech which allows them to efficiently learn the distinctions that matter (Baese-Berk, Bradlow, & Wright, 2013; Lively et al., 1993; Nygaard & Pisoni, 1998). Thus, if we were to draw analogy to the child speech intelligibility literature, those who interact regularly with a large variety of children (e.g., early childhood educators) as opposed to just a few (e.g., mothers) may be more successful in generalizing experience-dependent adaptions to child speech across novel children. Parents themselves consistently report that educators understand their typically developing child better than unfamiliar strangers (McLeod et al., 2012; 2015; Ng et al., 2014), however no direct evidence exists for this claim. Indeed, the only study that has investigated educators' ability to understand children they did not already know involved recordings produced by speech-delayed children (Flipsen Jr, 1995). If it is the case that adaptation to child speech is aided by experience interacting with many children, educators might outperform mothers when understanding novel children.

On the other hand, given the extensive inter- and intra-talker variability in child speech, it might be much more challenging to learn the patterns in child speech than it is to learn the patterns of an unfamiliar accent from just experiencing the speech of multiple exemplars. That is, perhaps routine exposure to multiple children is not enough to facilitate adaptation to the speech of all children. Rather, it is possible that additional specialized knowledge, for example, formal training about the known patterns in child speech productions, is required to successfully understand the speech of novel children. In this case, individuals like speech-language pathologists (SLPs), may be particularly equipped to contend with this challenge. Indeed, SLPs have been shown to be more sensitive to phonetic detail in child speech compared to untrained listeners (Munson et al., 2012). Therefore, given their specialized knowledge of the patterns typically observed in child speech, SLPs might be at an even greater advantage when tasked with understanding all children compared to primary caregivers for young children (e.g., mothers), and perhaps even educators.

To summarize, the current literature leaves many questions unresolved regarding the types of experience, if any, that may facilitate a general improvement in the processing of child speech. In the current study, we carry out two experiments examining these questions by testing adult listeners on their ability to understand words produced by familiar and unfamiliar children. In Experiment 1, we explore whether a general child speech advantage can be acquired via regular exposure to children. We presented child speech to undergraduates with little interaction with children, mothers of young children, early childhood educators, and SLPs, and asked them to transcribe the utterances as accurately as possible. If the intelligibility of child speech is driven by experience-dependent adaptation, listeners who frequently interact with children (e.g., mothers, early childhood educators, SLPs) should outperform undergraduates who report little interaction with children. Also, if the type of experience (e.g., number of children, formal training) one has with children differentially influences how intelligible one finds child speech, we might see additional performance differences between the experienced listener groups. Then in Experiment 2, we examine whether the ability to understand child speech is tied to the listener's familiarity with the talker by having mothers transcribe child speech produced by their own child as well as another child who matched in age and sex assigned at birth. Here, we expect to replicate past studies and to find a talker familiarity benefit, such that mothers will perform better with their own child's speech than another child's speech (as well as with their own speech over another adult's speech). Taken together, our findings offer clarity on the question of who, if anyone, 'speaks kid'. In turn, this clarity has implications for our general understanding of human speech processing and talker adaptation.

Experiment 1

What type of experience with children, if any, best enables adults to understand speech produced by typically developing children? We address this question in Experiment 1 by asking undergraduates, mothers, early childhood educators, and SLPs to complete a transcription task featuring spoken single words produced by a set of 2.5- to 6-year-old children and adults. We included a range of child talker ages, with the expectation that any child speech processing advantages that we observe might be more apparent with younger children who, relative to older children, are typically more difficult for adults to understand. If the ability to understand child speech is an experience-driven ability related to a general attunement to the manner in which children speak, then those with greater regular experience with children (e.g., mothers, educators) should demonstrate greater transcription accuracy than those with limited experience with children (e.g., undergraduates). In addition, if the type of experience held by the listener matters, we expect to also see differences in performance between experienced listener groups (i.e., SLPs and educators may perform better than mothers of young children). Although our key predictions all focused on child speech, we included adult speech trials to ensure that any observed listener-related advantages in performance were related to child speech specifically, rather than to differences in motivation or general task performance advantages.

Method

Participants

Four types of listeners were tested: 1) undergraduates who reported interacting with children no more than once per month (N = 48; M_{age} = 19.7 years; 30 female), 2) mothers of children between the ages of 2.5 to 6 (N = 48; M_{age} = 37.0 years), 3) female early childhood educators of children between the ages of 2.5 to 6 (N = 48; M_{age} = 41.3 years), and 4) female speech-language pathologists (N = 48; M_{age} = 34.2 years; also referred to as speech and language therapists in the UK) who reported regular experience with children between the ages of 2 to 6. All participants were recruited from Southern Ontario, learned English before the age of six in North America, and reported English was their dominant language. All participants reported normal hearing at the time of testing and were compensated for their participation with either course credit or $5. Data from all participants were included in the final analysis.

The number of permutations of experimental conditions necessary to counterbalance the stimuli required us to run at least 48 participants per listener group. Previous work examining how experience with a particular type of speech improves later recognition of words produced within that same type of speech reports effect sizes on the order of Cohen's d = 0.5 – 1.1 (Clarke & Garrett, 2004; Maye et al., 2008). Correspondingly our current sample size allows for 80% power to detect effect sizes of $d \geq 0.51$.

Participants in the mother listener group had an average of 2.56 children each (range = 1 - 6 children). Mothers who were also educators or SLPs were not included in the mother listener group since they had routine exposure to large numbers of children. Participants in the early childhood educator listener group each had an average classroom size of about 25.35 children (range = 10 - 31 children), and all were current full-time educators who had taught for at least one year prior to the time of testing. Participants in the SLP listener group reported regular experience with an average of 17.29 children per week (range = 1 - 31 children) and were required to be current full-time SLPs who had held their position for at least one year prior to the time of testing. In this experiment, the adults who recorded the stimuli for this study were not the same adults who participated in the listening task.

Stimuli

Speech stimuli were drawn from a corpus of recordings of typically developing Canadian English-learning children and their native English-speaking mothers (see Cooper et al., 2018, for more details on these longitudinal recordings). At the time of the recording, all children had normal hearing and none had ever received speech therapy. In the current study, speech consisted of 32 words (see Appendix) produced by each of 24 children (12 female) at each of three different age points: 30- to 36-months, 42- to 48-months, and 60- to 72-months, and by twelve female adults. Child productions were spontaneously elicited by prompting them to name an image on a screen. By eliciting productions in this fashion, these recordings avoided potential complications associated with imitated speech (e.g., Johannisson, 2014; Ingram, 1989; Weiss, 1982). Children who failed to initially produce the target label were prompted to produce the intended label (e.g., if a child said ‘cellphone’ instead of ‘phone’, they were asked if they could say that word minus the ‘cell’). Only the recordings of children who produced the intended label were used in the study.

Children's production accuracies by age, as well as some general acoustic measures of the stimuli, are presented in the online supplementary materials. Adult productions were included to ensure that any differences observed by listener groups were specific to child speech; our key predictions were centered on the child speech trials.

Stimuli were normalized for root mean square amplitude in Praat 6.0.22 (Boersma & Weenik, 2013), and were embedded in steady-state pink noise at 0 dB signal-to-noise ratio (SNR). Pink noise is an energetic masker commonly used to assess speech perception in noise given its similar frequency balance as speech (e.g., Kostić et al., 2020; Lam et al., 2017; Xie et al., 2014) and this particular noise level has been used in a number of speech-in-noise-tasks to avoid ceiling effects (e.g. Meador et al., 2000; Miller et al., 1951; see Morini & Newman, 2021 for discussion).

Design

In the transcription task, each participant transcribed 64 words produced across seven different talkers (i.e., six different children and one of twelve adult females). Each set of talkers included two children from each of the three age points. Each participant heard the full set of 32 words twice; each child talker produced eight different words and each adult talker produced sixteen different words. Order of trial presentation was randomized. The presentation of word productions by each talker at every age point was fully counterbalanced across participants.

Procedure

The transcription task was presented using the online survey software Qualtrics. Participants were instructed to use headphones and to complete the study in a quiet environment. Each trial consisted of an individually presented word-in-noise and participants were asked to transcribe each word they heard by typing their response into an empty text box. Participants were not told how many different talkers or words they would hear, only that they would hear some words produced by a variety of different talkers. Once a transcription was entered, participants were able to advance to the next trial by pressing the enter key. No feedback on their transcription accuracy was provided and participants were encouraged to use their best guess if unsure. A single experimental session lasted approximately 10 minutes.

Analysis

Each transcription was manually coded as either correct (entered as ‘1’) or incorrect (entered as ‘0’) based on whether it matched the target word exactly. Transcriptions that were homophonous variants (e.g., ‘babee’ instead of ‘baby’), or deviated from the target word by no more than one letter and could not be considered a distinctly different word (e.g., ‘squirel’ instead of ‘squirrel’), were identified as misspellings and typos and were subsequently hand-corrected by a coder. For ease of visualization, the number of words a participant got correct was divided by the total number of words presented, resulting in a mean proportion correct score: participants were given a score of 1 if they transcribed all of the words correctly, a score of 0 if they transcribed all of the words incorrectly, and, for example, a score of 0.625 if they transcribed 20 out of 32 words correctly.

Transparency and Openness

We report how we determined our sample size, all participant inclusion criteria, all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). All data and analysis code are available at [https://osf.io/9fmdr/]. In line with our ethics protocol, approved by the University of Toronto’s Social Sciences Humanities and Education Research Ethics Board, research materials (e.g., speech recordings) are not made publicly available in order to protect participant confidentiality. Data were analyzed using R, version 3.6.2 (R Core Team, 2019) and the package ggplot2, version 3.2.1 (Wickham, 2016). Although not pre-registered, the study's design, sample size, and analyses were planned ahead of data collection.

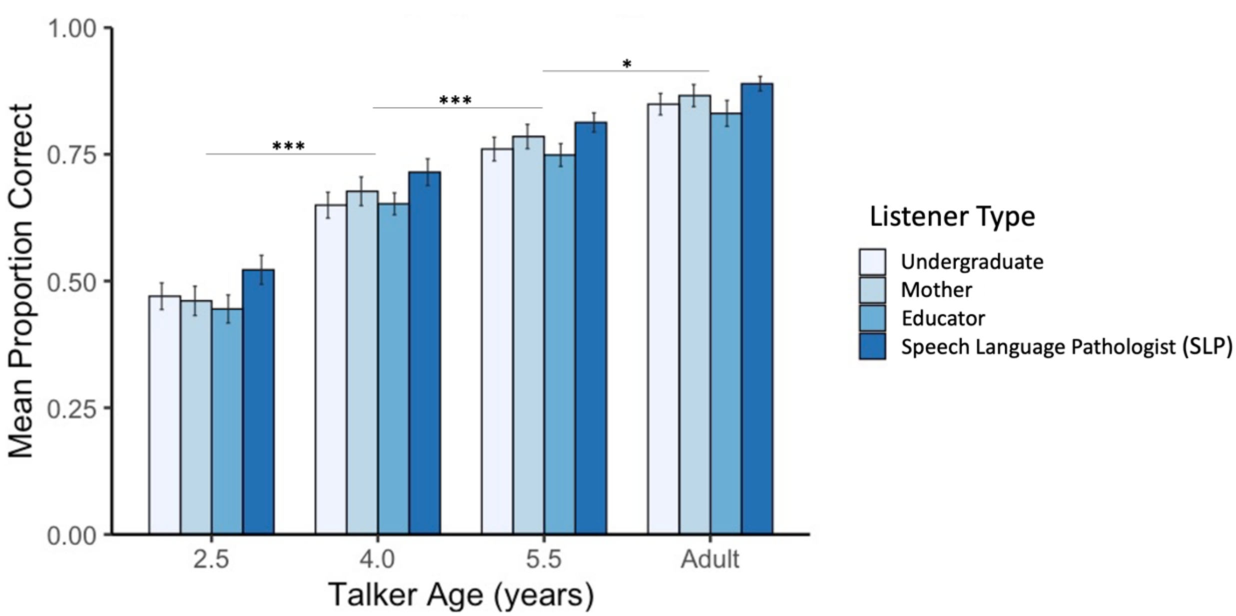
Results and Discussion

We computed average transcription accuracy for all talker age conditions for each of the four listener groups (see Figure 1). In order to compare the effect of listener type across talker age, we constructed a generalized logistic mixed-effects model using the *glmer* function in the *lme4* package Version 1.1-21 (Bates, Mächler, Boler, & Walker, 2015) in R. The model included transcription accuracy (1 = correct response) as the binary response variable, and listener type, talker age, and the interaction between listener type and talker age, as independent variables. Listener type was coded with Helmert contrasts to allow us to compare transcription accuracy between: 1) undergraduates vs. mothers, educators, and SLPs, 2) mothers vs. educators and SLPs, and 3) educators vs. SLPs. Also, since we expected that performance would be more accurate with older than younger talkers, talker age was forward difference-coded to allow for adjacent comparisons: 1) 2.5-year-olds vs. 4-year-olds, 2) 4-year-olds vs. 5.5-year-olds, and 3) 5.5-year-olds vs. adults. The maximal random effects structure that would converge was implemented and included random intercepts for participants, word, talker by talker age, and a random by-participant slope for talker age. The β-coefficient corresponding to each effect represents the difference in log odds of a correct response between the two relevant levels.

As predicted, performance by all listeners improved as the age of the talker increased. Transcription accuracy was less accurate with 2.5-year-olds ($M = 0.47$) than with 4-year-olds ($M = 0.67$), $\beta = -1.07$, $SE = 0.21$, $z = -5.10$, $p < .001$, less accurate with 4-year-olds than with 5.5-year-olds ($M = 0.78$), $\beta = -0.64$, $SE = 0.21$, $z = -3.04$, $p < .005$, and less accurate with 5.5-year-olds than with adults ($M = 0.86$), $\beta = -0.72$, $SE = 0.26$, $z = -2.79$, $p < .01$. Performance by undergraduates (*M* = 0.682) did not significantly differ from listeners who had regular experience with children, β = -0.12, SE = 0.12, z = -1.02, *p* = .31, and performance also did not differ between mothers (*M* = 0.697) compared to educators and SLPs combined, β = -0.02, SE = 0.13, z = -0.15, *p* = .88. There was a difference however, between the transcription accuracy of educators (*M* = 0.669) and SLPs (*M* = 0.735), β = -0.42, SE = 0.15, z = -2.82, $p < .005$. Importantly, there was no interaction effects of listener type comparisons with any of the talker ages (all *p*s > .05). Given that SLPs directly outperformed educators, follow up two-sample *t*-tests were performed between SLPs and the other listener types. Results showed that SLPs, on average, made significantly more accurate transcriptions than undergraduates, $t(94) = 2.35$, $p < .05$, Cohen's *d* = 0.5, but not mothers (*M* = 0.697), $t(94) = 1.56$, *p* = 0.12, Cohen's *d* = 0.3. All other types of listeners performed similarly to each other (all *p*s > .05).

Figure 1

Transcription Accuracy by Listener Type and Talker Age



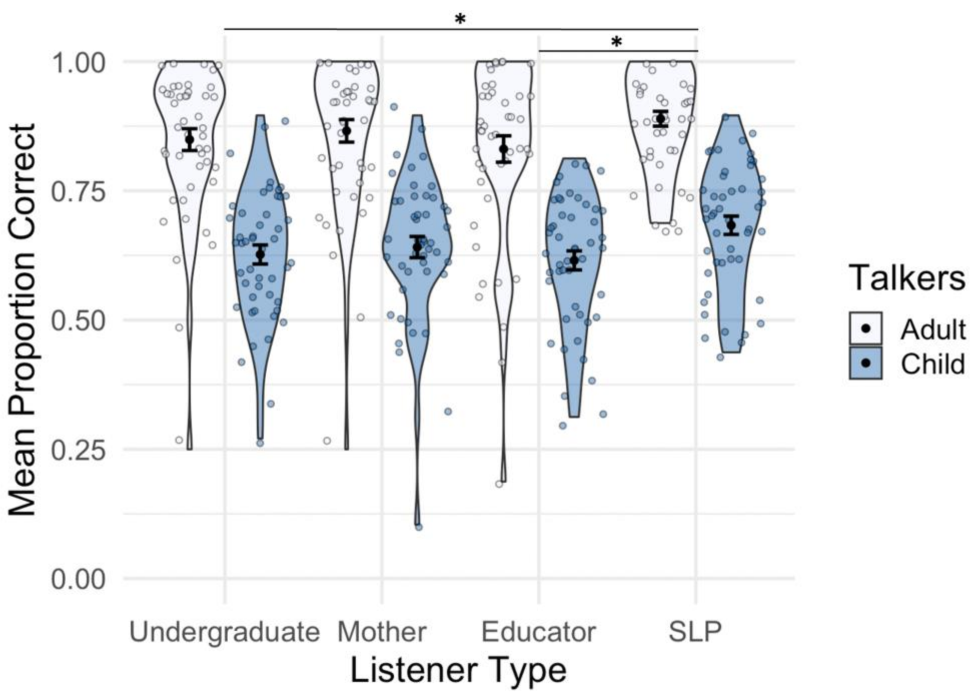
Note. Mean proportion correct transcriptions by each listener type with adult and children speech at three age points (error bars indicate by-participant SE). As expected, listener performance improved with talker age (i.e., 2.5-year-olds were the least intelligible and adults were the most intelligible). Importantly, there were no intergroup listener differences in performance with child speech at any particular age.

Since we were mainly interested in exploring listener-type-related performance differences with child speech and expected that there would be no differences in performance with adult speech, we ran a second analysis using the same model above to confirm that our findings would look the same when adult speech was excluded from the analysis (see Figure 2 for overall listener type performance with child speech versus adult speech). As in the previous analysis, all listeners' performance significantly improved as the age of talkers increased. Accuracy was lower with 2.5-year-olds than with 4-year-olds, $\beta = -1.07$, $SE = 0.21$, $z = -5.08$, $p < .001$, and lower with 4-year-olds than with 5.5-year-olds, $\beta = -0.65$, $SE = 0.21$, $z = -3.04$, $p < .005$. We confirmed that the observed listener-type-related performance differences remained consistent even when adult speech was excluded from the analysis; undergraduates ($M = 0.627$) did not significantly differ from listeners with regular exposure to children, $\beta = -0.11$, SE = 0.12, z = -0.96, $p = .34$, mothers ($M = 0.641$) did not differ from educators and SLPs combined, $\beta = -0.04$, SE = 0.13, z = -0.35, $p = .73$, but again, educators ($M = 0.615$) and SLPs ($M = 0.683$) did differ, $\beta = -0.41$, SE = 0.15, z = -2.84, $p < .005$. There was no interaction effects of listener type comparisons with any of the talker ages (all *p*s > .05). Follow up *t*-tests showed once again that in addition to educators, SLPs also outperformed undergraduates, $t(94) = 2.21$, $p < .05$, Cohen's $d = 0.5$, but not mothers, $t(94) = 1.56$, $p = 0.12$, Cohen's $d = 0.3$. All other types of listeners performed similarly to each other (all *p*s > .05).

To further confirm our finding that no listener-type related differences are observed in the transcriptions of child speech, it is important to determine whether there is indeed no true difference between different listeners' ability to transcribe child speech (i.e., the null hypothesis is true), or whether a difference does exist (i.e., the alternative hypothesis is true), but the present experiment simply lacks sufficient power to detect it. To investigate these two possibilities, we conducted a Bayesian mixed factor ANOVA on the data in the statistics software JASP (version 0.16.3, JASP Team, 2022) with listener type as a between-subjects factor and talker age as a within-subjects factor (see the online supplementary materials for the full analyses and model comparisons). In summary, the Bayes factor analyses confirmed the results of the generalized logistic regression mixed-effects model and supports that the ability to transcribe child speech is not modulated by listener experience.

Figure 2

Transcription Accuracy by Listener Type for Adult and Child Talkers



Note. Mean proportion correct transcriptions by each listener type (error bars indicate by-participant SE). In this figure, data for child speech is collapsed across all three child talker ages. SLPs outperformed undergraduates and educators with all talkers, including adults. Mothers, undergraduates, and educators did not differ in performance.

In sum, contrary to our expectations, we found that routine interactions with children did not appear to predict better ability to transcribe child speech as overall performance did not differ between the listeners who reported limited experience (e.g., undergraduates) and the listeners who reported having regular experience (e.g., mothers, educators, and SLPs) with young children. Also, while all listeners were more accurate at transcribing speech by older children than younger children, we did not find evidence in support of a general child speech advantage in mothers, educators, or SLPs — not even for the speech of the 2.5-year-olds who are the most difficult to understand. Although SLPs, on average, did outperform undergraduates and educators, their performance was not specific to child speech as they demonstrated the same pattern of findings with adult speech as well, indicating that SLPs may just be particularly skilled in understanding speech-in-noise in general. That is, the direction of the SLP effect in our data is impossible to determine; perhaps SLPs performed better in the transcription task because they have more experience listening to speech, or perhaps those who become SLPs just have better speech processing abilities to begin with. We discuss this further in the General Discussion.

Experiment 2

In past studies, mothers' ability to understand their own child was often compared to nonparents' ability to understand unfamiliar children, confounding talker age with talker familiarity. In Experiment 1, we disentangled these factors by testing four types of listeners who varied in the extent of routine exposure they had with child speech on their ability to transcribe utterances produced by unfamiliar children. Using this design, we found no evidence that any group (neither mothers, early childhood educators, nor SLPs) finds child speech more intelligible than nonparents who typically spend little time around children.

How can we reconcile these findings with the existing literature on experiential effects on the intelligibility of child speech? As mentioned above, one possibility is that since past studies often confounded talker familiarity with familiarity with child speech (by testing how intelligible mothers find the speech of their own children), a run-of-the-mill talker familiarity effect (rather than a general attunement to child speech) could explain much of the existing evidence for the maternal advantage in child speech processing. If this is the case, then we should observe an own child speech advantage if we present primary caregivers (i.e., mothers in the current study) with the recordings made by their own children. Thus, in Experiment 2, we test this prediction by asking mothers to transcribe the speech produced by their own child, another child, their own self, and another mother. To avoid ceiling effects and maximize the odds of observing an own child speech advantage, we embedded speech in noise and used the productions of the most difficult-to-understand age group included in Experiment 1 (i.e., 2.5-year-olds). Just as with the child speech, both familiar (i.e., mother's own voice) and unfamiliar (another mother's voice) adult voices were included in the stimulus set. We predicted that mothers would perform better with adult voices than child voices, and better with familiar voices than unfamiliar voices.

Method

Participants

Fifty additional mothers of young children (M_{age} = 36.1 years) were tested. Mothers had an average of 1.82 children each (range = 1 - 4 children). All participants were recruited using the same criteria as in Experiment 1. All participants were different from the mothers who had participated in Experiment 1. Although in theory we could have tested fathers or other types of primary caregivers, all children were accompanied by their mothers, so for that reason, we tested mothers. Participants received participation compensation in the form of $10. Data from all participants were included in the final analyses.

The sample size was determined by the counterbalance of experimental stimuli and is larger than most of the prior studies of familiar children's intelligibility reported by caregivers. Based on previous research on the familiar talker benefit in speech comprehension, we expect mothers' familiarity with child talkers to affect transcription accuracy on the order of Cohen's d = 0.7 - 1.7 (Holmes et al., 2018; Johnsrude et al., 2013; Levi et al., 2011). Correspondingly, with N = 50, we have 80% power to detect effect sizes of $d \geq 0.36$.

Stimuli

The stimuli were drawn from the same corpus of adult and child speech recordings as in Experiment 1, such that some tokens presented here are identical to the tokens presented in Experiment 1. Stimuli consisted of the same set of 32 words produced by each of the 50 mothers themselves and their children at 30- to 36-months.

Design

Participants returned to complete the transcription task one to five weeks (M = 19.7 days, range = 7-35 days) after their recording sessions. Each participant heard eight unique productions from themselves, their own child, another mother, and another age- and sex-matched child, such that they heard the full set of 32 words from four different talkers. Both children and adults were matched by their sex assigned at birth. No same word was heard more than once by a single participant.

Procedure

The transcription task was similar to that in Experiment 1. Participants were tested individually in a quiet testing room and wore Sennheiser HD 280 PRO headphones at a constant, comfortable listening level. Participants were told that they would hear words produced by themselves, their own child, another mother, and another child.

Stimuli were normalized and embedded in noise in the same manner as in Experiment 1. Each participant transcribed 32 words in random order, and a single experimental session lasted approximately 5 minutes.

Analysis

Transcription accuracy was coded in the same manner as in Experiment 1.

Transparency and Openness

Transparency and openness are the same as in Experiment 1.

Results and Discussion

We computed average transcription accuracy by each mother across the four talker groups (see Figure 3). To compare the effects of talker age and talker familiarity on performance, we fit a generalized logistic mixed-effects model to our data using the *glmer* function in the *lme4* package in R. The model included transcription accuracy (1 = correct response) as the binary response variable, and talker age (adult, child), talker familiarity (familiar, unfamiliar), and their interaction, as simple coded independent variables. The maximal random effects structure that would converge was implemented, and included a random intercept for participant, word, and talker, and random by-participant slopes for talker age and talker familiarity.

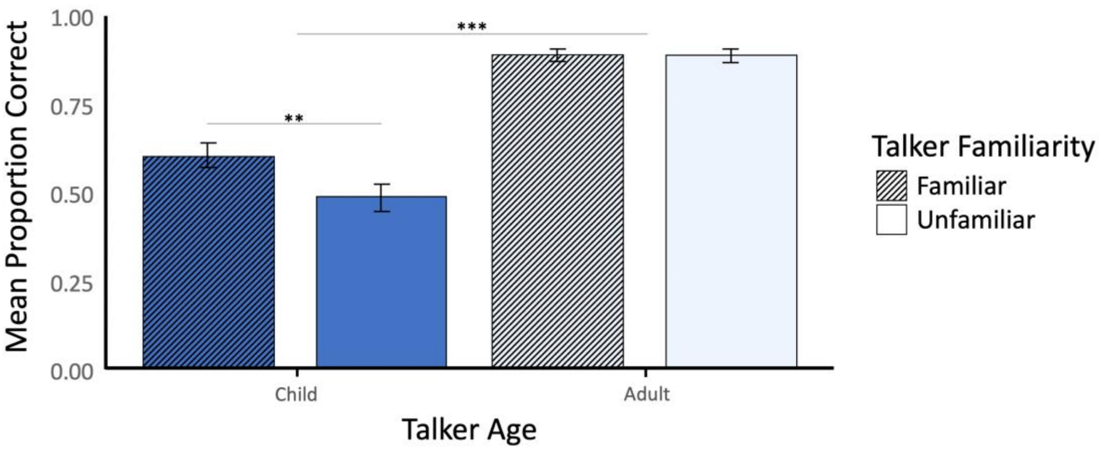
As predicted, the model revealed a significant effect of talker age, with mothers demonstrating more accurate transcriptions for adult productions than for child productions, $\beta = -2.29$, $SE = 0.27$, $z = -8.57$, $p < .001$. Also as predicted, there was a significant effect of talker familiarity, in which transcriptions were more accurate with familiar than unfamiliar talkers, $\beta =$

-0.34, *SE* = 0.16, *z* = -2.09, $p < .05$. Additionally, the interaction between talker age and talker familiarity was significant, β = -0.70, *SE* = 0.32, *z* = -2.18, $p < .05$.

To further investigate the effect of the interaction between talker age and talker familiarity, we performed follow-up tests via paired *t*-tests. There was no effect of talker familiarity with adult talkers, as mothers were equally accurate at transcribing speech by familiar (*M* = 0.90) and unfamiliar adult talkers (*M* = 0.898), *t*(49) = 0.09, *p* = .93, Cohen's *d* = 0.01, however there was a significant effect of talker familiarity with child talkers, as they were significantly more accurate at transcribing speech by familiar child talkers (*M* = 0.613) than unfamiliar child talkers (*M* = 0.494), *t*(49) = 2.85, $p < .01$. Cohen's *d* = 0.4.

Figure 3

Mothers' Transcription Accuracy by Talker Age and Talker Familiarity



Note. Mean proportion correct transcriptions by mothers with speech produced by their own 2.5-year-old child, another 2.5-year-old child, their own self, and another mother (error bars indicate by-participant SE). A talker familiarity benefit was observed by mothers who found their own child's speech more intelligible than another child's speech.

This pattern of results demonstrates that adult listeners found it substantially more difficult to process child speech than adult speech, and it indicates that talker familiarity effects do play a role in child speech processing. Contrary to previous work (e.g., Eger & Reinisch, 2019), however, we found no evidence for an own voice advantage; this is likely due to a ceiling effect with the adult voices in our study or perhaps failure by mothers to recognize their own voice, as bone conduction causes one's voice to sound different when hearing it while speaking versus hearing it played back in a recording. In sum, the results provide clear support for the argument that mothers' intelligibility of child speech is influenced by a talker familiarity benefit, such that they are more accurate at transcribing speech by their own child relative to an unfamiliar child.

General Discussion

Do mothers really know best when it comes to understanding children, as has been claimed? In this paper, we argue that if primary caregivers and others who routinely interact with children do indeed 'speak kid', then this ability should not only be apparent when adults are tested on familiar children's speech (as has been shown in past studies) – but also when they are presented with the speech of unfamiliar children (something not carefully tested in previous work). And that if experience with children can drive adaptation to child speech in the same way experience with an accent can drive adaptation to that accent, then we might see a stronger child speech processing advantage in adults – such as early childhood educators – who interact with many children on a regular basis, in much the same way we sometimes see better adaptation to novel accents when listeners are exposed to multiple talkers of that accent (e.g., Bradlow & Bent, 2008; Kraljic & Samuel, 2007; Reinisch & Holt, 2014; Potter & Saffran, 2017). We also consider the possibility that explicit training on how children's speech typically deviates from adult speech might offer additional benefits in the processing of child speech above and beyond the benefits that may be offered by routine interaction with children (i.e., SLPs might show an intelligibility benefit specifically for child speech due to their knowledge of how child speech typically patterns). But we found no convincing evidence for a general child speech processing advantage in any of the listener groups tested, not even early childhood educators who routinely interact with large numbers of children. We conclude that adult listeners may not adapt to child speech as they adapt to accented speech, and we suggest that evidence supporting claims that mothers ‘speak kid’ may not indicate a maternal advantage at all, but instead, indicate a talker familiarity effect often observed in the adult speech processing literature.

In Experiment 1, we tested four listener groups on their ability to transcribe speech produced by children with the prediction that mothers, educators, and SLPs should be more accurate than undergraduates (who report minimal experience interacting with young children) on their transcription of child – but not adult – productions. While all listeners’ transcription accuracy improved as the age of the child talkers increased (as one would predict), mothers and educators were not better than undergraduates at understanding words produced by children. Interestingly, only SLPs demonstrated greater transcription accuracy with child speech relative to the other types of listeners, however, this better performance was not exclusive to child speech as SLPs also outperformed the other listeners with adult speech. Interacting with multiple children on a routine basis also did not make a difference in listeners’ performance as educators and mothers did not differ in their understanding of child speech. Taken together, these findings indicate that simply having more regular routine interactions with children, no matter the number, does not lead to improved ability to understand of children’s speech productions.

The results from Experiment 1 imply that the child speech intelligibility advantage by mothers in past work may be best explained within the framework of talker familiarity effects. We investigated this possibility further by using a similar transcription task in Experiment 2 to test mothers' understanding of speech produced by themselves, another mother, their own child, and another age- and sex-matched child. Indeed, while mothers found child speech less intelligible compared to adult speech overall, they demonstrated a strong own child speech advantage, such that their understanding of child speech was largely improved when speech was by their own child. This suggests that although mothers' experience with children does not enable them to understand child speech better than those with less experience with children, it does allow them to better understand the particular children they have experience with. We argue then, that the purported mother advantage in past studies may be largely attributed to a run-of-the-mill talker familiarity benefit, such that since test stimuli often included speech by the mothers' own children (e.g., Flipsen Jr, 1995; Weist & Kruppe, 1977; Weist & Stebbins, 1972), their familiarity with the child talker aided processing of child speech.

But if child speech contains systematic deviations from adult forms (much like how accent varieties can systematically deviate in predictable ways), then why didn't we find evidence for adaptation to child speech in listeners who were routinely exposed to large numbers of child talkers? Given current conceptualizations of how talker and accent adaptation work, it seems only logical that listeners who routinely interact with children would show an advantage in understanding child speech over those who do not routinely interact with children. As previously mentioned, we recognize that compared to adult speech, child speech also features substantially higher variation not only across different children, but within the speech of the same child talker (e.g., Lee et al. 1999; MacDonald et al., 2012), and this may have introduced an additional challenge in adapting to child speech. This is further supported by recent work in the accent adaptation literature that reports that listeners are more likely to generalize across talkers that are acoustically similar (Xie & Myers, 2017). This suggests that cross-talker adaptation may not be based in the development of an accent-wide schema, but rather, in listeners' attunement to the specific talkers they hear. If child speech is processed in a similar manner, this could explain why experience-dependent general adaptation to child speech was absent from the present study. That said, we readily admit that the ecological validity of our conclusions could be strengthened by further exploring this issue using target words produced in speech phrases without noise (e.g., see Stringer & Iverson, 2019). It could also be illuminating to examine the potential impact of visual context on our results. For instance, it may be that in the real world, those with regular experience interacting with children also make use of the visual information (e.g., viseme information) to aid their processing of child speech.

Interestingly, although we did not find any evidence that SLPs were particularly good at processing speech by children specifically, we did find evidence that SLPs possessed a general advantage in processing speech by all types of talkers – both children and adults alike. What could explain this? As previously noted, SLPs are more reliable and accurate than untrained listeners at perceiving subtle speech cues within speech (Munson et al., 2012, see, however, Meyer & Munson, 2021). Past work also shows that SLPs are better at transcribing disordered speech (Doyle et al., 1989). It is possible, then, that SLPs utilize their specialized knowledge to help them process speech within challenging listening contexts (e.g., in noise). Just as prior research observes advantages in speech perception for those with musical training compared to nonmusicians (e.g., Kraus et al., 2009; Parbery-Clark et al., 2009), it is potentially the case that SLPs performed well on the current task as a function of improved auditory perceptual skills gained via training. This would crucially suggest that understanding speech is a skill that can be improved upon. In addition, it is also possible that our findings are an indication that individuals who self-select speech pathology as a career are people with a predisposition to acute auditory perception. Indeed, research has found that people with specialized training in phonetics differ from untrained people in the size of the left pars opercularis, a brain structure whose morphology is established in utero (Golestani et al., 2011). Future work should disentangle these possibilities by testing SLPs at different stages of their careers.

At the outset of this study, we set out to answer a simple question: who speaks ‘kid’? Past research has suggested that mothers were uniquely skilled in ‘speaking kid’; however, we found no evidence for this assertion. Moreover, we found no evidence that even educators — who have extensive experience interacting with large numbers of children — are any better at understanding unfamiliar children that undergraduates who rarely interact with children. Thus, our findings suggest that ‘speaking kid’ might not be an ability that is gained by any typical listener through regular interactional experience. At best, those who have received explicit formal training and have experience with hearing nonstandard speech (e.g., SLPs) may be more capable than others in transcribing speech within challenging contexts, that includes, but is not limited to, child speech. We conclude that conceptualizing child speech as something akin to an accent that listeners can simply adapt to may not be appropriate; rather, child speech (in all its variable manifestations) may be best conceptualized as highly idiosyncratic and varying along many of the same sorts of dimensions as adult speech. Thus, primary caregivers do not show a superior ability to understand their children because they possess a generalized child speech processing advantage, but because they are best at processing the speech of children they know well — just like adults are superior at processing the speech of familiar adults.

References

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. The Journal of the Acoustical Society of America, 133(3), EL174–EL180. https://doi.org/10.1121/1.4789864

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Baudonck, N. L. H., Buekers, R., Gillebert, S., & Van Lierde, K. M. (2009). Speech intelligibility of Flemish children as judged by their parents. Folia Phoniatrica et Logopaedica, 61(5), 288-295.

Boersma, P., & Weenink D. (2013). Praat: Doing phonetics by computer. 5.3.47. http://www.praat.org/.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Clarke, C. M., & Garrett, M. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647-58 .

Cooper, A., Fecher, N., & Johnson, E.K. (2020). Identifying children's voices. *The Journal of the Acoustical Society of America, 148,* 324-333. doi.org/10.1121/10.0001576

Cooper, A., Fecher, N., & Johnson, E. (2018). Toddlers' comprehension of adult and child talkers: adult targets versus vocal tract similarity. *Cognition, 173,* 16-20. doi.org/10.1016/j.cognition.2017.12.013

Cooper, A., Paquette-Smith, M., Bordignon, C., & Johnson, E.K. (2022). The influence of accent distance on perceptual adaptation in toddlers and adults. *Language Learning and Development.*

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 704–728. https://doi.org/10.1037/a0017449

Drager, K. D. R., & Reichle, J. E. (2001). Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC. *Journal of Speech, Language, and Hearing Research, 44*(5), 1052-1057.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914–927. https://doi.org/10.1037/0096-1523.23.3.914

Doyle, P. C., Swift, E. R., & Haaf, R. G. (1989). Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. *Journal of Communication Disorders*, *22*(2), 105-113.

Eger, N. A., & Reinisch, E. (2019). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 552-571. doi: 10.1037/xlm0000599

Fikkert, P. (2010). Developing representations and the emergence of phonology: Evidence from perception and production. In C. Fougeron, B. Kühnert, M. D'Imperio & N. Vallée (Ed.), *Laboratory Phonology 10* (pp. 227-260). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110224917.3.227

Flipsen Jr., P. (1995). Speaker-listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders, 28*(1), 3-19.

Fraser, C., & Roberts, N. (1975). Mothers' speech to children of four different ages. *Journal of Psycholingistic Research, 4(1),* 9-16.

Goehl, H., & Martin, P. (1987, November). Intelligibility and familiarity: Is experience enough. In *Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.*

Golestani, N., Price, C., & Scott, S. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience, 31,* 4213–4220.

Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar Voices Are More Intelligible, Even if They Are Not Recognized as Familiar. *Psychological Science, 29*(10), 1575–1583. https://doi.org/10.1177/0956797618779083

Ingram, D. (1989). *Phonological disability in children.* London, Whurr Publishers.

JASP Team (2022). JASP (Version 0.16.3) [Computer software].

Jeon, M., & De Boeck, P. (2017). Decision qualities of Bayes factor and p value-based hypothesis testing. *Psychological Methods, 22*(2), 340.

Johannisson, T. B., Lohmander, A., & Persson, C. (2014). Assessing intelligibility by single words, sentences and spontaneous speech: a methodological study of the speech production of 10-year-olds. *Logopedics Phoniatrics Vocology, 39*(4), 159-168.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological science, 24*(10), 1995-2004.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1-2. http://dx.doi.org/10.1037/amp0000263

Kostić, D., Milivojević, M., & Milivojević, Z. (2020). Evaluation of the influence of Pink Noise (PN) on the intelligibility of speech in the Serbian language using Serbian Matrix Sentence Test (SMST) base and STOI algorithm. In UNITEH.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*, 1–15.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009) Experience-induced malleability in neural encoding of pitch, timbre, and timing: Implications for language and music. *Annals of the New York Academy of Sciences,* 1169: 543–557. doi:10.1111/j.1749-6632.2009.04549.x.

Lam, B. P., Xie, Z., Tessmer, R., & Chandrasekaran, B. (2017). The downside of greater lexical influences: Selectively poorer speech perception in noise. Journal of Speech, Language, and Hearing Research, 60(6), 1662-1673.

Lee, S., Potamianos, A., and Narayanan, S. (1999). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *Journal of the Acoustical Society of America, 105,* 1455–1468

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: whose familiar voices are more intelligible?. *The Journal of the Acoustical Society of America, 130*(6), 4053–4062. https://doi.org/10.1121/1.3651816

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3 Pt 1), 1242–1255. https://doi.org/10.1121/1.408177

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., and Munhall, K. G. (2012). "Children's development of self-regulation in speech production," *Current Biology, 22*(2), 113–117

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognitive science, 32*(3), 543–562. https://doi.org/10.1080/03640210802035357

Meador, D., Flege, J., & Mackay, I. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3*(1), 55-67. doi:10.1017/S1366728900000134

Meyer, M. K., & Munson, B. (2021). Clinical experience and categorical perception of children's speech. *International Journal of Language & Communication Disorders, 56*(2), 374-388.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The intelligibility in Context Scale: validity and reliability of a subjective rating measure. *Journal of speech, language, and hearing research : JSLHR*, *55*(2), 648–656. https://doi.org/10.1044/1092-4388(2011/10-0130)

McLeod, S., Crowe, K., & Shahaeian, A. (2015). Intelligibility in Context Scale: Normative and validation data for English-speaking preschoolers. *Language, Speech, and Hearing Services in Schools*, *46*(3), 266-276.

Miller, G., Heise, G., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the speech materials. *Journal of Experimental Psychology, 41*(5), 329–335. https://doi.org/10.1037/h0062491

Moerk, E. (1974). Changes in verbal child-mother interactions with increasing language skills of the child. *Journal of Psycholinguistic Research*, *3*(2), 101-116

Morini, G. & Newman, R. S. (2021). A comparison of monolingual and bilingual toddlers' word recognition in noise. *International Journal of Bilingualism*. doi:10.1177/13670069211028664.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379–390. https://doi.org/10.3758/BF03210878

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: a comparison between speech-language pathologists and clinically untrained listeners. *American journal of speech-language pathology, 21*(2), 124–139. https://doi.org/10.1044/1058-0360(2011/11-0009)

Ng, K. Y. M., To, C. K. S., & Mcleod, S. (2014). Validation of the Intelligibility in Context Scale as a screening tool for preschoolers in Hong Kong. *Clinical linguistics & phonetics, 28*(5), 316-328.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology, 47*(2), 204–238. https://doi.org/10.1016/s0010-0285(03)00006-9

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & psychophysics, 60*(3), 355–376. https://doi.org/10.3758/bf03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009) Musician enhancement for speech-in-noise. *Ear and Hearing, 30*(6): 653–661. doi:10.1097/AUD.0b013e3181b412e9.

Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition, 166*, 67–72. https://doi.org/10.1016/j.cognition.2017.05.031

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 539–555. https://doi.org/10.1037/a0034409

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E. J. (2017). Bayesian analysis of factorial designs. *Psychological Methods, 22*(2), 304.

Shivakumar, P. G., & Georgiou, P. (2020). Transfer Learning from Adult to Children for Speech Recognition: Evaluation, Analysis and Recommendations. Computer speech & language, 63, 101077. https://doi.org/10.1016/j.csl.2020.101077

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology*, *24*(08), 689-700.

Stringer, L., & Iverson, P. (2019). Accent intelligibility differences in noise across native and nonnative accents: Effects of talker–listener pairing at acoustic–phonetic and lexical levels. *Journal of Speech, Language, and Hearing Research, 62*(7), 2213-2226.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child development*, 549-565.

Vihman, M. (1993). Variable Paths to Early Word Production. *Journal of Phonetics, 21*, 61-82.

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics, 39*(4), 479-491.

Weiss, C. E. (1982). *Weiss intelligibility test.* CC Publications.

Weist, R. & Kruppe, B. (1977). Parent and sibling comprehension of children's speech. *Journal of Psycholinguistic Research, 6(1)*, 49-58.

Weist, R. & Stebbins, P. (1972). Adult perception of children's speech. *Psychonomic Science, 27(6),* 359-360.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Xie, X., & Myers, E. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language, 97,* 30-46.

Xie, Z., Yi, H. G., & Chandrasekaran, B. (2014). Nonnative audiovisual speech perception in noise: dissociable effects of the speaker and listener. PloS one, 9(12), e114439. https://doi.org/10.1371/journal.pone.0114439

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging, 15*(1), 88–99. https://doi.org/10.1037/0882-7974.15.1.88

Appendix

A list of the 32 words used in Experiment 1 and 2

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

Supplementary Materials

Children's proportion of deviations from adult forms broken down by age (Experiment 1)

Deviations	***2.5-years-old***	***4.0-years-old***	***5.5-years-old***
0	0.3697	0.6263	0.8463
1	0.4088	0.2825	0.1328
1+	0.2213	0.0911	0.0208

Coders classified children's productions as featuring either no deviations, one deviation, or multiple deviations from the adult form of the word. Each of the 2304 productions were coded by two different independent coders. Intercoder agreement was 97.96% and any discrepancies in coding were resolved by a third independent coder.

Average acoustic measures, with standard deviations in parentheses (Experiment 1)

Measure	***2.5-year-olds***	***4.0-year-olds***	***5.5-year-olds***	***Adults***
Average f0 (Hz)	314.57 (50.90)	290.76 (51.01)	265.98 (46.33)	200.57 (40.42)
SD f0 (Hz)	64.00 (29.54)	52.13 (29.30)	43.17 (23.03)	46.65 (25.74)
Duration (s)	0.86 (0.29)	0.81 (0.26)	0.75 (0.22)	0.61 (0.16)

Note. SD = standard deviation; Duration = average token duration; Hz = Hertz; s = second.

Bayesian Repeated Measures ANOVA ▾

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject and random slopes)	0.625	2.791e−63	1.117e−62	1.000	
Age	0.125	0.673	8.226	2.410e+62	1.315
Age + ListenerType	0.063	0.326	1.933	1.167e+62	2.699
Age + ListenerType + Age ∗ ListenerType	0.063	0.001	0.005	4.744e+59	1.186
ListenerType	0.125	5.142e−64	2.057e−63	0.184	0.596

Note. All models include subject, and random slopes for all repeated measures factors.

Post Hoc Tests

Post Hoc Comparisons - Age

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
2.5	4.0	0.587	2.889e+21	4.919e+21	4.051e−24
	5.5	0.587	1.573e+49	2.677e+49	6.085e−54
4.0	5.5	0.587	9.895e+10	1.685e+11	2.308e−18

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Post Hoc Comparisons - ListenerType

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
Mother	SLP	0.414	0.189	0.455	0.036
	Teacher	0.414	0.086	0.208	0.070
	Undergrad	0.414	0.062	0.150	0.092
SLP	Teacher	0.414	1.945	4.696	0.005
	Undergrad	0.414	0.639	1.542	0.013
Teacher	Undergrad	0.414	0.059	0.144	0.095

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Examination of the Q-Q plots suggested that the assumption of normality was not violated. We used the default JASP prior for fixed effects (i.e., r scale prior width = 0.5). A Bayes factor greater than 3 indicates support in favor of the alternative hypothesis, a value lower than .33

indicates support for the null hypothesis, and values in between suggest the data is inconclusive (Rouder et al., 2017; Jeon & De Boeck, 2017). Here, the data were best represented by the model that included age as the only main factor. The Bayes factor (BF_{10}) was $2.41*10^{62}$, indicating decisive evidence in favor of this model over the null hypothesis. This age-only model was 508 times more likely than the full model (i.e., the model with talker age, listener type, and the interaction between the two). Comparatively, the model that included listener type as the only main factor yielded a Bayes factor of 0.184, indicating support in favor of the null hypothesis. To further investigate the effect of talker age and, importantly, to follow up on the null effect of listener type, post-hoc comparisons (Bayesian t-tests controlled for multiplicity) between the levels of transcription accuracy by each factor were subsequently performed. Posterior odds of < 1 indicate no difference. For age, the adjusted posteriors show decisive evidence for differences in transcription accuracy between all talker ages (all posterior odds > $9.90*10^{10}$). Post-hoc comparisons between listener type indicated moderate evidence (posterior odds = 1.94) that performance differed between SLPs and educators. There was strong support for the null hypothesis (all posterior odds < 1) for all other listener type comparisons.