Do you speak 'kid'? The role of experience in comprehending child speech

Madeleine E. Yu (madeleine.yu@mail.utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON, L5L 1C6, Canada

Angela Cooper (angela.cooper@utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON, L5L 1C6, Canada

Elizabeth K. Johnson (elizabeth.johnson@utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON, L5L 1C6, Canada

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 comprehending children? In Experiment 1, we explore this possibility by testing three types of participants with variable experience interacting with children: undergraduates with minimal experience with children (N=48), mothers of young children (N=48), and early childcare educators (N=36). Participants transcribed single-word utterances produced by the same set of children at 2.5-, 4-, and 5.5-years-old. In Experiment 2, mothers (N=50) completed a similar transcription task that featured speech by their own, and another, 2.5-year-old child. Participants performed similarly regardless of their experience with children, while mothers demonstrated a Familiar Talker Advantage with their own child's speech. Our findings suggest that while experience with children may not facilitate improved comprehension of child speech in general, it may lead to enhanced comprehension of those children in particular.

Keywords: word recognition; speech processing; child speech; comprehension; Familiar Talker Advantage

Introduction

Young children's speech is notoriously difficult to understand. For example, a word such as *spaghetti* can be produced as *getti*, *shark* as *sock*, and *baby* as *baba*. For those who study child speech, these realizations may be unsurprising because they represent known patterns in child productions — clusters are often simplified, easier-to-produce sounds are substituted for more difficult sounds, and stressed syllables can be reduplicated (Vihman, 1993). But what about other types of listeners who don't study child speech, but who still interact with children regularly?

There is ample evidence that indicates speech is better recognized and understood when produced by a familiar talker (Familiar Talker Advantage; Levi, Winters, & Pisoni, 2011; Mullennix & Pisoni, 1990; Nygaard, Sommers, & Pisoni, 1994; Yonan & Sommers, 2000). This advantage is thought to be driven by frequent exposure to the utterances by that individual which, by nature, involves numerous speech productions within many different contexts. In line with this body of work, there is evidence that mothers are superior to nonparents in comprehending their own children (e.g., Flipsen Jr, 1995; Weist & Kruppe, 1977; Weist & Stebbins, 1972). The improved performance we observe in mothers' comprehension of their own children may indicate that mothers recognize the specific speech patterns of their own child. At the same time, all studies to date that demonstrate this own child speech processing advantage in mothers have compared performances between mothers and other adults who likely interact with children to a lesser extent (e.g., undergraduates). Thus it is possible that this own child advantage is actually a general advantage for processing all children's speech. To distinguish between these possibilities (i.e., that mothers tested in past studies demonstrated an own child speech processing advantage versus a general child speech processing advantage), it is essential to compare mothers' ability to understand their own child versus other children. It would also be useful to examine whether other adults who frequently interact with children (e.g., early childhood educators) also show a general child speech processing advantage not seen in adults who interact with children infrequently (e.g., the typical undergraduate student).

Here, we investigate whether regular experience with children facilitates not just an advantage with a specific child, but also a general advantage with all children. Why might we expect a general advantage? Well, given that child speech consists of systematic deviations from adult forms, one could consider it to be somewhat analogous to accented speech. In all initial encounters with nonstandard talkers (e.g., child, accented), listeners typically experience greater perceptual difficulty (e.g. Dupoux & Green, 1997; Munro & Derwing, 1995). Yet, comprehension tends to improve as listeners gain more experience with the particular type of speech variation, leading to more efficient recognition of words produced in that accent (e.g., Clarke & Garrett, 2004; Dahan & Mead, 2010; Norris, McQueen, & Cutler, 2003; Maye, Aslin, & Tanenhaus, 2008). It has been argued that listeners adapt to nonstandard speech at the pre-lexical level, in a potentially lexically-guided fashion (e.g., Norris, McQueen, & Cutler,

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. It would follow then, that those who interact with children on a daily basis would outperform those who interact with children less when comprehending child speech.

Furthermore, while listeners demonstrate experience-dependent adaptation by successfully learning the patterns of deviations within accented speech by a specific talker (e.g., Bradlow & Bent, 2008; Dahan & Mead, 2010; Norris et al., 2003; Kraljic & Samuel, 2007; Xie & Myers, 2017), evidence also suggests that exposure to multiple accented talkers, may facilitate more successful generalization across novel talkers of the same accent (e.g., Bradlow & Bent, 2008; Kraljic & Samuel, 2007; Reinisch & Holt, 2013; Potter & Saffran, 2017). Some propose that the increased variability in the form of multiple talkers draws attention to relevant dimensions in nonstandard speech and allows listeners to efficiently learn what distinctions are likely to matter (e.g., Baese-Berk, Bradlow, & Wright, 2013; Lively, Logan, & Pisoni, 1993; Nygaard & Pisoni, 1998). Our second prediction then, is that those who interact regularly with multiple children, as opposed to just a few, may more successfully generalize their experience-dependent adaptions to child speech across novel children.

To summarize, it can be difficult to understand speech produced by young children. In the current study, we examined how adult listeners contend with this challenge, by exploring how their experience facilitates their ability to process child speech. We tested three groups of adult listeners who vary in their experience interacting with young children: undergraduates who reported minimal exposure to children, mothers of children under 6 who interact mainly with their own children, and early childhood educators who are routinely exposed to multiple children under 6, on their ability to recognize and transcribe spoken, single word utterances produced by the same typically developing children at 2.5, 4.0, and 5.5 years of age. We predicted that both early child educators and mothers of young children would outperform the undergraduates in the recognition of words produced by children, with the greatest advantage potentially appearing with the youngest age group who are likely to be the most difficult to comprehend, 2.5-year-olds. In addition, if the number of different children one interacts with on a regular basis facilitates better adaptation to child speech in general, we expected that early childhood educators would outperform mothers in the task.

Experiment 1

In Experiment 1, the ability to accurately transcribe spoken single word productions made by both children and adults was compared between undergraduates, mothers, and early childcare educators. We predicted that if adults are able to adapt towards child speech as they do with accented speech, those with greater regular experience with young children would demonstrate greater transcription accuracy.

Method

Participants Three types of listeners were tested: (1) undergraduates with limited experience around children (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), and (3) female early childhood educators of children between the ages of 2.5 to 6 (current N=36, target N=48; M_{age} = 39.8 years).

Mothers, on average, had two children each (range=1-6 children). Early childhood educators, on average, each had a classroom size of about 27 children (range=16-31 children), and were required to be current full-time educators who had taught for at least one year prior to the time of testing. 26 educators reported having experience with 1- to 6-year-old children, 7 reported having experience with 3- to 6-year-old children, and 3 reported having experience with 4- to 6-year-old children. Some of the early childhood educators we tested were also mothers of young children, but were only considered as part of the educator listener group given their routine experience with multiple children. All participants were recruited from the Greater Toronto Area, learned English before the age of six in North America, and English was their dominant language. All of the participants reported no hearing or vision impairments at the time of testing. All participants received compensation for their participation in the form of either course credit or $5.

Stimuli and Design Stimuli were drawn from a corpus of recordings of Canadian English-learning children and their native English-speaking mothers producing a set of 32 words (Cooper et al., 2018) that was elicited by using an experimenter-controlled video game in which an image of a referent of a target word was displayed on a computer screen (e.g., *strawberry, duck*). The child was prompted to name the image, and the child's mother was prompted to name the image after their child. Participants were encouraged to produce these words in citation form and were asked to repeat naming as necessary. Stimuli consisted of words typically known by children as young as 30-months of age, as indexed by an average word production rate of 95% according to Wordbank vocabulary norms (Frank et al., 2016), and varied by word frequency (M=81.5-per-million; range=0.88-514), and by phonological neighborhood density (M=17.9; range=0-46) and length (M=4.1; range=2-8; CLEARPOND database; Marian et al., 2012).

In the present study, the full set of 32 words was included. All words were produced by the same subset of 24 Canadian English-learning 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 adult females. Stimuli were normalized for root mean square amplitude in Praat 6.0.22 (Boersma & Weenik, 2013). Speech stimuli was embedded in speech-shaped noise at 0 dB signal-to-noise ratio (SNR) to increase task-difficulty and to counter potential ceiling effects that may mask differences in performance between listener types across different talker age points, especially as talkers get older. Adult speech was included to provide a

comparison of performance in transcription accuracy between adult speech child speech at three age points.

In the transcription task, each listener heard the full set of 32 words produced twice across seven different talkers (i.e., six different children and one of the twelve adult females) in noise. Eight productions were by each of the six child talkers and sixteen productions were by the adult talker. No same word was heard as being produced more than once by particular talker, and no more than twice in an experimental session by a single participant. All recordings by all child talkers at each age point and all adult talkers were heard once across participants.

Procedure The transcription task was presented using the online survey software Qualtrics. Participants were asked 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. 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. Each participant transcribed 64 words-in-noise in random order. A single experimental session lasted approximately 10 minutes.

Results and Discussion

Average transcription accuracy of the speech by child talkers at three age points and the adult speech by each of the different types of listeners were compared (see Figure 1).

To compare effects of voice age and listener type on performance, we fit a generalized linear mixed-effects model to our data using the *glmer* function in the *lme4* package Version 1.1-21 (Bates, Mächler, Boler, & Walker, 2015) in R. The model included the binary response variable, transcription Accuracy (1 = correct response). The independent variables, Listener Type, voice Age, and the interaction between Listener Type and voice Age, were entered as fixed effects. Listener Type was simple-coded (with mothers as the reference level), and because we expected listeners to be more accurate as voice Age increased, we coded voice Age with Helmert contrasts: 1) 2.5-year-olds vs. 4-year-olds, 5-year-olds, and adults combined, 2) 4-year-olds vs. 5-year-olds and adults combined, and 3) 5.5-year-olds vs. adults. The maximal random effects structure that would converge was implemented, including random intercepts for participant, Word, and Talker by Age, and a random by-participant slope for voice Age. The β-coefficient corresponding to each effect represents the difference in log odds of a correct response between the two levels of that comparison, collapsed over all levels of the other factor.

The model revealed a significant effect of voice Age, such that all listeners' performances differed significantly between the 2.5-year-olds and the older talkers, β = -1.75, SE = 0.18, z = -9.92, p < .001, between the 4-year-olds and the 5-year-olds and adults combined, β = -1.03, SE = 0.20, z = -5.23, p < .001, and between the 5-year-olds and the adults, β = -0.74, SE = 0.26, z = -2.90, p < .01. There was no significant effect for either Listener Type or Listener Type x voice Age interaction (ps > .05).

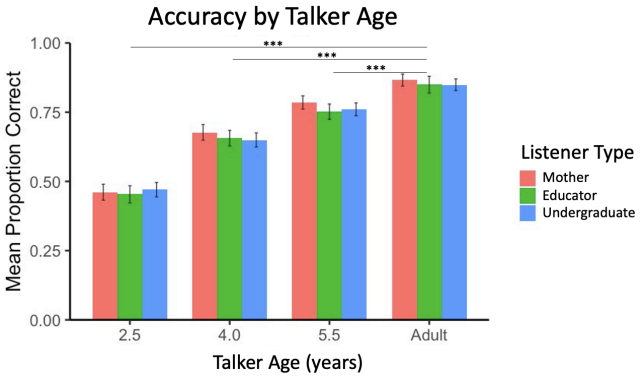

Figure 1: Mean proportion correct transcriptions by each listener type with adult and children speech at three age points (error bars indicate SE).

Thus, though listeners were more accurate at transcribing child speech as children got older, there were no differences in transcription performance between different types of listeners with child speech at any age, contrary to predictions. We therefore found no evidence for our hypothesis that greater experience interacting with children in daily life would facilitate better comprehension of child speech in general.

Experiment 2

Prior work reports that mothers are better at comprehending child speech compared to nonparents, though in all reported cases, speech was produced by the mothers' own children. It is therefore unknown whether those who regularly interact with children would show a generalized advantage for unfamiliar children as well. In Experiement 1, we found no evidence of this, which suggests that the advantages reported in previous studies may have been due in part to talker familiarity effects similarly observed in adult speech processing literature. But the paradigm used in Experiment 1 is quite different from that used in earlier studies that demonstrate an own child advantage. Thus in Experiment 2, we used a similar paradigm as in Experiment 1 to explore whether mothers' familiarity with the talker influences performance on a similar word recognition task when hearing speech produced by themselves, a different mother, their own child, and a different child. We predicted that mothers would demonstrate better word recognition for speech by 1) adults relative to children regardless of familiarity and by 2) their own child relative to an unfamiliar child. If mothers demonstrated a Familiar Talker Advantage with their own child, this would suggest that processing child speech may be tied to listeners' experience with a particular child or set of children.

Method

Participants 50 additional mothers of young children (M_{age} = 36.1 years) were tested. Mothers, on average, had 2 children each (range=1-4 children). All participants were recruited from the Greater Toronto Area, learned English before the age of six in North America, and English was their dominant language. Participants received participation compensation in the form of $10.

Stimuli and Design The stimuli was drawn from the same corpus of adult and child speech recordings as in Experiment 1, and consisted of the same set of 32 words produced by each of the 50 mothers themselves, and their children at 30- to 36-months. Recordings from every mother-child dyad was paired with a gender-matched dyad to ensure that each dyad's productions would be heard by another participant.

Each participant heard 8 unique productions from themselves, their own child, another mother, and another child, such that they heard the full set of 32 words from 4 different talkers. No same word was heard more than once by a single participant.

Procedure The presentation of the transcription task was the same as 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.

Each participant transcribed 32 words embedded in speech-shaped noise at 0 dB signal-to-noise ratio (SNR) in random order. A single experimental session lasted approximately 5 minutes.

Results and Discussion

Average transcription accuracy by each participant was compared between each voice type (see Figure 2).

To compare effects of voice age and familiarity on performance, we fit a generalized linear mixed-effects model to our data using the *glmer* function in the *lme4* package in R. The model included the binary response variable, transcription accuracy (1 = correct response). The independent variables, voice Age (adult, child), voice Type (familiar, unfamiliar), and their interaction, were entered as simple-coded fixed-effects. 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 voice Age and voice Type. The model revealed a significant effect of voice Age, with mothers demonstrating more accurate transcriptions for adult productions than for child productions, β = -2.29, SE = 0.27, z = -8.57, p < .001. There was also a significant effect of voice Type, in which transcriptions were more accurate with familiar than unfamiliar voices, β = -0.34, SE = 0.16, z = -2.09, p < .05. Additionally, the interaction between voice Age and voice Type was significant, β = -0.70, SE = 0.32, z = -2.18, p < .05.

Follow-up tests revealed that while mothers were equally accurate at comprehending familiar and unfamiliar adult voices, β = 0.02, SE = 0.28, z = 0.06, p = .95, they were significantly more accurate at comprehending familiar child voices than unfamiliar child voices, β = -0.75, SE = 0.18, z = -4.16, p < .001.

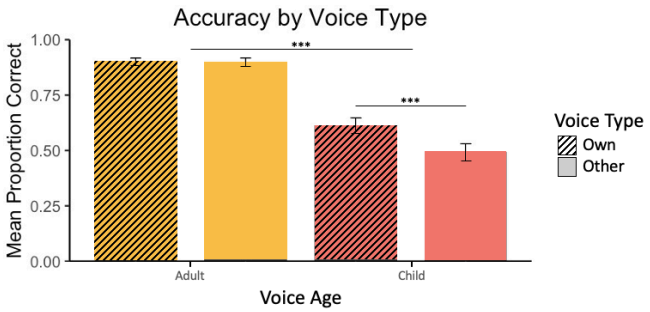

Figure 2: Mean proportion correct transcriptions by mothers with speech produced by themselves, a different mother, their own child, and a different child (error bars indicate SE).

In sum, the results supported our predictions that mothers would demonstrate a Familiar Talker Advantage for their own child relative to an unfamiliar child. The absence of this advantage when mothers transcribed speech by produced by themselves is likely due to performance being at ceiling with adult voices. This indicates first, that mothers process adult speech differently than child speech, and second, we found clear evidence that mothers show an advantage in child speech processing with their own child, which is supported by existing research on talker familiarity effects, but not with other children.

General Discussion

The current study used a speech-in-noise transcription task to investigate how experience interacting with children impacts listeners' comprehension of child speech in general. Given the systematic deviations between child speech and adult speech (Vihman, 1993), we hypothesized that child speech may be processed in a similar manner as accented speech (e.g. Norris et al., 2003), such that greater experience hearing child speech facilitates better recognition of words produced by child talkers. Additionally, we predicted that experience with multiple children on a regular basis would generate more successful comprehension of novel child talkers in general.

In Experiment 1, we found that transcription accuracy improved as the age of the child talkers increased, although performance with the oldest child speech, 5.5-year-olds, remained far below performance with adult speech. Importantly, this supports work that observes age-related voice characteristics greatly affect judgments of speech (Dilley et al., 2013). Contrary to predictions, we found no differences in the transcription accuracy between the three different types of listeners. That is, the varying amounts of experience each listener had interacting with young children did not affect performance.

In order to investigate whether child speech processing advantages by mothers reported in prior work are specific to the speech of familiar children, we used a similar paradigm as in Experiment 1 to test how familiarity with talkers influences performance by mothers of 30- to 36-month-olds in Experiment 2. When tasked with transcribing speech produced by themselves, another mother, their own child, and another age- and gender-matched child, we found first, that child speech was more difficult to transcribe than adult speech overall, and second, that familiar child talkers are much more comprehensible to mothers than unfamiliar child talkers.

Indeed, we reinforce the observation that adults process adult and child speech differently, and this is likely due in part to the systematic ways in which speech by children differs from expected representations of word forms. Yet, if child speech is expected to be processed akin to other types of nonstandard speech, why is it that listeners with regular experience interacting with children only demonstrated greater comprehension with speech by those specific children, but not with all child speech in general? Based on past findings (e.g., Clarke & Garrett, 2004; Dahan & Mead, 2010; Norris, McQueen, & Cutler, 2003; Maye, Aslin, & Tanenhaus, 2008; Van Heugten & Johnson, 2014), one would predict that listeners should use implicitly acquired knowledge about the systematic deviations between the speech they are accustomed to hearing (e.g., adult speech) and nonstandard speech (e.g., child speech) to adapt to novel talkers of the same nonstandard speech. But even within the adult accent adaptation literature, adults do not always demonstrate successful generalization across all talkers (e.g., see Clarke, 2000; Wade et al., 2007).

So how can we account for the absence of general experience-dependent adaptation towards child speech by listeners who routinely interact with children? Is it because any benefit mothers and educators may have in comprehending children is tied to the speech patterns of the particular children they interact with? Or is it because our task was not designed to appropriately tap into any general advantages these listeners might have with young children?

It is possible that the cues adult listeners typically utilize to comprehend novel child talkers in the real world were not accounted for in the current task. Rather, any advantages listeners gain from their regular experience interacting with children might not come from implicit knowledge of the systematic phonological deviations in child speech, but from a higher level, such as familiarity with the sentence structure typically used by child talkers. If this is the case, this would explain why generalization did not occur with comprehension of single-word utterances. Note, that the use of isolated word productions, rather than multi-word utterances, in the current study was necessitated by the practical difficulties associated with obtaining well-controlled high-quality in-lab recordings of 2.5-year-old children. Therefore, this practical challenge in recording child speech must first be addressed before it can be determined whether generalization of learned patterns within child speech occurs at a higher level. Word productions were also embedded in speech-shaped noise to increase overall task difficulty; however, we acknowledge that although children's speech can be heard in noisy environments (e.g., classrooms), the conditions may not be exactly comparable to real world experiences. It is possible that the addition of noise differentially impacted speech perception across talker ages, and perhaps masked crucial information in the speech signal that experienced listeners make use of to optimally comprehend children's speech. Thus, future work could also determine the cases in which successful generalization to novel child talkers should be likely to occur, by testing, for instance, listeners on speech without noise, or even those who have explicit knowledge of the existing patterns within child speech (i.e., speech language pathologists perceive phonetic detail in children's speech better than untrained listeners; e.g., Munson et al., 2012).

Our findings from Experiment 2 are in line with claims that mothers demonstrate superior comprehension with their own child's speech (e.g., Flipsen Jr, 1995; Weist & Kruppe, 1977; Weist & Stebbins, 1972). Additionally, by testing a much larger than typical sample size per listener group (48 in Experiment 1 and 50 in Experiment 2 vs. 4, 12 in other studies) and by directly comparing mothers' comprehension between their own child's speech versus another child's, we provide the first direct evidence to date that child speech processing abilities may be linked to the specific children that one routinely interacts with. One could argue, though, that since mothers were told that they would hear words recorded by themselves and their own child prior to test, this may have influenced them to employ different strategies when attending to familiar talkers which thereby facilitated better word comprehension. Indeed, parents percieve their child as being best understood by those who are more familiar to them, including themselves (Van Doornik et al., 2018), and in fact, brain activity differs depending on whether listeners attend to speech content or talker identity (von Kriegstein et al., 2005). However, evidence shows that the Familiar Talker Advantage persists even in cases when listeners do not explicitly recognize the voices presented to them (e.g., Holmes et al., 2018). Thus it is unlikely that the comprehension advantage demonstrated by mothers with their own child's speech was driven solely by awareness of talker identity.

Note that mothers comprehended their own child's speech at lower rates than what may be expected; however, this is in line with observations from existing work that report parents comprehend their own child's speech with about 55.5% accuracy (Weist & Stebbins, 1972). This is likely a low estimate of mothers' actual capacity to understand their own child and more of a reflection the nature of word recognition tasks, in which linguistic and nonlinguistic context is absent. Thus, future work should examine how such context might influence performance between listeners. Similarly, it should be investigated whether this own child advantage observed with mothers exists with listeners in other interactional contexts (e.g., early childcare educators hearing children

from their own class vs. children from a different class), and just how much experience then, is needed to generate a Familiar Talker Advantage in child speech. Future work could also implement additional, or alternate, measures (e.g., confidence ratings, training blocks, feedback, etc.) to better capture experiential differences in the performance with children's speech beyond just transcription accuracy, such as timecourse diffences in adaptation to speech, or the extent of individual variability between different types of listeners.

To summarize, past work on child speech processing suggests that those who routinely interact with children demonstrate an advantage in comprehending child speech. However, those studies confounded the observation of a general advantage with that of an own child advantage given that they evaluated mothers' and nonparents' comprehension abilities with child speech produced by the mothers' own children. The present study aimed to disentangle these two possibilities by directly examining whether listeners' experience interacting with children not only facilitates a comprehension advantage with those children in particular, but with all children in general. While we found no evidence to support a general advantage with child speech, we did find, as evidenced by mothers, that child speech processing appears to be tied to one's experience with specific child talkers.

Acknowledgments

We would like to thank Lisa Hotson, Thomas St. Pierre, Sofia Mirzazada, Arshnoor Khaira, and Gauri Chaudhari, as well 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 the Natural Sciences and Engineering Research Council.

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.

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

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

Clarke, C. M. (2000). Perceptual adjustment to foreign-accented English. *The Journal of the Acoustical Society of America, 107,* 2856.

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

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

Dilley, L. C., Wieland, E. A., Gamache, J. L., McAuley, J. D., & Redford, M. A. (2013). Age-related changes to spectral voice characteristics affect judgments of prosodic, segmental, and talker attributes for child and adult speech. *J Speech Lang Hear Res., 56*(1), 159–177.

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

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

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: A Repository For Vocabulary Data. *Journal of Child Language, 18,* 1–18.

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

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

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

Marian, V. Bartolotti J. Chabal S., & Shook A.(2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLOS ONE, 7*(8), e43230.

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

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)

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

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

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

Van Doornik, A., Gerrits, E., McLeod, S., & Terband, H. (2018). Impact of communication partner familiarity and speech accuracy on parents' ratings of their child for the Intelligibility in Context Scale: Dutch. *International journal of speech-language pathology, 20*(3), 350-360.

Van Heugten, M. & Johnson, E.K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General, 143,* 340-350. dx.doi.org/10.1037/a0032192

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

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of cognitive neuroscience, 17*(3), 367–376. https://doi.org/10.1162/0898929053279577

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native accented speech sounds, *Phonetica 6*4(2–3), 122–144.10.1159/000107913

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.

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.

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