JSLHR

Research Note

Does Implicit Voice Learning Improve Spoken Language Processing? Implications for Clinical Practice

Julie Case,[a] Scott Seyfarth,[b] and Susannah V. Levi[a]

Purpose: In typical interactions with other speakers, including a clinical environment, listeners become familiar with voices through implicit learning. Previous studies have found evidence for a Familiar Talker Advantage (better speech perception and spoken language processing for familiar voices) following explicit voice learning. The current study examined whether a Familiar Talker Advantage would result from implicit voice learning.
Method: Thirty-three adults and 16 second graders were familiarized with 1 of 2 talkers' voices over 2 days through live interactions as 1 of 2 experimenters administered standardized tests and interacted with the listeners. To assess whether this implicit voice learning would generate a Familiar Talker Advantage, listeners completed a baseline sentence recognition task and a post-learning sentence recognition task with both the familiar talker and the unfamiliar talker.
Results: No significant effect of voice familiarity was found for either the children or the adults following implicit voice learning. Effect size estimates suggest that familiarity with the voice may benefit some listeners, despite the lack of an overall effect of familiarity.
Discussion: We discuss possible clinical implications of this finding and directions for future research.

Every year, nearly a million and a half children receive speech-language therapy under the Individuals with Disabilities Education Act (2004). To qualify for services, children are typically assessed using data from multiple sources, such as standardized language testing and nonstandardized assessment of language (Ireland & Conrad, 2016). Performance across these tasks determines whether children qualify for speech-language intervention services under state and federal guidelines and whether they will continue to receive services as a result of periodic reevaluations. Because of the significant impact that performance on assessment tasks could have on the educational resources available to a child, it is essential to consider factors other than the actual language skills of the child that could influence performance on evaluations and reevaluations of speech and language skills.

Federal, state, and city regulations typically mandate the frequency of reevaluations, ranging from every 6 months (New York City Early Intervention System, 2014) to once every 3 years (New York City Department of Education, 2009). Despite the explicitness of when evaluations should occur, these same regulations do not specify other factors that have been shown to affect performance on standardized assessments. For instance, the particular language assessment tool can impact performance (Peña & Quinn, 1997), in addition to the clinical environment, where children perform better in a quiet setting than in a noisy classroom environment (Nelson, Kohnert, Sabur, & Shaw, 2005; Smith & Riccomini, 2013). The client–clinician relationship can also influence performance (Ebert, 2017; Hoffman, 2014). These factors could result in either better or worse performance on standardized assessments.

An additional factor that could affect performance on standardized and nonstandardized assessment tasks is with the voice of the evaluator, an element typically not discussed in the state and federal guidelines (Individuals with Disabilities Education Act, 2004). Research has shown that listeners are sensitive to talker-specific productions and process speech in a talker contingent manner (Allen & Miller, 2004; Eisner & McQueen, 2005; Kraljic & Samuel, 2007; McQueen, Cutler, & Norris, 2006; Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009; Theodore

[a]Department of Communicative Sciences and Disorders, New York University, New York
[b]Department of Linguistics and Office of Academic Affairs, Ohio State University, Columbus
Correspondence to Julie Case: julie.case@nyu.edu

Editor-in-Chief: Sean Redmond
Editor: Lizbeth Finestack
Received August 10, 2017
Revision received December 2, 2017
Accepted January 19, 2018
https://doi.org/10.1044/2018_JSLHR-L-17-0298

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

& Miller, 2010; Theodore, Miller, & DeSteno, 2009; Trude & Brown-Schmidt, 2012). Furthermore, it has been found that both child and adult listeners are better at processing spoken language by a familiar than an unfamiliar talker (Levi, 2015; Levi, Winters, & Pisoni, 2011; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Souza, Gehani, Wright, & McCloy, 2013; Yonan & Sommers, 2000), a phenomenon known as the "Familiar Talker Advantage." The majority of studies of the Familiar Talker Advantage have used an explicit voice learning task to generate familiarity with a voice (Levi, 2015; Levi et al., 2011; Nygaard & Pisoni, 1998; Nygaard et al., 1994). In these studies, listeners hear a word or sentence, identify the voice of the talker, and receive feedback. The Familiar Talker Advantage has most commonly been examined in young adult listeners (Levi et al., 2011; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Yonan & Sommers, 2000). However, this phenomenon is robust and has been found in both older adult listeners (Yonan & Sommers, 2000) and in school-age children (Levi, 2015). Levi (2015) also found that children with the lowest baseline word recognition scores showed the most benefit from talker familiarization.

Studies of the Familiar Talker Advantage have typically tested spoken language processing following explicit learning of a talker's voice in the laboratory setting (Levi, 2015; Levi et al., 2011; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Yonan & Sommers, 2000). The learning paradigm used in Experiment 2 of Yonan and Sommers (2000) used an incidental voice learning task where listeners first heard sentences and were asked to decide whether the final word had one or two definitions. Then, listeners heard a new set of sentences and were asked whether the voice of the talker had been used in the previous experiment. Although learning was incidental, the procedure still directed listeners' attention to the voice of the talker by asking whether it was a new or old voice. In normal spoken interactions, familiarization to a talker's voice occurs implicitly and does not involve direct attention (whether explicit or incidental) to the talker's voice. One recent study examined the Familiar Talker Advantage with naturally familiar voices (Souza et al., 2013). In this study, older adults with hearing impairment completed a sentence recognition task with the voice of a familiar (spouse or close friend) or an unfamiliar person. Listeners performed better with familiar talkers, showing that familiarity through natural, implicit learning also results in a Familiar Talker Advantage. This study examined highly familiar talkers, where familiarity occurred over many years. Taken together, these studies provide evidence for the Familiar Talker Advantage in the absence of explicit voice training (Yonan & Sommers, 2000) and with natural exposure to a voice (Souza et al., 2013).

In the clinical environment, clients naturally become familiarized to their clinician's voice through implicit learning across repeated interactions with the clinician during therapy. It remains unknown if this natural—or implicit—voice learning also leads to a Familiar Talker Advantage with shorter exposure, which would then have the potential to influence performance on language assessments. Although it would be challenging to control for many of the previously mentioned environmental and interpersonal/intrapersonal factors, the child's familiarity with the voice of the evaluator is something that could be controlled. Furthermore, if voice familiarity results in more accurate spoken language processing, it could mean that children perform better on spoken language assessments when listening to a familiar voice than an unfamiliar voice. Currently, there is no consensus regarding who should perform reevaluations of children's progress in speech-language therapy (i.e., treating clinician vs. novel clinician), a factor that should be considered if a familiar voice facilitates spoken language processing.

The goal of the current study was to investigate whether listeners demonstrate improvements in spoken language processing as a result of implicit learning of a talker's voice after short-term exposure (2 days). To address this goal, voice familiarization occurred through in-person interactions to reflect how listeners become familiar with a voice naturally during spoken exchanges. Sentences, rather than words, were used on the spoken language recognition task to make stimuli more natural and more consistent with the language used in nonstandardized and standardized speech-language assessments. Two groups of listeners were examined in the current study. The first group consisted of young adults, as most previous research on the Familiar Talker Advantage has examined perception in this population. In addition, a second group of school-age children similar in age to Levi (2015) completed the study, as our primary interest was possible implications for clinical evaluations. Listeners were divided into two talker groups (JEC, the first author, and SSL) and familiarized to one of the two talkers' voices. Given the previous research on the Familiar Talker Advantage, we expected listeners to display more improvement on a sentence recognition task following implicit learning of a talker's voice.

Method

Participants

All participants were native speakers of Standard American English with no reported history of speech-language or hearing impairments and passed a hearing screening at 25dB SPL at 500, 1000, 2000, and 4000 Hz using a portable Earscan3 Screening Audiometer. Thirty-three adults (16 in the JEC group, 12 women and four men, and 17 in the SSL group, 12 women and five men), 18–28 years of age (mean age = 20;10 years), participated in the study. An additional 12 participants were not included in the study because of history of speech-language impairment (n = 3), living outside the United States before the age of 1 (n = 1), being a nonnative speaker of American English (n = 2), not attending the second day of the study (n = 1), experimenter error (n = 2), and scoring 1 *SD* below the mean on the Recalling Sentences subtest (n = 3) of the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4; Semel, Wiig, & Secord, 2004).

Sixteen second graders between 6;10 and 8;4 years of age (mean age = 7;7 years) participated in the study (eight in the JEC group, four girls and four boys, and eight in the SSL group, three girls and five boys). Four additional children were not included in analyses for failing the hearing screening (n = 1), being a nonnative speaker of American English (n = 1), and for having articulation errors too severe to ensure consistent coding (n = 2). At the time of testing, all children had typical speech and language development as measured by a Core Language composite score greater than or equal to 85 on the CELF-4 (Semel et al., 2004). The composite scores on this test are normed to have a mean of 100 and a standard deviation of 15. The Core Language composite scores on the CELF-4 ranged from 104 to 132. The groups did not differ in age, $t(14) = .09$, $p = .923$, or in Core Language scores, $t(14) = 1.49$, $p = .156$. Speech production skills were assessed using the Preschool Language Scale–Fourth Edition articulation screening test (Zimmerman, Steiner, & Pond, 2002) and in connected speech using a story retell task.

Stimuli

Stimuli were recorded by two female native speakers of American English, one from New Jersey (JEC) and one from Maryland (SSL), in a sound-attenuated IAC booth. Recordings were made using a Shure 10A head-mounted microphone and a Fostex FR-2LE recorder. All sound files were normalized to have a uniform root-mean-square amplitude and down-sampled to 22050 Hz using Praat (Boersma & Weenink, 2016). Stimuli were mixed with signal-dependent noise (Benkı , 2003; Schroeder, 1968) using a MATLAB script (Felty, 2007). Signal-dependent noise masks each segment to the same degree, rather than adding a uniform level of noise across the entire sentence. Adults completed the sentence repetition task with a signal-to-noise ratio of −5 dB, and the children completed the task with a signal-to-noise ratio of −2.5 dB. These were selected to avoid ceiling and floor effects. A third female speaker (SVL, the third author) produced four nursery rhymes and eight sentences similar to the experimental stimuli that were used as practice. These were mixed with the same signal-dependent noise as the test stimuli. All speakers in the current study were women because (a) a high percentage (94%) of speech-language pathologists are women (American Speech-Language-Hearing Association, 1997) and (b) perceptual discrimination for female voices begins early in development (Kisilevsky et al., 2003).

Stimuli consisted of 60 high and 60 low predictability sentences created by Stelmachowicz, Hoover, Lewis, Kortekaas, and Pittman (2000). High predictability sentences were syntactically and semantically appropriate (e.g., Pour me more tea), whereas low predictability sentences were syntactically correct but semantically anomolous (e.g., Most birds knock tea). Both the high and low predictability sentences contained the same set of monosyllabic words. Each sentence contained four words that were at a vocabulary level familiar to children ages 4 and above. The practice stimuli did not contain any words that were used in the experimental stimuli.

Acoustic measures for the stimuli are presented in Table 1. Separate mixed-effects models were fit to each measure in the table, with fixed effects for talker (JEC, SSL), predictability (high, low), and their interaction, and random intercepts for sentence type. There were significant overall effects of talker (all ps < .001), but no other predictors were significant.

Procedure

Participants completed two sessions 1 week apart (±2 days), as voice information has been shown to be stored in memory for at least 1 week (Goldinger, 1996). Each session was conducted in a quiet testing room in the Department of Communicative Sciences and Disorders at New York University. All participants completed a baseline sentence recognition task, a series of implicit voice learning tasks designed to familiarize participants with one of the voices (JEC or SSL), and a post-learning sentence recognition task.

Baseline Sentence Recognition Task

On Day 1, participants first completed a self-paced (approximately 20 minute) sentence recognition test. Following the practice block, participants completed the sentence recognition task, which consisted of 30 low predictability sentences and 30 high predictability sentences randomly selected from the larger set, half spoken by each of the two talkers. The task was run using E-Prime 2.0 Professional (Schneider, Eschman, & Zuccoloto, 2007) and presented on a Panasonic Toughbook CF-52 laptop. Stimuli were presented binaurally over Sennheiser HD-280 circumaural headphones. Before each trial, the name of the speaker (Julie [for talker JEC] or Stephanie [for talker SSL]) was presented on the screen. Participants were asked to repeat each sentence, even if it did not make sense. A third experimenter (not JEC or SSL) administered the sentence recognition task.

Implicit Voice Learning Tasks

Both children and adults completed several tasks designed to provide naturalistic learning of a talker's voice. Listeners assigned to JEC's group had all tasks administered by JEC, whereas those in SSL's group had all tasks administered by SSL. There were two implicit learning sessions, one immediately following the baseline sentence recognition task on Day 1 and a second session approximately 1 week later at the beginning of Day 2. All tasks selected for the learning portion required a large amount of speaking by the familiar talker. In both sessions, the familiar talker wore a name tag to ensure that participants were aware of her name.

Adults. On Day 1, the familiar talker (JEC or SSL) verbally administered the participant questionnaire and three subtests of the CELF-4 (Recalling Sentences, Word Definitions, Understanding Spoken Paragraphs). On Day 2,

Table 1. Acoustics of experimental stimuli.

Talker	Predictability	Duration (s)	Rate (phonemes/s)	Mean f0	f0 variability
JEC	High	1.751	0.1315	219	58.2
	Low	1.811	0.1345	221	53.9
SSL	High	1.519	0.1136	178	39.2
	Low	1.581	0.1176	181	43.1

Note. JEC and SSL are the initials of the talker from each familiarization group.

adults completed the Semantic Relationships subtest of the CELF-4 and a hearing screening. Listeners received approximately 40 minutes of exposure to the talker's voice on Day 1 and 15 minutes on Day 2.

Children. On Day 1, children completed two subtests of the CELF-4 (Concepts & Following Directions, Word Structure) and the Preschool Language Scale–Fourth Edition articulation screening test. On Day 2, children completed the Recalling Sentences subtest of the CELF-4, a hearing screening, and a story retell of a wordless picture book. Children also completed the Formulated Sentences subtest of the CELF-4 to generate the Core Language composite score, but this task was administered following the post-learning sentence recognition task to reduce fatigue effects and because it involves little exposure to the familiar talker's voice. Prior to the post-learning sentence recognition task, children received approximately 40 minutes of exposure to the familiar talker's voice on Day 1 and 20 minutes on Day 2.

Post-learning Sentence Recognition

Following the implicit voice learning tasks on Day 2, the remaining 60 sentences (30 high predictability, 30 low predictability) were presented to the listeners with half produced by each of the talkers. The procedure was identical to the baseline sentence recognition task.

Coding and Analysis

Responses were coded in one of two ways. First, responses were coded for whole-sentence accuracy, similar to the approach used by Stelmachowicz et al. (2000). Responses were considered to be correct if all of the four target words appeared in the participant's response. Thus, if participants produced extraneous words, the sentence was still coded as correct (e.g., response "pick up this dirty room" for target "pick up this room"). Second, individual word accuracy was coded for each of the four words in the sentence to provide a more fine-grained level of analysis. The following coding conventions from Stelmachowicz et al. (2000) were also used: (a) a word was coded as correct even when in the wrong order, and (b) if an inflectional suffix (third person –s, plural –s, or regular past –ed) was added or deleted, the word was still coded as correct. In addition, an individual word was considered to be correct if the participant produced a word containing the target word (e.g., response "chocolate," which contains target "chalk").

Logistic mixed-effects models were fit to the data from the adults and the children separately. Each model included fixed-effects for four categorical predictors: time (baseline, post-learning), talker type (familiar, unfamiliar), talker identity (JEC, SSL), and predictability (high, low), plus all interactions (two-way, three-way, and four-way). Each predictor was sum-coded, with the first level of each variable (in alphabetical order) coded as 0.5 and the second level coded as −0.5. The models also included by-subject and by-sentence intercepts, as well as by-subject and by-sentence slopes for time, talker type, and their interaction. The crucial predictor is the interaction between time and talker type. We hypothesize that the effect of time on response accuracy should be moderated by whether a subject is hearing a familiar or an unfamiliar talker, such that there will be a larger difference between baseline and post-learning sentence accuracy for familiar talkers.

Models were fit in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2016). To facilitate interpretability and comparisons among the interacting factors, the lsmeans R package (Lenth, 2016) was used to predict marginal means from each model, averaging over the predictor levels not involved in each comparison. Thus, the crucial comparisons average over talker identity and predictability. Effect sizes were estimated by comparing marginal means, and p values were calculated from the effect size and standard error of the difference (or difference of differences) based on the Wald z statistic. In cases where a family of multiple tests is reported (such as testing the improvement for each talker individually), p values are adjusted using Holm's method. Adjusted p values are marked as p^* in the text.

In addition to the comparisons reported in the text, Appendixes C and D provide a table of parameter estimates for all models. Because of the coding scheme that was used, some parameter coefficients correspond to marginal contrasts reported below. Effect size estimates ($\hat{\beta}$) reported below are increases or decreases in the log-odds of an accurate response.

Results

Sentence-Level Analysis

For adult participants, there was an overall effect of time, such that participants were more accurate in the post-learning sentence recognition task than in the baseline task ($\hat{\beta} = 0.36$, $z = 2.94$, $p < .01$). The effect of time was

significant for sentences produced by both familiar talkers ($\hat{\beta}$ = 0.39, z = 2.69, p^* < .02) and unfamiliar talkers ($\hat{\beta}$ = 0.34, z = 2, p^* < .05). The crucial difference in improvement was not significant: There was no evidence that subjects' accuracy improved more for familiar talkers than unfamiliar talkers ($\hat{\beta}$ = 0.06, z = 0.29, p > .77). In addition, there was an overall effect of talker identity, where sentences produced by JEC were more accurately identified than sentences produced by SSL ($\hat{\beta}$ = 0.97, z = 9.35, p < .01), as well as an overall effect of predictability, where high predictability sentences were more accurately identified than low predictability sentences ($\hat{\beta}$ = 2.12, z = 7.28, p < .01). Figure 1 shows predicted marginal means for the four combinations of talker identity and predictability, and Table 2 lists raw averages for the key comparisons.

The results for children were qualitatively similar to the results for adults. There was an overall improvement in accuracy from baseline to post-learning ($\hat{\beta}$= 0.49, z = 2.81, p < .01), and this improvement was significant for sentences produced by unfamiliar talkers ($\hat{\beta}$ = 0.59, z = 2.56, p^* < .03), although not for familiar talkers after correcting for multiple comparisons ($\hat{\beta}$= 0.39, z = 1.74, p^* > .08). However, the crucial difference in improvement for sentences produced by familiar talkers compared to those produced by unfamiliar talkers was not significant ($\hat{\beta}$= −0.20, z = −0.7, p > .48). As with adults, there was also an overall effect of talker identity ($\hat{\beta}$= 1.00, z = 7.29, p < .01) and predictability ($\hat{\beta}$= 1.43, z = 5.48, p < .01). Figure 1 shows predicted marginal means, and Table 2 lists raw averages. Full model results for both adults and children are reported in Appendixes A and B.

Word-Level Analysis

As mentioned above, we also conducted a planned analysis of accuracy at the word level to allow for a more fine-grained analysis of accuracy. The model procedure was the same as described above for sentence-length accuracy, with the following exceptions. For these models, the dependent variable was whether each target word was correctly identified. Thus, because each sentence was exactly four words long, each trial contributed four responses to the model. The fixed effects and coding scheme were the same as the sentence accuracy models, but the random effects included intercepts and slopes (for time, talker type, and the interaction) for word nested within sentence, as well as by-subject intercepts and slopes for time, talker type, and the interaction. For the word analyses, the crucial interaction was in the predicted direction; however, similar to results of sentence-level analyses, it was not significant for either adults or children (all ps > .3). Full model results are reported in Appendixes C and D.

Additional Analyses

Gender

In an exploratory analysis, we modeled the effect of subject gender on whole-sentence accuracy and on individual word accuracy for adults and for children. Gender was coded as 0.5 (female) or −0.5 (male) and included in each model as an additional fixed effect, as well as all interactions with the other four predictors (time, talker type,

Figure 1. Probability of correct sentence identification (transformed from log-odds model predictions) at each testing time, for adults and children. Panels show the four subgroups for each combination of predictability (horizontal panels) and talker (vertical panels). JEC and SSL are the initials of the talker from each familiarization group. Error bars show confidence intervals for the predicted mean probability.

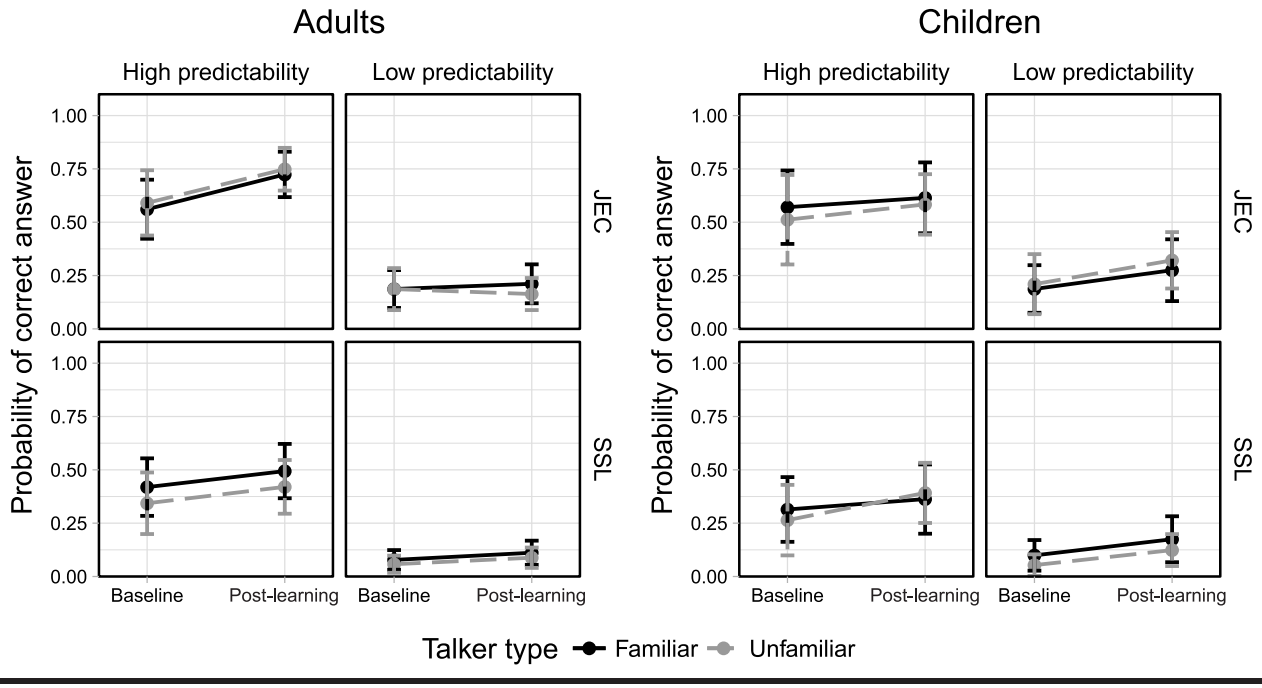


Table 2. Raw average percent correct for whole sentences, collapsed over high and low predictability.

Population	Training	Talker	Baseline % correct	Post % correct	Change
Adults	Trained on JEC	JEC	42	46	4
Adults	Trained on SSL	JEC	44	45	1
Adults	Trained on JEC	SSL	25	32	7
Adults	Trained on SSL	SSL	28	35	7
Children	Trained on JEC	JEC	39	47	8
Children	Trained on SSL	JEC	39	47	8
Children	Trained on JEC	SSL	24	30	6
Children	Trained on SSL	SSL	26	32	6

Note. JEC and SSL are the initials of the talker from each familiarization group.

talker identity, and predictability) up to the five-way interaction. For adults, there was an overall effect of gender for both sentence-level accuracy ($\hat{\beta}$= 0.42, $z = 2.34$, $p < .02$) and word-level accuracy ($\hat{\beta}$= 0.34, $z = 2.11$, $p < .04$), such that male subjects were more accurate than female subjects. However, gender did not moderate the crucial interaction (sentences: $p > .56$, words: $p > .57$). Although there were several significant three- and four-way interactions involving gender, the inclusion of gender did not qualitatively change any of the results for the adults reported above.

For children, there was no overall effect of gender (sentences: $\hat{\beta} = -0.47$, $z = -1.56$, $p > .11$, words: $\hat{\beta}$= −0.42, $z = -1.63$, $p > .10$). In the sentence-level analysis, there was a significant three-way interaction between gender, time, and talker identity ($p < .01$), but the results were not qualitatively different when the additional gender parameters were included in the analysis.

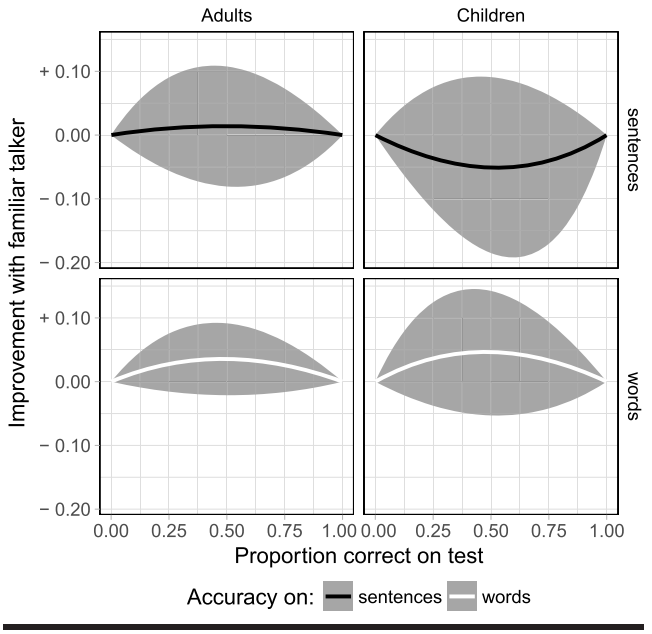
Effect Size Estimates

The hypothesized effect was not significant in any of the four tests (adults or children, for sentence accuracy or for word accuracy). However, the confidence intervals for the crucial difference in improvements were wide: For adults, the confidence intervals (all in log-odds; the hypothesized effect is in the positive direction) were [−0.33, 0.44] for sentence accuracy and [−0.09, 0.37] for word accuracy. For children, the confidence intervals were [−0.78, 0.37] for sentences and [−0.21, 0.59] for words.

An effect in this range may have clinical relevance, although future work is needed to provide narrower estimates and to evaluate whether the effect is likely to be nonzero overall. Figure 2 shows the expected benefit that a listener would enjoy on sentence or word accuracy tasks if evaluated by a familiar talker. Because the results were modeled in log-odds, the benefit is slightly different depending on the listener's expected score on the assessment task. Expected scores are shown on the *x* axis (from 0% to 100% correct on a test scored as correct/incorrect), and the *y* axis shows how much the score would increase if evaluated by a familiar talker, based on the estimates from this article. Note that the shaded confidence intervals include negative values, especially for the children; although these values are included in the model estimates, from a practical perspective it seems unlikely that a familiar talker would worsen performance.

We also note that there was high between-subject variability with respect to the crucial effect. Although the average slope for adults was −0.06 for sentence accuracy and −0.14 for word accuracy (the effect sizes reported above; note that because of the coding scheme, a negative slope is in the predicted direction), the slopes for individual subjects ranged from −0.41 to 0.37 for sentence accuracy and from −0.36 to 0.08 for word accuracy. For individual children, the slopes ranged from −0.72 to 0.67 for sentences and −1.14 to 0.36 for words, with some children showing a large effect in the predicted direction for word-level accuracy. The intersubject variability also suggests that future work is likely to require a large number of subjects or a task

Figure 2. Predicted increase in the proportion of correct responses if tested by a familiar talker (*y* axis), which varies based on the proportion of correct responses that the adult would have when tested by an unfamiliar talker (*x* axis). Shaded regions show 95% confidence intervals for the predictions.



that exaggerates the difference between familiar and unfamiliar talkers in order to evaluate the significance of the Familiar Talker Advantage in a clinical setting.

Discussion

The goal of this experiment was to test whether implicit voice learning, as occurs in typical interactions with other speakers, would lead to a Familiar Talker Advantage in children and adults. Both groups of listeners completed a baseline sentence recognition task with two unfamiliar talkers. They were then familiarized to one of these two talkers voices through live, in-person interactions. Following implicit voice learning, listeners completed a post-learning sentence recognition task. All listeners improved between baseline and post-learning in all conditions. Contrary to our initial prediction, neither child nor adult listeners demonstrated significantly more improvement on the sentence recognition task for the familiar talker than the unfamiliar talker. Examination of the intersubject variability in effect size estimates suggests that there may be a benefit to familiarity for some listeners, although an overall benefit was not found. These findings suggest that the implicit voice learning task used here may not have been sufficient for listeners to acquire talker-specific voice information to facilitate spoken language processing and that what listeners learned was how to process speech in this degraded (noise) condition as in Huyck, Smith, Hawkins, and Johnsrude (2017). More generally, the results of this study support that sentence recognition tests are robust whether administered by a familiar or unfamiliar talker.

The lack of a Familiar Talker Advantage could result from a variety of differences between the design used here and that used in previous studies. One possibility is that either short-term explicit voice training is necessary (Levi, 2015; Levi et al., 2011; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Yonan & Sommers, 2000) or that extended implicit exposure over years is necessary (Souza et al., 2013) to truly encode talker-specific voice information as seen in other studies of talker-voice familiarity. Alternatively, the lack of a Familiar Talker Advantage could result from a difference in how voice information was encoded during the learning phase and how it was retrieved during the post-learning sentence recognition task. Research on encoding specificity shows the importance of similarity between encoding and retrieval/testing conditions (Tulving & Thomson, 1973); thus, it is possible that differences in how voice information was learned (i.e., in-person) versus retrieved during the experimental task (i.e., over headphones) may have attenuated any benefit of talker familiarity.

Voice Familiarity in the Clinical Environment

In the clinical setting, it is likely that the same clinician administers therapy and performs assessments—or reevaluations of client progress—within similar environmental conditions. Thus, the client may be encoding voice attributes of his or her clinician in the same setting where he or she would be evaluated. Our findings indicate that familiar voice information may not influence performance, at least with short exposure, but the effect size estimate analysis suggests that there may be a small effect for some people. It is important to emphasize that there are a multitude of factors that could affect a client's performance on assessments. Talker familiarity is one consideration among many possible influences.

Despite the null results found here, this line of research raises the question of "who" should be administering assessments when a client has been enrolled in treatment for a time with the same clinician. Although federal and state legislation offer guidelines for "when" assessments take place, these do not address "who" is the most appropriate clinician to administer assessments. In the clinical setting, assessments play a critical role in determining whether a child continues to receive needed service or if he or she should be discharged from treatment. If factors related to the person administering assessments of client progress could impact performance, increased specifications related to the evaluator may be warranted.

Moreover, it is important to consider the overarching goal of the assessment. If the clinician wants to see a client's optimal performance under the most supportive conditions, it would be advisable for a familiar clinician—or a familiar voice—to administer assessment tasks. However, if the goal of the assessment is to assess how well a client is generalizing skills learned in therapy to a novel context with less support, then it would be more informative to have an unfamiliar person (or "voice") administer the assessment. Again, we acknowledge that voice familiarity is just one factor that may contribute to client performance.

Furthermore, there are many logistical factors that impact availability of clinicians. Restrictions related to scheduling and balancing large caseloads could certainly impact the ability to bring in a new clinician to administer testing. Certain settings also may not offer the flexibility for an outside clinician to perform reevaluations. The intent of the current study is not to say that voice (un)familiarity delegitimizes client performance on spoken language processing tasks administered by a familiar clinician. Rather, we hope that these findings will inspire future studies examining external factors that could influence performance and also discussions on the overarching goal of assessment.

Acknowledgments

This work was supported in part by a grant from the NIH-NIDCD: 1R03DC009851-01 (Levi). We would like to thank Gabrielle Alfano, Stephanie Lee, Maddy Lippman, Rebecca Piper, and Ashley Quinto for help with data collection and the children and families for their participation. Portions of this work were presented at the annual convention of the American Speech-Language-Hearing Association (2016) and at the Symposium on Research in Child Language Disorders (2017).

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 115*(6), 3171–3183.

American Speech-Language-Hearing Association. (1997). *Omnibus survey results*. Rockville, MD: Author.

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

Benkı , J. R. (2003). Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition. *The Journal of the Acoustical Society of America, 113*(3), 1689–1705.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.23) [Computer program]. Retrieved from: http://www.praat.org

Ebert, K. D. (2017). Measuring clinician–client relationships in speech-language treatment for school-age children. *American Journal of Speech-Language Pathology, 26*(1), 146–152.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics, 67*(2), 224–238.

Felty, R. A. (2007). *Context effects in spoken word recognition of English and German by native and non-native listeners*. East Lansing, MI: Michigan State University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166–1183.

Hoffman, L. (2014). Prologue: Improving clinical practice from the inside out. *Language, Speech, and Hearing Services in Schools, 45*(2), 89–91.

Huyck, J. J., Smith, R. H., Hawkins, S., & Johnsrude, I. S. (2017). Generalization of perceptual learning of degraded speech across talkers. *Journal of Speech, Language, and Hearing Research, 60*(11), 3334–3341.

Individuals with Disabilities Education Act, 20 U.S.C § 1400 (2004).

Ireland, M., & Conrad, B. J. (2016). Evaluation and eligibility for speech-language services in schools. *Perspectives of the ASHA Special Interest Groups, 1*(16), 78–90.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., . . . Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14*(3), 220–224.

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

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language, 42*(4), 843–872.

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

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*(6), 1113–1126.

Nelson, P., Kohnert, K., Sabur, S., & Shaw, D. (2005). Classroom noise and children learning through a second language: Double jeopardy? *Language, Speech, and Hearing Services in Schools, 36*(3), 219–229.

New York City Department of Education. (2009). *Standard operating procedures manual: The referral, evaluation, and placement of school-age students with disabilities*. New York, NY: Author.

New York City Early Intervention System. (2014). *Policy and Procedure Manual*. New York, NY: Author.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1), 42–46.

Peña, E. D., & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools, 28*(4), 323–332.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Retrieved from http://www.r-project.org/

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207–1218.

Schneider, W., Eschman, A., & Zuccoloto, A. (2007). *E-Prime 2.0 Professional*. Pittsburgh, PA: Psychological Software Tools, Inc.

Schroeder, M. (1968). Reference signal for signal quality studies. *The Journal of the Acoustical Society of America, 44*(6), 1735–1736.

Semel, E. M., Wiig, E. H., & Secord, W. (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. Bloomington, MN: Pearson.

Smith, G. W., & Riccomini, P. J. (2013). The effect of a noise reducing test accommodation on elementary students with learning disabilities. *Learning Disabilities Research & Practice, 28*(2), 89–95.

Souza, P., Gehani, N., Wright, R., & McCloy, D. (2013). The advantage of knowing the talker. *Journal of the American Academy of Audiology, 24*(8), 689–700.

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W., & Pittman, A. L. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43*(4), 902–914.

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail a. *The Journal of the Acoustical Society of America, 128*(4), 2090–2099.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences a. *The Journal of the Acoustical Society of America, 125*(6), 3974–3982.

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*(7–8), 979–1001.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373.

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.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale–Fourth Edition (PLS-4)*. San Antonio, TX: Psychological Corporation.

Appendix A

Fixed-Effect Parameter Estimates (in Log-Odds) for Logistic Mixed-Effects Model for Each Data Set, for Whole-Sentence Accuracy Analyses

Parameter	Adults	Children
Intercept	–0.89 (0.16)***	–0.92 (0.2)***
Time	–0.36 (0.12)**	–0.49 (0.17)**
Talker type	0.16 (0.11)	0.15 (0.13)
Talker	0.97 (0.1)***	1 (0.14)***
Predictability	2.12 (0.29)***	1.43 (0.26)***
Time × Talker type	–0.06 (0.19)	0.2 (0.29)
Time × Talker	0.01 (0.21)	0.21 (0.3)
Talker type × Talker	–0.28 (0.35)	–0.3 (0.64)
Time × Predictability	–0.31 (0.24)	0.34 (0.32)
Talker type × Predictability	–0.13 (0.21)	–0.06 (0.25)
Talker × Predictability	0.05 (0.2)	–0.04 (0.27)
Time × Talker type × Talker	–0.2 (0.38)	–0.22 (0.55)
Time × Talker type × Predictability	0.14 (0.38)	0.07 (0.52)
Time × Talker × Predictability	–0.84 (0.4)*	–0.08 (0.55)
Talker type × Talker × Predictability	–0.3 (0.35)	0.84 (0.48)†
Time × Talker type × Talker × Predictability	0.36 (0.71)	–0.07 (0.96)

Note. Standard errors are given in parentheses.

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$ based on the Wald z statistic.

Appendix B

Estimates for Standard Deviation of Random-Effects Parameters for Logistic Mixed-Effects Model for Each Data Set, for Whole-Sentence Accuracy Analyses

Group	Parameter	σ (Adults)	σ (Children)
By subject	Intercepts	0.44	0.59
	Slopes for time	0.19	0.23
	Slopes for talker type	0.17	0.04
	Slopes for Time × Talker type	0.26	0.42
By sentence	Intercepts	1.48	1.23
	Slopes for time	0.64	0.90
	Slopes for talker type	0.48	0.15
	Slopes for Time × Talker type	0.28	0.85

Appendix C

Fixed-Effect Parameter Estimates (in Log-Odds) for Logistic Mixed-Effects Model for Each Data Set, for Individual Word Accuracy Analyses

Parameter	Adults	Children
Intercept	1.3 (0.13)***	1.07 (0.16)***
Time	–0.35 (0.1)***	–0.45 (0.13)***
Talker type	0.11 (0.08)	0.02 (0.09)
Talker	0.7 (0.07)***	0.74 (0.08)***
Predictability	1.16 (0.21)***	0.81 (0.18)***
Time × Talker type	–0.14 (0.12)	–0.19 (0.2)
Time × Talker	–0.36 (0.11)***	0.05 (0.18)
Talker type × Talker	–0.01 (0.3)	0.22 (0.55)
Time × Predictability	–0.2 (0.17)	0.12 (0.19)
Talker type × Predictability	0.2 (0.12)*	0.00 (0.15)
Talker × Predictability	0.29 (0.1)**	–0.03 (0.14)
Time × Talker type × Talker	0.07 (0.26)	0.14 (0.43)
Time × Talker type × Predictability	–0.08 (0.21)	–0.19 (0.32)
Time × Talker × Predictability	–0.48 (0.21)*	0.46 (0.28)
Talker type × Talker × Predictability	0.11 (0.18)	0.6 (0.26)*
Time × Talker type × Talker × Predictability	–0.25 (0.37)	–0.17 (0.51)

Note. Standard errors are given in parentheses.

***$p < .001$, **$p < .01$, *$p < .05$ based on the Wald z statistic.

Appendix D

Estimates for Standard Deviation of Random-Effects Parameters for Logistic Mixed-Effects Model for Each Data Set, for Individual Word Accuracy Analyses

Group	Parameter	σ (Adults)	σ (Children)
By subject	Intercepts	0.42	0.54
	Slopes for time	0.25	0.34
	Slopes for talker type	0.27	0.13
	Slopes for Time × Talker type	0.13	0.45
By sentence	Intercepts	0.88	0.74
	Slopes for time	0.62	0.70
	Slopes for talker type	0.39	0.43
	Slopes for Time × Talker type	0.48	1.07
By word in sentence	Intercepts	1.31	1.13
	Slopes for time	0.56	0.43
	Slopes for talker type	0.08	0.03
	Slopes for Time × Talker type	0.01	0.22