

Research Article

Individual Differences in Distributional Learning for Speech: What's Ideal for Ideal Observers?

Rachel M. Theodore,^{a,b} Nicholas R. Monto,^{a,b} and Stephen Graham^{a,b}

Purpose: Speech perception is facilitated by listeners' ability to dynamically modify the mapping to speech sounds given systematic variation in speech input. For example, the degree to which listeners show categorical perception of speech input changes as a function of distributional variability in the input, with perception becoming less categorical as the input, becomes more variable. Here, we test the hypothesis that higher level receptive language ability is linked to the ability to adapt to low-level distributional cues in speech input.

Method: Listeners ($n = 58$) completed a distributional learning task consisting of 2 blocks of phonetic categorization for words beginning with /g/ and /k/. In 1 block, the distributions of voice onset time values specifying /g/ and /k/ had narrow variances (i.e., minimal variability). In the other block, the distributions of voice onset times specifying /g/ and /k/ had wider variances (i.e., increased variability).

In addition, all listeners completed an assessment battery for receptive language, nonverbal intelligence, and reading fluency.

Results: As predicted by an ideal observer computational framework, the participants in aggregate showed identification responses that were more categorical for consistent compared to inconsistent input, indicative of distributional learning. However, the magnitude of learning across participants showed wide individual variability, which was predicted by receptive language ability but not by nonverbal intelligence or by reading fluency.

Conclusion: The results suggest that individual differences in distributional learning for speech are linked, at least in part, to receptive language ability, reflecting a decreased ability among those with weaker receptive language to capitalize on consistent input distributions.

There is no one-to-one relationship between speech acoustics and a given speech sound; instead, different acoustic forms may be produced for the same speech sound, and the same acoustic form may be produced for different speech sounds. Variation arises in the speech signal due to a host of factors, including dialect (Byrd, 1992), speaking rate (Miller & Baer, 1983; Theodore, Miller, & DeSteno, 2009), speaking register (Picheny, Durlach, & Braida, 1986), and even individual differences in pronunciation across talkers (Hillenbrand, Getty, Clark, & Wheeler, 1995; Newman, Clouse, & Burnham, 2001; Theodore et al., 2009). Given this variability, listeners must solve the lack of invariance problem in order to map the acoustic signal to

representations for individual speech sounds. As a consequence, speech perception can be viewed as a process in which listeners make inferences regarding talkers' intended speech sounds from a signal that is implicitly uncertain (Kleinschmidt & Jaeger, 2015; Toscano & McMurray, 2010).

Despite the lack of invariance in the acoustic speech signal, some variability in speech acoustics is highly structured. Consider just one acoustic-phonetic property of speech, voice onset time (VOT). VOT is a temporal property of stop consonants that reflects the time between the release of the occlusion necessary for stop consonant production and the subsequent onset of vocal fold vibration (Lisker & Abramson, 1964). On any given day, listeners will hear a wide range of VOTs produced for stops consonants. However, the input is structured such that the VOTs produced for voiced stops will be shorter than those produced for voiceless stops (e.g., Lisker & Abramson, 1964), VOTs produced for labial stops will be shorter than those produced for velar stops (e.g., Cho & Ladefoged, 1999), and VOTs produced at a fast speaking rate will be shorter than those produced for a slow speaking rate (e.g., Volaitis & Miller, 1992). Furthermore, individual talkers also show stable differences in

^aDepartment of Speech, Language, and Hearing Sciences, University of Connecticut, Storrs

^bConnecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs

Correspondence to Rachel M. Theodore: rachel.theodore@uconn.edu
Editor-in-Chief: Bharath Chandrasekaran

Editor: Chao-Yang Lee

Received April 7, 2019

Revision received June 10, 2019

Accepted July 2, 2019

https://doi.org/10.1044/2019_JSLHR-S-19-0152

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

their characteristic VOT productions such that some talkers have shorter VOTs than other talkers, even when controlling for other contextual influences (Chodroff & Wilson, 2017; Theodore et al., 2009).

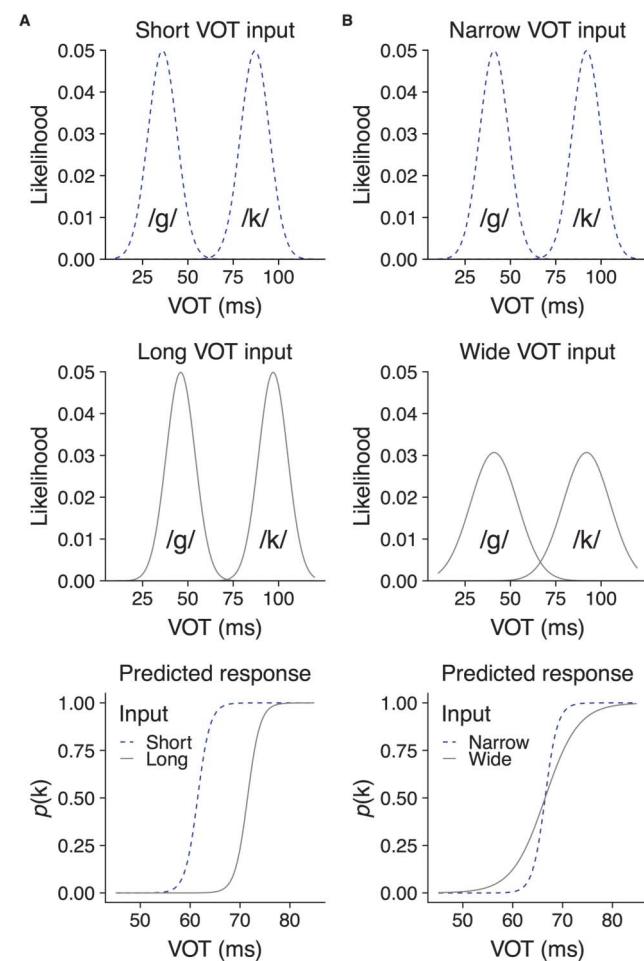
There is now a large evidence base demonstrating that listeners use structured phonetic variation to facilitate the mapping to speech sounds. Indeed, the ability to track structured phonetic variation in the speech input supports acquisition of the linguistic sound structure during development (Maye, Werker, & Gerken, 2002). For example, sensitivity to distributional input could allow the infant in a Spanish-speaking environment to learn that voiced stops are cued by negative VOTs (i.e., “prevoicing”) and voiceless stops are cued by VOTs near 0 ms and also allow the infant in an English-speaking environment to learn that VOTs near 0 ms are associated with voiced stops and long-lag VOTs are used to cue voiceless stops. Sensitivity to distributional variation in the input does not cease after the infant has acquired the phonetic inventory of a language (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Theodore & Monto, 2019). Instead, functional plasticity is observed across the life span such that listeners dynamically modify the mapping to speech sounds in line with statistical distributions of acoustic-phonetic cues in the input (e.g., Colby, Clayards, & Baum, 2018; Norris, McQueen, & Cutler, 2003; Theodore & Monto, 2019).

As an illustration, Figure 1A shows two sets of VOT input distributions that cue the voicing contrast for /g/ and /k/. The two sets of distributions differ in terms of the modal VOTs produced for /g/ and /k/, which are relatively shorter for one set (i.e., short VOT input) compared to the other set (i.e., long VOT input). The voicing contrast is clearly cued in both sets of distributions given the minimal overlap between VOTs specifying the /g/ and /k/ categories. However, the specific VOT that optimally marks the voicing contrast differs between the two sets of distributions. If listeners were to apply the same perceptual boundary to both sets of input distributions, then this would result in less accurate recovery of the intended speech sounds. Instead, optimal phonetic identification for these sets of distributions would entail an adjustment to the perceptual boundary in line with the distributional input. The “ideal” response can be predicted within an ideal observer computational framework according to Bayes’ theorem shown in Equation 1, simplified to reflect the assumption that the prior probabilities are equal. As shown in the bottom panel of Figure 1A, the predicted categorization response functions for the two sets of input distributions according to this equation differ in terms of the predicted category boundary that distinguishes /g/ and /k/; it is located at a shorter VOT for the short VOT compared to the long VOT input distributions.

$$p(k|VOT) = \frac{p(VOT|k)}{p(VOT|k) + p(VOT|g)} \quad (1)$$

The acoustic-phonetic input can also vary in terms of the consistency in which a cue is used to mark a phonetic

Figure 1. Panel A shows input distributions that differ in terms of modal voice onset time (VOT) values for the /g/ and /k/ categories, which are relatively shorter (top panel) or longer (middle panel), and the predicted categorization functions for each set of input distributions (bottom panel) according to Equation 1. Panel B shows input distributions that differ in terms of the variance of the VOT values for the /g/ and /k/ categories, which are relatively narrower (top panel) or wider (middle panel), and the predicted categorization functions for each set of input distributions (bottom panel) according to Equation 1. The input distributions presented in the current study are those shown in Panel B.



contrast. This situation is illustrated in Figure 1B. The two sets of input distributions are identical with respect to the modal VOTs produced for /g/ and /k/ but differ in terms of the variance of the /g/ and /k/ distributions, such that the distributions show either minimal variability (narrow VOT input) or relatively more variability (wide VOT input) around the modal VOTs. Functionally, this type of input variability could schematize a typical speaker (narrow VOT input) versus a speaker with a motor speech disorder (wide VOT input), or variability as a function of speaking style, such as when a speaker uses a clear speech register (narrow VOT input) and then changes to a more casual speech register (wide VOT input). The predicted categorization response

functions for the narrow and wide input distributions are shown in the bottom panel of Figure 1B. In contrast to those derived for the input distributions shown in Figure 1A, the predicted response functions differ with respect to the slope of the identification function. The ideal observer framework used here predicts that responses will be more categorical for the consistent compared to the inconsistent input, and thus, the predicted response function shows a steeper identification slope for the narrow versus wide input distributions.¹

Previous research has shown that listeners' behavioral responses in distributional learning tasks follow the predictions of ideal observer models (e.g., Clayards et al., 2008; Kleinschmidt & Jaeger, 2016; Nixon & Best, 2018; Nixon, van Rij, Mok, Baayen, & Chen, 2016). For example, Clayards et al. (2008) presented one group of listeners with VOTs that formed narrow distributions specifying /b/ and /p/ and a different group of listeners with VOTs that formed wider distributions specifying the same speech sounds. The results showed that the slope of the identification function was steeper for those who heard the narrow compared to the wide input. Sensitivity to variability of the distributional input has also been observed in a within-subject design, demonstrating dynamic adaptation to changes in the distributional input (Theodore & Monto, 2019). When listeners are first presented with narrow input distributions and then presented with wide input distributions, the slope of the identification function moves from steeper to shallower even in the course of a single experimental session (Theodore & Monto, 2019). The observed dynamic adaptation followed the predicted response patterns generated by computational simulations with the Bayesian belief-updating model of speech adaptation (Kleinschmidt & Jaeger, 2015), and it suggests that online identification reflects a cumulative integration of statistical experience with the talker's input distributions (Theodore & Monto, 2019).

As with most behavioral measures of human performance, the patterns that are observed at the group level for distributional learning for speech often exhibit wide individual variability among participants. In contrast to traditional approaches where individual variability in the sample is

considered noise with respect to characterizing group-level patterns, there is a growing body of literature that specifically seeks to identify and explain factors that drive individual variability in learning. *Distributional learning* or *statistical learning* is a broad term used to describe a change in behavior as a function of exposure to statistical regularities in the input. For language processing, this term has been used to describe the mechanisms by which listeners modify the mapping to speech sounds (Clayards et al., 2008; Theodore & Monto, 2019), learn to extract novel words given short-term adjacencies between syllables (e.g., Saffran, Johnson, Aslin, & Newport, 1999), and learn to extract higher levels of the grammar given long-term adjacencies among words (e.g., Hall, Owen Van Horne, McGregor, & Farmer, 2017). Outside language processing, statistical learning has referred to improved performance given statistical regularities in motor tasks (Lum, Conti-Ramsden, Morgan, & Ullman, 2014) and increased memory span for visual patterns that contain a redundant statistical structure (Conway, Bauernschmidt, Huang, & Pisoni, 2010). As outlined by Siegelman, Bogaerts, Christiansen, and Frost (2017), there are challenges to the view that statistical learning is a unified theoretical construct and that all statistical learning tasks are interchangeable. However, and of interest to the current work, there are findings showing stable relationships between language processing and statistical learning. For example, individual differences in statistical learning of adjacent and nonadjacent dependencies predict online comprehension of sentences (Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010). Differences in statistical learning ability have been examined between individuals with developmental language disorder (DLD) and peers with typical language abilities. As reviewed by Hall et al. (2017), the most robust evidence of a link between statistical learning and language processing concerns the finding that individuals with DLD show reduced learning in serial reaction time tasks. In these tasks, the learning effect manifests as a facilitated motor response for button presses that are predicted by a sequential statistical pattern (e.g., Lum et al., 2014). Hall et al. examined whether adults with DLD would show deficits in using statistical regularities for a different task, which was to learn grammatical categories in an artificial language. The specific statistical manipulation assessed was the ability to form grammatical categories from distributions of words presented in an artificial language; thus, this study measured participants' ability to use distributional information to generate grammatical categories instead of tracking sequential statistical sequences. Strikingly, Hall et al. found no evidence indicating reduced learning in those with DLD compared to control participants, suggesting that the ability to use distributional cues to learn a grammar is intact in individuals with DLD (Hall et al., 2017).

Relatively less is known about factors that influence individual differences in distributional learning for the earliest stages of language comprehension, including the stage in which listeners map speech acoustics to consonants and vowels. Colby et al. (2018) recently examined whether individual differences in receptive vocabulary, working memory,

¹In the current work, qualitative predictions for distributional learning are informed by the "straight Bayes rule" model from Clayards et al. (2008), which predicts that the slope of the identification function will be steeper for the narrow input distributions compared to the wide input distributions. Other ideal observer models that take into account uncertainty about the current distributions exist, including the Bayesian belief-updating model of Kleinschmidt and Jaeger (2015). The specific quantitative (and qualitative) predictions generated by ideal observer models may vary, depending on which changes they are open to and how they deal with the changes in distributional input. In the study of Theodore and Monto (2019), simulations were performed with the Bayesian belief-updating model of Kleinschmidt and Jaeger, setting the model to cumulatively update prior beliefs in response to the narrow input followed by the wide input. These simulations lead to the same qualitative predictions generated here, namely, that the slope of the identification function will move from steeper to shallower across the exposure period.

and attention-switching control predicted perceptual learning among two age cohorts, younger adults and older adults. All participants completed two perceptual learning tasks. In one task, lexical information was available as a learning signal for potentially ambiguous acoustic–phonetic input. In the other task, lexical information was not available; instead, the putative learning signal was differences in the input distributions of formant patterns specifying the /ɛ/–/ɪ/ contrast. The input distributions were formed so that formant patterns were typical of that expected for one category (e.g., /ɛ/), but for the other category (e.g., /ɪ/), the formant patterns reflected those ambiguous between /ɛ/ and /ɪ/. Across both age cohorts, perceptual learning was predicted by individual differences in receptive vocabulary, but not by individual differences in working memory or attention-switching control. Colby et al. suggest that this relationship may be the consequence of a facilitative effect of lexical knowledge on the ability to adapt to ambiguous input, regardless of whether the learning task specifically recruits lexical knowledge, which is consistent with other findings showing that speech recognition in noise is facilitated in those with larger receptive vocabularies (Baese-Berk, Bent, Borrie, & McKee, 2015).

The results of Colby et al. (2018) provide a key finding for understanding individual differences in perceptual learning, suggesting that there may be a specific relationship between low-level adaptation to distributional speech cues and receptive language ability, given that neither working memory nor attention-switching control reliably influenced the magnitude of perceptual learning. Here, we provide a further test of this hypothesis. All listeners ($n = 58$) completed two blocks of phonetic categorization in which they were presented with VOTs specifying word-initial /g/ and /k/. In the first block, VOTs formed two distributions, each with a narrow variance, and thus reflect a speaker who is extremely consistent in his or her use of VOT as a cue to the stop voicing contrast. In the second block, the VOTs formed two distributions with a wider variance, and thus, the speaker became less consistent in how VOT cued the voicing contrast. This represents a different distributional manipulation than was examined previously. In Colby et al., learning the input distributions required modifying the perceptual boundary between /ɛ/ and /ɪ/. In the predictions for the current work, generated by Equation 1, learning was not specific to the VOT voicing boundary. Namely, the ideal observer framework used here predicts that distributional learning will manifest as a change in the slope of the identification function relating VOT to voiceless responses. Specifically, the slope of the identification function in the narrow block will be steeper than the slope of the identification function in the wide block, indicating that listeners capitalized on the consistent input initially and then modified the mapping when the input changed to be less consistent.

In addition to the distributional learning task, all listeners completed an assessment battery to measure receptive language, nonverbal intelligence, and reading fluency. If the ability to dynamically adjust the mapping to speech sounds

in line with structured phonetic variation reflects individual differences in receptive language ability, then we predict that the degree to which the identification slope changes across the blocks will be graded such that those with the highest receptive language scores show the largest change compared to those with the lower receptive language scores. Moreover, if individual differences in distributional learning for speech reflect a specific relationship to receptive language ability, then nonverbal intelligence and reading fluency will not predict individual differences in learning.

Method

Participants

The participants were 58 adults (15 men, 43 women) between 18 and 30 years of age ($M = 20.9$ years, $SD = 2.6$ years) who were recruited from the University of Connecticut community. To recruit individuals with a wide range of language abilities, separate recruitment materials targeted individuals with no history of language disorder and individuals specifically with a history of language disorder. All participants were monolingual speakers of American English and passed a pure-tone hearing screen administered at 25 dB for octave frequencies between 500 and 4000 Hz on the day of testing.

All participants completed a distributional learning task (described below) in addition to assessments of receptive language, nonverbal intelligence, and reading fluency. Thirty of the participants completed the distributional learning task as part of the Narrow–Wide order group reported in the study of Theodore and Monto (2019); the other 28 participants did not participate in that study. For all participants, receptive language was measured using the receptive language battery developed by Fidler, Plante, and Vance (2011). This battery consists of a 15-word spelling test and a modified version of the Token Test (Morice & McNicol, 1985). The raw scores on these two tasks are used to derive a weighted composite measure of receptive language according to Equation 2.

$$\text{Composite} = 6.5727 + (-0.2184 * \text{Spelling Score}) + (-0.1298 * \text{Token Test Score}) \quad (2)$$

The weighted composite measure is a continuous score that varies between –2.4145 (ceiling performance on the spelling and modified token tests) and 6.5727 (floor performance on the spelling and modified token tests). Note that lower scores on the continuous composite measure are associated with stronger receptive language ability and higher scores on the continuous composite measure are associated with weaker receptive language ability. A discriminant analysis of the continuous composite measure (i.e., positive composite scores indicate DLD; negative composite scores indicate typical performance) shows 80% sensitivity and 87% specificity for the identification of childhood DLD. We selected this measure given its growing

use in the research domain for identifying DLD in adulthood (e.g., Earle, Landi, & Myers, 2018; Hall et al., 2017). Nonverbal intelligence was assessed using the standard score obtained from administration of the Test of Nonverbal Intelligence–Fourth Edition (TONI; Brown, Sherbenou, & Johnsen, 2010). Reading fluency was assessed using the Test of Word Reading Efficiency–Second Edition (TOWRE; Torgesen, Wagner, & Rashotte, 2012) in terms of the TOWRE Index Score, a standard score derived from performance on the Sight Word Efficiency and Phonological Decoding subtests of the TOWRE, which assess reading fluency for real words and nonwords, respectively.² For both the TONI and the TOWRE, standard scores reflect a population mean of 100 ($SD = 15$).

The receptive language, nonverbal intelligence, and reading fluency scores of the current sample are shown in Figure 2, along with the relationships among the three measures. The color mapping in Figure 2 reflects receptive language scores in ascending order. Recall that, for the receptive language composite, positive scores are associated with weaker receptive language abilities; thus, the color map ranges from the strongest receptive language score (red) to the weakest receptive language score (blue). The sample shows wide individual variability for all three measures. Eight individuals met criterion for DLD (i.e., a positive receptive language composite score). All individuals scored at or above 1 SD of the population mean (≥ 85) on the TONI, and all but two individuals scored at or above 1 SD of the population mean (≥ 85) on the TOWRE.

Moderate relationships were observed between receptive language and nonverbal intelligence ($r = -.32$, $p = .014$) and between receptive language and reading fluency ($r = -.35$, $p = .006$). Though the direction of the correlation is negative, these relationships represent positive associations between receptive language and both nonverbal intelligence and reading fluency given that lower scores on the composite measure are associated with stronger receptive language. No relationship was observed between nonverbal intelligence and reading fluency ($r = -.03$, $p = .824$).

Stimuli

The stimuli consisted of auditory tokens of *goal*, *coal*, *gain*, and *cane* that varied in word-initial VOT. The stimuli (also used in the study of Theodore & Monto, 2019) were drawn from two VOT continua, a *goal-coal* continuum and a *gain-cane* continuum. The continua were created using a

²Six of the 58 participants were beyond the oldest age (24;11 [years; months]) provided for the standard score conversion of the TOWRE performance. As a consequence, the raw score to standard score conversion for these participants was made using the oldest age provided for the conversion, which is sensible given that the oldest age bracket represents a maturational end state for reading fluency. However, all analyses conducted with the TOWRE standard score were also conducted using the TOWRE raw score, with parallel results observed in all cases. These analyses can be viewed at the OSF repository associated with this article: <https://osf.io/tsnx4/>.

naturally produced token as the voiced-initial end point following the procedure outlined in the study of Allen and Miller (2004), to which the reader is referred for comprehensive details on stimulus creation. In brief, productions of *gain* and *goal* with equivalent word durations (568 and 569 ms, respectively) were obtained from a native female speaker of American English to serve as the voiced end points. For each voiced end point, the linear predictive coding-based speech synthesizer in the ASL software package (Kay Elemetrics) was used to successively increase word-initial VOT in 4- to 5-ms increments by systematically changing parameters of the linear predictive coding analysis and synthesizing new tokens using the modified parameters. This procedure resulted in VOTs that perceptually ranged from /g/ to /k/ across each continuum. Representative spectrograms can be viewed in Figure 3.

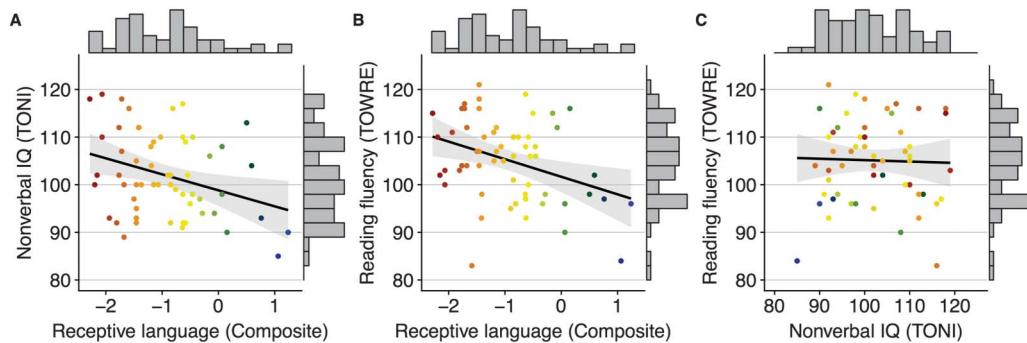
Twelve tokens were selected from each continuum for further use consisting of VOTs that ranged from 11 to 119 ms in approximately 10-ms increments. The selected tokens were arranged into two sets, one for the narrow block and one for the wide block, to form input distributions that were more consistent to less consistent, respectively. As shown in Table 1, the two sets differed with respect to the frequency in which each VOT was presented. The mean VOT for the /g/ and /k/ distributions (40 and 92 ms, respectively) was identical between the narrow and wide stimulus sets. The critical difference between the two stimulus sets was the standard deviation of the /g/ and /k/ distributions, which was 8 ms in the narrow set and 13 ms in the wide set. Figure 1B shows the probability density functions for the /g/ and /k/ distributions in each stimulus set.

Procedure

All testing took place in a sound-attenuated booth. Participants were seated at a table that contained a computer monitor and a response box. Auditory stimuli were presented via headphones (Sony MDR-7506) at a comfortable listening level that was held constant across participants. Stimulus presentation and response collection were controlled using SuperLab 4.5 running on a Mac OS X system.

Participants completed two blocks of phonetic categorization (472 trials in total), one for the narrow stimulus set and one for the wide stimulus set. All participants completed the narrow block followed by the wide block. In each block, the 236 tokens that formed the /g/ and /k/ distributions were presented in randomized order. On each trial, participants were asked to identify each token as either *goal*, *coal*, *gain*, or *cane* by pressing an appropriately labeled button on the response box. Participants were instructed to make their decision as quickly as possible without sacrificing accuracy and to guess if they were unsure. The interstimulus interval was 2000 ms, timed from the participant's response. Prior to the start of the first block, participants completed 12 practice trials consisting of three repetitions of *gain*, *cane*, *goal*, and *coal* with VOTs matching the modes of the

Figure 2. Scatter plots illustrating the relationship between receptive language and nonverbal IQ (A), receptive language and reading fluency (B), and nonverbal IQ and reading fluency (C). In each panel, the shaded region depicts the 95% confidence level interval for a linear regression and marginal histograms show the distribution of scores for each variable. The color map is constant across panels to reflect receptive language composite score. TONI = Test of Nonverbal Intelligence; TOWRE = Test of Word Reading Efficiency.



/g/ and /k/ distributions. Participants were given a brief break between the two blocks, and the entire procedure lasted approximately 30 min.

Results

Two sets of analyses were performed. The primary analyses were conducted to test the hypothesis that distributional learning for speech is linked to receptive language ability. The second set of analyses was performed for the 28 participants who did not also participate in the Narrow-Wide condition of Theodore and Monto (2019) in order to assess replication of the previous finding. The raw data and analysis scripts can be retrieved at <https://osf.io/tsnx4/>; analysis scripts operate on the raw data to reproduce

all results presented here, in addition to generating all figures.

Primary Analyses

Responses on the distributional learning task were coded as either voiced (i.e., responses of *gain* and *goal*) or voiceless (i.e., responses of *cane* and *coal*). Trials for which no response was provided were excluded from further analysis (185 of 27,376 trials, representing < 1% of the total trials). To visualize performance, mean proportion of voiceless responses was first calculated for each participant for each VOT in each block and was then averaged across the 58 participants. As shown in Figure 4A, the participants in the aggregate show the expected categorical relationship

Figure 3. Spectrograms of the tokens corresponding to the mean voice onset times of the /g/ (*gain*, *goal*) and /k/ (*cane*, *coal*) input distributions.

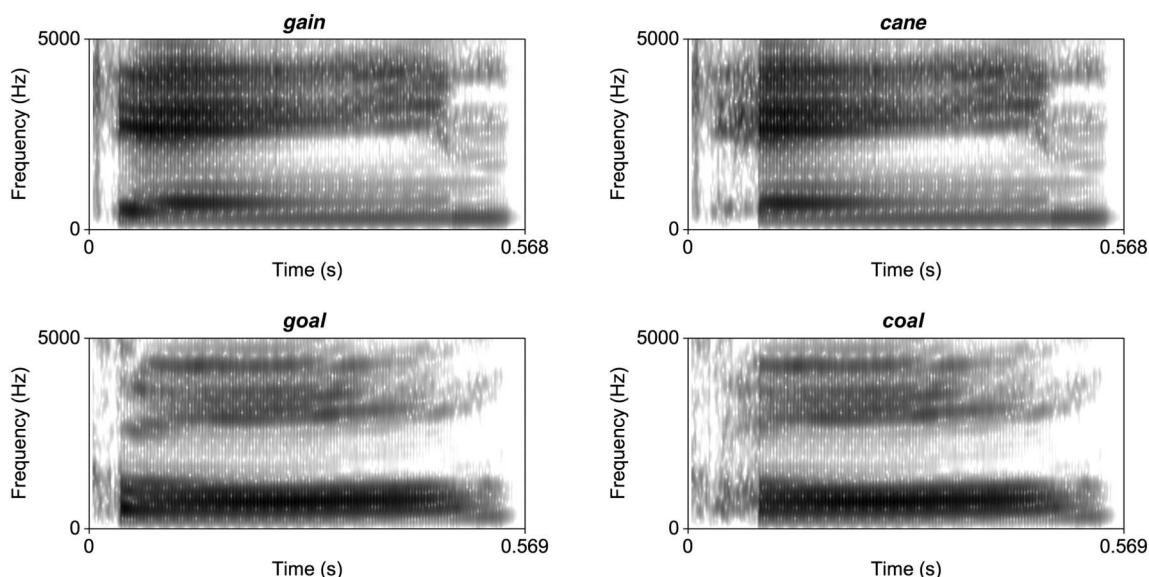


Table 1. Number of tokens for each voice onset time (ms) in the narrow and wide experimental blocks.

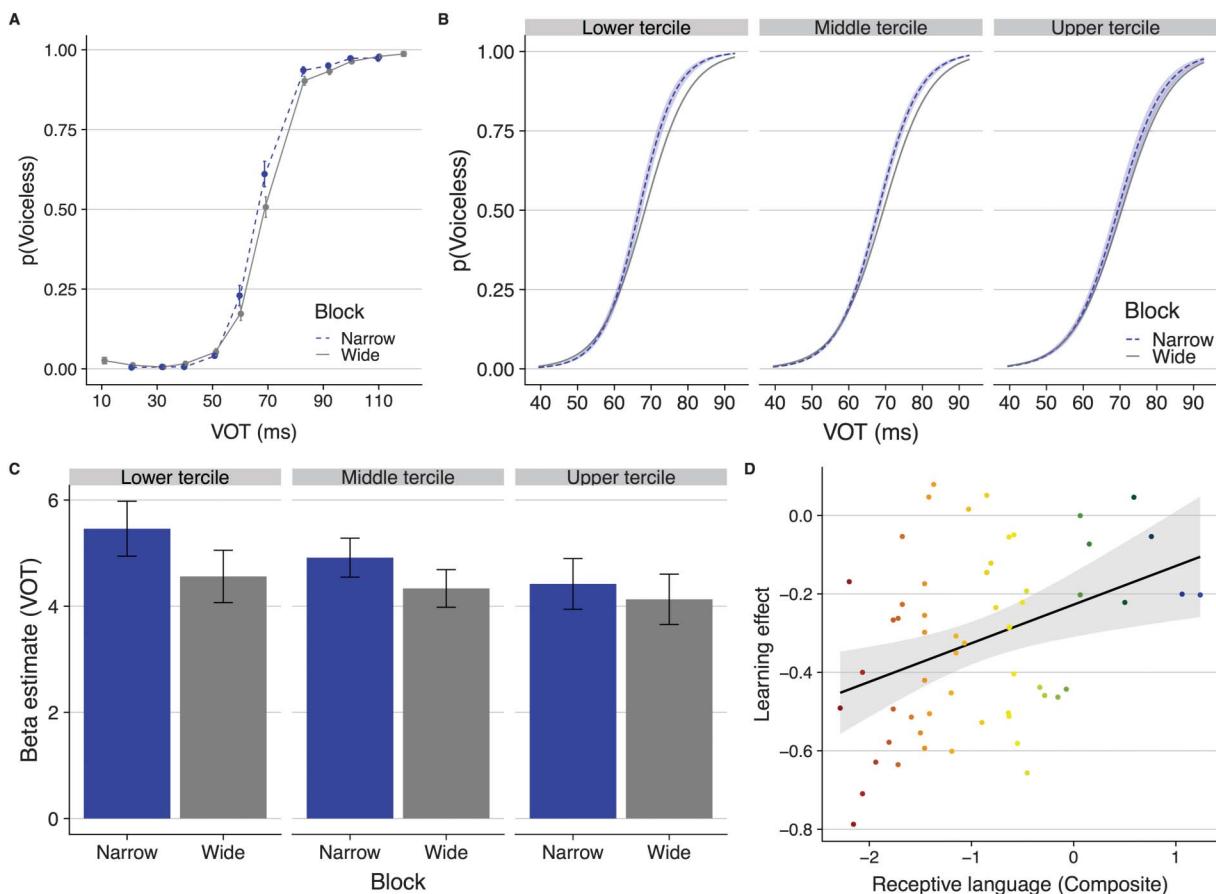
Block	11	21	32	40	51	60	69	83	92	100	110	119
Narrow	0	4	28	54	28	4	4	28	54	28	4	0
Wide	4	12	28	30	28	16	16	28	30	28	12	4

between VOT and voiceless responses in each block. Furthermore, the slope of the function relating VOT to voiceless responses appears to be steeper in the narrow compared to the wide block, indicative of distributional learning across the two input blocks.

To examine this pattern statistically and the degree to which it may be influenced by receptive language, non-verbal intelligence, and reading fluency, trial-level responses

(0 = *voiced*, 1 = *voiceless*) were fit to a generalized linear mixed-effects model (GLMM) using the *glmer()* function with the binomial response family as implemented in the *lme4* package in R (Bates et al., 2014). All test statistics reflect those reported by the *lme4* package. The fixed effects included VOT, block, receptive language composite, TONI, and TOWRE. The fixed effects also included the interaction between VOT and block and all interactions between VOT, block, and each of the three individual difference measures. Here and throughout, VOT, receptive language composite, TONI, and TOWRE were entered into the model as continuous variables, each scaled and centered around the mean; block was contrast coded (narrow = -1, wide = 1). The random effects structure consisted of random intercepts by participant and random slopes by participant for both VOT and block.

Figure 4. Panel A shows the mean proportion of voiceless responses as a function of voice onset time (VOT); error bars indicate standard error of the mean. Panel B shows the effect of VOT on voiceless responses in each block for three levels of receptive language (corresponding to the median of each receptive language composite tercile) as derived from the fixed effects of the model reported in Table 4. To promote visualization, the abscissa spans the intermediate VOTs of the input distributions. Panel C shows the simple slope (beta estimate) for VOT in each block for each composite tercile; error bars indicate the 95% confidence interval for the beta estimate. Higher beta estimates indicate steeper identification slopes. Panel D shows the relationship between the distributional learning effect and receptive language composite score across the 58 participants; the shaded region depicts the 95% confidence level interval for a linear regression. We note that the regression line is provided for visualization purposes only. As described in the main text, negative learning effect values are associated with increased learning (i.e., a larger change in slope between the narrow and wide blocks), and lower composite scores are indicative of stronger receptive language.



The results of the model are shown in Table 2. In the model, the fixed effect of VOT reflects the slope of the identification function. There was a main effect of VOT, indicating that voiceless responses increased as did VOT ($\beta = 4.645$, $SE = 0.178$, $z = 26.157$, $p < .001$). There was also a main effect of block ($\beta = -0.093$, $SE = 0.045$, $z = -2.050$, $p = .040$), indicating more /k/ responses in the narrow compared to the wide block. As expected, there was an interaction between VOT and block ($\beta = -0.299$, $SE = 0.052$, $z = -5.721$, $p < .001$), with the direction of the beta estimate for the interaction indicating that the rate at which voiceless responses increased given an increase in VOT (i.e., the identification slope) was higher for the narrow compared to the wide block. This interaction confirms that participants in the aggregate showed a steeper identification function for the narrow compared to the wide input distributions, as predicted by the ideal observer computational framework. Critically, the model also showed a significant interaction between VOT, block, and receptive language composite score ($\beta = 0.179$, $SE = 0.054$, $z = 3.327$, $p = .001$). No other main effect or interaction was reliable ($p \geq .055$ in all cases).

The results of the omnibus model suggest that distributional learning is influenced by receptive language, but not reading fluency and nonverbal intelligence. To examine this possibility more directly, four successively complex models were compared using likelihood ratio tests. Model 1 included the fixed effects of VOT, block, and their interaction. Model 2 added the fixed effect of receptive language composite, including all interactions with VOT and block. To the structure of Model 2, Model 3 added the fixed effect of reading fluency, including all interactions with VOT and block. Model 4 is the omnibus model (see Table 2) and thus included all three individual difference measures as fixed effects, including all interactions with VOT and block for each measure. The random effects structure was identical across all four models, consisting of random intercepts by

participant and random slopes for VOT and block by participant.

The results of the model comparisons are shown in Table 3. Compared to the initial model (Model 1), there was a significant change to goodness of fit when receptive language was added as a fixed effect, $\chi^2(4) = 17.92$, $p = .001$. However, there was no further change to the goodness of fit by the successive inclusion of reading fluency, $\chi^2(4) = 3.41$, $p = .491$, and nonverbal intelligence scores, $\chi^2(4) = 0.46$, $p = .977$. Though there is a statistically significant change in goodness of fit when receptive language is added to the initial model, the inclusion of composite score leads to only a slight increase in the R^2 for the fixed effects ($R^2 = .802$ and .794, respectively), indicative of a small effect size.

Table 4 shows the results of the best-fitting model, which included the fixed effects of VOT, block, and receptive language composite. As observed for the omnibus model (see Table 2), this model confirmed the presence of the three-way interaction between VOT, block, and composite score ($\beta = 0.159$, $SE = 0.046$, $z = 3.468$, $p = .001$), indicating that the degree to which the slope of the identification function changed across blocks was influenced by receptive language composite score. The model is visualized in Figure 4B in terms of the fixed effects of VOT, block, and composite score, with the latter shown by the composite scores corresponding to the median of each composite tercile. Inspection of this plot shows that the degree to which the identification slope changes between the narrow and wide blocks is largest for those with lower composite scores (reflecting stronger receptive language) and smallest for those with higher composite scores (reflecting weaker receptive language).

To further illustrate this interaction, a simple slope analysis was performed using the interactions package in R (Long, 2019) in order to extract the VOT beta estimate (i.e., the identification slope) in each block for three levels of the receptive language composite score, representing the

Table 2. Results of the generalized linear mixed-effects model for voiceless responses that included voice onset time (VOT), block, receptive language composite score, Test of Nonverbal Intelligence (TONI), and Test of Word Reading Efficiency (TOWRE) as fixed effects.

Fixed effect	β	SE	95% CI	z	p
(Intercept)	-0.427	0.099	[-0.62, -0.23]	-4.324	< .001
VOT	4.645	0.178	[4.30, 4.99]	26.157	< .001
Block	-0.093	0.045	[-0.18, -0.00]	-2.050	.040
Composite	-0.214	0.112	[-0.43, 0.00]	-1.919	.055
TOWRE	-0.117	0.107	[-0.33, 0.09]	-1.095	.274
TONI	0.025	0.105	[-0.18, 0.23]	0.239	.811
VOT \times Block	-0.299	0.052	[-0.40, -0.20]	-5.721	< .001
VOT \times Composite	-0.377	0.198	[-0.77, 0.01]	-1.904	.057
Block \times Composite	0.017	0.049	[-0.08, 0.11]	0.346	.729
VOT \times TOWRE	0.068	0.191	[-0.44, 0.31]	0.355	.722
Block \times TOWRE	-0.039	0.048	[-0.13, 0.06]	-0.813	.416
VOT \times TONI	0.116	0.188	[-0.25, 0.48]	0.616	.538
Block \times TONI	-0.006	0.047	[-0.10, 0.09]	-0.128	.898
VOT \times Block \times Composite	0.179	0.054	[0.07, 0.28]	3.327	.001
VOT \times Block \times TOWRE	0.066	0.054	[-0.04, 0.17]	1.206	.228
VOT \times Block \times TONI	-0.016	0.050	[-0.11, 0.08]	-0.318	.751

Table 3. Results of the likelihood ratio tests for model comparisons.

Model	Fixed effects	df	R ²		logLik	Deviance	Likelihood ratio test		
			Fixed	Total			χ ²	df	p
1	VOT × Block	10	.794	.878	-4655.6	9311.3	—	—	—
2	+ VOT × Block × Composite	14	.802	.878	-4646.7	9293.4	17.92	4	.001
3	+ VOT × Block × TOWRE	18	.803	.878	-4645.0	9289.9	3.41	4	.491
4	+ VOT × Block × TONI	22	.803	.879	-4644.7	9289.5	0.46	4	.977

Note. As described in the main text, the initial model included voice onset time (VOT), block, and their interaction as fixed effects. Comparison models successively added the fixed effects of receptive language (Composite), reading fluency (Test of Word Reading Efficiency [TOWRE] Index), and nonverbal intelligence (Test of Nonverbal Intelligence [TONI]), including all interactions with VOT and block for each individual difference measure. As described in the main text, the random effects structure was identical across models. The full results for the omnibus model (Model 4) are shown in Table 2. The full results of Model 2 are shown in Table 4.

median composite score for the lower, middle, and upper terciles; this is shown in Figure 4C. The three-way interaction can be observed by comparing the degree to which the identification slope (i.e., the VOT beta estimate) differs between the narrow and wide blocks as a function of receptive language score; lower composite scores (indicative of stronger receptive language) show the largest distributional learning effect, and higher language scores (indicative of weaker receptive language) show a minimal distributional learning effect.

In addition to showing that the distributional learning effect was larger for those with stronger compared to weaker receptive language ability, inspection of Figure 4C suggests that this interaction was driven by a stronger association between receptive language and identification slope in the narrow block compared to the wide block. To test this possibility, additional GLMMs were constructed in order to examine the effect of receptive language composite score in each block. In both models, the fixed and random effects structure followed that outlined previously except for removing the fixed effect of block. For the narrow block, there was a main effect of VOT ($\hat{\beta} = 4.728$, $SE = 0.180$, $z = 26.261$, $p < .001$), a main effect of receptive language composite ($\hat{\beta} = -0.235$, $SE = 0.105$, $z = -2.233$, $p = .026$), and an interaction between these two factors ($\hat{\beta} = -0.534$, $SE = 0.167$, $z = -3.191$, $p = .001$), indicating steeper identification slopes for stronger compared to weaker composite scores. For the wide block, there was a main effect of VOT ($\hat{\beta} = 4.422$, $SE = 0.174$, $z = 25.471$, $p < .001$), but no effect

of composite ($\hat{\beta} = -0.135$, $SE = 0.103$, $z = -1.311$, $p = .190$) nor an interaction between VOT and composite score ($\hat{\beta} = -0.218$, $SE = 0.166$, $z = -1.314$, $p = .189$). These results indicate that the locus of the interaction between VOT, block, and receptive language composite in the full model reflects a more limited ability among those with weaker receptive language scores to capitalize on the consistent input distributions.

A final analysis was performed to visualize the distributional learning effect at the level of individual participants. To quantify the learning effect for each participant, we constructed a GLMM on trial-level responses with VOT, block, and their interaction as fixed effects; random intercepts by participant; and random slopes by participant for the interaction between VOT and block. With this structure, the coefficients of the random slopes for the VOT × Block interaction can serve as a measure of the distributional learning effect for each participant. In terms of interpreting the coefficients, negative values indicate that the VOT slope decreased from the narrow to wide block, and values of 0 indicate no change in the VOT slope between the two blocks. Figure 4D shows the distributional learning effect and receptive language composite score for each participant.

Replication Analyses

Recall that 30 of the 58 participants completed the distributional learning task as part of the Narrow-Wide

Table 4. Results of the generalized linear mixed-effects model for voiceless responses that included voice onset time (VOT), block, and receptive language composite score as fixed effects.

Fixed effect	$\hat{\beta}$	SE	95% CI	z	p
(Intercept)	-0.428	0.100	[-0.62, -0.23]	-4.286	< .001
VOT	4.639	0.177	[4.29, 4.99]	26.187	< .001
Block	-0.091	0.045	[-0.18, 0.00]	-2.015	.044
Composite	-0.180	0.099	[-0.37, 0.01]	-1.830	.067
VOT × Block	-0.296	0.052	[-0.40, -0.19]	-5.675	< .001
VOT × Composite	-0.385	0.172	[-0.72, -0.05]	-2.241	.025
Block × Composite	0.031	0.042	[-0.05, 0.11]	0.725	.468
VOT × Block × Composite	0.159	0.046	[0.07, 0.25]	3.468	.001

condition reported in the study of Theodore and Monto (2019), in addition to completing the assessment battery for inclusion in the current study. Because of this, the distributional learning effects presented above cannot be considered as a replication of the previous study given that approximately half of the participants were included in both studies. In order to assess replication of the distributional learning effect reported in the study of Theodore and Monto, we conducted an analysis with only the 28 participants unique to the current sample. For this analysis, trial-level voiceless responses were submitted to a GLMM with the fixed effects of block (narrow = -1, wide = 1) and VOT (scaled/centered around the mean). The model also included random intercepts by participant and random slopes by participant for VOT and block. The results of this model showed a main effect of VOT ($\beta = 4.279$, $SE = 0.213$, $z = 20.064$, $p < .001$), no main effect of block ($\beta = -0.075$, $SE = 0.060$, $z = -1.257$, $p = .209$), and a significant interaction between VOT and block ($\beta = -0.293$, $SE = 0.065$, $z = -4.477$, $p < .001$). The significant interaction indicates that the slope of the identification function relating VOT to voiceless responses is steeper in the narrow compared to the wide block, replicating the previous finding. A final model compared the magnitude of the VOT \times Block interaction between participants unique to the current study and those who participated in both studies by adding sample as a fixed effect (unique = -1, both = 1) to the model described above. The interaction between VOT, block, and sample was not reliable ($\beta = 0.050$, $SE = 0.047$, $z = 1.063$, $p = .288$).

Discussion

Listeners must accommodate wide variability in the acoustic speech signal in order to map the speech signal to the speech sound representations that support language comprehension. One mechanism that supports this process is distributional learning for speech, wherein adaptation can be viewed as the process of dynamically modifying the mapping to speech sounds to optimize phonetic categorization for specific input distributions (Clayards et al., 2008; Kleinschmidt & Jaeger, 2015; Theodore & Monto, 2019). As predicted by ideal observer frameworks, listeners' phonetic identification responses reflect variability of the speech input, with perception more categorical for consistent compared to inconsistent input distributions (Clayards et al., 2008; Nixon et al., 2016; Theodore & Monto, 2019). Recent research suggests that the ability to dynamically modify the acoustic-phonetic boundary between speech sound categories as a consequence of exposure to structured phonetic variability may reflect individual differences in receptive language ability (Colby et al., 2018). The goal of the current work was to provide an additional test of this hypothesis. Specifically, we examined whether the ability to modify the mapping to speech sounds as a function of changes to the consistency of an acoustic-phonetic cue would be linked to receptive language ability and, if so, whether it would also be linked to nonverbal intelligence and reading fluency. We predicted

that young adults who have weaker receptive language abilities would demonstrate a reduced ability to modify their mapping in response to variable acoustic information, manifesting in no difference in the slopes of their identification functions for more versus less consistent input distributions.

Robust distributional learning was observed in our sample as a whole, with steeper identification slopes observed for narrow compared to wide input distributions, providing further evidence that distributional learning reflects rapid, dynamic adaptation to cumulative input statistics. Moreover, individual variation in receptive language ability influenced the magnitude of distributional learning; individuals with stronger receptive language abilities showed the largest distributional learning effect, with weaker learning effects observed among those with weaker receptive language ability. The attenuated learning across test blocks for those with weaker receptive language was driven by the failure to capitalize on the consistent input distributions presented in the narrow test block. Analysis of performance within each test block showed that stronger receptive language was associated with steeper identification slopes in the narrow block, but no such relationship was observed in the wide block. Thus, it appears that individuals with weaker receptive language failed to take advantage of the consistency provided in the narrow block, consistent with previous findings demonstrating that individuals with deficits in language processing abilities show poor adaptability to structured variation when engaging in statistical learning of nonadjacent dependencies (Misyak et al., 2010).

The results of the present investigation converge with those of Colby et al. (2018), who found that perceptual learning through both bottom-up and top-down learning mechanisms was influenced by individual differences in receptive language ability, as measured by receptive vocabulary. The current work extends these findings in four ways. First, receptive language ability in the current sample was measured using the receptive language composite measure of Fidler et al. (2011) instead of using the Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997). Reliable relationships between perceptual learning and receptive language were observed with both measures of receptive language, demonstrating generalization across the specific measures used to assess receptive language as a construct. Second, the current work examined distributional learning for a temporal acoustic-phonetic cue instead of a spectral cue, thus demonstrating that the relationship between receptive language and distributional learning is not limited to a specific acoustic-phonetic property. Third, the current work assessed learning for input distributions that differed in the consistency in which the acoustic-phonetic property was used to cue the two phonetic categories. According to the ideal observer model used here, optimal adaptation to the input distributions required a change in the slope of the identification function over time, as opposed to a shift in the perceptual boundary between the two phonetic categories, as examined previously. Thus, the current results demonstrate that the relationship between receptive language and

distributional learning generalizes to other statistical cues, including those indicative of variability in speech input.

Fourth, Colby et al. (2018) found no evidence to suggest that distributional learning was linked to individual differences in attention-switching control, working memory, or speech perception in noise, in contrast to the reliable relationship that was observed between distributional learning and receptive vocabulary. This finding suggests that the relationship between distributional learning and receptive language ability may not reflect general cognitive ability but rather is more indicative of relationships within the language architecture. In the current study, we provided a further test of this hypothesis by examining the relationship between distributional learning and two measures of linguistic ability, receptive language and reading fluency, in addition to nonverbal intelligence. As in Colby et al., we observed no relationship between distributional learning and general cognitive ability (i.e., nonverbal intelligence). Moreover, we observed no relationship between distributional learning and reading fluency. Given the robust relationship that was observed between distributional learning and receptive language, the results of the current study provide further evidence of a specific relationship between distributional learning and receptive language ability.

We conclude by considering implications of the current investigation for individuals with language impairment, noting that only 14% of the current sample met criterion for DLD and thus the current data are not sufficient to describe patterns between those who meet criterion for DLD and those who do not. Past research has shown that individuals with DLD demonstrate impairments in statistical learning (Lum et al., 2014) and categorical perception (e.g., Robertson, Joanisse, Desroches, & Ng, 2009), the latter of which may reflect specific characteristics of the stimuli and task (Coady, Evans, Mainela-Arnold, & Kluender, 2007; Coady, Kluender, & Evans, 2005). However, we know little about whether individuals with DLD are able to modify their representation of phonetic category structure in response to variability in speech input. Poor adaptability could lead to impairments in efficient processing and comprehension of speech sounds and language (Misyak et al., 2010; Wanrooij, Escudero, & Raijmakers, 2013). Though DLD is characterized by marked deficits in acquiring aspects of language including the sound structure, grammatical morphology, and syntactic rules that govern word order (Bird & Bishop, 1992; Leonard, 2014; van der Lely, 1996), the specific etiology of DLD is unknown. The locus of language impairment has traditionally been described as impairments in the representation of grammar (e.g., van der Lely & Stollwerck, 1996). However, some findings suggest that the grammatical language deficits observed in this population may stem from earlier deficits in the processing stream, including auditory processing (Bishop & McArthur, 2004; McArthur & Bishop, 2004) and speech perception abilities (Joanisse & Seidenberg, 2003), and may reflect deficits that are not language specific (Montgomery, 1995; Spaulding, Plante, & Vance, 2008). In particular, one hypothesis suggests that an inability to attend to fine-grained differences in speech sounds may lead to

impairment in the acquisition of grammatical morphemes that are less salient (e.g., the word-final /t/ signaling the past tense morpheme in *jumped*; Joanisse & Seidenberg, 2003). On this view, deficits in speech perception can lead to broad deficits in language impairment, including impairments in word learning and grammatical morphology (Joanisse & Seidenberg, 1998, 2003; Ziegler, Pech-Georgel, George, Alario, & Lorenzi, 2005). Children with DLD show deficits in forming categories for nonspeech sounds, which is consistent with the possibility that this population has difficulties creating and organizing auditory information into structured perceptual categories (Coady et al., 2007; Nittrouer, Shune, & Lowenstein, 2011).

Indeed, processing-based accounts of DLD have been motivated in light of these findings. Two etiological accounts of language impairment that do account for speech perception ability are the statistical learning deficit hypothesis (Hsu & Bishop, 2014) and the procedural deficit hypothesis (Ullman & Pierpont, 2005). Hsu and Bishop (2014) propose that language impairment manifests as a deficit in statistical learning of grammatical forms and not a deficit in learning grammatical rules. Similarly, Ullman and Pierpont (2005) implicate deficits in the procedural memory system as the etiology of DLD. The procedural memory system establishes and facilitates activation of new sensorimotor plans, such as coordination and motoric functioning, manipulation of visual-spatial imagery, and performance on tasks of working memory. Ullman and Pierpont suggest that deficits in the procedural memory system can explain the linguistic and—critically—nonlinguistic deficits in individuals with DLD. Our results are consistent with both of these hypotheses, as distributional learning is a task that may be mediated by procedural memory and statistical learning abilities. However, future research that examines distributional learning of low-level acoustic-phonetic cues with larger samples sizes of individuals with DLD is needed to test this possibility.

In conclusion, listeners show an exquisite ability to modify the mapping to speech sounds to accommodate statistical cues in speech input throughout the life span. The results of the current investigation point toward a link between adaptation to distributional variation in acoustic-phonetic input and receptive language ability but found no evidence of a similar association between distributional learning and either nonverbal intelligence or reading fluency. These results contribute to a theoretical framework that can account for individual variation in spoken language processing, which will help to inform the role of low-level speech perception abilities as an etiological locus of language impairment.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R21DC016141 to R. M. T. and by the Raymond H. Stetson Scholarship in Phonetics and Speech Science from the Acoustical Society of America to N. R. M. The views expressed here reflect those of the authors and not the National Institutes of Health or the National Institute on

Deafness and Other Communication Disorders. A pilot study for this work was completed as a master's thesis by the third author under the direction of the first author. Portions of this study were presented at the 177th meeting of the Acoustical Society of America.

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.

Baese-Berk, M., Bent, T., Borrie, S., & McKee, M. (2015). Individual differences in perception of unfamiliar speech. *Proceedings of the 18th International Congress of the Phonetic Sciences*, 0460, 1–5.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Green, P. (2014). Package 'lme4.' [Computer software]. Retrieved from <https://www.r-project.org/>

Bird, J., & Bishop, D. (1992). Perception and awareness of phonemes in phonologically impaired children. *European Journal of Disorders of Communication*, 27(4), 289–311.

Bishop, D. V. M., & McArthur, G. M. (2004). Immature cortical responses to auditory stimuli in specific language impairment: Evidence from ERPs to rapid tone sequences. *Developmental Science*, 7(4), F11–F18.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Non-verbal Intelligence—Fourth Edition (TONI-4)*. Austin, TX: Pro-Ed.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *The Journal of the Acoustical Society of America*, 92(1), 593–596.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–229.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, 61, 30–47.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809.

Coady, J. A., Evans, J. L., Mainela-Arnold, E., & Kluender, K. R. (2007). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *Journal of Speech, Language, and Hearing Research*, 50, 41–57.

Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research*, 48, 944–959.

Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research*, 61(8), 1855–1874.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. <https://doi.org/10.1016/j.cognition.2009.10.009>

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—III (PPVT-III)*. Shoreview, MN: AGS.

Earle, F. S., Landi, N., & Myers, E. B. (2018). Adults with specific language impairment fail to consolidate speech sounds during sleep. *Neuroscience Letters*, 666, 58–63.

Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, 20(1), 2–13.

Hall, J., Owen Van Horne, A., McGregor, K. K., & Farmer, T. (2017). Distributional learning in college students with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 60(11), 3270–3283.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5), 3099–3111.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, 17(3), 352–365.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences*, 2(7), 240–247.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, 86(1), 40–56.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203.

Kleinschmidt, D. F., & Jaeger, T. F. (2016). *What do you expect from an unfamiliar talker?* Paper presented at the Proceedings of the 38th Annual Meeting of the Cognitive Science Society, Austin, TX.

Leonard, L. B. (2014). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions* (Version R package Version 1.0.0). Retrieved from <https://cran.r-project.org/package=interactions>

Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, 51, 1–10.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111. [https://doi.org/10.1016/S0010-0277\(01\)00157-3](https://doi.org/10.1016/S0010-0277(01)00157-3)

McArthur, G. M., & Bishop, D. V. M. (2004). Frequency discrimination deficits in people with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 527–541.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, 73(5), 1751–1755.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302–331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, 31.

Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research*, 38(1), 187–199.

Morice, R., & McNicol, D. (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex*, 21(4), 567–580.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America*, 109(3), 1181–1196.

Nittrouer, S., Shune, S., & Lowenstein, J. H. (2011). What is the deficit in phonological processing deficits: Auditory sensitivity, masking, or category formation? *Journal of Experimental Child Psychology*, 108(4), 762–785.

Nixon, J. S., & Best, C. T. (2018). *Acoustic cue variability affects eye movement behaviour during non-native speech perception*. Paper presented at the 9th International Conference on Speech Prosody, Poznan, Poland.

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: Eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, 90, 103–125.

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](https://doi.org/10.1016/S0010-0285(03)00006-9)

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29(4), 434–446.

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science*, 12(5), 753–767. <https://doi.org/10.1111/j.1467-7687.2009.00806.x>

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160059.

Spaulding, T. J., Plante, E., & Vance, R. (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. *Journal of Speech, Language, and Hearing Research*, 51, 16–34.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America*, 125(6), 3974–3982. <https://doi.org/10.1121/1.3106131>

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, 26(3), 985–992.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency—Second Edition*. Austin, TX: Pro-Ed.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34(3), 434–464. <https://doi.org/10.1111/j.1551-6709.2009.01077.x>

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399–433.

van der Lely, H. K. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua*, 98(4), 243–272.

van der Lely, H. K., & Stollwerck, L. (1996). A grammatical specific language impairment in children: An autosomal dominant inheritance? *Brain and Language*, 52(3), 484–504.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, 92(2), 723–735.

Wanrooij, K., Escudero, P., & Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, 41(5), 307–319.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 14110–14115.