Individual differences in the use of acoustic-phonetic versus lexical cues for speech perception

1 **Nikole Giovannone[1], Rachel M. Theodore[2*]**

2 [1]Department Speech, Language, and Hearing Sciences, University of Connecticut, Storrs,
3 Connecticut, USA

4 [2]Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, Storrs,
5 Connecticut, USA

6 *** Correspondence:**
7 Rachel M. Theodore
8 rachel.theodore@uconn.edu

9 **Keywords: keyword$_1$, keyword$_2$, keyword$_3$, keyword$_4$, keyword$_5$. (Min.5-Max. 8)**

10 **Abstract**

11 Previous research suggests that individuals with weaker receptive language show increased reliance
12 on lexical information for speech perception relative to individuals with stronger receptive language,
13 which may reflect a difference in how acoustic-phonetic and lexical cues are weighted for speech
14 processing. Here we examined whether this relationship is the consequence of conflict between
15 acoustic-phonetic and lexical cues in speech input, which has been found to mediate lexical reliance
16 in sentential contexts. Two groups of participants completed standardized measures of language
17 ability and a phonetic identification task to assess lexical recruitment (i.e., a Ganong task). In the
18 high conflict group, the stimulus input distribution removed natural correlations between acoustic-
19 phonetic and lexical cues, thus placing the two cues in high competition with each other; in the low
20 conflict group, these correlations were present and thus competition was reduced as in natural speech.
21 The results showed that (1) the Ganong effect was larger in the low compared to the high conflict
22 condition in single-word contexts, suggesting that cue conflict dynamically influences online speech
23 perception, (2) the Ganong effect was larger for those with weaker compared to stronger receptive
24 language, and (3) the relationship between the Ganong effect and receptive language was not
25 mediated by the degree to which acoustic-phonetic and lexical cues conflicted in the input. These
26 results suggest that listeners with weaker language ability down-weight acoustic-phonetic cues and
27 rely more heavily on lexical knowledge, even when stimulus input distributions reflect characteristics
28 of natural speech input.

29 1 Introduction

30 In order to successfully comprehend the speech stream, listeners must map variable acoustic
31 productions to the same phonemic category. This poses a computational challenge for speech
32 perception because there is no one-to-one mapping between speech acoustics and any given speech
33 sound (Hillenbrand et al., 1995; Lisker & Abramson, 1964; Newman et al., 2001; Theodore, Miller,
34 & DeSteno, 2009) Lexical knowledge can help listeners map the speech signal to meaning, especially
35 when the input is potentially ambiguous between speech sound categories (e.g., Ganong, 1980). For
36 example, if an acoustic-phonetic variant ambiguous between /g/ and /k/ is followed by the context -

/ift/, listeners are more likely to perceive the variant as *gift* than *kift*, as the former is consistent with what listeners know to be a word. However, if the exact same variant was instead followed by the context -/iss/, then listeners are more likely to perceive the lexically consistent form *kiss* than the nonword *giss*.

Though listeners use both acoustic-phonetic and lexical cues in speech perception, these sources of information may at times be in competition with each other. This competition may occur naturally in everyday speech (e.g., clear speech vs. casual speech) and in more difficult listening conditions (e.g., nonnative speech). For example, a native Spanish speaker may produce the phoneme /p/ with a shorter voice-onset-time (VOT) than is typical of English /p/. Based on acoustic-phonetic cues alone, an English listener may perceive this production of a /p/ as more /b/-like because the VOT of this /p/ production more closely maps to VOTs associated with /b/ productions in English (Abramson & Lisker, 1973). If this production is followed by a lexical context consistent with /p/, such as *-anda*, then acoustic-phonetic and lexical information are in competition: while the acoustics may be consistent with /b/ to an English listener's ear, the lexical context argues that the production must have been a /p/ because *panda* is an English word and *banda* is not.

This phenomenon is illustrated through an example of the classic paradigm used to examine the Ganong effect (e.g., Ganong, 1980). In this paradigm, listeners complete a phonetic identification task for tokens drawn from two speech continua. For example, a continuum of word-initial VOTs perceptually ranging from an exemplar /g/ to an exemplar /k/ are appended to two different lexical contexts: *-ift* and *-iss*. Each VOT onset is appended to each context, creating two continua, one that perceptually ranges from *gift* to *kift* and one that perceptually ranges from *giss* to *kiss*. In this paradigm, acoustic-phonetic and lexical information conflict with each other when, for example, tokens from the /g/ end of the continuum are presented in the lexically inconsistent context -/iss/. Though the token is a clear /g/ based on the phonetic cue, it is presented in a context that is inconsistent with lexical knowledge. In this case, bottom-up processing usually prevails; listeners are more likely to categorize continuum endpoints based on acoustic-phonetic cues than on lexical context. In contrast, listeners appear to rely more heavily on lexical information to categorize more ambiguous VOTs; specifically, categorization of the midpoint tokens differs between the two continua in line with lexical context (i.e., more /k/ responses for the *giss-kiss* continuum compared to the *gift-kift* continuum).

Though this lexical effect on speech categorization is robust at the group level, wide individual variability in the Ganong task has been observed (e.g., Giovannone & Theodore, 2021; Schwartz, Scheffler, & Lopez, 2013). A growing body of literature suggests that these individual differences may be driven in part by differences in how listeners integrate acoustic-phonetic cues and lexical knowledge during speech perception. For example, some individuals rely more highly on lexical information to scaffold speech perception than others (Giovannone & Theodore, 2021; Ishida, Samuel, & Arai, 2016; Schwartz et al., 2013). Recent research suggests that a potential mechanism driving individual differences in reliance on the lexicon may be receptive language ability. Receptive language ability, which refers to the ability to comprehend language, is a broad construct related to understanding phonological, lexical, syntactic, and semantic knowledge. Deficits in receptive language ability are linked to many language impairments, including developmental language disorder and specific language impairment. Schwartz and colleagues (2013) found that children with specific language impairment (which is associated with receptive language deficits) show a larger Ganong effect than their peers, suggesting increased reliance on lexical cues relative to their typically-developing peers. A similar pattern has been observed in adult listeners; adults with weaker receptive language ability show a larger Ganong effect compared to adults with stronger receptive

language ability (Giovannone & Theodore, 2021). It is possible that higher reliance on lexical cues in individuals with weaker receptive language ability is the consequence of a decreased weighting of lower-level acoustic-phonetic cues during speech perception. Indeed, some theories of developmental language disorder suggest that the higher-level linguistic deficits associated with this diagnosis may stem from early auditory processing impairments (e.g., Joanisse & Seidenberg, 2003; McArthur & Bishop, 2004). On this view, the increased weighting of top-down lexical cues in this population may be the consequence of decreased access to or saliency of bottom-up phonetic cues in speech input.

However, previous tests (Schwartz et al., 2013; Giovannone & Theodore, 2021) artificially inflated the conflict between acoustic-phonetic and lexical cues in speech input, which makes it difficult to determine the extent to which individual differences in receptive language ability may be linked to differences in phonetic and lexical weighting during more natural speech processing conditions. For example, in the typical Ganong paradigm, every step of a VOT continuum ranging from /g/ to /k/ is presented an equal number of times in each of the *-ift* and *-iss* lexical contexts. This creates extreme conflict between acoustic-phonetic and lexical cues because the natural correlation between VOT and lexical context has been removed. That is, in natural speech, listeners generally hear shorter VOTs for /g/-initial words and longer VOTs for /k/-initial words, and rarely hear exemplar tokens in inconsistent lexical contexts (e.g., listeners rarely hear a clear *kift* when the word *gift* was intended). Yet, in the typical Ganong task, listeners hear unambiguous /g/ and /k/ tokens equally in both lexical contexts, causing acoustic-phonetic cues to conflict with lexical cues to a higher extent than they do naturally.

Recent findings demonstrate that the correlation between acoustic-phonetic and lexical cues in the input influences the magnitude of the Ganong effect (Bushong & Jaeger, 2019). In this experiment, the critical stimuli consisted of a word-to-word VOT continuum ranging from *dent* to *tent*. On a given trial, a token from this continuum was presented in a sentence that provided disambiguating lexical context. For example, the token was followed by "in the fender…" to bias interpretation towards *dent* or "in the forest…" to bias interpretation towards *tent*. The key manipulation in this experiment concerned the distribution of VOTs across biasing contexts, which differed between two listener groups. For the high conflict group, each VOT was presented the same number of times in each of the two biasing contexts, as in the standard Ganong paradigm. For the low conflict group, VOTs were presented more frequently in contexts that preserved the natural correlation between VOT and lexical cues (e.g., short VOTs were most often presented in a *dent*-biasing context and long VOTs were most often presented in a *tent*-biasing context). Thus, conflict between acoustic-phonetic and lexical cues was manipulated by either removing (high conflict) or preserving (low conflict) a natural correlation between these two cues in the stimulus input distributions.

Bushong and Jaeger (2019) reported two key findings. First, the magnitude of the lexical effect was larger in the low conflict compared to the high conflict condition. This finding demonstrates that the lexical influence on phonetic identification is graded in response to the correlation between phonetic and lexical cues in the input. Moreover, this finding suggests that the standard Ganong paradigm may *underestimate* the influence of lexical context for phonetic categorization, given that the standard manipulation in this paradigm removes the natural correlation between acoustic-phonetic and lexical cues. Second, the low conflict and high conflict groups showed differences in the dynamic reweighting of cue usage during the course of the experiment. In the low conflict condition, a robust lexical effect on phonetic identification responses was observed for early trials and the magnitude of the lexical influence remained constant throughout the experiment. In the high conflict condition, a robust lexical effect was observed for early trials, but the magnitude of this effect diminished throughout the experiment such that no lexical influence on identification responses was observed for

trials near the end of the experiment. These results suggest that listeners dynamically adjusted the relative weight of acoustic-phonetic and lexical cues throughout the experiment in response to the input distributions. Specifically, when conflict between the two cues was low, listeners weighted lexical cues more highly than phonetic cues throughout the entire experiment. However, when conflict between the two cues was high, listeners down-weighted their reliance on lexical information to rely more strongly on acoustic-phonetic information over time (Bushong & Jaeger, 2019).

This study (Bushong and Jaeger, 2019) provides a unique framework for interpreting individual differences in lexical recruitment (i.e., the use of lexical information to facilitate speech perception), including evidence that individuals with weaker receptive language exhibit a larger Ganong effect than those with stronger receptive language (Giovannone & Theodore, 2021, Schwartz et al., 2013). In Bushong and Jaeger's (2019) low conflict group, listeners maintained strong use of lexical cues over time relative to acoustic-phonetic cues, resulting in a larger Ganong effect in the low conflict group than in the high conflict group. It is possible that individuals with weaker receptive language are in a constant low-conflict state due to deficits in perceptual analysis at early levels of speech processing. That is, reducing the saliency of acoustic-phonetic cues in the low conflict group in Bushong and Jaeger (2019) yielded a larger Ganong effect, and a larger Ganong effect is also observed in individuals with weaker language abilities characteristic of developmental language disorder (Giovannone & Theodore, 2021), for whom general auditory processing may be impaired (Joanisse & Seidenberg, 2003; McArthur & Bishop, 2004). If individuals with weaker receptive language ability have less access to acoustic information due to courser perceptual analysis at the acoustic level, then it is possible that lexical cues serve as the more informative cue for phonetic categorization, even when acoustic-phonetic and lexical cues are in high conflict (as in the typical Ganong paradigm used in Giovannone & Theodore, 2021).

However, drawing this parallel is challenged given that Bushong and Jaeger's (2019) paradigm used sentence-length stimuli to provide disambiguating lexical context, and as a consequence, lexical context was temporally displaced from the token to-be-categorized. Because of this, it is unclear whether their findings reflect a reweighting of cues for *online* phonetic identification or, rather, a reweighting of how these cues are used to inform post-perceptual decisions. If these findings do indeed reflect online perceptual processing, then it is possible that acoustic-phonetic and lexical cue conflict could be a mechanism that explains individual differences in lexical recruitment. When sentence-length stimuli are used (e.g., "the *-ent* in the fender"), the temporal distance between the critical phonetic cue and disambiguating context is maximized relative to when disambiguating lexical context is contain within a single word (e.g., "*-ift*" or *"-iss*"), which makes it more probable that phonetic identification responses are influenced by post-perceptual decisions in the former compared to the latter. Examining sensitivity to conflict between phonetic and lexical cues for single-word input distributions, where the phonetic cue and disambiguating lexical context are temporally more immediate, would shed light on whether dynamic reweighting of acoustic-phonetic and lexical cues occurs perceptually or post-perceptually. In addition to contributing to the understanding of the mechanisms that support individual differences in lexical recruitment, this question also potentially bears on the long-lasting debate on when and how lexical information interacts with speech perception (e.g., either perceptually, as in the TRACE model; McClelland & Elman, 1987, or post-perceptually, as in the Merge model; Norris et al., 2000).

The current study extends past research in two ways. First, we assess whether the conflict effect observed by Bushong and Jaeger (2019) emerges for a more standard Ganong task that uses single-word stimuli, which will shed light on whether cue conflict may influence online speech perception as opposed to (or in addition to) post-perceptual decision processes. To address this question, we

compare the magnitude of the Ganong effect in two groups: a group exposed to a high-conflict, single-word Ganong distribution and a group exposed to a low-conflict, single-word Ganong distribution. If Bushong and Jaeger's (2019) finding is contingent on post-perceptual decision-making processes afforded by temporally displaced lexical context, then we will not observe a difference in the magnitude of the Ganong effect across groups. However, if the conflict effect reflects more online use of distributional information for perceptual processing, then the Ganong effect will be larger in the low conflict group than in the high conflict group. Second, we examine whether the relationship between the Ganong effect and receptive language ability (e.g., Giovannone and Theodore, 2021; Schwartz et al., 2013) is mediated by cue conflict. If the relationship between receptive language ability and the magnitude of the Ganong effect is not influenced by conflict, then we expect a similar effect of receptive language ability to emerge in both the high and low conflict conditions; namely, that individuals with weaker receptive language will show increased reliance on lexical information compared to those with stronger receptive language. Such a result would suggest that individuals with weaker receptive language ability give higher weight to lexical cues than individuals with stronger receptive language ability, even when more naturalistic correlations between phonetic and lexical cues are preserved in the input. However, if the effect of receptive language ability is only observed in the high conflict condition, then this pattern of results would suggest that previous evidence for increased reliance on lexical cues among those with weaker receptive language may be the consequence of artificially inflating conflict between phonetic and lexical cues in the standard Ganong paradigm.

2 Materials and Methods

2.1 Participants

The participants (n = 129) were native speakers of American English who were recruited from the University of Connecticut community. Participants were assigned to either the high conflict or the low conflict group. The high conflict group consisted of 70 individuals (20 men, 50 women) between 18 and 26 years of age (*mean* = 20 years, *SD* = 2 years) who completed this task as part of a larger experiment (reported in Giovannone & Theodore, 2021). The low conflict group consisted of 59 individuals (23 men, 36 women) between 18 and 31 years of age (*mean* = 20 years, *SD* = 3 years). The target sample size for each group was 70; however, data collection was halted early due to the COVID-19 pandemic. Sixty-five participants (31 high conflict, 34 low conflict) reported experience with a second language, with self-reported proficiencies of novice (n = 35; 18 high conflict, 17 low conflict), intermediate (n = 21; 11 high conflict, 10 low conflict), or advanced (n = 8; 2 high conflict, 6 low conflict). All participants passed a pure tone hearing screen administered bilaterally at 25 dB for octave frequencies between 500 and 4000 Hz.

2.2 Stimuli

Stimuli consisted of two eight-step voice-onset-time (VOT) continua, one that perceptually ranged from *gift* to *kift* and one that perceptually ranged from *giss* to *kiss*. These stimuli were used in Giovannone and Theodore (2021), to which the reader is referred for a comprehensive reporting of stimulus creation and validation testing. The continua were created using the Praat software (Boersma, 2002) using tokens produced by a male native speaker of American English. Lexical contexts consisted of /ɪs/ and /ɪft/ portions that were extracted from natural productions of *kiss* and *gift*, respectively. Eight different VOTs (17, 21, 27, 37, 46, 51, 59, and 71 ms) were created by successively removing energy from the aspiration region of a natural *kiss* production. The shortest VOT consisted of the burst plus the first quasi-periodic pitch period; each subsequent VOT contained

this burst in addition to aspiration energy that increased across continuum steps. The /ɪs/ (374 ms) and /ɪft/ (371 ms) portions were then appended to each of the eight VOTs, and all stimuli were equated in amplitude. Given this procedure, steps within each continuum differed only by their word-initial VOT and, across continua, any given step differed only in its lexical context.

2.3 Procedure

All participants completed a standardized assessment battery (to measure expressive and receptive language ability) and a phonetic identification task. Each component is described in turn below. The duration of the experimental session was approximately two hours. Participants were given partial course credit or monetary compensation (at a rate of $10 per hour) as incentive for participation.

2.3.1 Standardized assessment battery

All participants completed four subtests from the Clinical Evaluation of Language Fundamentals – 5th Edition (CELF-5; Wiig et al., 2013) in order to assess expressive and receptive language ability. Participants also completed the Test of Nonverbal Intelligence – 4th Edition (TONI-4; Brown et al., 2010) to assess nonverbal intelligence. All testing took place in a quiet laboratory and scoring was performed as specified in the CELF-5 and TONI-4 administration manuals. Twenty-five of the 129 participants (12 high conflict, 13 low conflict) were beyond the oldest age (21 years) provided for the standard score conversion of the CELF-5; calculation of standard scores for these participants was made using the oldest age provided for the conversion, which is sensible given that this age bracket represents a maturational end-state.

Expressive language ability was assessed using the standard scores of the Formulated Sentences and Recalling Sentences CELF-5 subtests. In Formulated Sentences, participants must create a sentence based on a picture using one or two words provided by the test administrator. For example, an appropriate sentence in response to a picture of a mother and two children at the zoo using the provided words *or* and *and* would be "The mother *and* children can go see the elephants *or* the lions." In Recalling Sentences, participants must repeat sentence spoken aloud by the test administrator verbatim. For example, a moderately complex item is "The class that sells the most tickets to the dance will win a prize." While the Recalling Sentences subtest also requires contributions of perception and memory, this subtest is characterized as an expressive language subtest in the CELF-5 manual. Receptive language ability was assessed using the standard scores of the Understanding Spoken Paragraphs and Semantic Relationships CELF-5 subtests. In Understanding Spoken Paragraphs, the administrator reads paragraphs aloud to participants. Participants must then answer verbally administered, open-ended comprehension questions. For example, after hearing a passage about hurricanes that specifies the beginning and end of hurricane season, the participant is asked 3-4 questions, including "When is hurricane season?". In the Semantic Relationships subtest, the administrator reads aloud a short word problem that probes semantic knowledge; participants must select the two correct answers from a set of four displayed in text in the administration booklet. For example, a participant would hear the question "Jan saw Pedro. Pedro saw Francis. Who was seen?" and would be shown Jan, Dwayne, Pedro, and Francis as response options (with the correct answers being Pedro and Francis).

These four subtests were administered in order to assess convergence between measures associated with the same construct (e.g., the Understanding Spoken Paragraphs and Semantic Relationships subtests both assess receptive language ability) and to examine whether any observed relationships between language ability and performance in the phonetic identification task were specific to either expressive or receptive language ability. Figure 1 shows the distribution of standard scores on each

subtest for the two conflict groups[1]. To facilitate interpretation of standard scores, note that each subtest is normed to have a mean score of 10 (*SD* = 3); thus, a score of 10 reflects performance at the 50th percentile. A range of standard scores between 4 and 15 on the Formulated Sentences subtest reflects a range in performance from the 2nd to the 95th percentiles; on the Recalling Sentences subtest, a standard score range from 7 to 18 reflects performance from the 16th to 99.6th percentiles; for Understanding Spoken Paragraphs, a standard score range from 3 to 12 corresponds to performance between the 1st and 75th percentiles; finally, standard scores between 8 and 15 on the Semantic Relationships subtest reflect performance ranging from the 25th to 95th percentiles. While these scores do not reflect the entire range of expressive and receptive language ability possible for each subtest, the participants do span a wide range including below average and above average performance. Mean standard score between the low and high conflict groups did not differ for the Recalling Sentences [$t(108) = 0.384$, $p = 0.702$], Understanding Spoken Paragraphs [$t(127) = 1.018$, $p = 0.311$], and Semantic Relationships [$t(120) = 1.089$, $p = 0.278$] subtests. Mean standard score did differ between the two conflict groups for the Formulated Sentences subtest [$t(93) = 4.360$, $p < 0.001$], reflecting a lower mean standard score for the low conflict (*mean* = 9.5, *SD* = 2.7) compared to the high conflict (*mean* = 11.7, *SD* = 2.2) group.

All participants showed nonverbal intelligence within normal limits (*mean* = 103, *SD* = 9, *range* = 83 – 122) as measured by the TONI-4 standard score. Mean standard score for the TONI-4 did not differ between the low and high conflict groups [$t(127) = 0.197$, $p = 0.844$].

2.3.2 Phonetic identification task

All participants completed the phonetic identification (i.e., Ganong) task individually in a sound-attenuated booth. Stimuli were presented via headphones (Sony MDR-7506) at a comfortable listening level that was held constant across participants. For both the high conflict and low conflict groups, the task consisted of 160 trials that were presented in a different randomized order for each participant. On each trial, participants were directed to indicate whether the stimulus began with /g/ or /k/ by pressing an appropriately labeled button on a button box (Cedrus RB-740). Stimulus presentation and response collection were controlled via SuperLab 4.5 running on a Mac OS X operating system. Participants were instructed to respond as quickly as possible without sacrificing accuracy and to guess if they were unsure.

Trial composition differed between the two participant groups as shown in Figure 2. With this manipulation, modeled after the conflict manipulation in Bushong and Jaeger (2019), listeners in both the high conflict and low conflict groups heard each VOT and each lexical context the same number of times. However, the input distributions specific to each group created high conflict between acoustic-phonetic and lexical cues in the high conflict group relative to the low conflict group. For the high conflict group, each VOT was presented 10 times in each of the two lexical contexts; that is, each of the eight VOT steps from the two continua was presented an equal number of times. This flat frequency input distribution is consistent with the standard Ganong paradigm and creates high conflict between acoustic-phonetic and lexical cues given that listeners hear VOTs

[1] Due to an error in implementing the reversal rule during CELF-5 administration, the number of participants for which CELF-5 subtests scores were available varies slightly across subtests. Of the full sample ($n = 129$), 95 participants had scores for Formulated Sentences (54 high conflict, 41 low conflict), 110 participants had scores for Recalling Sentences (58 high conflict, 52 low conflict), 129 participants had scores for Understanding Spoken Paragraphs (70 high conflict, 59 low conflict), and 122 participants has score for Semantic Relationships (63 high conflict, 59 low conflict).

typical of /g/ in an *–iss* context (i.e., step one of the *giss-kiss* continuum) and VOTs typical of /k/ in an *–ift* context (e.g., step eight of the *gift-kift* continuum).

In contrast, conflict was reduced for listeners in the low conflict condition by structuring the input distributions to be more consistent with lexical knowledge. As shown in Figure 2, each VOT was presented 20 times (as in the high conflict group), but the number of presentations from each continuum varied across the eight VOTs. The endpoint VOTs (i.e., step one and step eight) were always presented in a lexically-consistent context. That is, the most /g/-like VOT (i.e., step one) was always drawn from the *gift-kift* continuum and the most /k/-like VOT (i.e., step eight) was always drawn from the *giss-kiss* continuum. Presentations of steps two and three were weighted towards the lexically-consistent context *gift* (i.e., 15 presentations from the *gift-kift* continuum and 5 presentations from the *giss-kiss* continuum), and presentations of steps six and seven were weighted towards the lexically-consistent context *kiss* (i.e., 5 presentations from the *gift-kift* continuum and 15 presentations from the *giss-kiss* continuum). Presentations of steps four and five, the continuum midpoints, were drawn equally from the *gift-kift* and *giss-kiss* continua. Compared to the high conflict group, listeners in the low conflict group heard input distributions that approximate natural correlations between VOT and lexical context, and thus there was minimal conflict between acoustic-phonetic and lexical cues in the input distributions.

3 Results

Two sets of analyses were conducted. The first analysis was conducted in order to examine whether the effect of cue conflict for single-word stimuli follows the patterns observed in Bushong and Jaeger (2019) for sentence-length stimuli. The second analysis was conducted in order to examine the relationship between cue conflict and performance on the language ability measures. Each is presented in turn.

3.1 Effect of cue conflict for single-word stimuli

Responses on the Ganong task were coded as either /g/ (0) or /k/ (1). Trials for which no response was provided were excluded (< 1% of the total trials). To visualize performance in the aggregate, mean proportion /k/ responses was calculated for each participant for each step of the two continua. Responses were then averaged across participants separately for each conflict condition and are shown in Figure 3, panel (A). Visual inspection of this figure suggests, as expected, the presence of a Ganong effect for both conflict conditions. Specifically, proportion /k/ responses across the range of VOTs are higher for the *giss–kiss* continuum compared to the *gift–kift* continuum, consistent with a lexical influence on perceptual categorization. Moreover, visual inspection of this figure suggests that the magnitude of the Ganong effect is larger for the low conflict compared to the high conflict condition, suggesting that the lexical effect is graded to reflect the correlation between phonetic and lexical cues in the input, as was observed in Bushong and Jaeger (2019).

To examine this pattern statistically, trial-level responses (0 = /g/, 1 = /k/) were fit to a generalized linear mixed-effects model (GLMM) using the glmer() function with the binomial response family (i.e., a logistic regression) as implemented in the lme4 package (Bates et al., 2015) in R. The fixed effects included VOT, continuum, conflict, trial, and all interactions among these factors. VOT was entered into the model as continuous variable, scaled and centered around the mean. Continuum (*giss–kiss* = 1, *gift–kift* = -1) and conflict (high conflict = -1, low conflict = 1) were sum-coded. Trial was first log-transformed (as in Bushong & Jaeger, 2019) and then entered into the model as a continuous variable, scaled and centered around the mean. The random effects structure consisted of random intercepts by participant and random slopes by participant for VOT, continuum, and trial.

The model results are shown in Table 1. As expected, the model showed a significant effect of VOT ($\hat{\beta}$ = 3.329, *SE* = 0.120, *z* = 27.804, *p* < .001), indicating that /k/ responses increased as VOT increased. There was a significant effect of continuum ($\hat{\beta}$ = 1.314, *SE* = 0.094, *z* = 14.048, *p* < .001), with the direction of the beta estimate indicating increased /k/ responses in the *giss–kiss* compared to the *gift–kift* continuum. There was also an interaction between VOT and continuum ($\hat{\beta}$ = 0.457, *SE* = 0.058, *z* = 7.858, *p* < .001), indicating that the lexical effect differed across continuum steps. Critically, there was a robust interaction between continuum and conflict ($\hat{\beta}$ = 0.580, *SE* = 0.092, *z* = 6.297, *p* < .001), indicative of a larger Ganong effect (i.e., effect of continuum) in the low conflict compared to the high conflict condition. This interpretation was confirmed by analysis of simple slopes; there were more /k/ responses in the low compared to the high conflict condition for the *giss–kiss* continuum ($\hat{\beta}$ = 0.386, *SE* = 0.119, *z* = 3.235, *p* = .001), and fewer /k/ responses in the low compared to the high conflict condition for the *gift–kift* continuum ($\hat{\beta}$ = -0.612, *SE* = 0.144, *z* = -4.260, *p* < .001).

There was no significant interaction between continuum, conflict, and trial ($\hat{\beta}$ = -0.012, *SE* = 0.038, *z* = -0.319, *p* = .750), nor between VOT, continuum, conflict, and trial ($\hat{\beta}$ = 0.045, *SE* = 0.053, *z* = 0.862, *p* = .389). These results suggest that – in contrast to the finding of Bushong and Jaeger (2019) – the conflict effect was present during early trials and was maintained over time[2]. As displayed in Figure 3, panel (B), the magnitude of the Ganong effect (i.e., the difference in /k/ responses between the two continua) was relatively stable across trials within each of the high conflict and low conflict conditions.

3.2 Relationships with language measures

The next set of analyses was conducted in order to examine whether the relationship between the Ganong effect and performance on the language measures differed between the high conflict and low conflict conditions. To visualize performance, mean percent /k/ responses was calculated for each participant for each continuum (collapsing over VOT). Figure 4 shows the distribution of proportion /k/ responses across participants separately for the high conflict and low conflict conditions and, critically, split into lower and upper halves based on performance for each of the CELF-5 subtests. Consider first performance for the high conflict condition. As expected (given that these participants reflect a subset of those reported in Giovannone & Theodore, 2021), participants with scores in the lower half of the subtest distribution for the Recalling Sentences, Understanding Spoken Paragraphs, and Semantic Relationships subtests show a larger Ganong effect (i.e., difference in /k/ responses between the two continua) compared to those with scores in the upper half of the subtest distribution, an effect that is attenuated for the Formulated Sentences subtest. Now consider performance for the low conflict condition. As expected based on the analysis reported in 3.1, the magnitude of the Ganong effect is overall larger for the low conflict compared to the high conflict condition. In addition, the effect of language ability on the Ganong effect for the low conflict condition appears to track with that observed for the high conflict condition. Specifically, those with weaker performance

[2] A parallel model was fit using raw trial instead of log-transformed trial as the fixed effect, and comparable results were obtained. In addition, a generalized additive mixed model (GAMM) was performed to detect possible non-linear changes in cue weights across trials following the methodology used in Bushong and Jaeger (2019). The results of the GAMM were consistent with those observed for the model reported in the main text. These supplementary analyses can be viewed on the OSF repository associated with this manuscript (https://osf.io/ubek4/).

on the CELF-5 (i.e., subtest scores in the lower half of the distribution) appear to show a larger Ganong effect in the low conflict condition for all subtests except Formulated Sentences.

To analyze these patterns statistically, trial-level data (0 = /g/, 1 = /k/) were fit to a series of mixed effects models, one for each CELF-5 subtest. Subtests were analyzed in separate models due to collinearity among predictors. In each model, the fixed effects consisted of VOT, continuum, conflict, the CELF-5 subtest, and all interactions among predictors. VOT and CELF-5 subtest were entered into the model as scaled/centered continuous variables. Continuum and conflict were sum-coded as described previously. The random effects structure consisted of random intercepts by participant and random slopes by participant for VOT and continuum. In all models, evidence that cue conflict in the input mediates the relationship between the Ganong effect and language ability would manifest as an interaction between continuum, conflict, and subtest.

The results of the models that included the expressive language measures (i.e., Formulated Sentences, Recalling Sentences) are shown in Table 3. The results of the models that included the receptive language measures (i.e., Understanding Spoken Paragraphs, Semantic Relationships) are shown in Table 4. The two key fixed effects of interest for each model are (1) the interaction between continuum and subtest and (2) the interaction between continuum, conflict, and subtest. As can be seen in Tables 3 and 4, the continuum by subtest interaction was reliable for the Recalling Sentences ($\hat{\beta}$ = -0.299, *SE* = 0.091, *z* = -3.285, *p* = .001), Understanding Spoken Paragraphs ($\hat{\beta}$ = -0.204, *SE* = 0.087, *z* = -2.350, *p* = .019), and Semantic Relationships ($\hat{\beta}$ = -0.259, *SE* = 0.085, *z* = -3.000, *p* = .003) models; it was not reliable for the Formulated Sentences model ($\hat{\beta}$ = -0.016, *SE* = 0.106, *z* = -0.152, *p* = .879). These results confirm that the magnitude of the Ganong effect was related to performance on three of the four CELF-5 subtests; those with lower scores showed a larger Ganong effect compared to those with higher scores. None of the models showed a significant three-way interaction between continuum, conflict, and subtest ($p \geq 0.346$ in all cases); nor were any of the four-way interactions reliable ($p \geq 0.331$ in all cases). These results provide no evidence to suggest that the relationship between the Ganong effect and language ability was mediated by cue conflict in the input; instead, they are consistent with the interpretation that those with weaker language ability show increase reliance on lexical information for speech perception, even for input distributions that preserve natural correlations between acoustic-phonetic and lexical cues.

4 Discussion

Previous research shows that individuals with weaker receptive language ability rely more heavily on lexical information to facilitate speech perception (e.g., Giovannone & Theodore, 2021; Schwartz et al., 2013); however, the mechanisms that drive this difference are not yet known. Differences in acoustic-phonetic and lexical cue weighting for speech perception is one potential mechanism that could explain this observation. To explore this possibility, the current study assessed two main questions. First, we examined whether the conflict effect reported in Bushong and Jaeger (2019) can be elicited using single-word stimuli, rather than sentence stimuli. Second, we assessed whether the relationship between the lexical effect and receptive language ability reported in Giovannone and Theodore (2021) is influenced by conflict between acoustic-phonetic and lexical cues in the input.

To address our first question, we compared the size of the lexical effect (elicited with single-word stimuli) for high conflict and low conflict input distributions. Our results showed a larger lexical effect for the low conflict input compared to the high conflict input, replicating Bushong and Jaeger's (2019) conflict effect using single-word stimuli. This finding provides further evidence of listeners' dynamic sensitivity to competing cues in speech input. Moreover, this finding suggests that input-

driven learning may impact online speech perception rather than post-perceptual decision making, consistent with recent electrophysiological evidence that the lexical effect reflects online processing (Getz & Toscano, 2019; Noe & Fischer-Baum, 2020). The current results are thus most consistent with models of speech perception that posit direct interaction between lexical information and online speech perception (e.g. TRACE; McClelland & Elman, 1987) and less consistent with models that posit a purely modular feed-forward architecture (e.g., Merge; Norris et al, 2000).

In their experiment, Bushong and Jaeger (2019) found that the magnitude of the lexical effect decreased over time in response to the high conflict distribution, but remained steady over time in response to the low conflict distribution. Of note, we did not observe the same pattern in the current study. That is, the results showed that the lexical effect remained steady over time for both the high and low conflict groups, in contrast to Bushong and Jaeger (2019) who observed a diminishing lexical effect over time for their high conflict group. We hypothesize that this difference is the result of using single-word stimuli rather than sentence-length stimuli. In the case of single-word stimuli, the disambiguating context comes earlier within each trial than it does within sentence-length stimuli. The temporal proximity of the ambiguous phoneme and the disambiguating context in our single-word stimuli may have facilitated faster adaptation to the input distribution statistics than for the sentence-length stimuli used in previous work. Thus, we hypothesize that adaptation to the current high conflict distribution occurred early and persisted over the course of experimental exposure, relative to past work using sentence-length stimuli.

With regard to our second question, our results suggest that the relationship between language ability and the lexical context is not mediated by cue conflict. Instead, we found a relationship between receptive language ability and lexical context for both the low conflict and high conflict input distributions. As in previous findings (e.g. Giovannone & Theodore, 2021; Schwartz et al., 2013) individuals with weaker receptive language ability demonstrated a larger lexical effect. Across both the high and low conflict distributions, the magnitude of the lexical effect was predicted by both measures of receptive language ability (Understanding Spoken Paragraphs and Semantic Relationships) and one measure of expressive language ability (Recalling Sentences). While the Recalling Sentences subtest is categorized as an expressive language subtest in the CELF-5 manual, successful completion of this task (repetition of sentences that were read aloud by an experimenter) also requires many aspects of receptive language. The fact that the Recalling Sentences subtest tracks with the two receptive measures is therefore somewhat unsurprising.

For each of the three subtests that showed a relationship with the lexical effect, this relationship did not differ as a function of conflict group. For both the high conflict and low conflict input distributions, individuals with weaker receptive language ability show a larger lexical effect than individuals with stronger receptive language ability. This result informs models of language impairment, including developmental language disorder. In the current work, individuals with weaker receptive language ability relied more highly on lexical information than individuals with stronger receptive language ability, even when the stimulus input distribution was made more naturalistic in the low cue conflict group. This pattern of results suggests that lower weighting of acoustic-phonetic cues relative to lexical cues may be typical of individuals with language impairment, even outside of a laboratory context.

While this study's results contribute to theoretical debates of speech perception and clinical etiologies of language impairment, there are three key limitations that must be acknowledged. First, though our participants do not span the full possible range of CELF subtest score, the sample does show wide variability (from as low as the 1st percentile to as high as the 99.6th percentile across subtests). Future

research sampling individuals with formal language impairment diagnoses can further assess performance at the lower end of the receptive language ability spectrum. Second, we did not observe the expected three-way interaction of cue conflict, continuum, and trial found by Bushong and Jaeger (2019). As previously mentioned, this lack of interaction is likely due to faster adaptation to the single-word stimuli than to sentence-length stimuli. However, this lack of a by-trial effect precludes us from speaking definitively about the ways in which cue usage might change over time in individuals with weaker receptive language ability. Finally, although the results of this study suggest that individuals with weaker receptive language ability weight lexical information more highly than acoustic-phonetic information during speech perception, it does not directly address potential causal mechanisms for this difference. Discovering the underlying mechanisms that drive differential cue usage across language ability is essential to fully understanding the etiology of language impairments such as developmental language disorder.

The findings of this experiment yield three main conclusions. First, the results suggest that distributional information from the input is used to facilitate online speech perception, rather than to inform post-perceptual decision making. Second, and consistent with Bushong and Jaeger (2019), the magnitude of the lexical effect is larger when more naturalistic correlations between acoustic-phonetic and lexical cues are maintained in the stimulus input distribution. Third, individuals with weaker receptive language ability demonstrate a larger lexical effect than those with stronger receptive language ability, regardless of the level of cue conflict in the input. Increased reliance on lexical information in response to the low conflict distribution suggests that listeners with weaker language abilities down-weight acoustic-phonetic cues and rely highly on lexical knowledge even in more naturalistic speech settings.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

NG and RMT conceptualized and designed the study. NG programmed the experiment. NG and RMT performed statistical analyses. NG wrote the first draft of the manuscript, and RMT supplied revisions. The final draft of the manuscript was revised, read, and approved by both NG and RMT.

7 Funding

This work was supported by NIH NIDCD grant R21DC016141 to RMT, NSF grants DGE-1747486 and DGE-1144399 to the University of Connecticut, and by the Jorgensen Fellowship (University of Connecticut) to NG. The views expressed here reflect those of the authors and not the NIH, the NIDCD, or the NSF.

8 Acknowledgments

Portions of this study were presented at the 179th meeting of the Acoustical Society of America. We extend gratitude to Lee Drown and Andrew Pine for assistance with administration and scoring of the assessment battery.

9 References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence: TONI-4*. Pro-Ed.

Bushong, W., & Jaeger, T. F. (2019). Dynamic re-weighting of acoustic and contextual cues in spoken word recognition. *The Journal of the Acoustical Society of America*, *146*(2), EL135–EL140. https://doi.org/10.1121/1.5119271

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Getz, L. M., & Toscano, J. C. (2019). Electrophysiological Evidence for Top-Down Lexical Influences on Early Speech Perception. *Psychological Science*, 095679761984181. https://doi.org/10.1177/0956797619841813

Giovannone, N., & Theodore, R. M. (2021). Individual differences in lexical contributions to speech perception. *Journal of Speech, Language, and Hearing Research*, *64*(3), 707-724.

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. https://doi.org/10.1121/1.411872

Ishida, M., Samuel, A. G., & Arai, T. (2016). Some people are "more lexical" than others. *Cognition*, *151*, 68–75. https://doi.org/10.1016/j.cognition.2016.03.008

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, *86*, 40–56. https://doi.org/10.1016/S0093-934X(02)00533-3

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

McArthur, G. M., & Bishop, D. V. M. (2004). Which People with Specific Language Impairment have Auditory Processing Deficits? *Cognitive Neuropsychology*, *21*(1), 79–94. https://doi.org/10.1080/02643290342000087

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

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. https://doi.org/10.1121/1.1348009

Noe, C., & Fischer-Baum, S. (2020). Early lexical influences on sublexical processing in speech perception: Evidence from electrophysiology. *Cognition*, *197*, 104162. https://doi.org/10.1016/j.cognition.2019.104162

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325. https://doi.org/10.1017/S0140525X00003241

Schwartz, R. G., Scheffler, F. L. V., & Lopez, K. (2013). Speech perception and lexical effects in specific language impairment. *Clinical Linguistics & Phonetics*, *27*(5), 339–354. https://doi.org/10.3109/02699206.2013.763386

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

Wiig, E. H., Semel, E., & Secord, W. (2013). Clinical Evaluation of Language Fundamentals—Fifth Edition. Bloomington, MN: Pearson.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

10 Data Availability Statement

Trial-level data and an analysis script (in R) for this study can be found on the Open Science Framework at https://osf.io/ubek4/. The script operates on trial-level data to reproduce all analyses reported in this manuscript in addition to generating all figures.

11 Tables

Table 1. Results of the mixed effects model for /k/ responses that included fixed effects of VOT, continuum, conflict, and trial.

Fixed effect	$\widehat{\beta}$	*SE*	*z*	*p*
(Intercept)	-1.223	0.097	-12.658	< 0.001
VOT	3.329	0.120	27.804	< 0.001
Continuum	1.314	0.094	14.048	< 0.001
Conflict	-0.100	0.095	-1.058	0.290
Log Trial	0.283	0.053	5.354	< 0.001
VOT x Continuum	0.457	0.058	7.858	< 0.001
VOT x Conflict	-0.281	0.115	-2.437	0.015
Continuum x Conflict	0.580	0.092	6.297	< 0.001
VOT x Trial	0.059	0.055	1.075	0.282
Continuum x Trial	-0.209	0.039	-5.317	< 0.001
Conflict x Trial	-0.096	0.052	-1.859	0.063
VOT x Continuum x Conflict	-0.062	0.056	-1.102	0.270
VOT x Continuum x Trial	0.105	0.053	1.968	0.049
VOT x Conflict x Trial	-0.045	0.053	-0.860	0.390
Continuum x Conflict x Trial	-0.012	0.038	-0.319	0.750
VOT x Continuum x Conflict x Trial	0.045	0.053	0.862	0.389

Table 2. Results of the mixed effects models for /k/ responses that included fixed effects of VOT, continuum, conflict, and CELF-5 expressive language subtest (each in a separate model). Bold font is used to mark the key interactions of interest.

Model	Fixed effect	$\widehat{\beta}$	*SE*	*z*	*p*
Formulated Sentences (FS)	(Intercept)	-1.256	0.122	-10.291	<0.001
	VOT	3.167	0.128	24.760	<0.001
	Continuum	1.150	0.109	10.543	<0.001
	Conflict	-0.154	0.121	-1.273	0.203
	FS	-0.228	0.119	-1.921	0.055
	VOT x Continuum	0.482	0.072	6.684	<0.001
	VOT x Conflict	-0.174	0.125	-1.391	0.164
	Continuum x Conflict	0.533	0.108	4.937	<0.001
	VOT x FS	0.217	0.124	1.753	0.080
	Continuum x FS	**-0.016**	**0.106**	**-0.152**	**0.879**
	Conflict x FS	-0.160	0.119	-1.344	0.179
	VOT x Continuum x Conflict	-0.061	0.070	-0.870	0.384
	VOT x Continuum x FS	0.065	0.067	0.960	0.337
	VOT x Conflict x FS	-0.185	0.124	-1.492	0.136
	Continuum x Conflict x FS	**0.029**	**0.106**	**0.272**	**0.786**
	VOT x Continuum x Conflict x FS	-0.022	0.067	-0.333	0.739
Recalling Sentences (RS)	(Intercept)	-1.119	0.097	-11.527	<0.001
	VOT	3.248	0.113	28.729	<0.001
	Continuum	1.313	0.092	14.204	<0.001
	Conflict	-0.125	0.096	-1.312	0.189
	RS	-0.073	0.095	-0.769	0.442
	VOT x Continuum	0.412	0.061	6.765	<0.001
	VOT x Conflict	-0.273	0.109	-2.494	0.013
	Continuum x Conflict	0.576	0.091	6.300	<0.001
	VOT x RS	0.471	0.111	4.254	<0.001
	Continuum x RS	**-0.299**	**0.091**	**-3.285**	**0.001**
	Conflict x RS	-0.101	0.095	-1.061	0.289
	VOT x Continuum x Conflict	-0.085	0.059	-1.433	0.152
	VOT x Continuum x RS	-0.141	0.061	-2.334	0.020
	VOT x Conflict x RS	0.053	0.111	0.481	0.630
	Continuum x Conflict x RS	**-0.086**	**0.091**	**-0.942**	**0.346**
	VOT x Continuum x Conflict x RS	-0.059	0.061	-0.971	0.331

Table 3. Results of the mixed effects models for /k/ responses that included fixed effects of VOT, continuum, conflict, and CELF-5 receptive language subtest (each in a separate model). Bold font is used to mark the key interactions of interest.

Model	Fixed effect	$\widehat{\beta}$	*SE*	*z*	*p*

Understanding Spoken Paragraphs (USP)	(Intercept)	-1.149	0.091	-12.672	<0.001
	VOT	3.184	0.108	29.443	<0.001
	Continuum	1.249	0.087	14.388	<0.001
	Conflict	-0.102	0.089	-1.143	0.253
	USP	0.041	0.091	0.451	0.652
	VOT x Continuum	0.475	0.054	8.764	<0.001
	VOT x Conflict	-0.217	0.105	-2.071	0.038
	Continuum x Conflict	0.548	0.087	6.388	<0.001
	VOT x USP	0.244	0.105	2.316	0.021
	Continuum x USP	**-0.204**	**0.087**	**-2.350**	**0.019**
	Conflict x USP	0.101	0.091	1.117	0.264
	VOT x Continuum x Conflict	-0.058	0.052	-1.108	0.268
	VOT x Continuum x USP	0.010	0.053	0.180	0.857
	VOT x Conflict x USP	0.054	0.105	0.511	0.610
	Continuum x Conflict x USP	**0.009**	**0.087**	**0.098**	**0.922**
	VOT x Continuum x Conflict x USP	0.027	0.053	0.505	0.614
Semantic Relationships (SR)	(Intercept)	-1.144	0.089	-12.904	<0.001
	VOT	3.173	0.110	28.762	<0.001
	Continuum	1.207	0.088	13.715	<0.001
	Conflict	-0.108	0.087	-1.241	0.215
	SR	-0.083	0.087	-0.960	0.337
	VOT x Continuum	0.491	0.055	8.868	<0.001
	VOT x Conflict	-0.275	0.107	-2.568	0.010
	Continuum x Conflict	0.620	0.087	7.110	<0.001
	VOT x SR	0.309	0.106	2.908	0.004
	Continuum x SR	**-0.259**	**0.086**	**-3.000**	**0.003**
	Conflict x SR	0.093	0.087	1.070	0.285
	VOT x Continuum x Conflict	-0.076	0.054	-1.409	0.159
	VOT x Continuum x SR	0.126	0.053	2.397	0.017
	VOT x Conflict x SR	0.116	0.106	1.087	0.277
	Continuum x Conflict x SR	**-0.018**	**0.086**	**-0.210**	**0.833**
	VOT x Continuum x Conflict x SR	0.005	0.053	0.103	0.918

12 Figure Captions

Figure 1. Beeswarm plots showing individual variation of standard scores in each conflict condition for the four CELF-5 subtests. Expressive language measures are shown in blue; receptive language measures are shown in gray. Points are jittered along the x-axis to promote visualization of overlapping scores.

Figure 2. Histograms showing input distributions (i.e., number of presentations of each VOT for each of the two continua contexts) for the high conflict and low conflict conditions.

Figure 3. Panel (A) shows mean proportion /k/ responses at each voice-onset-time (VOT) for each continuum separately for the high conflict and low conflict conditions. Means reflect grand means calculated over by-subject averages. Error bars indicate standard error of the mean. Panel (B) shows smoothed conditional means (reflecting a generalized additive model fit to trial-level raw data using the geom_smooth() function in the ggplot2 package in R) for proportion /k/ responses over time (i.e., log-transformed trial number) for each continuum separately for the high conflict and low conflict conditions. The shaded region indicates the 95% confidence interval.

Figure 4. Each panel shows the distribution of proportion /k/ responses for each continuum for listeners in the high conflict and low conflict conditions, grouped by performance (i.e., lower half vs. upper half) on a specific CELF-5 subtest. As noted in the main text, this grouping is for visualization purposes only; performance on each subtest was entered into the analysis models as a continuous variable. Performance is split by Formulated Sentences in (A), Recalling Sentences in (B), Understanding Spoken Paragraphs in (C), and Semantic Relationships in (D).