Children's syntactic parsing and sentence comprehension with a degraded auditory signal

Isabel A. Martin, Matthew J. Goupell and Yi Ting Huang

Citation: The Journal of the Acoustical Society of America **151**, 699 (2022); doi: 10.1121/10.0009271

View online: https://doi.org/10.1121/10.0009271

View Table of Contents: https://asa.scitation.org/toc/jas/151/2

Published by the Acoustical Society of America

ARTICLES YOU MAY BE INTERESTED IN

Making sense of periodicity glimpses in a prediction-update-loop—A computational model of attentive voice tracking
The Journal of the Acoustical Society of America **151**, 712 (2022); https://doi.org/10.1121/10.0009337

Perception difference for approaching and receding sound sources of a listener in motion in architectural sequential spaces
The Journal of the Acoustical Society of America **151**, 685 (2022); https://doi.org/10.1121/10.0009231

The physics of knocking over LEGO minifigures with time reversal focused vibrations for use in a museum exhibit
The Journal of the Acoustical Society of America **151**, 738 (2022); https://doi.org/10.1121/10.0009364

Reviewers of Manuscripts, 2021
The Journal of the Acoustical Society of America **151**, 677 (2022); https://doi.org/10.1121/10.0009353

REVIEWS OF ACOUSTICAL PATENTS
The Journal of the Acoustical Society of America **151**, 663 (2022); https://doi.org/10.1121/10.0009375

Interferometric processing of hydroacoustic signals for the purpose of source localization
The Journal of the Acoustical Society of America **151**, 666 (2022); https://doi.org/10.1121/10.0009381

JASA ARTICLE

Children's syntactic parsing and sentence comprehension with a degraded auditory signal

Isabel A. Martin, Matthew J. Goupell,[a] and Yi Ting Huang[b]
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742, USA

ABSTRACT:
During sentence comprehension, young children anticipate syntactic structures using early-arriving words and have difficulties revising incorrect predictions using late-arriving words. However, nearly all work to date has focused on syntactic parsing in idealized speech environments, and little is known about how children's strategies for predicting and revising meanings are affected by signal degradation. This study compares comprehension of active and passive sentences in natural and vocoded speech. In a word-interpretation task, 5-year-olds inferred the meanings of novel words in sentences that (1) encouraged agent-first predictions (e.g., *The blicket is eating the seal* implies *The blicket* is the agent), (2) required revising predictions (e.g., *The blicket is eaten by the seal* implies *The blicket* is the theme), or (3) weakened predictions by placing familiar nouns in sentence-initial position (e.g., *The seal is eating/eaten by the blicket*). When novel words promoted agent-first predictions, children misinterpreted passives as actives, and errors increased with vocoded compared to natural speech. However, when familiar words were sentence-initial that weakened agent-first predictions, children accurately interpreted passives, with no signal-degradation effects. This demonstrates that signal quality interacts with interpretive processes during sentence comprehension, and the impacts of speech degradation are greatest when late-arriving information conflicts with predictions.

https://doi.org/10.1121/10.0009271

(Received 2 December 2020; revised 15 October 2021; accepted 16 December 2021; published online 3 February 2022)

[Editor: Deniz Başkent] Pages: 699–711

I. INTRODUCTION

Natural speech unfolds at roughly 2.5 words per second, and this requires listeners to anticipate future words to keep pace with a rapidly unfolding signal. During sentence interpretation, listeners recruit information from word order (e.g., who is the grammatical subject?), word meanings (e.g., who is a plausible actor in events?), and morphology (e.g., which event occurred in the past?) to determine who did what to whom. Since spoken sentences unfold on a word-by-word basis, listeners must calculate likely meanings using early-arriving words and revise their interpretations when their initial predictions are incorrect.

(1) a. Active: The blicket is eating the seal. [infer that *the blicket* is the agent]
b. Passive: The blicket is eaten by the seal. [infer that *the blicket* is the theme]

For example, when listeners hear first noun phrases (NP1s) in sentences like (1), they often assume that NP1s are agents (e.g., *The blicket* is the performer of the action) and anticipate that upcoming second noun phrases (NP2s) will be themes (e.g., *The seal* is the recipient of the action). Agent-first predictions are correct for active sentences like (1a), but incorrect for passive sentences like (1b), where the order of roles is reversed. After listeners encounter late-arriving cues like the past participle (e.g., *-en* in *eaten*) and *by*-phrase (e.g., *by the seal*), they must revise initial predictions and reinterpret NP1s as themes and NP2s as agents (Bever, 1970; Huang *et al.*, 2017a; Huang and Hollister, 2019; Abbot-Smith *et al.*, 2017; Deen *et al.*, 2018; Messenger and Fisher, 2018).

While effects of syntactic parsing on sentence meaning are well understood, nearly all work to date has studied real-time interpretation using signals of ideal quality. Yet, real-world speech information is rarely pristine, and this raises questions of how listeners interpret sentence cues (e.g., word order, meanings, morphology) when signals are degraded. Likewise, listeners predict and revise sentence meanings by learning from past experiences with cues, but it remains unknown how signal degradation impacts children, who have less linguistic knowledge from which to draw. In clear speech environments, children predict sentence meanings based on early-arriving words, but encounter difficulties revising incorrect predictions (Trueswell *et al.*, 1999; Lidz *et al.*, 2017; Huang *et al.*, 2013; Omaki *et al.*, 2014; Qi *et al.*, 2020; Kidd *et al.*, 2011). To identify how children's syntactic parsing is affected by signal degradation, the current study compares the comprehension of actives and passives in natural and vocoded speech. Since children interpret sentences to learn and communicate, understanding the impacts of signal quality will advance our knowledge of speech comprehension and promote interventions for children with congenital deafness who receive cochlear implants (CIs) to provide partial access to sound (e.g., Leibold, 2017; Iglehart, 2016; von Koss Torkildsen *et al.*, 2019).

---

[a] ORCID: 0000-0003-2662-3057.
[b] Electronic mail: ythuang1@umd.edu

A. Effects of signal degradation on children's sentence comprehension

Research on signal-degradation effects on children's speech perception has focused on phoneme identification (e.g., Nozza *et al.*, 1991; Polka *et al.*, 2008) and word recognition (e.g., Leibold *et al.*, 2013; Newman, 2009), with far less work examining sentence comprehension. Studies have manipulated signal degradation by varying levels of background noise (e.g., signal-to-noise ratio) or amount of reverberation, and measured accuracy in tasks involving picture matching, instruction completion, or sentence repetition. Negative impacts of background noise are greater in sentence comprehension than word recognition (Klatte *et al.*, 2010; McCreery *et al.*, 2017; Schiller *et al.*, 2020), and in younger compared to older children (Klatte *et al.*, 2010; Youngdahl *et al.*, 2018; Wightman and Kistler, 2005; Buss *et al.*, 2018). Relative to adults, children are more susceptible to reverberation, and require greater signal-to-noise ratios to achieve similar comprehension accuracy (Valente *et al.*, 2012; Klatte *et al.*, 2010; Neuman *et al.*, 2010; Wróblewski *et al.*, 2012; Iglehart, 2016).

An extreme example of signal degradation can be found in children with CIs, who not only experience a degraded signal now, but also have never previously heard an intact acoustic signal and thus, may have developed weaker or different linguistic representations. To understand the impacts of signal degradation on sentence comprehension within a subject, past work presented normal-hearing (NH) children with simulations of CI speech (Eisenberg *et al.*, 2000) through a channel vocoder (Shannon *et al.*, 1995). In sentence-recall tasks, 3- to 11-year-olds required more channels to achieve similar comprehension levels as adolescents and adults (Dorman *et al.*, 2000; Eisenberg *et al.*, 2000; Waked *et al.*, 2017; Huyck, 2018). In eye-tracking measures, children as young as 2 years recognize familiar words in sentences with eight-channel vocoded speech (Newman and Chatterjee, 2013; Newman *et al.*, 2015), but have difficulty learning new words in similar contexts (Newman *et al.*, 2020). This suggests that the intelligibility of vocoded speech varies with age, task, and level of degradation.

Recent work suggests that linguistic knowledge may support sentence comprehension in adverse listening environments. Across different types of background noise (e.g., speech-shaped noise, amplitude-modulated noise, two-talker babble), 5- to 6-year-olds with larger vocabulary sizes achieve lower speech-recognition thresholds for natural speech, defined as the signal-to-noise ratio associated with 50% comprehension accuracy (McCreery *et al.*, 2020). Negative impacts of background noise are smaller when the syntactic and semantic relations of words in sentences scaffold interpretation (e.g., better performance for sentences like *The jaws giggle at the frosty tractor* compared to word lists like *Ghost four smart tooth*) (McCreery *et al.*, 2017; Prodi *et al.*, 2019; O'Neill *et al.*, 2019). However, children's ability to leverage linguistic relations also varies with sentence complexity and listener age (Başkent *et al.*, 2013). In a picture-matching task, all children accurately interpret simple reflexive sentences with vocoded speech (e.g., one-participant events like *The penguin is hitting himself with a pan*), but 5- and 8-year-olds experience greater difficulty compared to 10- and 11-year-olds with more complex sentences that required pronoun interpretation (e.g., two-participant events like *The penguin is hitting him with a pan*).

Together, this suggests that syntactic information in sentences may facilitate children's interpretation in adverse listening environments (i.e., background noise, reverberation, hearing loss, CI processing), but parsing procedures may be subject to immature linguistic knowledge that introduces greater susceptibility to signal degradation. However, since past studies often compare stimuli with highly disparate properties (e.g., implausible sentences, word lists), it remains unclear what cognitive processes enable children to benefit from syntactic information during sentence comprehension. Do degraded signals alter initial predictions of upcoming words, revision of incorrect interpretations, or both?

B. Effects of signal degradation on prediction and revision

To address this question, the current study compares how young children interpret active and passive sentences in natural and vocoded speech. In natural speech, children often generate agent-first predictions in sentences like (1), leading to more accurate interpretations of actives compared to passives (Bever, 1970; Huang *et al.*, 2017a; Huang and Hollister, 2019; Abbot-Smith *et al.*, 2017; Deen *et al.*, 2018; Messenger and Fisher, 2018). Importantly, children are also sensitive to early-arriving cues that allow them to avoid making agent-first predictions in sentence contexts like (2). Here, the argument order reverses, and NP1s are now familiar words (e.g., *the seal*) while NP2s are novel ones (e.g., *the blicket*). In past research (Huang and Arnold, 2016; Huang and Ovans, 2021), referring to familiar entities early in sentences decreases interpretive uncertainty, and this leads children to wait for more reliable cues to role assignment to arrive later on the main verb (e.g., *eating* or *eaten*). This delay in processing is beneficial for interpreting passives since children no longer need to revise incorrect predictions in order to assign roles. Under these circumstances, they accurately interpret both constructions.

(2) a. Active: The seal is eating the blicket. [infer that *the blicket* is the theme]
b. Passive: The seal is eaten by the blicket. [infer that *the blicket* is the agent]

Sentences (1) and (2) create well-matched contexts (e.g., identical words in different orders) where children probabilistically parse sentences based on early- and late-arriving cues in natural speech. This offers opportunities to examine how interpretive strategies may vary with signal degradation. In this study, we draw connections across three literatures, and consider the following hypotheses (Table I).

TABLE I. With natural speech, children make agent-first predictions after novel NP1s, but have difficulty revising interpretations after late-arriving cues. This leads to accurate comprehension of actives but not passives. After familiar NP1s, children do not make agent-first predictions, and this promotes use of late-arriving cues to distinguish actives and passives. This leads to accurate comprehension of actives and passives. Relative to natural speech, we test three possible effects of vocoded speech on sentence comprehension.

Hypotheses	Novel NP1 (e.g.,*"The blicket will be…"*)	Familiar NP1 (e.g.,*"The seal will be…"*)
Hypothesis 1. Signal degradation reduces predictions	Fewer agent-first predictions compared to natural speech; passive accuracy improves	Same as natural speech
Hypothesis 2. Signal degradation promotes predictions	Same as natural speech	More agent-first predictions compared to natural speech; passive accuracy worsens
Hypothesis 3. Signal degradation reduces revision	Less revision compared to natural speech; passive accuracy worsens	Same as natural speech

These accounts share common mechanisms (e.g., procedures for word recognition, cue-based syntactic parsing) and are not mutually exclusive. However, in order to make clearer connections between theories and study predictions, we focus on distinct facets of three prominent accounts.

1. Hypothesis 1: Slower processing leads to less prediction

Signal degradation decreases the speed and accuracy of word recognition of CI listeners relative to NH peers (Grieco-Calub *et al.*, 2009; Huang *et al.*, 2017b; McMurray *et al.*, 2017). Delayed rejection of lexical competitors may reflect a strategy for processing ambiguous signals, whereby listeners maintain early-arriving alternatives until decisive cues emerge later in words and sentences (i.e., "wait and see" approach) (see McMurray *et al.*, 2017; Conway *et al.*, 2014; Pisoni *et al.*, 2016). This hypothesis predicts that signal degradation may limit children's sensitivity to early-arriving cues in sentences and decrease agent-first predictions in contexts that previously prompted them [e.g., less likely to interpret novel NP1s as agents in sentence (1)]. Since delaying predictions would remove the need to revise initially incorrect interpretations for passives, signal degradation may lead to accurate comprehension across constructions, for all sentence contexts.

2. Hypothesis 2: Greater uncertainty leads to more prediction

NH adults often generate agent-first predictions for passives when sentences are implausible (e.g., misinterpret NP1 as the agent in *The girl was kicked by the ball*) (Ferreira, 2003; MacWhinney *et al.*, 1984; Ryskin *et al.*, 2018). This is consistent with Bayesian frameworks that describe how listeners combine expectations of what is likely to be said (prior) with what is actually said (likelihood). When faced with ambiguous signals, listeners maximize the probability of accurate interpretation by down-weighting signal properties and relying on typical parsing cues instead (Gibson *et al.*, 2013; Levy *et al.*, 2009; O'Neill *et al.*, 2019). Since adverse listening environments make signals more ambiguous, this hypothesis predicts that signal degradation may promote agent-first predictions in contexts that did not previously elicit them [e.g., more likely to interpret familiar NP1s as agents in sentence (2)]. This would require children to revise incorrect interpretations for passives, and worsen comprehension compared to actives, for all sentence contexts.

3. Hypothesis 3: Uncertainty impacts revision not prediction

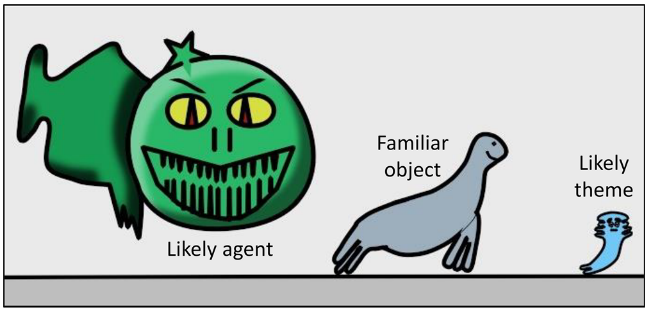
For NH children, revising incorrect predictions is related to language experience and knowledge. Across languages, they revise agent-first predictions more consistently when passive cues are more frequent (Huang *et al.*, 2013; Huang *et al.*, 2017a; Ehrenhofer *et al.*, 2018). Within English, language development increases the magnitude of error signals for prompting revision (Ovans *et al.*, 2020), and children with larger vocabulary sizes are more adept at resolving late-arriving conflicts than those with smaller vocabularies (Huang *et al.*, 2017a; Leech *et al.*, 2017; Huang and Holister, 2019). By analogy, signal degradation temporarily reduces children's ability to perceive relevant linguistic information in sentences (e.g., hearing the difference between *eating* versus *eaten by*), and this may result in poorer revision. If signal degradation reduces recognition of late-arriving cues that are necessary for revising agent-first predictions, this would worsen children's comprehension of passives, specifically in contexts where early-arriving cues promote predictions [e.g., more likely to interpret novel NP1s as agents in passive sentence (1b)].

C. The current study

To test these alternatives, we paired simulations of CI speech processing using a channel vocoder (Shannon *et al.*, 1995) with a word-interpretation and eye-tracking paradigm (Huang and Arnold, 2016). We measured NH 5-year-olds' assignment of agent versus theme roles to arguments as sentence cues unfolded in real time, and this allowed us to identify how signal degradation affected prediction and revision of meaning and altered final sentence interpretation. On each trial, children saw brief animations of two unfamiliar objects interacting with a familiar object (e.g., the seal). The likely agent (e.g., large monster-like predator) acted on the familiar object (e.g., chases a seal), and the familiar object acted on the likely theme (e.g., the seal chases a small, non-threatening prey). Next, children heard active and passive sentences that featured novel words, and their eye-movements were measured to the unfamiliar objects in the display (Fig. 1). After each sentence, children were asked to select the object that corresponded to the novel word (e.g., *Click on the blicket!*).

To comprehend sentences accurately, children must use late-arriving cues on verbs (e.g., *eating* in actives, *eaten by* in passives) to assign roles to familiar nouns (e.g., is *the seal* an agent or theme?), infer the roles for novel words (e.g., is *the blicket* an agent or theme?), and select plausible referents on this basis (e.g., which novel object is the seal likely to eat or is likely to eat the seal?). In novel-NP1 sentences like (1), active cues imply that familiar NP2s are themes, and so novel NP1s are likely agents. Conversely, passive cues imply that familiar nouns are agents, and so novel NP1s are likely themes. In familiar-NP1 sentences like (2), the order of novel and familiar words reverse, so active cues now imply that novel NP2s are likely themes while passive cues imply that they are likely agents. In addition to off-line judgments of novel-word meanings, children's eye-movements during the sentences provide a metric of how agent-first predictions are incorporated with late-arriving verbal cues in real time.

Our critical comparisons of interest were interpretations of late-arriving cues on passive sentences, and how they differ from baseline active sentences. If signal degradation reduces agent-first predictions (McMurray *et al.*, 2017; Conway *et al.*, 2014; Pisoni *et al.*, 2016), children who hear vocoded speech should consistently distinguish passives from actives in their eye-movements and actions (i.e., construction effects but no sentence-context effects). If, however, signal degradation promotes agent-first predictions (Gibson *et al.*, 2013; Levy *et al.*, 2009; O'Neill *et al.*, 2019), children who hear vocoded speech should consistently fail to distinguish passives from actives across contexts (i.e., no construction or sentence-context effect). Finally, if signal degradation hinders revision but not prediction (Huang and Ovans, 2021; Ovans *et al.*, 2020), children who hear vocoded speech should be less sensitive to late-arriving conflicts in novel-NP1 sentences, and this may lead to less differentiation of constructions compared to familiar-NP1 sentences (i.e., an interaction between construction and sentence context).



Sentence context	Construction	Sentence	Correct target
Novel NP1	Active	The blicket will be quickly eat***ing*** the seal	Likely agent
	Passive	The blicket will be quickly eat***en by*** the seal	Likely theme
Familiar NP1	Active	The seal will be quickly eat***ing*** the blicket	Likely theme
	Passive	The seal will be quickly eat***en by*** the blicket	Likely agent

FIG. 1. (Color online): In the word-interpretation task, a sample display featuring the familiar object (e.g., seal), likely agent (e.g., large, novel creature), and likely theme (e.g., small, novel creature) in critical trials. In sentences, the identity of the novel word is disambiguated at the late-arriving verb morphology, which is underlined in sentences. The correct target refers to the identity of the novel word based on disambiguating cues in the construction.

II. EXPERIMENT

A. Participants

Forty-seven children were recruited from the District of Columbia metropolitan area through partnerships with local private schools and the Infant and Child Studies Database at the University of Maryland. Six were excluded because of challenges with eye-tracking calibration, and one because of poor performance on the vocoded-speech training. The mean age of the remaining 40 participants was 5;4 (SD = 0;4, range = 4;10 to 5;11). This age range was chosen because children produce passives in spoken production but experience difficulties in comprehension (Huang *et al.*, 2017a; Huang and Arnold, 2016; Deen *et al.*, 2018). Current participants are drawn from a similar age range as children tested by Huang and Arnold (2016), who performed the same task using natural speech and served as the basis for comparison (M = 5;5, SD = 0;3, range = 5;0 to 5;11). Across speech conditions, there was no significant difference in age, and parental reports all indicated hearing within normal limits and English as the primary language.

B. Procedures and materials

Based on pilot work, we developed a training task to examine sentence comprehension when word-recognition challenges were reduced. The vocoded-speech training and word-interpretation task were presented in sequential order in a single session. Short breaks were included as needed. We adopted eight-channel vocoded speech since it is commonly used in studies (e.g., Dorman *et al.*, 2000; Başkent *et al.*, 2013), and is a level that matches the maximum benefit adult CI users routinely get from their devices (e.g., Friesen *et al.*, 2001; Goupell *et al.*, 2008) but see recent work suggesting larger benefits for some array types (Berg *et al.*, 2019, 2021; Croghan *et al.*, 2017). Eight-channel vocoded speech is also associated with successful word recognition in young children, although their impacts on interpreting novel words are more variable (Newman and Chatterjee, 2013; Newman *et al.*, 2015; Newman *et al.*, 2020).

Vocoded-speech stimuli were created by recording natural speech using a slow and fluent pace. Then these stimuli were passed through an eight-channel bandpass filterbank that had corner frequencies that were logarithmically spaced between 200 and 8 000 Hz. The filterbank utilized second-order forward-backward Butterworth bandpass filters. For each channel, envelopes were extracted using the Hilbert transform that was low-pass filtered with a third-order forward-backward Butterworth filter at 400 Hz. Envelopes were used to modulate narrowband noises with the same bandwidth as the analysis filters, and the modulated narrowband noises were combined into a single waveform. Vocoded-speech stimuli were normalized to have the same overall energy as natural speech.

1. Vocoded-speech training

The vocoded-speech training was divided into two blocks. During the familiarization block, children were told that they would hear sentences to get used to a robot's voice. Based on training in adults (Davis *et al.*, 2005), each sentence trial (e.g., *The boy is in the garden*) was presented first as vocoded speech, then natural speech, and vocoded speech again. Up to 12 familiarization sentences were presented in a fixed order, with approximately 2 s between natural- and vocoded-speech versions. Sentences were arranged in six pairs, with sentence length matched within a pair and increasing across pairs. To promote recognition, sentences were presented alongside related pictures (e.g., ocean scene), and included nouns that later appeared in the word-interpretation task (e.g., *the seal*). Trial presentation was dynamically adjusted based on responses to questions about pictures (e.g., *What do you see?*), recall probes after the vocoded sentence (e.g., *What did the robot say?*), and engagement questions (e.g., *Do you understand the robot?*). When responses suggested limited comprehension of the vocoded sentence, a new picture was presented and the second sentence in the pair was played. When they demonstrated understanding of sentences, children moved on to the next pair.

During the test block, children were presented with six new sentence pairs and pictures. On each trial, vocoded speech was played alongside a picture, and children were asked to repeat the sentence verbatim. Omissions or alterations of content words (e.g., *polar bear*, *standing*, *ice*) were considered incorrect while errors in function words (e.g., *the*, *a*) were not. This was followed by the natural-speech recording as feedback. Test sentences were arranged in six pairs, with increasing sentence length. If children repeated the sentence incorrectly, they heard the second sentence in a pair. If they correctly repeated the second sentence, they moved to the next sentence pair. Our dependent variable was the number of sentences needed to achieve accurate repetition, where fewer indicated higher accuracy. Overall, children performed well and were presented on average 7 out of 12 total sentences (SD = 1, range = 6–10). Only children who accurately repeated the final vocoded test sentences moved to the word-interpretation task. One child was excluded on this basis. The vocoded-speech training took approximately 10 min.

2. Word-interpretation task

During the word-interpretation task, children sat in front of a computer connected to an EyeLink 1000 desktop eye tracker (SR Research, Mississauga, Ontario, Canada). All speech stimuli were vocoded, including critical trial sentences, selection commands, and filler trials. At the start of the task, the experimenter explained that a robot would describe what is happening on the screen and would sometimes use silly words to talk about new creatures. Children were told to listen to sentences and select objects in the display. Each trial was divided into two parts. During the familiarization phase, children were introduced to a familiar object (e.g., seal), likely agent (e.g., large, scary-looking creature), and likely theme (e.g., small, non-threatening creature). A short animation showed the familiar object by itself (e.g., children heard *Look at the seal!*), followed by the likely agent chasing the familiar object, and the familiar object chasing the likely theme (i.e., children heard *Look at these!*). These scenes established role relationships between the familiar object, likely agent (e.g., something that can plausibly act on the familiar object), and likely theme (e.g., something that the familiar object can plausibly act on).

During the test phase, children saw a static image of the novel objects on either side of the familiar object and heard a sentence describing an action (Fig. 1). In critical trials, test sentences paired a familiar noun (e.g., *the seal*) with a two-syllable novel noun (e.g., *the blicket*; see Table II), similar to (1) and (2). After each test sentence (e.g., *The blicket will be quickly eating the seal)*, children were asked to select the object corresponding to the novel word (e.g., *Click on the blicket!*). Children then used the computer mouse to select either the familiar object (e.g., seal), likely agent (e.g., large creature), or likely theme (e.g., small creature). Once they did so, the trial ended, and the next trial began. Children heard each novel word twice during the test phase (i.e., once in the test sentence, again in the selection command). Real-time sentence interpretation was measured by continuously sampling fixations to display locations from the start of the test sentence to the start of object selection. Final interpretation of novel words was measured by children's mouse clicks to a novel object after the selection command.

TABLE II. The critical trials of the word-interpretation task presented familiar nouns and novel words together with transitive verbs in active and passive sentences (e.g., *The blicket will be quickly eating the seal* or *The seal will be quickly eating the blicket*).

Familiar nouns	Novel words	Transitive verbs
Seal	Blicket	Eat
Cat	Nedoke	Scare
Dog	Coopa	Chase
Boy	Hantil	Kick
Rabbit	Leepo	Eat
Frog	Daylon	Catch
Rock	Tayvak	Smash
Girl	Chowvag	Lift
Mouse	Vaychip	Grab
Car	Noytoff	Squish
Fox	Bellwer	Chase
Monkey	Furpin	Scare

Speech type was manipulated between subjects. Children in the current study heard vocoded speech during the word-interpretation task, and their performance was compared to another group of children, who heard natural speech. Sentence context was also manipulated between subjects, with half of the children randomly assigned to the novel-NP1 condition, and the other half to the familiar-NP1 condition. Sentence construction (i.e., active or passive) was manipulated within subjects and varied randomly across trials. Children saw 12 critical items, and each item was a unique novel word and display. We adopted a Latin Square design (e.g., Hinkelmann and Kempthorne, 2008), whereby four versions of each critical item (represented by the sentences shown in Fig. 1) were distributed across four presentation lists, and each item appeared once in every list. Each list contained sentences from one context (i.e., novel NP1 or familiar NP1) and contained six items in each construction (i.e., active or passive). In each list, six filler trials were included to provide opportunities for successful interpretation and maintain children's engagement with the task. Filler sentences included familiar nouns in active sentences (e.g., *The sheep will be slowly eating the grass. Click on the sheep!*), and were paired with new displays (e.g., picture of a sheep, grass, wolf). Children heard a total of 18 sentences. The word-interpretation task took approximately 20 min.

C. Coding

Approximately 2.1% of trials in the word-interpretation task were excluded from analyses because of experimenter error or equipment track loss. The remaining data were coded in the following manner. First, eye-movements were coded as looks to one of three objects (i.e., familiar object, likely agent, likely theme) or missing due to looks away from these interest areas (e.g., other display areas, blinking). Missing looks accounted for 17.8% of fixations. Second, object selection after the instruction was coded as actions to likely agents or likely themes and recoded based on accuracy for the trial condition. *Correct actions* used verb morphology to select the target novel object (Fig. 1). For novel-NP1 sentences, active cues indicated that novel words were agents while passive cues indicated they were themes. This reversed for familiar-NP1 sentences, such that correct actions involved likely themes for active cues and likely agents for passive cues. *Incorrect actions* involved selecting the other novel object. Actions to familiar objects accounted for 3.1% of trials and were not analyzed.

III. RESULTS

We compared current performance with vocoded speech to previous effects with natural speech (Huang and Arnold, 2016) in three ways. First, we examined eye movements during critical sentences to assess how children used active and passive cues to distinguish referents as sentences unfolded. Next, we examined actions after sentences were finished to assess the accuracy of object selection and sensitivity to early- and late-arriving sentence cues. Finally, we used Bayes Factor analysis to evaluate the strength of evidence for hypotheses about how signal-degradation effects vary across constructions and sentence contexts.

Eye-movements and actions were analyzed using linear or logistic mixed-effects models, depending on whether the data were continuous or categorical. Our analytical strategy was based on best practices outlined in Barr *et al.* (2013), and included fixed and random effects that were justified by the study design. Separate analyses were conducted for sentence context (novel-NP1 versus familiar-NP1) where construction (active versus passive) and speech type (natural versus vocoded) were included as fixed-effects variables. Within fixed-effects levels, deviation coding compared condition means to the grand mean. Subjects and items were simultaneously modeled as random-effects variables, with random slopes and intercepts for subjects and items. If full models failed to converge, simpler ones were adopted with random intercepts only. To test for possible age effects, follow-up analyses used likelihood ratios to compare models with and without age (in months) as a fixed effect. Since age never improved model fit (all p-values >0.40), it was omitted from our final report. Analyses were implemented through the lme4 software package 1.1–27 in R (Bates *et al.*, 2014).[1]

A. Eye-movements during sentences

To assess online sensitivity to sentence cues, we focused on fixations during a 1150 ms window from the disambiguating verb morphology (e.g., *-ing* in *eating* versus *-en* in *eaten*) to sentence offset. Based on Huang *et al.* (2013), time windows were shifted by 400 s to account for the time it takes children to generate saccadic eye movement during syntactic-parsing tasks. Across natural and vocoded speech, fixations to familiar objects were greater in novel-NP1 sentences compared to familiar-NP1 sentences (Fig. 2). This pattern was expected since familiar objects were mentioned after disambiguation in the novel-NP1 condition [e.g., *…(by) the seal*] but not in the familiar-NP1 condition [e.g., *…(by) the blicket*]. Eye-movements in the familiar-NP1 condition revealed rapid convergence to correct referents for natural and vocoded speech. After disambiguation, there was an expected preference for likely themes in active sentences (2a) and likely agents in passive sentences (2b). In the novel-NP1 condition, fixations correctly shifted to likely agents in active sentences (1a), but remained equivocal in passive sentences (1b). This was true for both natural and vocoded speech.

To compare fixations across conditions, we focused on looks to the novel objects, which accounted for 72.0% of sampled fixations. We calculated a preference score that tracked fixations after hearing active or passive cues. This measure allowed us to account for children's visual preferences by comparing fixations to the same object for both active and passive sentences. For the novel-NP1 condition, we subtracted average looks to the likely themes minus likely agents. For the familiar-NP1 condition, we subtracted

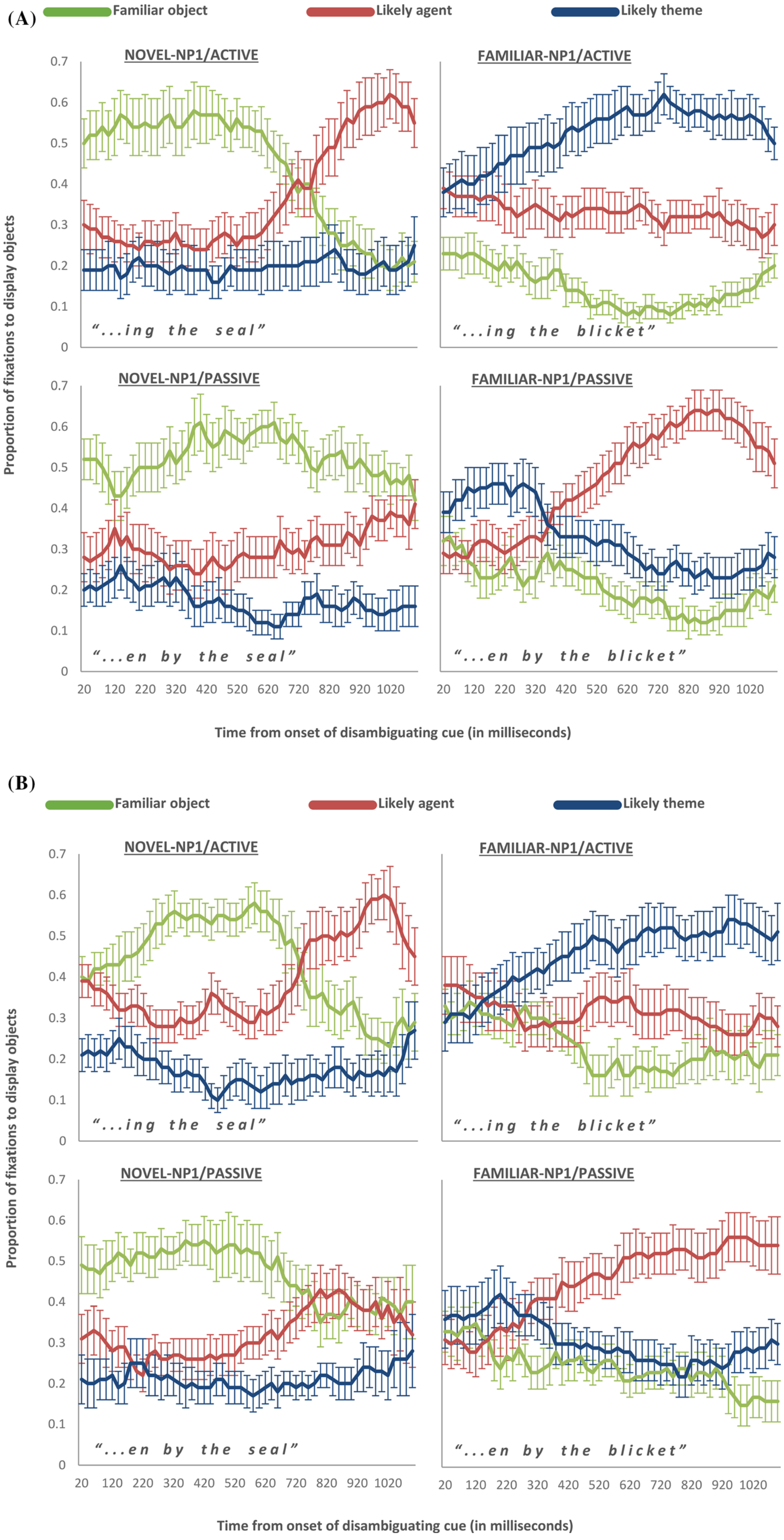


FIG. 2. (Color online): Eye-movements after the disambiguating sentence cue in the natural (panel A) and vocoded speech (panel B) conditions. Correct referents are likely agents in the novel-NP1/active condition (upper left of each panel A and B), likely themes in the novel-NP1/passive condition (lower left of each panel A and B), likely themes in the familiar-NP1/active condition (upper right of each panel A and B), and likely agents in the familiar-NP1/passive condition (lower right of each panel A and B).

average looks to the likely agents minus likely themes. Across conditions, positive values indicated greater sensitivity to passive cues, and negative values indicated greater sensitivity to active cues. Thus, differences in preference scores across passives versus actives track the extent to which children used sentence cues to distinguish constructions and inferred different referents for novel words. Preference scores were analyzed using linear mixed-effects

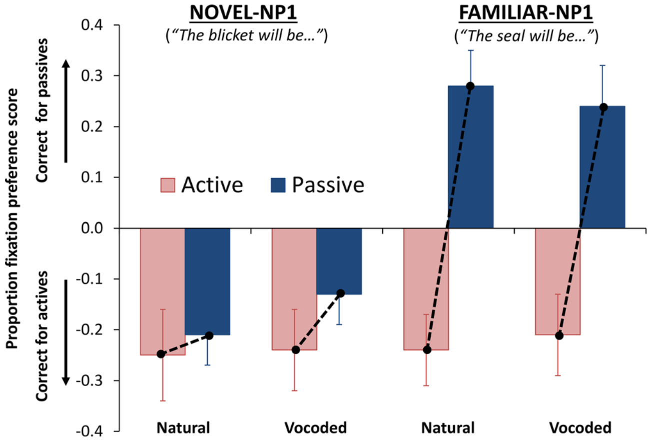


FIG. 3. (Color online): Preference scores of eye-movements from the disambiguating sentence cue to sentence offset. Positive scores indicate correct fixations in passive trials and negative scores indicate correct fixations in active trials. Dashed lines highlight the fixation differences that result from hearing active and passive sentences. Steeper lines indicate greater sensitivity to disambiguating sentence cues.

models, and parameter-specific p-values were estimated through Satterthwaite approximation (Luke, 2017).

In the novel-NP1 condition, Fig. 3 illustrates that children preferred to look at likely agents irrespective of whether late-arriving cues implied active (red bars) or passive sentences (blue bars). This led to negative scores across construction (active versus passive) and speech type (natural versus vocoded), which corresponded to correct fixations for actives but incorrect fixations for passives. While scores were numerically higher for passives compared to actives, statistical comparisons revealed no reliable effects of construction [$\chi^2(1, N = 40) = 2.17, p > 0.10$] or speech type [$\chi^2(1, N = 40) = 0.37, p > 0.50$]. This suggests that children initially interpreted all novel-NP1 sentences based on agent-first predictions. In contrast, in the familiar-NP1 condition, children predicted correct referents for active (red bars, leading to negative scores) and passive sentences (blue bars, leading to positive scores), and in natural and vocoded speech (Table III). This led to a main effect of construction [$\chi^2(1, N = 40) = 48.17, p < 0.001$] and no effect of speech type [$\chi^2(1, N = 40) = 0.01, p > 0.90$].

Together, children's fixations reveal agent-first predictions after novel NP1s but not familiar NP1s. When children committed to agent-first predictions, they were less sensitive to late-arriving cues that distinguished active and passive constructions (e.g., *eating* versus *eaten by*), across both natural and vocoded speech. In contrast, when agent-first predictions were absent, children readily used the *same* late-arriving cues to fixate on appropriate referents for actives and passives. This was true for both natural and vocoded speech.

TABLE III. In eye-movement analyses, fixed effects (construction × speech type) in a linear mixed-effects regression model for the preference scores.

	Novel-NP1				Familiar-NP1			
Fixed effects	β	SE	t	p	β	SE	t	p
Intercept	0.22	0.04	5.30	0.01*	0.02	0.07	0.25	0.81
Construction	0.04	0.03	1.51	0.13	0.25	0.03	7.13	0.01*
Speech	0.03	0.04	0.61	0.55	0.01	0.04	0.05	0.96
Construction × Speech	0.01	0.03	0.47	0.64	0.01	0.03	0.21	0.83

B. Actions after sentences

To examine how signal degradation impacted final interpretation of novel words, we first focused on the accuracy of object selection after sentences. Trial-level accuracy was binary, and was analyzed using logistic mixed-effects models (Jaeger, 2008). Figure 4(A) illustrates consistent performance across speech type. In the novel-NP1 condition, children accurately inferred novel words in actives (red bars), but were less accurate with passives that required revision (blue bars). This led to a main effect of construction [$\chi^2(1, N = 40) = 119.54, p < 0.001$] and no effect of speech type [$\chi^2(1, N = 40) = 1.68, p > 0.15$] [Table IV(A)]. This

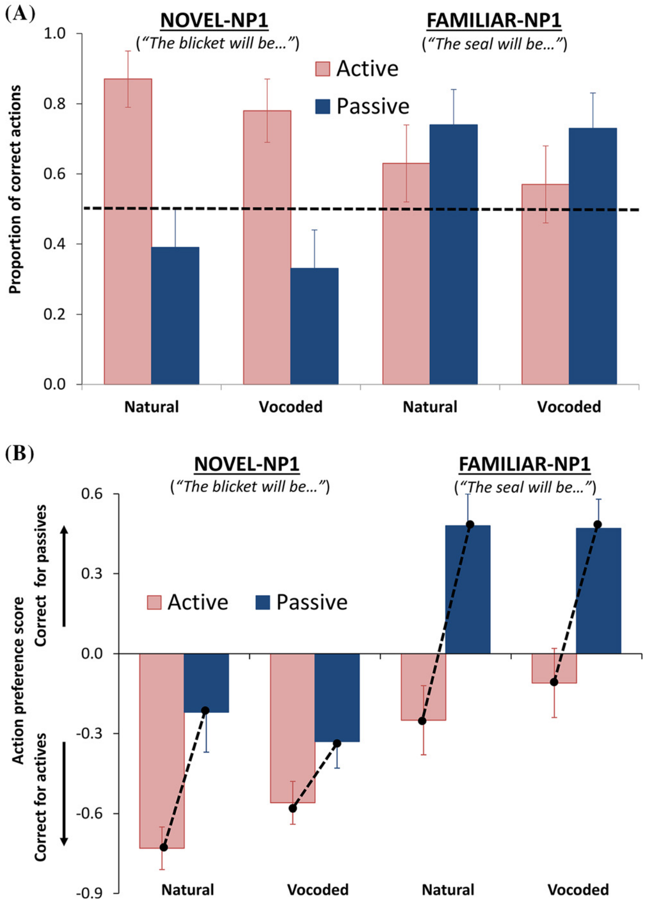


FIG. 4. (Color online): For actions after sentence completion, responses coded based on accuracy and preference scores. For accuracy (panel A), higher values correspond to more correct actions. Since there are two novel objects, the dash line highlights chance performance in interpreting the meaning of the novel word. For preference score (panel B), positive values correspond to correct actions in passive trials and negative scores correspond to correct actions in active trials. Dashed lines highlight the differences in fixations that result from hearing active and passive sentences. Steeper lines indicate greater sensitivity to disambiguating cues.

pattern reversed in the familiar-NP1 condition, such that children were unexpectedly *more* accurate for passives (blue bars) compared to actives (red bars). This unexpected result led to a main effect of construction [$\chi^2(1, N = 40) = 10.86$, $p < 0.001$] and no effect of speech type [$\chi^2(1, N = 40) = 0.33$, $p > 0.50$]. It suggests that object selections were not only affected by sentence properties (i.e., hearing active versus passive cues) but also by visual displays (i.e., seeing a large versus small creature). Since likely agents were more visually interesting than likely themes, they may have induced a preference that inflated the accuracy of passives in the familiar-NP1 condition and the accuracy of actives in the novel-NP1 condition.

To assess final interpretations using a more neutral measure, we returned to preference scores from the eye-movement analyses. Since these scores track the same novel objects across constructions (e.g., likelihood of selecting likely agents for actives and passives), they are useful for comparing how sentence cues distinguish constructions and whether this varies with speech type. Figure 4(B) illustrates that similar to eye-movement analyses, preference scores in the novel-NP1 condition were consistently negative across constructions and speech type, providing additional evidence that actives (red bars) were often misinterpreted as passives (blue bars). However, unlike eye-movement patterns, children's actions correctly distinguished passives from actives in natural ($\beta = 0.52$, $SE = 0.10$, $t = 5.24$, $p < 0.001$) and vocoded speech ($\beta = 0.23$, $SE = 0.11$, $t = 2.05$, $p < 0.05$). This difference was larger for natural compared to vocoded speech, leading a main effect of construction [$\chi^2(1, N = 40) = 24.40$, $p < 0.001$] and trending towards a significant interaction between speech type and construction [$\chi^2(1, N = 40) = 3.55$, $p < 0.06$] [Table IV(B)]. In the familiar-NP1 condition, preference scores were appropriately negative for actives (red bars) and positive for passives (blue bars), in natural and vocoded speech. This led to a main effect of construction [$\chi^2(1, N = 40) = 66.13$, $p < 0.001$], with no effect of speech type [$\chi^2(1, N = 40) = 0.13$, $p > 0.70$].

TABLE IV. In action analyses, fixed effects (construction × speech type) in (A) a logistic mixed-effects regression model of accuracy and (B) a linear mixed-effects regression model for the preference score.

(A) Accuracy

	Novel-NP1				Familiar-NP1			
Fixed effects	β	SE	z	p	β	SE	z	p
Intercept	0.57	0.18	3.16	0.01*	0.77	0.15	5.19	0.01*
Construction	1.21	0.13	9.49	0.01*	0.34	0.10	3.27	0.01*
Speech	0.24	0.17	1.42	0.16	0.07	0.14	0.55	0.58
Construction × speech	0.11	0.12	0.94	0.35	0.05	0.10	0.45	0.65

(B) Preference score

	Novel-NP1				Familiar-NP1			
Fixed effects	β	SE	t	p	β	SE	t	p
Intercept	0.46	0.08	6.06	0.01*	0.14	0.08	1.65	0.11
Construction	0.19	0.04	5.00	0.01*	0.33	0.04	8.44	0.01*
Speech	0.01	0.05	0.26	0.80	0.02	0.06	0.37	0.71
Construction × speech	0.07	0.04	1.89	0.06*	0.03	0.04	0.82	0.41

Taken together, children's actions demonstrate that they made agent-first predictions after novel NP1s but not familiar NP1s, and this was true across natural and vocoded speech. When children generated agent-first predictions, they were less sensitive to late-arriving passive cues compared to active cues, particularly when speech stimuli were vocoded. However, when agent-first predictions were absent, children consistently used active and passive cues (e.g., *eating* versus *eaten by*) to correctly infer novel-word meanings, suggesting that perceiving late-arriving cues was not entirely limited by signal degradation.

C. Bayes factor analysis of actions

The absence of widespread effects of signal degradation on eye-movements and actions raises questions about whether children's comprehension is fairly robust to signal quality or whether the current study lacked power to detect sizable effects. The latter is difficult to resolve methodologically, given restrictions on in-person data collection due to Covid-19. The former highlights the need for analytical strategies that quantify the strength of evidence for disparate hypotheses. We turned to Bayesian inference, which has gained ground as a viable alternative to null-hypothesis significance testing (Jarosz and Wiley, 2014; Lee and Wagenmakers, 2014; van Doorn *et al.*, 2020). Unlike traditional p-values, this approach calculates the relative odds of two hypotheses by estimating beliefs about the hypotheses before the current data (priors) and updates to beliefs after the data (likelihoods). Recall that unlike Hypotheses 1 and 2, Hypothesis 3 specifically predicts that signal-degradation effects would interact with sentence context (Table I). Using the BayesFactor package 0.9.2 in R (Morey and Rouder, 2011), we analyzed preference scores from actions and calculated the relative odds of a hypothesis that sentence context (familiar-NP1 versus novel-NP1) affects comprehension (alternative) compared to a hypothesis where sentence context does not matter (null). Comparisons of Bayes Factors across construction (actives versus passives) reveals how sentence context affects sensitivity to late-arriving cues and how this process changes with speech properties (natural versus vocoded).

For active sentences, sentence-context effects on preference scores were similar across signal quality. In natural speech, the current data were 14.3 times more likely to occur under a model that included sentence context compared with one without it. In vocoded speech, these data were 13.4 times more likely in a model that included context compared to without. Based on guidelines for interpreting Bayes factors (Jeffreys, 1961; van Doorn *et al.*, 2020), these values provide strong evidence that interpreting active cues is affected by the extent to which sentence contexts promote agent-first predictions. Signal quality does not appear to alter this relationship. In contrast, estimates of Bayes factor for passives revealed divergent effects of sentence context across signal properties. In natural speech, the current data were 26.2 times more likely to occur under a model that included sentence context compared with one without it. In vocoded speech, however, these data were 5,944.6 times more likely in a model that included context compared to without. Together, this provides evidence that interpreting passives is influenced by the presence of late-arriving conflicts, and this effect increases in the presence of signal degradation.

IV. GENERAL DISCUSSION

To interpret sentences, children must apply their limited linguistic knowledge to predict and revise meanings based on rapidly unfolding speech and in noisy real-world environments. This study examined how one type of signal degradation (eight-channel vocoding) impacts comprehension processes by comparing real-time and final interpretations of active and passive sentences. Do degraded signals alter initial predictions of upcoming words, subsequent revision of incorrect interpretations, or both? Our study indicates that signal degradation decreases recognition of late-arriving cues and reduces sentence revision (Huang and Ovans, 2021; Ovans *et al.*, 2020). When early-arriving cues promoted agent-first predictions (e.g., novel NP1s like *The blicket*), on-line interpretations were less sensitive to late-arriving conflicts (e.g., passive cues like *eaten by*) (see Fig. 3). Bayes factor analyses suggest that signal degradation enhances these challenges by decreasing sensitivity to passive cues [see Fig. 4(b)]. However, when early-arriving cues reduced agent-first predictions (e.g., familiar NP1s like *The seal*), children readily interpreted late-arriving active and passive cues, with no signal-degradation effects.

If vocoded speech instead affected processing of early-arriving cues (Hypotheses 1 and 2; Table I), we expected to see evidence of this in earlier metrics such as eye-movements. For example, if vocoded speech reduced agent-first predictions, we expected greater differentiation of actives and passives in the novel-NP1 condition, similar to what emerged in the familiar-NP1 condition. If vocoded speech instead enhanced agent-first predictions in the familiar-NP1 condition, this may have resulted in fixation preferences that mirror the novel-NP1 condition and diverge from the natural speech condition. Instead, what we observe is no effect of signal degradation on eye-movements (Fig. 4). Our findings indicate that children's predictions were largely guided by sentence context (Hypothesis 3; Table I), leading to less differentiation of actives and passives when revision was necessary and more when role assignment relied on late-arriving cues. This suggests that initial predictions were influenced by parsing of early-arriving cues (e.g., whether NP1s are novel or familiar), which was not affected by signal degradation.

These results mirror prior work in clear speech environments, demonstrating that children interpret sentences on a word-by-word basis, but face difficulties revising incorrect predictions (Trueswell *et al.*, 1999; Lidz *et al.*, 2017; Huang *et al.*, 2013; Omaki *et al.*, 2014; Qi *et al.*, 2020; Kidd *et al.*, 2011). Likewise, we show that children adapt to signal degradation (Newman and Chatterjee, 2013; Newman *et al.*, 2015; Başkent *et al.*, 2013), and sentence interpretations arise from interactions between signal properties and interpretive processes (Huang *et al.*, 2017a; Huang and Ovans, 2021; Ovans *et al.*, 2020). However, it remains unknown the extent to which current patterns generalize to other forms of signal degradation, such as four-channel vocoded speech, background noise, or reverberation. For example, previous research demonstrates that sentence-context effects on children's pronoun interpretation are larger with eight- compared to four-channel vocoded speech signal (Başkent *et al.*, 2013), presumably because higher signal quality enables more faithful extraction of linguistic information. While a definitive test of this hypothesis awaits future research, our study suggests the need for finer-grained descriptions of parsing strategies that are based on specific tests of how linguistic cues are instantiated in the signal and what these cues are used for in sentence interpretation.

Children's overall aptitude with degraded signals in the current study raises questions about the causes of difficulty in previous studies (e.g., Klatte *et al.*, 2010; Buss *et al.*, 2018). We offer two suggestions for this apparent discrepancy. First, unlike prior work with environmental noise, our vocoded-speech training provided children with relevant experiences for adapting to degraded signals. This type of experience is useful for adults (Davis *et al.*, 2005; Kleinschmidt and Jaeger, 2015), but it may be more important for listeners with still-developing linguistic representations (Pajak *et al.*, 2016; Feldman *et al.*, 2013). Second, our word-interpretation task provided visual (e.g., referential display) and linguistic cues (e.g., syntactic structure, lexical semantics) to sentence interpretation, which may allow children to recruit past experiences to interpret current signals. These benefits offer a more nuanced understanding of developmental challenges with signal degradation. Rather than having immature systems for attenuating noise that lead to consistent effects across all losses of fidelity to the auditory signal (Leibold, 2017; Youngdahl *et al.*, 2018; Leibold and Buss, 2019), children may adopt a rational strategy to listen to the full signal to learn fine-grained phonetic categories. Critically, when properties of communicative contexts constrain likely meanings (e.g., visual, linguistic cues), children will combine this information with signal properties to generate appropriate sentence interpretations, even under adverse listening environments.

Our findings suggest that syntactic parsing may be a fundamental process for extracting sentence meanings, particularly when speech co-occurs with environmental noise and linguistic knowledge is limited during development. Under these circumstances, syntactic parsing is useful because it relies on information about sentence position (e.g., agent-first predictions), which does not require accessing detailed phonetic forms for accurate word recognition

(Grieco-Calub *et al.*, 2009; Huang *et al.*, 2017a; McMurray *et al.*, 2017). Moreover, while word recognition requires learning from a limited set of input (e.g., only sentences containing *cat* will be useful for learning about the word *cat*), syntactic predictions exploit regularities between word order (e.g., sentence-initial arguments), grammatical representations (e.g., subject), and role assignment (e.g., agents) across *all* sentences in parental input (MacWhinney *et al.*, 1984; Chang *et al.*, 2006). These broad-scale relations may generate stable sentence interpretations and provide the evidence base for acquiring fine-grained properties of individual words (e.g., phonetic forms, lexical biases) (Huang and Ovans, 2021). Support to this comes from children with developmental language disorders, whose difficulties with statistical learning lead to fewer agent-first predictions and more comprehension errors compared to typically developing peers (Oppenheimer *et al.*, 2020; Montgomery *et al.*, 2018). Future research may take an individual-differences approach to examining interactions between syntactic parsing, signal degradation, and linguistic knowledge (e.g., Huang *et al.*, 2017a; McCreery *et al.*, 2020; Blomquist *et al.*, 2021).

Finally, our findings raise questions about how sentence comprehension unfolds in children with congenital deafness or early-onset hearing loss who have CIs, who only have access to degraded signals in their input. While there are admittedly many differences between children with CIs and NH peers, recent work suggests some shared comprehension strategies. Similar to NH children, children with CIs leverage early-arriving word meanings (e.g., *He eats…*) to anticipate late-arriving words in sentences (e.g.,*…the sandwich*) (Holt *et al.*, 2016; Nittrouer *et al.*, 2015; Blomquist *et al.*, 2021). They generate agent-first predictions, which lead to more accurate comprehension of subject *wh*-questions (e.g., *Which bear is catching the dog?*) compared to object *wh*-questions (e.g., *Which bear is the dog catching?*) (Schouwenaars *et al.*, 2019). However, children with CIs are also unique in notable ways. Compared to vocabulary-matched NH peers, they demonstrate weaker lexical-semantic predictions (Blomquist *et al.*, 2021), suggesting that long-term effects of signal degradation persist among listeners with comparable linguistic knowledge. Compared to aged-matched NH peers, children with CIs are less likely to assign roles using acoustically subtle morphological cues (Schouwenaars *et al.*, 2019), suggesting that long-term exposure to signal degradation may affect general parsing strategies. Future work testing the generalizability of these effects will lead to more detailed models of interactions between signal properties, real-time comprehension, and year-to-year development, as well as offer a roadmap for how to maximize signal use in children from different speech backgrounds.

ACKNOWLEDGMENTS

This work has benefited from insightful conversations with Rochelle Newman, Jan Edwards, Jared Novick, Zoe Ovans, Kathleen Oppenheimer, and Christina Blomquist. We also thank Kathy Dow-Burger, Alison Arnold, Elinora Leonard, Allison Granados, and Allesondra Sanchez for their help with recruitment and data collection. This study would not have been possible without the staff and families at Bethesda Montessori School, Celebration Christian Academy, Child's Garden Center, Crestview Montessori School, Greenwood School, Paint Branch Montessori School, and the Center for Young Children. This work was supported by NSF-BCS Grant No. BCS-1844194 (Huang) and NIH-NIDCD Grant No. R01DC014948 (Goupell). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Portions of this work have been presented at the 42nd Boston University Conference on Language Development.

---

[1] See https://osf.io/nyte7/ for data and analysis code.

Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., and Pine, J. (**2017**). "Two and three year old children use an incremental first-NP-as-agent bias to process transitive sentences," PloS ONE **12**(10), e0186129–20.

Barr, D. J., Levy, R., Scheepers, C., and Tilly, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**(3), 255–278.

Başkent, D., Van Rij, J., Ng, Z. Y., Free, R., and Hendriks, P. (**2013**). "Perception of spectrally degraded reflexives and pronouns by children," J. Acoust. Soc. Am. **134**(5), 3844–3852.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2014**). "Fitting linear mixed-effects models using lme4," arXiv preprint arXiv:1406.5823.

Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., and Gifford, R. H. (**2019**). "Speech recognition as a function of the number of channels in perimodiolar electrode recipients," J. Acoust. Soc. Am. **145**, 1556–1564.

Berg, K. A., Noble, J. H., Dawant, B. M., Dwyer, R. T., Labadie, R. F., and Gifford, R. H. (**2021**). "Speech recognition as a function of the number of channels for an array with large inter-electrode distances," J. Acoust. Soc. Am. **149**, 2752–2763.

Bever, T. G. (**1970**). "The cognitive basis for linguistic structures," in *Cognition and the Development of Language*, edited by J. R. Hayes (John Wiley, New York), pp. 279–362.

Blomquist, C., Newman, R., Huang, Y., and Edwards, J. (**2021**). "Children with cochlear implants use semantic prediction to facilitate spoken word recognition," J. Speech Lang. Hear. Res. **64**, 1636–1649.

Buss, E., Leibold, L. J., and Lorenzi, C. (**2018**). "Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers," J. Acoust. Soc. Am. **143**, 1458–1466.

Chang, F., Dell, G. S., and Bock, K. (**2006**). "Becoming syntactic," Psychol. Rev. **113**(2), 234–272.

Conway, C. M., Deocampo, J. A., Walk, A. M., Anaya, E. M., and Pisoni, D. B. (**2014**). "Deaf children with cochlear implants do not appear to use sentence context to help recognize spoken words," J. Speech Lang, Hear. Res. **57**(6), 2174–2190.

Croghan, N. B. H., Duran, S. I., and Smith, Z. M. (**2017**). "Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility," J. Acoust. Soc. Am. **142**, EL537–EL543.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**(2), 222–241.

Deen, K., Bondoc, I., Camp, A., Estioca, S., Hwang, H., Shin, G. H., Takahashi, M., Zenker, F., and Zhong, J. (**2018**). "Repetition brings success: Revealing knowledge of the passive voice," in *Proceedings of the 42nd Annual Boston University Conference on Language Development*, Boston, Massachusetts, pp. 200–213.

Dorman, M. F., Loizou, P. C., Kemp, L. L., and Kirk, K. I. (**2000**). "Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants," Ear Hear. **21**(6), 590–596.

Ehrenhofer, L., Yatsushiro, K., Fritzsche, T., Höhle, B., Lidz, J., Phillips, C., and Huang, Y. (**2018**). "Verbs, not subjects, drive subject-as-agent misinterpretation in children's comprehension of passives," [Poster presentation] in *43rd Boston University Conference on Language Development*, 2–4 November 2018, Boston, MA.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **23**(5), 2704–2710.

Feldman, N. H., Griffiths, T. L., Goldwater, S., and Morgan, J. L. (**2013**). "A role for the developing lexicon in phonetic category acquisition," Psychol. Rev. **120**(4), 751–778.

Ferreira, F. (**2003**). "The misinterpretation of noncanonical sentences," Cogn. Psychol. **47**, 164–203.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**(2), 1150–1163.

Gibson, E., Bergen, L., and Piantadosi, S. T. (**2013**). "Rational integration of noisy evidence and prior semantic expectations in sentence interpretation," Proc. Natl. Acad. Sci. U.S.A. **110**(20), 8051–8056.

Goupell, M. J., Laback, B., Majdak, P., and Baumgartner, W. D. (**2008**). "Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation," J. Acoust. Soc. Am. **123**, 2295–2309.

Grieco-Calub, T. M., Saffran, J. R., and Litovsky, R. Y. (**2009**). "Spoken word recognition in toddlers who use cochlear implants," J. Speech Lang. Hear. Res. **52**(6), 1390–1400.

Hinkelmann, K., and Kempthorne, O. (**2008**). *Design and Analysis of Experiments. I and II*, 2nd ed. (John Wiley & Sons, Hoboken, New Jersey).

Holt, C. M., Demuth, K., and Yuen, I. (**2016**). "The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants," Ear Hear. **37**(4), e256–e262.

Huang, Y. T., and Arnold, A. R. (**2016**). "Word learning in linguistic context: Processing and memory effects," Cognition **156**, 71–87.

Huang, Y. T., and Hollister, E. (**2019**). "Developmental parsing and linguistic knowledge: Reexamining the role of cognitive control in the kindergarten path effect," J. Exp. Child Psychol. **184**, 210–219.

Huang, Y. T., Leech, K., and Rowe, M. L. (**2017a**). "Exploring socioeconomic differences in syntactic development through the lens of real-time processing," Cognition **159**, 61–75.

Huang, Y. T., Newman, R. S., Catalano, A., and Goupell, M. J. (**2017b**). "Using prosody to infer discourse prominence in cochlear-implant users and normal-hearing listeners," Cognition **166**, 184–200.

Huang, Y. T., and Ovans, Z. (**2021**). "Who 'it' is influences what 'it' does: Discourse effects on children's syntactic parsing," E-pub ahead of print.

Huang, Y. T., Zheng, X., Meng, X., and Snedeker, J. (**2013**). "Children's assignment of grammatical roles in the online processing of Mandarin passive sentences," J. Mem. Lang. **69**(4), 589–606.

Huyck, J. J. (**2018**). "Comprehension of degraded speech matures during adolescence," J. Speech Lang. Hear. Res. **61**(4), 1012–1022.

Iglehart, F. (**2016**). "Speech perception in classroom acoustics by children with cochlear implants and with typical hearing," Am. J. Audiol. **25**, 100–109.

Jaeger, T. F. (**2008**). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," J. Mem. Lang. **59**, 434–446.

Jarosz, A. F., and Wiley, J. (**2014**). "What are the odds? A practical guide to computing and reporting Bayes factors," J. Probl. Solving **7**(1), 2.

Jeffreys, H. (**1961**). *Theory of Probability*, 3rd ed. (Oxford University Press, Oxford, UK).

Kidd, E., Stewart, A. J., and Serratrice, L. (**2011**). "Children do not overcome lexical biases where adults do: The role of the referential scene in garden-path recovery," J. Child Lang. **38**, 222–234.

Klatte, M., Lachmann, T., and Meis, M. (**2010**). "Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting," Noise Health **12**, 270–282.

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

Lee, M. D., and Wagenmakers, E. J. (**2014**). *Bayesian Cognitive Modeling: A Practical Course* (Cambridge University Press, Cambridge, United Kingdom).

Leech, K., Rowe, M., and Huang, Y. T. (**2017**). "Variations in the recruitment of syntactic knowledge contribute to SES differences in syntactic development," J. Child Lang. **44**, 995–1009.

Leibold, L. J., Taylor, C. N., Hillock-Dunn, A., and Buss, E. (**2013**). "Effect of talker sex on infants' detection of spondee words in a two-talker or a speech-shaped noise masker," in *Proceedings of Meetings on Acoustics ICA2013*, Montreal, Canada (Acoustical Society of America), Vol. 19, No. 1, p. 060074.

Leibold, L. J. (**2017**). "Speech perception in complex acoustic environments: Developmental effects," J. Speech Lang. Hear. Res. **60**, 3001–3008.

Leibold, L. J., and Buss, E. (**2019**). "Masked speech recognition in school-age children," Front. Psychol. **10**, 1981.

Levy, R., Bicknell, K., Slattery, T., and Rayner, K. (**2009**). "Eye movement evidence that readers maintain and act on uncertainty about past linguistic input," Proc. Nat. Acad. Sci. U.S.A. **106**(50), 21086–21090.

Lidz, J., White, A. S., and Baier, R. (**2017**). "The role of incremental parsing in syntactically conditioned word learning," Cogn. Psychol. **97**, 62–78.

Luke, S. G. (**2017**). "Evaluating significance in linear mixed-effects models in R," Beh. Res. Methods **49**(4), 1494–1502.

MacWhinney, B., Bates, E., and Kliegl, R. (**1984**). "Cue validity and sentence interpretation in English, German, and Italian," J. Verb. Learn. Verb. Behav. **23**(2), 127–150.

McCreery, R. W., Spratford, M., Kirby, B., and Brennan, M. (**2017**). "Individual differences in language and working memory affect children's speech recognition in noise," Int. J. Audiol. **56**, 306–315.

McCreery, R. W., Miller, M. K., Buss, E., and Leibold, L. J. (**2020**). "Cognitive and linguistic contributions to masked speech recognition in children," J. Speech Lang. Hear. Res. **63**, 3525–3538.

McMurray, B., Farris-Trimble, A., and Rigler, H. (**2017**). "Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally," Cognition **169**, 147–164.

Messenger, K., and Fisher, C. (**2018**). "Mistakes weren't made: Three-year-olds' comprehension of novel-verb passives provides evidence for early abstract syntax," Cognition **178**, 118–132.

Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., and Gillam, R. B. (**2018**). "Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder," J. Speech Lang. Hear. Res. **61**(12), 2950–2976.

Morey, R. D., and Rouder, J. N. (**2011**). "Bayes factor approaches for testing interval null hypotheses," Psychol. Meth. **16**(4), 406–419.

Newman, R. S. (**2009**). "Infants' listening in multitalker environments: Effect of the number of background talkers," Atten. Percept. Psychophys. **71**(4), 822–836.

Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (**2010**). "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," Ear Hear. **31**(3), 336–344.

Newman, R., and Chatterjee, M. (**2013**). "Toddlers' recognition of noise-vocoded speech," J. Acoust. Soc. Am. **133**(1), 483–494.

Newman, R. S., Chatterjee, M., Morini, G., and Remez, R. E. (**2015**). "Toddlers' comprehension of degraded signals: Noise-vocoded versus sine-wave analogs," J. Acoust. Soc. Am. **138**(3), EL311–EL317.

Newman, R. S., Morini, G., Shroads, E., and Chatterjee, M. (**2020**). "Toddlers' fast-mapping from noise-vocoded speech," J. Acoust, Soc. Am. **147**(4), 2432–2441.

Nittrouer, S., Kuess, J., and Lowenstein, J. H. (**2015**). "Speech perception of sine-wave signals by children with cochlear implants," J. Acoust. Soc. Am. **137**(5), 2811–2822.

Nozza, R. J., Miller, S. L., Rossman, R. N., and Bond, L. C. (**1991**). "Reliability and validity of infant speech-sound discrimination-in-noise thresholds," J. Speech Lang. Hear. Res. **34**(3), 643–650.

Omaki, A., White, I., Goro, T., Lidz, J., and Phillips, C. (**2014**). "No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions," Lang. Learn. Dev. **10**, 206–233.

O'Neill, E. R., Kreft, H. A., and Oxenham, A. J. (**2019**). "Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions," J. Acoust. Soc. Am. **146**(1), 195–210.

Oppenheimer, K., Morini, G., Strother-Garcia, K., Huang, Y., and Owen Van Horne, A. (November, **2020**). "Online comprehension of passives by children with developmental language disorder and typically developing age-matched peers," [Poster presentation] *45th Boston University Conference on Language Development*, November 2018, Boston, MA.

Ovans, Z., Huang, Y., and Feldman, N. (**2020**). "The (un)surprising kindergarten path," Paper presented at the *42nd Annual Meeting of the Cognitive Science Society*, July, Toronto, Canada.

Pajak, B., Fine, A. B., Kleinschmidt, D. F., and Jaeger, T. F. (**2016**). "Learning additional languages as hierarchical probabilistic inference: Insights from first language processing," Lang. Learn. **66**(4), 900–944.

Pisoni, D. B., Kronenberger, W. G., Chandramouli, S. H., and Conway, C. M. (**2016**). "Learning and memory processes following cochlear implantation: The missing piece of the puzzle," Front. Psychol. **7**, 493.

Polka, L., Rvachew, S., and Molnar, M. (**2008**). "Speech perception by 6- to 8-month-olds in the presence of distracting sounds," Infancy **13**(5), 421–439.

Prodi, N., Visentin, C., Borella, E., Mammarella, I. C., and Di Domenico, A. (**2019**). "Noise, age, and gender effects on speech intelligibility and sentence comprehension for 11- to 13-year-old children in real classrooms," Front. Psychol. **10**, 2166.

Qi, Z., Love, J., Fisher, C., and Brown-Schmidt, S. (**2020**). "Referential context and executive functioning influence children's resolution of syntactic ambiguity," J. Exp. Psychol. Learn. Mem. Cogn. **46**(10), 1922–1947.

Ryskin, R., Futrell, R., Kiran, S., and Gibson, E. (**2018**). "Comprehenders model the nature of noise in the environment," Cognition **181**, 141–150.

Schiller, I. S., Morsomme, D., Kob, M., and Remacle, A. (**2020**). "Noise and a speaker's impaired voice quality disrupt spoken language processing in school-aged children: Evidence from performance and response time measures," J. Speech Lang. Hear. Res. **63**, 2115–2131.

Schouwenaars, A., Finke, M., Hendriks, P., and Ruigendijk, E. (**2019**). "Which questions do children with cochlear implants understand? An eye-tracking study," J. Speech Lang. Hear. Res. **62**, 387–409.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**(5234), 303–304.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. (**1999**). "The kindergarten-path effect: Studying on-line sentence processing in young children," Cognition **73**(2), 89–134.

Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., and Lewis, D. E. (**2012**). "Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children," J. Acoust. Soc. Am. **131**, 232–246.

van Doorn, J., Ly, A., Marsman, M., and Wagenmakers, E. J. (**2020**). "Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ," J. Appl. Stat. **47**(16), 2984–3006.

von Koss Torkildsen, J., Hitchins, A., Nyhrum, M., and Wie, O. B. (**2019**). "Speech-in-noise perception in children with cochlear implants, hearing aids, developmental language disorder and typical development: The effects of linguistic and cognitive abilities," Front. Psychol. **10**, 2530.

Waked, A., Dougherty, S., and Goupell, M. J. (**2017**). "Vocoded speech perception with simulated shallow insertion depths in adults and children," J. Acoust. Soc. Am. **141**(1), EL45–EL50.

Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.

Wróblewski, M., Lewis, D. E., Valente, D. L., and Stelmachowicz, P. G. (**2012**). "Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults," Ear Hear. **33**(6), 731–744.

Youngdahl, C. L., Healy, E. W., Yoho, S. E., Apoux, F., and Holt, R. F. (**2018**). "The effect of remote masking on the reception of speech by young school-age children," J. Speech Lang. Hear. Res. **61**, 420–427.