

©Copyright 2025

Erin Smith

Familiarization to Speech in Noise: The Roles of Phonological Processing and Audiovisual Integration

Erin Smith

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Erin Ingvalson

Sara Kover

Christina Zhao

Gabriel Cler

Program Authorized to Offer Degree:

Speech and Hearing Sciences

University of Washington

Abstract

Familiarization to Speech in Noise: The Roles of Phonological Processing and Audiovisual Integration

Erin Smith

Chair of the Supervisory Committee:

Erin Ingvalson

Department of Speech and Hearing Sciences

Phonological processing, or the ability to recognize and manipulate phonemes, is a key aspect of language comprehension. Accurate speech recognition relies on the listener's ability to rapidly decode phonemic information. It is known that audiovisual (AV) integration can facilitate the speech recognition process, where simultaneously presented visual cues of a speaker's articulators can calibrate perception of the acoustic characteristics of incoming speech signals. In typically developing listeners, speech recognition is more robust with AV stimulus integration compared to unimodal auditory presentation. However, it remains unclear whether AV integration similarly benefits speech recognition in listeners with poor phonological processing ability. The present study uses a high-level linguistic task to examine whether a listener's recognition of unfamiliar speech in noise is moderated by their phonological processing ability, given a familiarization passage with unimodal auditory or multimodal AV stimulus presentation. Results show that there was no benefit of passive AV exposure to the speaker for sentence recognition. Moreover, while phonological processing ability supports overall recognition of

speech in noise, it did not moderate the amount of gain received by AV integration. These findings indicate that as a listener-specific factor, phonological processing does not impact the effect of AV integration for the perceptual learning of unfamiliar speech.

Acknowledgements

Sharing the experience of writing this thesis with my community has been an immense privilege.

Gratitude is due first and foremost to my mentor, Dr. Ingvalson, for her guidance. Dr. Ingvalson, working with you has been a highlight of my graduate education and I am grateful for all of the advice and support you have provided. I would also like to voice appreciation for the other members of my committee: Dr. Kover, Dr. Zhao, and Dr. Cler, for their valuable insight and enthusiasm for this research.

To Anja Bullen and Noelle Abbott, thank you for all of your help; I've enjoyed the opportunity to work with you. Likewise, to my fellow Listen & Learn Lab members Madison Kendrick, Natalie Guisto, and Amanda Silber, thank you for your assistance with this project. To those who participated in this research, I am grateful for your time and your dedication to science. My experience as a graduate student in the SPHSC department at UW has been nothing short of wonderful, and I owe that entirely to the kind faculty, staff, and students that I have met here. Thank you for being so welcoming. I would especially like to recognize my fellow CoreSLP student Shannon Xu for her friendship and unwavering support. Shannon, I am so glad that we were able to do this together. I am also forever grateful for my family, who have cheered me on from near and far. Last but certainly not least, to my partner, Ben: you are my inspiration. Thank you for being the best.

Table of Contents

Abstract.....	iii
Acknowledgements.....	v
Introduction.....	1
Phonological Processing.....	1
Audiovisual Contributions to Speech Perception.....	3
Clinical Relevance.....	6
Research Questions.....	8
Methods.....	11
Participants.....	11
Materials.....	13
Procedure.....	13
Analysis.....	16
Results.....	17
Discussion.....	20
Extent of Speech Familiarization.....	21
Benefit of Audiovisual Cues for Sentence Recognition Accuracy.....	23
Effect of Phonological Processing Ability on Audiovisual Integration and Familiarization....	24
Limitations and Future Directions.....	27
Conclusion.....	28
References.....	30
Appendix A: Bamford-Kowal-Bench (BKB) Sentence Lists and Keywords.....	40
Appendix B: Zoo Passage.....	45
Appendix C: Demographic and Listening History Questionnaire.....	46

Introduction

Phonological Processing

Phonology is the study of how speech sounds are organized and manipulated to form meaningful units of language. Phonological processing, which is the ability to accurately map phonemes to meaning in oral and written stimuli, is thus a precursor and key component of successful expressive and receptive language. This ability plays a role in tasks such as spelling, rhyming, and auditory discrimination.

Phonological processing can be further broken down into components including phonological awareness, phonological working memory, and phonological retrieval (American Speech-Language Hearing Association, 2025). In more descriptive terms, a listener of spoken language must not only have prior knowledge of phonemes and recognize phonological components of speech, but they must do so online and on demand at a rapid pace during speech comprehension tasks, while simultaneously holding a wealth of phonological information related to the context at hand (American Speech-Language Hearing Association, 2025). While the term “phonological awareness” broadly encompasses the ability to manipulate speech sounds at all linguistic levels, the more specific “phonemic awareness” refers to knowledge and manipulation of individual phonemes and is often a point of interest for researchers seeking to understand skills such as reading. An individual with a deficit in phonemic awareness may struggle to discriminate between sound approximations such as “bat” and “pat”, or to assemble words based on their requisite phonemic parts in blending tasks, such as combining the phonemes /b/, /ʌ/, and /g/ to form the word “bug”. Functional implications of phonological processing deficits include increased difficulty with reading, writing, speech perception, and speech production, decreasing the accessibility of verbal and written communication.

Previous neuroimaging studies aiming to localize the neural structures involved in phonological processing during reading tasks point to brain activation in the left fronto-temporal region, and specifically the superior temporal gyrus (STG) and left opercular part of the inferior frontal gyrus

(opIFG), common to children (Mathur et al., 2020) and adults (Jobard et al., 2003). Note that the observed extent and specialization of activation likely depends on stimulus modality, as the left STG is found to be specialized for phonological processing in young children for auditory stimuli (Weiss et al., 2018), but not for visual stimuli (Mathur et al., 2020). This is in agreement with a systematic review of fMRI studies assessing age-related changes in neural activation for language processes, in which Weiss-Croft and Baldeweg (2015) concluded that while the neural basis of phonological processing develops across the lifespan, there is some variability attributable to the task demands, as well as the modality in which phonological processing is engaged.

Phonological processing skills, including phonemic awareness, are known to be integral to acquisition of reading ability (Wagner & Torgesen, 1987). In order for an emerging reader to meaningfully map orthographic representations to phonemes, they must have phonemic awareness. Impaired phonemic awareness is historically considered to be an underlying contributor to dyslexia (Felton et al., 1990; Pennington et al., 1990) and is referenced in contemporary literature citing the Phonological Deficit Hypothesis of dyslexia (PDH, Share 2021; but see also Zoccolotti 2022 for a discussion of limitations). Individuals with dyslexia demonstrate poor performance on assessments of broader phonological processing including blending words, rhyming, and elision. As such, phonological processing is identified by the International Dyslexia Association as an area of importance in the screening and comprehensive evaluation of children with suspected dyslexia (The International Dyslexia Association). Phonological processing impairments have also been noted in children with other disorders impacting communication ability, including autism spectrum disorder (ASD) (Nordgren, 2015; Wolk et al., 2016) and developmental language disorder (DLD) (Lara-Díaz et al., 2024).

Just as phonological processing is vital to a reader's ability to map phonemes to their orthographic representations, it also contributes to the successful decoding of oral language. For example, someone with intact phonemic awareness may recognize interdental tongue placement as a

strong indicator of the production of "th". This type of visual referent is known as a viseme. While typical listeners are generally not skilled at speech-reading, or relying only on visemes to recognize speech at the linguistic level, visemes can support accurate perception of speech by providing complementary and redundant information about an incoming auditory signal (Campbell, 2008).

Sparse phonological processing skills, such as those seen in the still-developing language systems of children, may hinder a person's ability to map known referents onto novel speech input, such as that of an unfamiliar accent. Indeed, in a study of typically developing (TD) monolingual listeners, Bent and Atagi (2016) found that while all participants suffered a decrease in task accuracy, children had poorer performance than adults on a speech recognition task when nonnative-accented talkers were introduced. This result is somewhat expected, as phonological processing ability is known to strengthen over the course of language development in TD individuals. Additionally, the authors suggested that phonological awareness was a stronger predictor for nonnative-accented speech recognition outcomes than phonological memory in adults and TD children, while the adverse was true for speech perception of native talkers. It may be the case that the requisite components of phonological processing are differentially activated depending on task-specific features such as linguistic level and the familiarity of the speaker.

Audiovisual Contributions to Speech Perception

In TD listeners, it is known that audiovisual (AV) stimulus integration benefits speech perception outcomes compared to auditory stimulus presentation alone (Skipper et al., 2007; Sumby & Pollack, 1954). While theories of the mechanism driving audiovisual speech perception vary, the cortical areas typically implicated in neuroimaging studies include the primary auditory cortex, the left posterior superior temporal sulcus (STS), and the premotor cortex (Peelle, 2019). In AV integration, inputs from each stimulus modality are fused into a perceptually unique construct. This is famously observed through the McGurk effect, in which mismatched audio and video presentation of simple phonemes

(e.g., auditory “pa” and visual “ka”) are perceptually identified as a third auditory exemplar (e.g., “ta”) (McGurk & MacDonald, 1976; Tiippana, 2014). When a speaker’s auditory and visual speech cues match, their integration can provide more robust speech perception than unimodal stimulus presentation. In a practical and ecologically valid context, this function is especially useful for speech recognition when the saliency of auditory cues is limited. For example, visemes serve to support listener comprehension in a noisy environment.

In addition to the saliency of the perceptual stimuli, the relative weights of incoming auditory and visual information to the formation of an AV representation are dependent on listener characteristics such as hearing status (Lalonde & McCreery, 2020; Sekiyama et al., 2014) and language history (Drijvers & Özyrűek, 2018; Sekiyama & Burnham, 2008; but see also Hazan, Kim, & Chen, 2010). Research in adults also shows that attention is known to affect AV stimulus processing for speech perception (Seijdel et al., 2024; Tiippana et al., 2004; Tuomainen et al., 2005). In general, AV integration is susceptible to individual variation and is modulated by differences in task and stimulus presentation.

In a recent study of the AV perception of degraded speech in adults with typical hearing, von Seth et al. (2025) observed that individual differences in benefit from AV cues were a result of perceptual ability, where participants with poorer hearing or greater lipreading ability showed a proportionally larger benefit from access to AV cues during perception of speech materials. This benefit was found to be stable across linguistic context (consonant, word, and sentence-level stimuli) and across time as assessed in multiple temporally separated sessions. Additionally, the authors found that individuals who showed greater sensitivity to perception of place and manner of articulation likewise showed greater individual benefit of AV cues for speech perception (von Seth et al., 2025). These findings further confirm the individual variability of AV integration and suggest some level of interaction with phonological processing for benefit in speech perception tasks.

There is limited research directly assessing the influence of phonological processing ability on AV integration for speech in adult populations. However, a relationship between AV integration and phonological processing is apparent from the early stages of development. Cross-modal association of auditory and visual speech signals is present from the first year of life (Hollich et al., 2005; Kushnerenko et al., 2008) and continues to develop through adolescence with age and exposure to speech stimuli (Danielson et al., 2017; Soto-Faraco et al., 2012). The strong associations between temporally congruent auditory and visual speech cues demonstrated by infants attenuates in young school-aged children before re-emerging to benefit speech perception in adolescence and adulthood, indicating a U-shaped course of development for AV speech perception related to the process of phonological reorganization in childhood (Jerger et al., 2009). This pattern reflects the trajectory of typical language development and may be associated with evolving phonological representations in early childhood.

In a study of AV speech perception at the levels of detection, discrimination, and recognition, adults relied more on the visual component of AV speech stimuli than 6- to 8-year-old children for tasks that require access to lexical representations stored in memory in addition to basic perceptual processing (Lalonde & Holt, 2016). In a recent review, Lalonde and Werner (2021) explain that while visual speech cues allow for improved speech discrimination on the basis of temporal processing in infants and young children, it may be the case that phonetic cues aren't integrated for the benefit of speech recognition until later childhood and adulthood due to the development of speech-specific AV integration mechanisms. As mentioned previously, visemes may support speech recognition accuracy, but it is assumed in this case that the listener has phonemic awareness sufficient to map the visual input (e.g., labiodental contact) to an auditory referent (e.g., the /f/ sound).

These findings point to the existence of an evolving AV speech integration mechanism that is affected by the maturation of an individual's phonemic processing ability and lexical access. Indeed, electrophysiological data indicate that while adults, 8-9 year-old children, and 11-12 year-old children all

benefited from the phonemic content comprising congruent AV cues in a behavioral speech-in-noise task, the N400 component only significantly predicted task performance for adults, while the late positive complex (LPC) amplitude only predicted task performance for older children (Kaganovich et al., 2019). As the authors argue that the N400 represented phonological processing in their study, while the LPC ERP component was associated with lexical access, their findings indicate that older adults could utilize phonological cues for lexical processing, whereas children could not use visemes to access lexical representations of the stimuli. Thus, the utility of AV cues for speech perception depends in part on the strength of a listener's phonological processing ability.

Clinical Relevance

It is currently unclear how the mechanism of phonological processing contributes to AV speech integration in populations with impaired phonological processing ability. The results of von Seth et al. (2025) would indicate that individuals with deficits in phonemic awareness should show less benefit from AV cues for speech perception; indeed, an absence of STS activation has been found in adults with developmental dyslexia (DD) during an AV integration task (Ye et al., 2017), pointing to a neurophysiological difference in AV processing in individuals with DD. However, evidence from behavioral research shows otherwise. While Leybaert et al. (2014) found that children with specific language impairment were resistant to the influence of visual speech in a McGurk-style task due to deficits in phonemic categorization, they demonstrated typical McGurk responses when given more distinguishable visual phonemic cues. In other words, while individuals with language disorders may show less robust processing of speech input than their TD peers across modalities, they can access AV speech integration at the perceptual level.

Furthermore, relevant work has suggested that even though language-impaired groups had lower overall performance on nonword repetition tasks than their TD peers, AV integration can similarly benefit performance for children with DLD, DD, concomitant DLD and DD, and their TD peers (Ehrhorn et

al., 2021). This finding suggests a potential benefit of AV integration for speech perception even in the context of poor phonological processing skills. With evidence from an experimental phonological skill training program, Heikkilä et al. (2018) have gone so far as to suggest that the incorporation of AV speech cues in phonological skill training could improve phonological representation outcomes in children with DLD as opposed to training with auditory cues alone.

In a recent study, Gijbels et al. (2024) assessed the performance of children with DD and their TD peers on an ecologically valid AV speech recognition task in varying levels of noise. They found that greater phonological processing ability was related to improved speech perception performance for TD children and children with DD. The DD group showed similar AV integration benefit for speech perception as their TD peers, though they weighed auditory information less heavily than TD children at more difficult signal-to-noise ratio (SNR) listening conditions. As Gijbels et al. (2024) argue that phonological processing is an auditory skill, this implies that children with DD have sufficient access to, and awareness of, visual phonemic information from which they derive benefit for functional understanding. The authors concluded that a deficit in phonological processing could impact speech-in-noise performance, though AV integration itself was not affected by poor phonological processing ability. The authors acknowledged potential differences in the weighting of auditory and visual cues during AV speech perception by children with DD versus their TD peers.

On the other hand, findings in other populations characterized by poor phonological processing ability suggest that there may be a relationship between the weakness of this skill and impedance of the potential benefit of AV speech cues for speech recognition. Research in children with ASD indicates a lack of AV integration for speech perception even when accounting for the characteristic reticence for eye contact during verbal communication noted in this population (Irwin & DiBlasi, 2017). The existing literature in respect to this process in DLD is inconclusive, with some researchers arguing that individuals with DLD are unable to properly integrate auditory and visual cues at a basic level due to differences in

other cognitive processes, such as attention (Pons et al., 2013; Pons et al., 2018). Others have suggested that children with DLD demonstrate multisensory phonemic processing deficits, as they do not demonstrate electrophysiological responses to incongruent AV patterns on par with their TD peers (Kaganovich et al., 2021) or utilize available visual cues during phoneme perception tasks even in degraded auditory conditions (Meronen et al., 2013). These findings are in disagreement with those presented above concerning AV integration in individuals with DLD. Moreover, if a deficit in phonological processing is considered to be a key mechanism underlying DLD, as in DD, these results are unexpected based on the above reported research on the use of AV cues for speech perception by children with DD (Gijbels et al., 2024).

Research Questions

In sum, the existing evidence is inconclusive as to how well individuals with poor phonological processing ability can access and utilize AV cues toward the benefit of speech recognition. The available literature varies widely in terms of subject-specific factors, such as age and language abilities, as well as task-specific factors, such as the ecological validity of the AV stimuli and the outcomes measured. While evidence points toward phonological processing ability as a prerequisite for the perceptual benefit of AV integration in high-level speech perception tasks, it is unclear how this relationship manifests in individuals with impaired phonological processing. Furthermore, there is conflicting evidence on the level of benefit that individuals with phonological processing deficits can derive from AV cues, with some literature pointing to a superior level of speech recognition gains, and others negating potential benefits. Finally, much of the research exploring the relationship between phonological processing ability and AV speech perception benefit has been conducted in pediatric populations. Given the differences in phonological processing and AV integration for speech recognition over the lifespan, it is important to determine whether these findings are similar in adult populations.

The present study examines the roles of phonological processing ability and AV integration during a sentence-level speech recognition task with adult participants. More specifically, does AV integration boost speech recognition accuracy during familiarization to a novel speaker? Additionally, does a listener's phonological processing ability moderate the level of benefit received by AV cues for speech recognition? Heeding the call of recent advocacy for more ecologically valid, and thus clinically relevant work- in other words, shifting the focus away from singular perceptual features such as the McGurk effect (Van Engen et al., 2017; Van Engen et al., 2022)- this research question may be investigated during a speech perception task in which keyword recognition is measured before and after passive exposure to a novel speaker of accented English reading a passage, presented through auditory-only or AV channels.

In this task, the stimuli are presented by a non-native speaker of English with a Mandarin accent. This choice of speaker serves to reduce the familiarity of the speech stimuli, as their speech production has phonetic and prosodic differences from American English dialects. Moreover, listeners have been shown to familiarize to accented-English speech given training through exposure (Bradlow & Bent, 2008). Thus, the use of accented speech provides a methodological opportunity for the listener to demonstrate speech recognition gains across time as they familiarize to the phonetic characteristics of the speaker. With the addition of an AV exposure block to this task structure, there is the possibility for visual cues to provide additional phonemic information about the speech signal, further supporting accented speech recognition outcomes in the listener (Zheng & Samuel, 2019).

Banks et al. (2015) conducted a task similar to what is described above, in which participants familiarized to a novel accent with active auditory or AV training, with or without noise, and with a visual-only or "speechreading" control condition. The authors found that there was a benefit of AV stimuli in noise, which aligns with existing theories that AV cues support speech recognition in difficult listening conditions. However, there was no benefit of AV stimulus presentation for familiarization to a

novel accent in quiet (Banks et al., 2015). The authors concluded that while AV cues supported speech recognition accuracy, there was no gain in perceptual adaptation to the speaker compared with the auditory-only condition. In other words, while AV cues boosted immediate speech recognition in noise, they did not support long-term familiarization to accented speech. Perceptual adaptation is attributed to listener-specific factors in this instance. To build upon this line of work, it is worthwhile to investigate whether phonological processing ability, which is one such listener-specific factor, moderates the amount of benefit derived from AV speech cues during a sentence-level task in noise and subsequently supports familiarization.

We suspect that AV cue integration, as well as phonological processing ability, will support recognition of unfamiliar speech. Moreover, individuals with relatively stronger phonological processing ability may demonstrate increased benefit from AV integration for speech recognition. This would indicate a relationship in which individuals with robust phonological processing ability are more sensitive to concurrently presented visual phonemic information. Likewise, individuals with poor phonological processing ability may have less access to the perceptual benefits of AV integration, resulting in no tangible difference in keyword recognition in accented-English speech following a period of AV exposure relative to an auditory-only condition.

This work contributes to the literature by directly observing the effects of phonological processing and AV integration on speech recognition in a high-level linguistic context, aiming to further elucidate our understanding of the role of phonological processing in face-to-face communication, and where this mechanism may differ in individuals with poor phonological processing ability. Furthering our knowledge of AV speech perception benefit in listeners with atypical or otherwise relatively poor phonological processing abilities is a clinically relevant endeavor. If individuals with phonological processing deficits are indeed able to benefit from AV integration during speech recognition, screening methods could be updated for increased sensitivity by incorporating AV rather than unimodal auditory

stimuli to better assess the functional profile of a listener's speech perception abilities. Likewise, if AV cues are found to be beneficial for understanding in linguistic tasks for individuals with language disorders, future interventions could adopt AV stimuli in clinical settings to support phonemic awareness and use.

Methods

Participants

Twenty-eight participants were recruited for this study. Participants with self-reported language difficulties were recruited along with their TD peers to include a diverse profile of phonological processing skills. Prior to the consent and experiment process, participants joined a videoconferencing call with the author to allow for questions and concerns to be addressed, as well as for troubleshooting purposes. Following the collection of informed consent, participants remained in the videoconference call throughout the study to have continued access to the researcher in case of technical difficulties. The experiment took 44 minutes to complete on average with optional breaks available. Participants were provided monetary compensation for completing the study.

Fifteen participants self-reported as being bilingual. Three participants reported some level of familiarity with Mandarin Chinese from two years of age or younger. One participant self-reported non-normal hearing. They were included in the analysis as they were able to set their volume to an accessible level and were observed to functionally engage with all task items. One participant who was above the age of 35 was excluded. The mean age of the final sample of participants ($N = 27$) was 22.67 ($SD = 3.14$), ranging from 18 to 31. Additional demographic information can be found in Table 1.

Table 1*Demographic Characteristics of Participants by Condition*

Characteristic	Audio		Audiovisual	
	<i>n</i>	%	<i>n</i>	%
Sex				
Male	4	29	2	15
Female	10	71	9	70
Non-Binary	0	0	2	15
Native Language ^a				
English	10	71	13	100
Mandarin	2	14	1	8
Japanese	1	7	1	8
Tagalog	0	0	1	8
Russian	1	7	0	0
Spanish	1	7	1	8
Vietnamese	1	7	0	0
Currently Bilingual				
No	5	38	7	54
Yes	9	62	6	46
Language Difficulty Status				
No	10	71	12	92
Yes	3	22	1	8
Prefer Not to Answer	1	7	0	0

Note. *N* = 27 (*n* = 14 in the Audio condition; *n* = 13 in the Audiovisual condition). Participants were on average 22.67 years old (*SD* = 3.14). ^a Reflects number and percentage of participants indicating that this was at least one of their native language[s]. Participants could report more than one native language.

Materials

During the experiment, participants completed a task of nonword repetition adapted for online use from the Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2; Wagner et al., 2013). Responses were recorded using the participants' hardware. Participants also completed the Test of Nonverbal Intelligence (TONI-3; Brown et al., 1997), adapted for online administration.

The speech familiarization task stimuli were recorded by a native Mandarin speaker in a quiet room using an Apple MacBook Pro (running MacOS Big Sur v.11.7.10) and version 3.4.2 of Audacity(R) recording and editing software. The AV stimuli were recorded on the same day using an Apple iPad. Each stimulus was manually trimmed at zero crossing points using Audacity(R); intensity normalized at 68dB using Praat (Boersma & Weenink, 2024) using an open-source script (Winn, 2013); RMS matched at 0.09 to prevent clipping; padded with 50ms of silence at beginning and end; and coupled with white Gaussian noise at a SNR of -2dB using Matlab (The MathWorks Inc., 2024).

The talker who recorded sentence stimuli for the speech familiarization task began learning English at age 5 in an academic context. At the time of recording, they were a student at the University of Washington and reported speaking English on a daily basis. Their total TOEFL score was 102, with subscores of 28 in reading; 25 in listening; 23 in speaking; and 26 in writing. During a pilot study with 5 TD adult participants with self-reported normal hearing, the average percent speech recognition accuracy when repeating a sample of 20 sentence stimuli presented at -2dB SNR was 82.21%.

Procedure

The experiment was hosted online using research platform Gorilla Experiment Builder (Anwyl-Irvine et al., 2018). Participants were asked to complete the study in a quiet room using a laptop or desktop with audio recording capability and the Google Chrome web browser. All tasks were administered using the Gorilla Experiment Builder platform. Microphone and audio capability were automatically tested by Gorilla Experiment Builder at the beginning of the session, serving as a screener of device compatibility.

Following the collection of informed consent, participants completed an additional system check that allowed them to set the volume of their headphones to a comfortable level adapted from the procedure outlined by Zhao et al. (2022). The procedure remained the same with two changes: first, the white noise in the original experiment was replaced with a 400ms sample of white Gaussian noise derived from the sentence stimuli created for the speech familiarization experiment. This noise sample was selected to better approximate the noise that participants would later hear in the experiment. Stimuli were buffered by 50ms cosine ramps at the beginning and end of the signal. Additionally, the white noise presented during this task was provided with a switch for the participants to control, rather than as a continuous pulsing noise.

During the nonword repetition task, nonwords of increasing length and complexity were presented to participants while a fixation cross was displayed on the screen. Participants were asked to repeat the nonwords out loud, with their utterances recorded. Response recordings were scored for accuracy offline. Scores were given according to the standardized instructions by a student with graduate training in speech-language pathology. Nonword repetition performance was collected as a proxy for participants' phonological abilities (Dollaghan & Campbell, 1998; Schwob et al., 2021).

During the adapted test of nonverbal intelligence, participants were presented with an incomplete visual pattern on the screen. They were asked to complete the pattern by using the numeric keys of their keyboard to select one of several options displayed along the bottom of the screen. Following an initial set of five practice trials with feedback provided for accuracy, participants proceeded through a test set of 45 items without feedback. The task ended once participants completed all items or responded incorrectly on three out of five consecutive trials. The raw score was calculated as the sum of correct items, then converted to a standard score following the procedure outlined in the examiner's manual. Participants with a standard score of 70 or above on the TONI-3 were considered not to have an

intellectual disability and were included in subsequent analysis (see Miller & Gilbert, 2008, for discussion).

Following the TONI-3, participants completed a speech familiarization task. They were randomly assigned to one of two task conditions: auditory-only or audiovisual. First, they heard sentences from the Bamford-Kowal-Bench (BKB) corpus (Bench et al., 1979) spoken in Mandarin-accented English (see Appendix A for the full list of sentences). Participants were asked to repeat the sentences aloud within a 5-second time limit on each trial. If participants were unsure of the sentences they heard, they were asked to make their best guess during the repetitions. Following three practice trials and one block of 32 sentence repetition trials, participants passively attended to the same speaker reciting the Zoo passage (Fletcher, 1972; see Appendix B). In the auditory-only condition, the passage was presented concurrently with a static image of the speaker's face accompanied by subtitles at the bottom of the screen. In the AV condition, participants observed a video of the speaker reading the same passage aloud with subtitles at the bottom of the screen, such that they received auditory and visual speech input. This block was intended to support familiarization to the speaker through passive exposure. Following the reading of the Zoo passage, participants completed two additional blocks of BKB sentence repetitions containing 32 sentences each, recorded by the same speaker. Each stimulus was presented once, for a total of 96 unique sentences. The final ratio of trials presented pre- and post-exposure passage was 1:2. This number of trials was anticipated to be sufficient for familiarization, as previous research on perceptual learning of unfamiliar speech has shown familiarization in as little as one minute (Borrie, McAuliffe, & Liss, 2012; Clarke & Garrett, 2004).

Responses made during the speech familiarization task were transcribed and scored for raw number of keywords correct per trial by trained graduate and undergraduate students in the Department of Speech & Hearing Sciences at the University of Washington. Transcribers were blinded to the task condition. Each sentence presented during the speech familiarization task had between two

and four predefined keywords. Keywords primarily consisted of nouns, verbs, and adjectives. Participants were randomly assigned to two possible sets of sentence lists, with 278 total keywords in the first version and 272 total keywords in the second version. Incorrect use of pronouns and the present and past tenses of the verbs “to be” and “to have” were not penalized. Likewise, there was no penalty for morphologic errors (e.g., incorrect tense), aligning with scoring procedures described in previous studies of speech recognition (Borrie, McAuliffe, Lisse, et al., 2012; Lansford et al., 2018).

For a conservative approach to ensuring data quality, trials with audio recordings that were damaged or otherwise incomplete due to technical failure were removed from the final analysis. All participants had at least 85% intact audio recordings. Recordings that were intact, but without a response attempt or with an unintelligible response due to unclear speech, were included and scored with 0 keywords correct.

Participants completed two questionnaires during the study. First, participants completed a demographics questionnaire to provide their demographic profile. They also completed a novel language history and exposure questionnaire that assessed their prior experiences with, and acquisition of, languages (see Appendix C). This included information regarding their hearing history and proficiency in languages they listen to, speak, read, or write. Responses to this questionnaire were provided through a combination of open-ended text fields and numeric scales.

Analysis

Analyses were completed using RStudio software (version 4.4.3; Posit team, 2025). The main variable of interest in this study was speech recognition, considered to be the correct identification and repetition of keywords. Accuracy per block was considered for a descriptive view of task performance, but as sentence presentations were randomized across participants, accuracy calculated within blocks does not fully capture the variability of the stimuli in the dataset. Thus, the raw number of keywords correct on a per-trial basis was considered the dependent variable of interest for a more robust analysis.

Linear mixed effects modeling was used to investigate the impact of fixed effects while accounting for random variation by participant, following recommendations for psycholinguistic analysis (Barr et al., 2013). Linear mixed effects modeling was accomplished using the lmer function from the lme4 package (Bates et al., 2015). The significance level was set to $\alpha = .05$ in all analyses.

Results

Descriptive statistics were calculated for each key variable across the entire sample and by task condition; see Table 2. Across all participants, the mean nonword repetition score was 13.19 (SD = 5.26; range 0-21) out of a maximum possible score of 30. The mean nonverbal intelligence score for all participants was 111.30 (SD = 14.4; range 83-135) out of a maximum possible score of 145. Across all participants, the mean accuracy during the sentence repetition task was 79.6% (SD = 13.6%) in the pre-exposure Block 1, 83.1% (SD = 14.0%) in Block 2, and 83.5% (SD = 12.9%) in Block 3; see Figure 1.

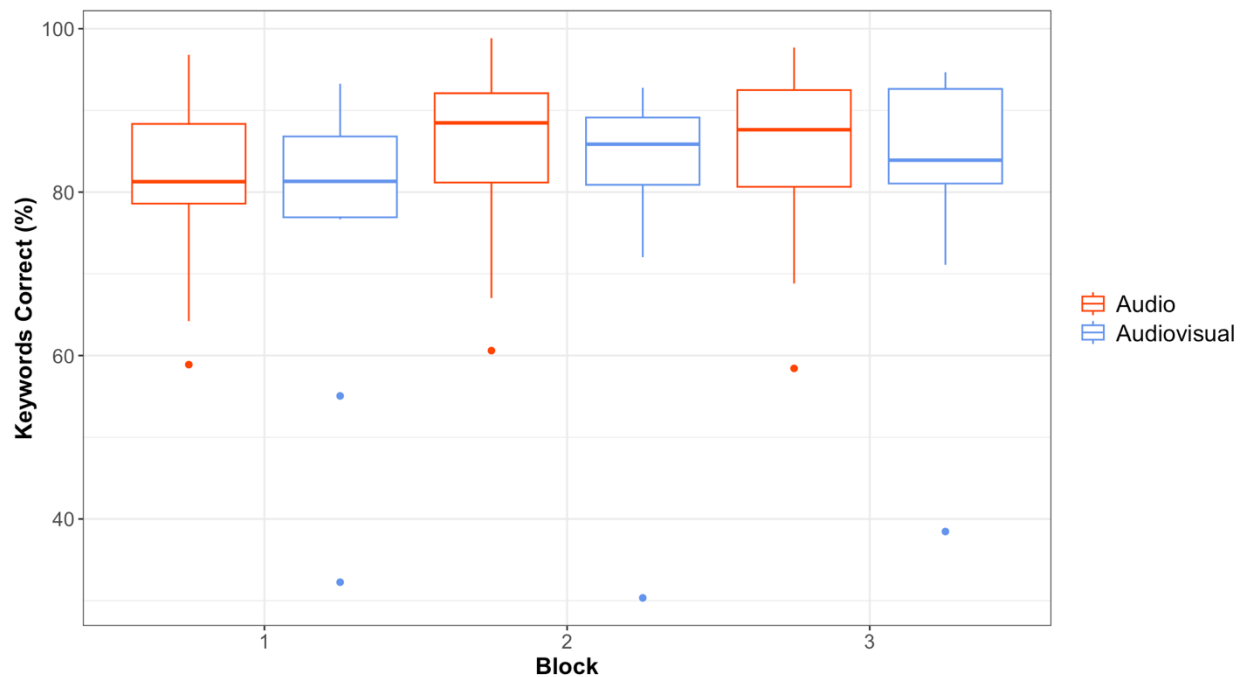
Table 2

Means and Standard Deviations of Scores on Study Measures

Task	Audio ($n = 14$)	Audiovisual ($n = 13$)
Nonword Repetition	13.7 (5.1)	12.6 (5.6)
Nonverbal Intelligence ^a	111 (14.5)	111 (14.9)
Sentence Repetition		
Block 1	81.5 (10.4)	77.5 (16.7)
Block 2	84.8 (11.6)	81.4 (16.4)
Block 3	85.0 (10.9)	82.0 (15.1)

Note. Standard deviations are presented in parentheses. The tests of nonword repetition and nonverbal intelligence were adapted for online administration from the CTOPP-2 and TONI-3, respectively.

^a Reflects the standardized score.

Figure 1*Keyword Recognition Accuracy by Block and Condition*

Note. Percentage of keywords correct were calculated per block within each condition.

Descriptive statistics were further analyzed to determine whether the inclusion of participants who self-reported familiarity with Mandarin ($n = 3$) influenced the results of the study. When participants who self-reported familiarity with Mandarin were excluded from analysis, the mean average Block 1 accuracy was 79.01% (SD = 14.2%). The difference in baseline sentence recognition accuracy when including participants who were familiar with Mandarin was less than 1%. As Block 1 was the block most likely to be impacted by previous familiarity with Mandarin (Bradlow & Bent, 2008), this relatively small difference gave us confidence to retain these listeners in the sample.

A linear mixed effects model with a three-way interaction between post-exposure passage status, condition, and nonword repetition score entered as a fixed effect was used to predict the raw number of keywords correctly recognized. Participant nested in condition was added as a random effect to account for between-participant variability. The model output is reported in Table 3.

Table 3*Results from Linear Mixed Effects Model Analysis of Study Effects on Keywords Correct*

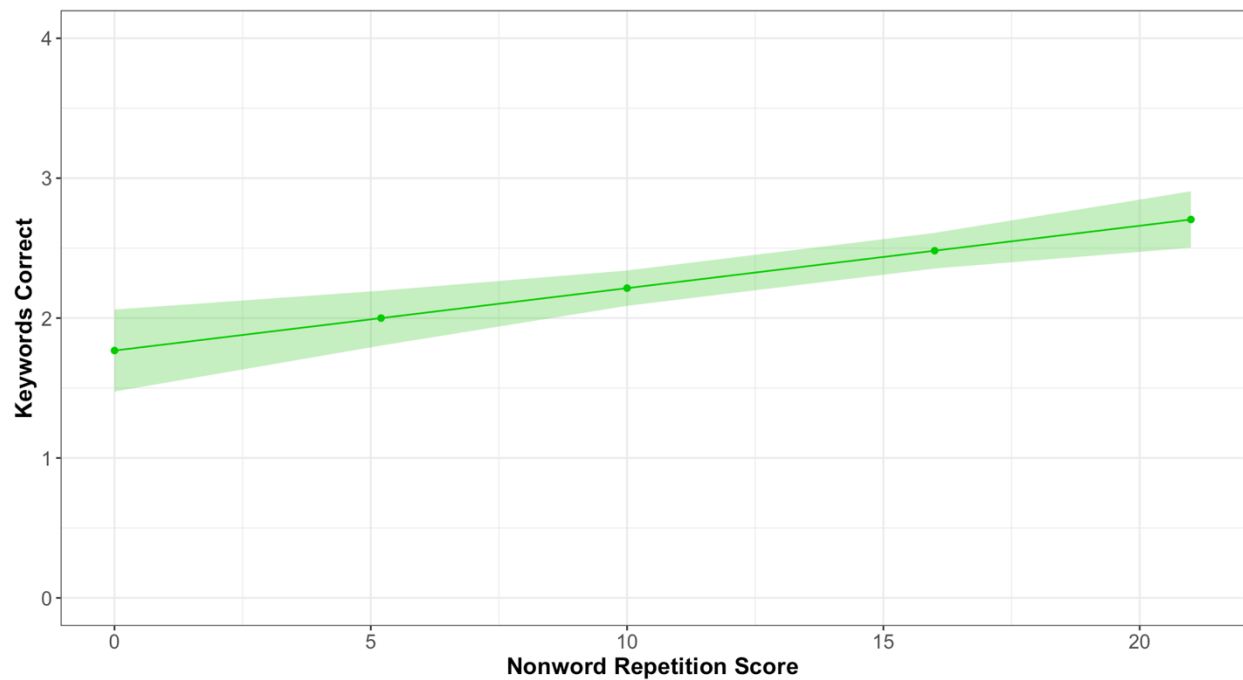
Fixed Effects	Estimate	SE	df	t	p
(Intercept)	1.90	0.19	22.85	9.89	< .001
PostExp	-0.07	0.15	2546.95	-0.48	.634
ConditionAV	-0.37	0.33	27.18	-1.10	.280
NWRepScore	0.03	0.01	22.85	2.47	.021
PostExp:ConditionAV	0.29	0.20	2546.97	1.42	.156
PostExp:NWRepScore	0.01	0.01	2546.95	1.08	.281
ConditionAV:NWRepScore	0.02	0.02	25.84	0.93	.361
PostExp:NWRepScore:ConditionAV	-0.02	0.01	2546.96	-1.32	.187
Random Effects	Name	Variance	SD	r	
ID	(Intercept)	0.03	0.19		
	ConditionAV	0.16	0.40	-.47	
σ^2		0.78	0.88		
Marginal R ² / Conditional R ²		0.07 / 0.15			

Note. Number of participants = 27, number of observations = 2578. *p* values < .05 are in bold. PostExp = post-exposure passage; ConditionAV = audiovisual condition; NWRepScore = nonword repetition score.

Adapted from "Visualizing Mixed-effects Models," by L. Mudge, 2019.

https://lmudge13.github.io/sample_code/mixed_effects.html

There was no significant effect of post-exposure status or of condition. Nonword repetition score was a significant predictor of keyword recognition; see Figure 2. There was no significant interaction between post-exposure status and condition, where participants in the AV condition and the auditory-only condition had comparable performance post-exposure. There was also no significant interaction between post-exposure status and nonword repetition score or between condition and nonword repetition score. Finally, there was no significant three-way interaction between post-exposure status, condition, and nonword repetition score.

Figure 2*Effect of Nonword Repetition Score on Keywords Correct*

Note. $N = 27$. Adapted from "Visualizing Mixed-effects Models," by L. Mudge, 2019.

https://lmudge13.github.io/sample_code/mixed_effects.html

Discussion

The aims of this study were to determine whether audiovisual (AV) stimulus integration boosts speech recognition when listening to unfamiliar speech, and whether an individual's phonological processing ability moderates the benefit of AV cues for speech recognition. These questions were investigated through a high-level linguistic task in which participants were asked to repeat accented-English sentences presented in noise, with randomization to passive auditory-only or AV exposure to the speaker. It was hypothesized that individuals who received AV exposure to a novel speaker would show greater familiarization, evidenced by a boost in speech recognition in subsequent trials. Moreover, individuals with more robust phonological processing were expected to show higher speech recognition throughout the experiment. Finally, the results were explored to determine whether individuals with

superior phonological processing ability experienced greater benefit from AV cues toward familiarization to accented speech.

Extent of Speech Familiarization

The results of this study show that there was no significant effect of the speaker exposure passage on later keyword recognition. This is indicative of an overall null effect of the passive exposure passage on familiarization to accented speech. However, qualitative observation of the results confirms that participants generally improved at recognizing keywords within the sentence stimuli (see Figure 1), as expected based on previous literature describing participant familiarization to novel speech (e.g., Bradlow & Bent, 2008). Across all participants, there was an average gain of 3.5% accuracy in keyword recognition from the first to the second blocks, with a passive exposure passage in between. From there, the average gain in keyword recognition accuracy was observed to increase by 0.4% from the second to the third block, trending towards a plateau in recognition gains.

There is more than one possible reason as to why the extent of familiarization to accented speech in this study, while observable, was not statistically significant. Previous literature on familiarization to novel speech points out that adaptation occurs very quickly, on the order of less than one minute of exposure to speech (Borrie, McAuliffe, & Liss, 2012; Clarke & Garrett, 2004). It may be the case that the bulk of the participants' familiarization to the accented speech occurred within the first block of the sentence repetition task. Participants demonstrated greater performance in the post-exposure blocks, but the extent to which this boost in accuracy is related to the exposure passage itself, versus time spent attending to the speech in noise, is unclear. In a future iteration of the study, inclusion of a control group that does not attend to an exposure passage would help to reveal potential effects on familiarization.

There are also likely to be factors intrinsic to the design of the sentence repetition task and its stimuli that affected the amount of familiarization to accented speech demonstrated by participants.

With the exception of a small number of underperforming outliers, the average accuracy of sentence recognition across participants was high at baseline (approximately 80%). This level of accuracy, which was reflected in original pilot testing, was assumed to be appropriate for recruitment of individuals with variable phonological processing abilities. However, it may have been too high to reveal effects of individual and task-specific factors on sentence recognition.

The stimuli themselves could have unintentionally supported high performance through cues such as the semantic predictability of the sentences. Previous literature has shown that contextual cues support accurate recognition of foreign-accented speech (Holt & Bent, 2017). Additionally, both native and non-native adult listeners have been found to benefit from semantic predictability during speech recognition (Bradlow & Alexander, 2007). In their study, there was an effect of experience, where non-native listeners only showed contextual predictability-related gains in speech recognition during a clear speech condition, while native listeners benefited from semantic predictability in clear and plain (i.e., relatively lower intelligibility) speech. In our sentence recognition study, while all sentences were presented in noise, they had high semantic predictability. While the participants had heterogeneous language backgrounds, if experience is an integral factor in a listener's ability to utilize contextual cues at the sentence level, there may be an interaction by which familiarization to accented speech further increases the top-down benefit received from predictable sentences. A future study could include two conditions varying in predictability to help isolate the benefit of semantic cues.

Finally, it may also be the case that other methodologies could be more sensitive for detecting familiarization. While familiarization was measured through the recognition of whole keywords in this case, similar research has focused on more subtle perceptual adaptation to novel speech with alternative methods. These range from the behavioral, such as measurement of reaction time to assess adaptation to accented speech (e.g., Janse & Adank, 2012), to the electrophysiological, such as the use of EEG recordings to detect accent adaptation (e.g., Bujok et al., 2021). As familiarization to unfamiliar

speech is a complex perceptual phenomenon, several forms of measurement could support our understanding of the underlying mechanisms.

Benefit of Audiovisual Cues for Sentence Recognition Accuracy

The availability of AV cues during a passive speaker exposure block was hypothesized to support greater familiarization to a novel speaker than auditory-only exposure. This hypothesis was based on relevant literature showing that AV integration supports accurate speech perception (Skipper et al., 2007; Sumbly & Pollack, 1954), including recognition of accented speech (Zheng & Samuel, 2019). Statistical analysis found that participants in the AV condition of the experiment did not show significantly greater gains in their keyword recognition accuracy following exposure compared to participants in the auditory-only condition. This is reflected in the average recognition accuracy between Blocks 1 and 2, where participants in the AV condition of the experiment showed a change in task accuracy immediately pre- and post-exposure to the speaker (+3.9%) that was comparable to those in the auditory-only condition (+3.3%), indicating no immediate benefit of passive exposure to AV cues for speech recognition.

The lack of an immediate boost on familiarization from AV cues is unexpected, given that an advantage of AV integration for initial foreign-accented speech learning has previously been reported (Janse & Adank, 2012). In a study of older adults adapting to a novel accent with AV or auditory-only cues, the authors found an early benefit for the AV group, who showed more rapid adaptation to the accented speech than their peers in the auditory-only group. However, as both groups demonstrated similar levels of overall adaptation as assessed by reaction time, the authors concluded that there was no benefit of AV cues for perceptual adaptation beyond the initial stage. As reaction time was not considered in our experiment, it's not clear whether passive AV exposure to the speaker benefited perceptual adaptation at the level of stimulus processing. In general, the finding that both groups

demonstrated equivalent familiarization to the speaker aligns with the lack of AV benefit for long-term perceptual adaptation described by Janse and Adank (2012).

Interestingly, participants in the AV condition of the present experiment performed more poorly on average than those in the auditory-only condition, though this difference was not statistically significant. Participants in the AV condition performed with 4% lower sentence recognition accuracy at baseline compared to those in the auditory-only condition. Participants were randomly assigned to condition, so with a relatively small sample size, an incidental assignment of poor performers to the AV condition seems to have deflated the group average. Likewise, performance was near ceiling for more than one participant in the auditory-only condition. A greater sample size, and additional testing with increased task difficulty, could provide more detail about the effects of the task when performance is equalized at baseline for the two groups.

Effect of Phonological Processing Ability on Audiovisual Integration and Familiarization

Phonological processing ability, assessed through nonword repetition, significantly predicted the number of keywords a listener could recognize in the speech recognition task. This is not surprising, as phonological awareness has been shown to predict nonnative-accented speech recognition performance in TD adults and children (Bent & Atagi, 2016). It is plausible that auditory phonological processing ability supported participants' performance throughout the task independently of the passive exposure to the speaker. This benefit was consistent across groups, as there was no significant interaction between nonword repetition score and condition, indicating that they were equivalent in phonological processing ability.

However, the lack of a significant interaction between post-exposure keyword recognition and nonword repetition score indicates that while phonological processing ability may have provided a boost in overall task performance, it did not support passive familiarization to a novel speaker. Moreover, the lack of a significant interaction between post-exposure keyword recognition, condition, and nonword

repetition score shows that phonological processing ability did not moderate the potential benefit of AV integration for familiarization to the speaker.

The initial hypothesis that phonological processing could moderate the benefits of AV speech cues stemmed from the proposed existence of a speech-specific AV integration mechanism (Baart et al., 2014; Lalonde & Werner, 2021). As phonological processing is key to decoding and manipulating verbal speech, a speech-specific mechanism for AV stimulus integration should theoretically interact with individual phonological processing ability. While there is limited research in this area in adult populations, evidence of this relationship is found in studies of AV speech integration in children, whose phonological processing systems are in a state of flux. Jerger et al. (2009) found that AV cues influenced younger (4 year-old) and older (10-14 year-old) children, but not 5-9 year-old children, on a picture-naming task in which auditory-only and AV phonological stimuli served as distractors. The authors reported that the relatively decreased sensitivity of AV cues in the middle age group was due to phonological reorganization within that developmental period. However, the present study results do not provide sufficient evidence of a robust interaction between phonological processing ability and a speech-specific AV integration mechanism in adults.

The results of this study align more closely with arguments that phonological skills do not moderate the benefit of AV integration for speech tasks. In a study of adaptation to accented speech with active AV or auditory-only training in adults, Banks et al. (2015) found that AV stimuli did not significantly support long-term speech recognition outcomes beyond auditory-only stimuli. Given that levels of perceptual adaptation were similar across AV and auditory-only conditions, they concluded that adaptation to accented speech could be influenced by listener-specific factors, rather than external factors such as the availability of AV cues. The results of the present study build upon this discussion by concluding that phonological processing ability did not drive greater perceptual learning of accented speech given AV versus auditory-only exposure. As a listener-specific factor, phonological processing

ability may not have an overt effect on the amount of benefit derived from AV cues for speech recognition.

In a relevant experiment, Gijbels et al. (2021) administered a closed-set speech recognition task to first graders and examined the effect of individual language skills, such as vocabulary knowledge and phonological awareness, on the extent of benefit derived from AV cues. The authors similarly found that phonological awareness did not impact the level of benefit participants gained from AV integration. In fact, they found that children with greater phonological awareness showed less benefit from AV cues, though this was statistically non-significant. Similar to how AV cues become more supportive toward speech perception in degraded auditory conditions (e.g., Sumbly & Pollack, 1954), it is possible that AV cues could be most beneficial for speech perception in the context of poor auditory phonological skills. However, additional research is needed to determine whether this is true, and under what environmental and task-specific circumstances.

Alternative findings have shown promise for the use of AV cues to support phonological skills in children with language disorders. Heikkilä et al. (2018) found that over a period of six weeks, children with specific language impairment assigned to an AV speech training condition showed greater improvement than their peers in auditory-only training on a later test of nonword repetition. They concluded that visual articulatory cues could reinforce phonological representations. The present study explored this relationship from a different angle, examining whether phonological processing skills could moderate the expected benefit of AV speech cues for familiarization to speech. While there was a positive effect of phonological processing ability on keyword recognition, phonological processing ability was not found to moderate the boost of AV stimuli for familiarization. This finding does not rule out the benefit of AV integration for phonological skill acquisition. Additionally, as we did not recruit participants with specific diagnoses of language disorders, the findings of our study do not necessarily generalize to populations with language disorders. Further investigation of the relationship between

these mechanisms will be informative for designing effective interventions for phonological skill development.

Limitations and Future Directions

There are limitations to note in the present study. A larger sample size would help to increase generalizability of the findings to a greater population. Additionally, while phonological processing ability was considered along a continuum in the analysis, it may be informative to more specifically recruit participants with diagnosed language impairments, or to otherwise separate participants by a binary measure of high and low nonword repetition performance for group comparison. This approach to analysis could be useful for investigating familiarization to speech in noise by language ability, informing clinical understanding.

Performance across the task was high for most participants. The SNR of noise presented during the sentence recognition task could be calibrated to each participant's listening threshold as described in Zhao et al. (2022) to ensure appropriate difficulty, helping to prevent a ceiling effect. Especially given the longstanding consensus that AV benefits for speech recognition are greatest in degraded auditory conditions, any existing effect of AV cues for speech familiarization may be more robust at particularly challenging signal-to-noise-ratios (Sumby & Pollack, 1954). It could be beneficial to more carefully calibrate the audio presentation within the task to maintain greater control over key perceptual factors.

Participants were screened for nonverbal intelligence, and their phonological processing ability as assessed via nonword repetition was included in the statistical analysis. However, there are many other listener-specific factors that could impact sentence recognition. For example, it is known that listening to speech in noise is cognitively taxing and draws on a complex network of functions including verbal working memory (Peelle, 2018). Listener attention and fatigue were not measured in the present study, but they may have modulated participant performance at the individual level. To account for listener-specific factors that may contribute to speech recognition, a battery of additional tests across

the domains of language and cognition should be administered to participants, including measures of vocabulary knowledge and working memory.

Finally, in the context of our sentence repetition task specifically, it could be informative to transcribe participant responses using the International Phonetic Alphabet to more closely observe the type and frequency of phonemic errors rather than whole-word omissions or substitutions. IPA transcription and phonemic error analysis could boost the sensitivity of the sentence repetition task for detecting differences in AV speech recognition at a high linguistic level on the basis of phonological processing ability.

Conclusion

In a speech recognition task assessing familiarization to accented speech in noise, adult participants generally became more accurate at reporting keywords with experience, though this was not driven by passive exposure to the speaker. Participants who were assigned to an audiovisual passive exposure condition showed similar levels of post-exposure recognition than their peers in an auditory-only condition. This finding aligns with previous research showing that while AV cues may increase intelligibility in noisy listening conditions, they do not support long-term familiarization to speech beyond auditory-only stimulus presentation. Finally, while phonological processing ability predicted overall speech recognition accuracy, it was not found to moderate the effect of AV integration on speech recognition. These findings contribute to the discussion of how listener-specific factors may play a role in AV integration for speech recognition.

While phonological processing ability influences several aspects of communication, it is only part of the greater constellation of cognitive and linguistic skills used to navigate auditory and written language. If phonological processing was the single key factor driving language ability, we would expect individuals with language disorders such as DD to have systematic challenges with verbal production due to poor phonological processing skills; however, this is not the case (Ramus & Ahissar, 2012). As such,

while an individual may show poor phonological awareness in a nonword or single-word phonological decoding task, this does not solely account for disorders such as DD, nor does it necessarily map to real-world functional linguistic aptitude. Further research is needed in this area to uncover exactly how AV integration for speech is utilized in individuals with language impairment, but existing work suggests a complex mechanism that may be supportive of success in non-laboratory settings.

The hope for this line of research, beyond the gain of theoretical knowledge, is to improve the quality of life of individuals with language disorders through more sensitive diagnosis procedures and targeted intervention programs. Examining listener-specific factors in isolation is one way of developing a greater understanding of how perceptual mechanisms support speech recognition. However, it is important to take a holistic view of the individual and their access to multidimensional language functions such as AV integration when trying to support their overall language ability. Future research with ecologically valid methods will guide our understanding of these processes.

References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2018). Gorilla Experiment Builder. <https://gorilla.sc/>
- Audacity Team (2024). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 3.4.2. Retrieved February 17, 2024 from <http://audacity.sourceforge.net/>
- Baart, M., Stekelenburg, J. J., & Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia*, 53, 115-121. doi: 10.1016/j.neuropsychologia.2013.11.011
- Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Audiovisual cues benefit recognition of accented speech in noise but not perceptual adaptation. *Frontiers in Human Neuroscience*, 9, 422. <https://doi.org/10.3389/fnhum.2015.00422>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.
- Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, 13(3), 108–112. <https://doi.org/10.3109/03005367909078884>
- Bent, T., & Atagi, E. (2016). Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: The role of phonological processing skills. *Language and Speech*, 60(1), 110–122. <https://doi.org/10.1177/0023830916645374>
- Boersma, P. & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Version 6.2, retrieved 17 February 2024 from <http://www.praat.org/>

- Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research*, 55(1), 290–305.
[https://doi.org/10.1044/1092-4388\(2011/10-0349\)](https://doi.org/10.1044/1092-4388(2011/10-0349))
- Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes*, 27(7-8), 1039–1055.
<https://doi.org/10.1080/01690965.2011.610596>
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. <https://doi.org/10.1121/1.2642103>
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>
- Brown, L., Sherbenou, R.J., & Johnsen, S.K. (1997). Test of Nonverbal Intelligence, Third Edition Examiner's Manual. Austin, TX: Pro-Ed.
- Bujok, R. Bultena, S., McQueen, J. M., & Broersma, M. (2021). Accent adaptation through error-based learning. Talk presented at EDLL 2021 - International Conference on Error-Driven Learning in Language. Tübingen, Germany.
https://pure.mpg.de/rest/items/item_3292779_5/component/file_3317265/content
- Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 363(1493), 1001–1010.
<https://doi.org/10.1098/rstb.2007.2155>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647–3658. <https://doi.org/10.1121/1.1815131>

- Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*, 42, 37–48. <https://doi.org/10.1016/j.cogdev.2017.02.004>
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146. <https://doi.org/10.1044/jslhr.4105.1136>
- Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language*, 177-178, 7–17. <https://doi.org/10.1016/j.bandl.2018.01.003>
- Ehrhorn, A., Adlof, S., Fogerty, D., & Laing, S. (2021). Probing phonological processing differences in nonword repetition for children with separate or co-occurring dyslexia and developmental language disorder. *Scientific Studies of Reading*, 25(6), 486-503. <https://doi.org/10.1080/10888438.2020.1849223>
- Felton, R. H., Naylor, C. E., & Wood, F. B. (1990). Neuropsychological profile of adult dyslexics. *Brain and Language*, 39(4), 485–497. [https://doi.org/10.1016/0093-934x\(90\)90157-c](https://doi.org/10.1016/0093-934x(90)90157-c)
- Fletcher, S. G. (1972). Contingencies for bioelectronic modification of nasality. *The Journal of Speech and Hearing Disorders*, 37(3), 329–346. <https://doi.org/10.1044/jshd.3703.329>
- Gijbels, L., Yeatman, J. D., Lalonde, K., & Lee, A. K. C. (2021). Audiovisual speech processing in relationship to phonological and vocabulary skills in first graders. *Journal of Speech, Language, and Hearing Research*, 64(12), 5022-5040. https://doi.org/10.1044/2021_JSLHR-21-00196
- Gijbels, L., Lee, A. K. C., & Yeatman, J. D. (2024). Children with developmental dyslexia have equivalent audiovisual speech perception performance but their perceptual weights differ. *Developmental Science*, 27(1), e13431. <https://doi.org/10.1111/desc.13431>

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speakers and listener effects. *Speech Communication*, 52(11), 996-1009.

<https://doi.org/10.1016/j.specom.2010.05.003>

Heikkilä, J., Lonka, E., Meronen, A., Tuovinen, S., Eronen, R., Leppänen, P. H., Richardson, U., Ahonen, T., & Tiippa, K. (2018). The effect of audiovisual speech training on the phonological skills of children with specific language impairment (SLI). *Child Language Teaching and Therapy*, 34(3), 269-287.

<https://doi.org/10.1177/0265659018793697>

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, 76(3), 598-613.

<https://doi.org/10.1111/j.1467-8624.2005.00866.x>

Holt, R. F., & Bent, T. (2017). Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research*, 60(1), 223-230.

https://doi.org/10.1044/2016_JSLHR-H-16-0014

International Dyslexia Association: <https://dyslexiaida.org/testing-and-evaluation/>

Irwin, J., & DiBlasi, L. (2017). Audiovisual speech perception: A new approach and implications for clinical populations. *Language and Linguistics Compass*, 11(3), 77-91.

<https://doi.org/10.1111/lnc3.12237>

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, 65(8), 1563-1585. <https://doi.org/10.1080/17470218.2012.658822>

Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: A new multimodal picture-word task. *Journal of Experimental Child Psychology*, 102(1), 40-59.

<https://doi.org/10.1016/j.jecp.2008.08.002>

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, 20(2), 693-712.

[https://doi.org/10.1016/S1053-8119\(03\)00343-4](https://doi.org/10.1016/S1053-8119(03)00343-4)

Kaganovich, N., & Ancel, E. (2019). Different neural processes underlie visual speech perception in school-age children and adults: An event-related potentials study. *Journal of experimental child psychology*, 184, 98–122. <https://doi.org/10.1016/j.jecp.2019.03.009>

Kaganovich, N., Schumaker, J., & Christ, S. (2021). Impaired audiovisual representation of phonemes in children with developmental language disorder. *Brain Sciences*, 11(4), 507.

<https://doi.org/10.3390/brainsci11040507>

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11442–11445.

<https://doi.org/10.1073/pnas.0804275105>

Lalonde, K., & Holt, R. F. (2016). Audiovisual speech perception development at varying levels of perceptual processing. *The Journal of the Acoustical Society of America*, 139(4), 1713.

<https://doi.org/10.1121/1.4945590>

Lalonde, K., & McCreery, R. W. (2020). Audiovisual enhancement of speech perception in noise by school-age children who are hard of hearing. *Ear and Hearing*, 41(4), 705–719.

<https://doi.org/10.1097/AUD.0000000000000830>

Lalonde, K., & Werner, L. A. (2021). Development of the mechanisms underlying audiovisual speech perception benefit. *Brain Sciences*, 11(1), 49. <https://doi.org/10.3390/brainsci11010049>

Lansford, K. L., Luhrsen, S., Ingvalson, E. M., & Borrie, S. A. (2018). Effects of familiarization on intelligibility of dysarthric speech in older adults with and without hearing loss. *American Journal of Speech-Language Pathology*, 27(1), 91–98. https://doi.org/10.1044/2017_AJSLP-17-0090

- Lara-Díaz, M. F., Beltrán Rojas, J. C., & Aponte Rippe, Y. (2024). Visual attention and phonological processing in children with developmental language disorder. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1386279>
- Leybaert, J., Macchi, L., Huyse, A., Champoux, F., Bayard, C., Colin, C., & Berthommier, F. (2014). Atypical audio-visual speech perception and McGurk effects in children with specific language impairment. *Frontiers in Psychology*, 5, 422. <https://doi.org/10.3389/fpsyg.2014.00422>
- Mathur, A., Schultz, D., & Wang, Y. (2020). Neural bases of phonological and semantic processing in early childhood. *Brain Connectivity*, 10(5), 212–223. <https://doi.org/10.1089/brain.2019.0728>
- The MathWorks Inc. (2024). MATLAB version: 23.2.0.2515942 (R2023b), Natick, Massachusetts: The MathWorks Inc. <https://www.mathworks.com>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Meronen, A., Tiippana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*, 56(1), 211–221. [https://doi.org/10.1044/1092-4388\(2012/11-0270\)](https://doi.org/10.1044/1092-4388(2012/11-0270))
- Miller, C. A., & Gilbert, E. (2008). Comparison of performance on two nonverbal intelligence tests by adolescents with and without language impairment. *Journal of Communication Disorders*, 41(4), 358–371. <https://doi.org/10.1016/j.jcomdis.2008.02.003>
- Mudge, L. (2019, September 12). *Visualizing mixed-effects models*. R Tutorials. https://lmudge13.github.io/sample_code/mixed_effects.html
- Nordgren, P. M. (2015). Phonological development in a child with autism spectrum condition: Case study of an intervention. *Journal of Interactional Research in Communication Disorders*, 6(1), 25-51. <https://doi.org/10.1558/jircd.v6i1.25>

- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*(2), 204–214.
<https://doi.org/10.1097/AUD.0000000000000494>
- Peelle, J. E. (2019). The neural basis for auditory and audiovisual speech perception. In *The Routledge Handbook of Phonetics*. Routledge.
- Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development, 61*(6), 1753–1778.
- Pons, F., Andreu, L., Sanz-Torrent, M., Buil-Legaz, L., & Lewkowicz, D. J. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *Journal of Child Language, 40*(3), 687–700.
<https://doi.org/10.1017/S0305000912000189>
- Pons, F., Sanz-Torrent, M., Ferinu, L., Birulés, J., & Andreu, L. (2018). Children with SLI can exhibit reduced attention to a talker's mouth. *Language Learning, 68*(1), 180-192.
<https://doi.org/10.1111/lang.12276>
- Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.
- Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology, 29*(1-2), 104–122. <https://doi.org/10.1080/02643294.2012.677420>
- Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(9), 3578-3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Seijdel, N., Schoffelen, J. M., Hagoort, P., & Drijvers, L. (2024). Attention drives visual processing and audiovisual integration during multimodal communication. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 44(10), e0870232023.

<https://doi.org/10.1523/JNEUROSCI.0870-23.2023>

Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. *Frontiers in Psychology*, 5, 323.

<https://doi.org/10.3389/fpsyg.2014.00323>

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, 11(2), 306–320. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-7687.2008.00677.x)

[7687.2008.00677.x](https://doi.org/10.1111/j.1467-7687.2008.00677.x)

Phonological processing. (n.d.). Retrieved April 23, 2025 from American Speech-Language-Hearing Association, Written Language Disorders. <https://www.asha.org/practice-portal/clinical-topics/written-language-disorders/phonological-processing/>

Share, D. L. (2021). Common misconceptions about the phonological deficit theory of dyslexia. *Brain Sciences*, 11(11), 1510. <https://doi.org/10.3390/brainsci11111510>

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception.

Cerebral Cortex, 17, 2387-2399. <https://doi:10.1093/cercor/bhl147>

Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J., & Lewkowicz, D. (2012). Development of audiovisual speech perception. In *Multisensory Dev* (pp. 207-228).

<https://doi.org/10.1093/acprof:oso/9780199586059.003.0009>

Sumby, W. H. & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212-215. <https://doi.org/10.1121/1.1907309>

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, 5.

<https://doi.org/10.3389/fpsyg.2014.00725>

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, 16(3), 457–472.

<https://doi.org/10.1080/09541440340000268>

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition*, 96(1), B13–B22. <https://doi.org/10.1016/j.cognition.2004.10.004>

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, 79(2), 396–403.

<https://doi.org/10.3758/s13414-016-1238-9>

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. *The Journal of the Acoustical Society of America*, 152(6), 3216.

<https://doi.org/10.1121/10.0015262>

von Seth, J., Aller, M., & Davis, M. H. (2025). Unimodal speech perception predicts stable individual differences in audiovisual benefit for phonemes, words and sentences. *The Journal of the Acoustical Society of America*, 157, 1554-1576. <https://doi.org/10.1121/10.0034846>

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192–212.

<https://doi.org/10.1037/0033-2909.101.2.192>

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). Comprehensive Test of Phonological Processing-Second Edition (CTOPP-2). Austin, TX: PRO-ED.

Weiss, Y., Cweigenberg, H. G., & Booth, J. R. (2018). Neural specialization of phonological and semantic processing in young children. *Human Brain Mapping*, 39(11), 4334–4348.

<https://doi.org/10.1002/hbm.24274>

Weiss-Croft, L. J., & Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22 years of functional MRI. *NeuroImage*, 123, 269–281.

<https://doi.org/10.1016/j.neuroimage.2015.07.046>

Winn, M. (2013). Scale intensity of all sounds in a directory. Praat Scripts,

<http://www.mattwinn.com/praat.html>

Wolk, L., Edwards, M. L., & Brennan, C. (2016). Phonological difficulties in children with autism: An overview. *Speech, Language and Hearing*, 19(2), 121–129.

<https://doi.org/10.1080/2050571X.2015.1133488>

Ye, Z., Rüsseler, J., Gerth, I., & Münte, T. F. (2017). Audiovisual speech integration in the superior temporal region is dysfunctional in dyslexia. *Neuroscience*, 356, 1–10.

<https://doi.org/10.1016/j.neuroscience.2017.05.017>

Zhao, S., Brown, C. A., Holt, L. L., & Dick, F. (2022). Robust and efficient online auditory psychophysics.

Trends in Hearing, 26, 1-24. <https://doi.org/10.1177/23312165221118792>

Zheng, Y., & Samuel, A. G. (2019). How much do visual cues help listeners in perceiving accented

speech? *Applied Psycholinguistics*, 40(1), 93–109. <https://doi.org/10.1017/S0142716418000462>

Zoccolotti, P. (2022). Success is not the entire story for a scientific theory: The case of the phonological

deficit theory of dyslexia. *Brain Sciences*, 12(4), 425. <https://doi.org/10.3390/brainsci12040425>

Appendix A

Bamford-Kowal-Bench (BKB) Sentence Lists and Keywords

List	Item	Sentence	Keyword 1	Keyword 2	Keyword 3	Keyword 4	# Keywords
1	1	The clown had a funny face .	clown	funny	face	--	3
1	2	The car engine's running .	car	engine	running	--	3
1	3	She cut with her knife .	cut	knife	--	--	2
1	4	Children like strawberries .	children	like	strawberries	--	3
1	5	The house had nine rooms .	house	nine	rooms	--	3
1	6	They're buying some bread .	buying	some	bread	--	3
1	7	The green tomatoes are small .	green	tomatoes	small	--	3
1	8	They're looking at the clock .	looking	clock	--	--	2
1	9	He played with his train .	played	train	--	--	2
1	10	The postman shut the gate .	postman	shut	gate	--	3
1	11	The bag bumps on the ground .	bag	bumps	ground	--	3
1	12	The boy did a handstand .	boy	did	handstand	--	3
1	13	A cat sits on the bed .	cat	sits	bed	--	3
1	14	The lorry carried fruit .	lorry	carried	fruit	--	3
1	15	The rain came down .	rain	came	down	--	3
1	16	The ice cream was pink .	ice	cream	pink	--	3
2	1	The ladder's near the door .	ladder	near	door	--	3
2	2	They had a lovely day .	lovely	day	--	--	2
2	3	The ball went into the goal .	ball	went	goal	--	3
2	4	The old gloves were dirty .	old	gloves	dirty	--	3
2	5	He cut his finger .	cut	finger	--	--	2
2	6	The thin dog was hungry .	thin	dog	hungry	--	3
2	7	The boy knew the game .	boy	knew	game	--	3
2	8	Snow falls at Christmas .	snow	falls	Christmas	--	3
2	9	She's taking her coat .	taking	coat	--	--	2
2	10	The police chased the car .	police	chased	car	--	3
2	11	A mouse ran down the hole .	mouse	ran	down	hole	4
2	12	The lady's making a toy .	lady	making	toy	--	3
2	13	Some sticks were under the tree .	Some	sticks	under	tree	4
2	14	The little baby sleeps .	little	baby	sleeps	--	3
2	15	They're watching the train .	watching	train	--	--	2
2	16	The school finished early .	school	finished	early	--	3
3	1	The glass bowl broke .	glass	bowl	broke	--	3
3	2	The dog played with a stick .	dog	played	stick	--	3
3	3	The kettle's quite hot .	kettle	quite	hot	--	3

3	4	The farmer keeps a bull .	farmer	keep	bull	--	3
3	5	They say some silly things .	say	some	silly	things	4
3	6	The lady wore a coat .	lady	wore	coat	--	3
3	7	The children are walking home .	children	walking	home	--	3
3	8	He needed his holiday .	needed	holiday	--	--	2
3	9	The milk came in a bottle .	milk	came	bottle	--	3
3	10	The man cleaned his shoes .	man	cleaned	shoes	--	3
3	11	They ate the lemon jelly .	ate	lemon	jelly	--	3
3	12	The boy's running away .	boy	running	away	--	3
3	13	Father looked at the book .	father	looked	book	--	3
3	14	She drinks from her cup .	drink	cup	--	--	2
3	15	The room's getting cold .	room	getting	cold	--	3
3	16	A girl kicked the table .	girl	kicked	table	--	3
4	1	The wife helped her husband .	wife	helped	husband	--	3
4	2	The machine was quite noisy .	machine	quite	noisy	--	3
4	3	The old man worries .	old	man	worries	--	3
4	4	A boy ran down the path .	boy	ran	down	path	4
4	5	The house had a nice garden .	house	nice	garden	--	3
4	6	She spoke to her son .	spoke	son	--	--	2
4	7	They're crossing the street .	crossing	street	--	--	2
4	8	Lemons grow on trees .	lemons	grow	trees	--	3
4	9	He found his brother .	found	brother	--	--	2
4	10	Some animals sleep on straw .	Some	animals	sleep	straw	4
4	11	The jam jar was full .	jam	jar	full	--	3
4	12	They're kneeling down .	kneeling	down	--	--	2
4	13	The girl lost her doll .	girl	lost	doll	--	3
4	14	The cook's making a cake .	cook	making	cake	--	3
4	15	The child grabs the toy .	child	grabs	toy	--	3
4	16	The mud stuck on his shoe .	mud	stuck	shoe	--	3
5	1	The bath towel was wet .	bath	towel	wet	--	3
5	2	The matches lie on the shelf .	matches	lie	shelf	--	3
5	3	They're running past the house .	running	past	house	--	3
5	4	The train had a bad crash .	train	bad	crash	--	3
5	5	The kitchen sink's empty .	kitchen	sink	empty	--	3
5	6	A boy fell from the window .	boy	fell	window	--	3
5	7	She used her spoon .	used	spoon	--	--	2
5	8	The park's near the road .	park	near	road	--	3
5	9	The cook cut some onions .	cook	cut	some	onions	4
5	10	The dog made an angry noise .	dog	made	angry	noise	4
5	11	He's washing his face .	washing	face	--	--	2
5	12	Somebody took the money .	somebody	took	money	--	3

5	13	The light went out.	light	went	out	--	3
5	14	They wanted some potatoes.	wanted	some	potatoes	--	3
5	15	The naughty girl's shouting.	naughty	girl	shouting	--	3
5	16	The cold milk's in a jug.	cold	milk	jug	--	3
6	1	The paint dripped on the ground.	paint	dripped	ground	--	3
6	2	The mother stirs the tea.	mother	stirs	tea	--	3
6	3	They laughed at his story.	laughed	story	--	--	2
6	4	Men wear long trousers.	men	wear	long	trousers	4
6	5	The small boy was asleep.	small	boy	asleep	--	3
6	6	The sun melted the snow.	sun	melted	snow	--	3
6	7	The father's coming home.	father	coming	home	--	3
6	8	She had her pocket money.	pocket	money	--	--	2
6	9	The lorry drove up the road.	lorry	drove	road	--	3
6	10	He's bringing his raincoat.	bringing	raincoat	--	--	2
6	11	The lady goes to the shop.	lady	goes	shop	--	3
6	12	A sharp knife's dangerous.	sharp	knife	dangerous	--	3
6	13	They took some food.	took	some	food	--	3
6	14	The clever girls are reading.	clever	girls	reading	--	3
6	15	The broom stood in the corner.	broom	stood	corner	--	3
6	16	The woman tidied her house.	woman	tidied	house	--	3
7	1	The children dropped the bag.	children	dropped	bag	--	3
7	2	The dog came back.	dog	came	back	--	3
7	3	The floor looked clean.	floor	looked	clean	--	3
7	4	She found her purse.	found	purse	--	--	2
7	5	The fruit lies on the ground.	fruit	lies	ground	--	3
7	6	Mother fetches a saucepan.	mother	fetches	saucepan	--	3
7	7	They washed in cold water.	washed	cold	water	--	3
7	8	The young people are dancing.	young	people	dancing	--	3
7	9	The bus went early.	bus	went	early	--	3
7	10	They had two empty bottles.	two	empty	bottles	--	3
7	11	A ball's bouncing along.	ball	bouncing	along	--	3
7	12	The father forgot the bread.	father	forgot	bread	--	3
7	13	The girl has a picture book.	girl	picture	book	--	3
7	14	The orange was quite sweet.	orange	quite	sweet	--	3
7	15	He's holding his nose.	holding	nose	--	--	2
7	16	The new road's on the map.	new	road	map	--	3
8	1	The boy forgot his book.	boy	forgot	book	--	3
8	2	A friend came for lunch.	friend	came	lunch	--	3
8	3	The match boxes are empty.	match	boxes	empty	--	3
8	4	He climbed his ladder.	climbed	ladder	--	--	2

8	5	They heard a funny noise .	heard	funny	noise	--	3
8	6	The jug stood on the shelf .	jug	stood	shelf	--	3
8	7	The ball broke the window .	ball	broke	window	--	3
8	8	They're shopping for cheese .	shopping	cheese	--	--	2
8	9	The family bought a house .	family	bought	house	--	3
8	10	The pond water's dirty .	pond	water	dirty	--	3
8	11	Police are clearing the road .	police	clearing	road	--	3
8	12	The bus stopped suddenly .	bus	stopped	suddenly	--	3
8	13	She writes to her brother .	writes	brother	--	--	2
8	14	The footballer lost a boot .	footballer	lost	boot	--	3
8	15	The three girls are listening .	three	girls	listening	--	3
8	16	The coat lies on a chair .	coat	lies	chair	--	3
9	1	The book tells a story .	book	tells	story	--	3
9	2	The young boy left home .	young	boy	left	home	4
9	3	They're climbing the tree .	climbing	tree	--	--	2
9	4	She stood near her window .	stood	near	window	--	3
9	5	The table has three legs .	table	three	legs	--	3
9	6	A letter fell on the mat .	letter	fell	mat	--	3
9	7	The five men are working .	five	men	working	--	3
9	8	The shoes were very dirty .	shoes	very	dirty	--	3
9	9	He listens to his father .	listens	father	--	--	2
9	10	They went on holiday .	went	holiday	--	--	2
9	11	Baby broke his mug .	baby	broke	mug	--	3
9	12	The lady packed her bag .	lady	packed	bag	--	3
9	13	The dinner plate's hot .	dinner	plate	hot	--	3
9	14	The train's moving fast .	train	moving	fast	--	3
9	15	The child drank some milk .	child	drank	some	milk	4
9	16	The car hit a wall .	car	hit	wall	--	3
10	1	A tea towel's by the sink .	tea	towel	sink	--	3
10	2	The cleaner used a broom .	cleaner	used	broom	--	3
10	3	She looked in her mirror .	looked	mirror	--	--	2
10	4	The good boy's helping .	good	boy	helping	--	3
10	5	They followed the path .	followed	path	--	--	2
10	6	The kitchen clock was wrong .	kitchen	clock	wrong	--	3
10	7	Someone's crossing the road .	someone	crossing	road	--	3
10	8	The postman brings a letter .	postman	bring	letter	--	3
10	9	The dog jumped on the chair .	dog	jumped	chair	--	3
10	10	They're cycling along .	cycling	along	--	--	2
10	11	He broke his leg .	broke	leg	--	--	2
10	12	The milk was by the front door .	milk	front	door	--	3

10	13	The shirts hang in the cupboard .	shirts	hang	cupboard	--	3
10	14	The ground was too hard .	ground	hard	--	--	2
10	15	The buckets hold water .	buckets	hold	water	--	3
10	16	The chicken laid some eggs .	chicken	laid	some	eggs	4
11	1	The family like fish .	family	like	fish	--	3
11	2	Sugar's very sweet .	sugar	very	sweet	--	3
11	3	The baby lay on a rug .	baby	lay	rug	--	3
11	4	The washing machine broke .	washing	machine	broke	--	3
11	5	They're clearing the table .	clearing	table	--	--	2
11	6	The cleaner swept the floor .	cleaner	swept	floor	--	3
11	7	A grocer sells butter .	grocer	sells	butter	--	3
11	8	The milkman drives a small van .	milkman	drives	small	van	4
11	9	The bath water was warm .	bath	water	warm	--	3
11	10	He's reaching for his spoon .	reaching	spoon	--	--	2
11	11	She hurt her hand .	hurt	hand	--	--	2
11	12	The boy slipped on the stairs .	boy	slipped	stairs	--	3
11	13	They're staying for supper .	staying	supper	--	--	2
11	14	The girl held a mirror .	girl	held	mirror	--	3
11	15	The cup stood on a saucer .	cup	stood	saucer	--	3
11	16	The cows went to market .	cows	went	market	--	3
12	1	The boy got into trouble .	boy	got	trouble	--	3
12	2	They're going out .	going	out	--	--	2
12	3	The football hit the goalpost .	football	hit	goalpost	--	3
12	4	He paid his bill .	paid	bill	--	--	2
12	5	The teacloth's quite wet .	teacloth	quite	wet	--	3
12	6	A cat jumped off the fence .	cat	jumped	fence	--	3
12	7	The baby has blue eyes .	baby	blue	eyes	--	3
12	8	Mother made some curtains .	mother	made	some	curtains	4
12	9	They sat on a wooden bench .	sat	wooden	bench	--	3
12	10	The oven's too hot .	oven	hot	--	--	2
12	11	The girl caught a cold .	girl	caught	cold	--	3
12	12	The raincoat's hanging up .	raincoat	hanging	up	--	3
12	13	She brushed her hair .	brushed	hair	--	--	2
12	14	The two children are laughing .	two	children	laughing	--	3
12	15	The man tied his scarf .	man	tied	scarf	--	3
12	16	The flower stands in a pot .	flower	stands	pot	--	3
<i>n</i>			192	192	153	13	550

Appendix B*Zoo Passage*

Look at this book with us. It's a story about a zoo. This is where bears go. Today it is very cold out of doors but we see a cloud overhead that's a pretty, white, fluffy shape. We heard that straw covers the floor of cages to keep the chill away; yet a deer walks through the trees with her head high. They feed seeds to birds so they're able to fly.

Appendix C*Demographic and Listening History Questionnaire*

Thank you for participating in our study. Please answer the following questions to the best of your ability.

Age

Sex

Race

Ethnicity

Next →

Country where you were born

Country/countries where you grew up

What is the first language(s) you learned?

Did you speak a language other than English growing up?

What language(s) did you use to communicate with your parents while growing up?

What language(s) did you use in school?

Do you speak more than one language now?

What language(s) do you use most often now?

Do you listen to music in more than one language?

Please list the language(s) of the music you listen to.

For example: English, Korean.

Do you watch or listen to movies, TV, or other videos in more than one language?

Please list the language(s) of the videos you watch.

For example: English, Spanish.

Next →

Please list ALL languages that you have studied (including English). Please indicate your proficiency in each language on a scale from 1 to 10, and the age when you started learning.

For example: Spanish, 5/10, 13 years old

Of the languages listed above, please list the languages that you can **LISTEN** to. Please indicate your proficiency in each language on a scale from 1 to 10, and the age when you started learning.

For example: Spanish, 5/10, 13 years old

Of the languages listed above, please list the languages that you can **SPEAK**. Please indicate your proficiency in each language on a scale from 1 to 10, and the age when you started learning.

For example: Spanish, 5/10, 13 years old

Of the languages listed above, please list the languages that you can **READ**. Please indicate your proficiency in each language on a scale from 1 to 10, and the age when you started learning.

For example: Spanish, 5/10, 13 years old

Of the languages listed above, please list the languages that you can **WRITE**. Please indicate your proficiency in each language on a scale from 1 to 10, and the age when you started learning.

For example: Spanish, 5/10, 13 years old

Of the languages listed above, please list where you learned them.
For example: Spanish classroom first, then Duolingo

Next →

Are you right or left handed?

Do you have any difficulties with language or reading?

Please describe your difficulties with language or reading.

As far as you know, is your hearing normal?

Have you ever had any music lessons (private or group lessons)?

Please indicate when you started, what instrument(s) you studied, the type of lesson, how often, how long you studied, and your current proficiency from 1 (not proficient) to 7 (extremely proficient).

Next →