How does our visual mind retain information through our ears? A study of suprasegmental features, aphantasia, and comprehension

Vanessa Silveira

Department of Speech, Language, and Hearing Sciences

Senior Honors Thesis

March 13, 2026

Committee Members

Thesis Advisor: Christine Brennan, Ph.D., CCC-SLP (SLHS)

Honors Council Representative: Allison Hilger, Ph.D., CCC-SLP (SLHS)

Outside Reader: Raichele Farrelly, Ph.D. (LING)

Abstract

The increasing use of accelerated playback speeds in audiobooks raises important questions about how temporal compression affects speech comprehension. Although faster playback enhances convenience and efficiency, prior research indicates that comprehension declines as speech rate increases. In spoken language, suprasegmental features (SSFs), including pitch, intonation, stress, rhythm, and pausing, are essential for signaling emphasis, emotional tone, and narrative structure. Compression of acoustic information at higher playback speeds may reduce access to prosodic cues that support higher-level interpretation. Despite evidence of general comprehension decline under acceleration, relatively little research has examined how playback speed specifically influences suprasegmental processing or how individual cognitive differences shape listeners' adaptation. Narrative comprehension often involves the construction of internal mental imagery; however, imagery vividness varies substantially across individuals, raising the possibility that imagery ability moderates' sensitivity to prosodic disruption. The present study investigated how audiobook playback speeds (1.0x, 1.4x, and 1.8x) affect comprehension of SSF, with visual imagery ability assessed using the Vividness of Visual Imagery Questionnaire (VVIQ). Results revealed a modest association between imagery vividness and accuracy on prosody-dependent questions, particularly at 1.4x speed. Whereas factual comprehension remained relatively stable across speeds, SSF-dependent comprehension was more susceptible to increased playback rate and imagery variability. These findings support a component-based model of auditory comprehension in which prosodic processing relies on cognitive resources that are especially vulnerable under heightened temporal demands.

Key words: Aphantasia, audiobook, visualization, comprehension, playback speed, memory, psychoacoustics

Introduction

Since the emergence of audiobooks in the 1990s, listeners have increasingly used adjustable playback speeds to fit personal schedules and preferences (Kruez, 2023). Comprehension in this format relies not only on words but also on suprasegmental features (SSF)- including stress, intonation, rhythm, tempo, and pausing- that signal emphasis, emotion, and narrative structure (Pierrehumbert 1990; Cutler et al. 1997). While listeners can often maintain overall understanding at moderate acceleration (Foulke and Sticht 1969; Wingfield et al. 2005), less is known about whether higher playback speeds compromise access to these prosodic cues or how individual differences mental imagery influence adaptation (Paivio 1991; Zwaan and Radvansky 1998). Some listeners form vivid mental images to support comprehension, while others may rely more heavily on SSF. Despite this variability, research has not yet clarified whether prosodic information remains accessible under accelerated conditions, leaving a gap in understanding how auditory and cognitive factors jointly shape narrative processing. The present study addresses this gap by examining how varying playback speeds (1.0x, 1.4x, 1.8x) affect comprehension of SSF and whether individual differences in mental imagery moderate this effect. It is hypothesized that higher playback speeds will reduce prosodic comprehension and that stronger imagery ability will buffer listeners against this decline.

Growth in Audiobook Popularity

Over the past decade audiobooks have grown rapidly as a digital medium, reflecting broader shifts in how individuals' access and engage with written narratives (Snelling 2021). Importantly, research indicates that listening and reading activate largely overlapping cognitive systems, suggesting comparable levels of comprehension across modalities (Wolf et al. 2019; Yorio 2019). However, individual differences moderate these effects. For example, individuals

with dyslexia may benefit from auditory narratives that reduce visual decoding demands (Wimmer and Mayringer 2002), while children and adults with ADHD may continue to experience comprehension challenges due to sustained attention deficits and processing inefficiencies (Wassenberg et al. 2010; Cain and Bignell 2014). These findings highlight that modality alone does not fully account for differences in engagement or understanding, pointing to the importance of audiobook-specific variables. One such variable is playback speed, which can shape processing efficiency and attention allocation. Additionally, narration- including tone, pacing, prosody, and emotional expression- represents a potentially critical factor influencing listener engagement, immersion, and interpretive comprehension.

Role of Audio Playback Speed in Listening Comprehension

Playback speed affects comprehension in a non-linear manner: moderate acceleration often preserves understanding, whereas excessive speed can overload cognitive processing. Studies examining accelerated audiobook playback report varying effects on comprehension. While some participants maintain understanding at speeds up to approximately 1.5x (Song et al. 2018), others show declines at moderately faster or slower rates (Ness et al. 2021; Chen et al. 2024). Beyond this range, comprehension consistently declines, with sharper decreases observed at speeds approaching or exceeding 2.0x (Goldhaber and Weaver 1968; Song et al. 2018; Chen et al. 2024). These findings align with the accelerated playback hypothesis, which proposes that listeners can compensate for increased speech rates up to a certain threshold, after which cognitive processing capacity is exceeded (Foulke and Sticht 1969). As speech rate increases, listeners initially adapt by allocating additional attentional and working memory resources; however, once processing demands surpass cognitive limits, comprehension deteriorates.

Despite the foundational work on playback speed, several limitations constrain the existing literature. First, studies vary in the speed at which comprehension begins to decline, likely due to differences in task demands, stimulus complexity, assessment methods, and participant characteristics (Song et al. 2018; Ness et al. 2021; Kıyak et al. 2023; Chen et al. 2024). Second, much research has focused on audiovisual materials, such as recorded lectures or video-based learning environments, rather than audio-only formats like audiobooks (Ritzhaupt et al. 2008; Pastore 2012; Murphy et al. 2021 Nov). Because visual cues in these materials can offset increased auditory processing demands, findings may not fully generalize to pure audiobook listening. Third, participant samples are often limited to undergraduate or neurotypical adult populations, restricting generalizability. Although some evidence suggests that individuals with attentional difficulties, such as ADHD, may prefer or benefit from altered playback speeds, research directly examining how attentional profiles interact with playback speed remains limited (Wassenberg et al. 2010; Karimi 2022). Finally, comprehension measures vary across studies, with many relying on short-term recall or multiple-choice formats, which may not fully capture deeper understanding or integration of narrative material (Ness et al. 2021).

Together, these limitations underscore the need for research that: (a) examines audio-only materials, (b) incorporates diverse participant populations, and (c) employs more comprehensive and standardized measures of comprehension. Addressing these gaps will clarify how cognitive resources interact with playback speed to support or hinder listening comprehension in audiobook contexts.

Role of Suprasegmental Features and Prosody in Listening Comprehension

In spoken narrative contexts such as audiobooks, prosody refers to the SSF of speech, including pitch, intonation, rhythm, duration, stress, and loudness. These features play a critical role in conveying emotional and linguistic information in spoken language (Cutler et al. 1997). At the sentence level, prosodic cues assist listeners in segmenting continuous speech, identifying syntactic boundaries, and resolving structural ambiguities (Streeter 1978; Christophe et al. 2003). At the discourse level, prosody also conveys emotional tone and speaker intent, which are essential for narrative comprehension and engagement (Bänziger and Scherer 2005). When prosodic features are reduced or distorted, listeners experience increased cognitive processing demands, relying more heavily on working memory and contextual inference, which can impair speech perception and comprehension (Cutler et al. 1997; Wingfield et al. 2005).

Despite the established importance of prosody in spoken language processing, limited research has examined how playback speed may influence access to suprasegmental cues. Most playback speed studies focus on overall comprehension accuracy, with less attention to how accelerated speech may compress temporal and acoustic features critical for stress rhythm, and intonation. Because audiobook narration inherently depends on expressive prosodic delivery, alterations in playback speed may interact with narrator performance in ways that influence listener engagement and understanding. Examining this interaction is therefore essential for clarifying the mechanisms underlying comprehension of differences in audiobook contexts.

Role of Narrator Performance in Listening Comprehension

Narrator performance plays a critical role in shaping listener engagement, emotional stimulation, and overall connection to audiobook content (Liu et al. 2020; Ji et al. 2024). Converging experimental evidence indicates that expressive vocal delivery enhances both affective and cognitive processing. For example, emotionally arousing narration has been shown

to improve listening comprehension regardless of valence, with positively valanced narration producing the strongest comprehension outcomes (Chung et al., 2015). Similarly, research comparing human and synthetic narration demonstrates that stories delivered by a human voice actor result in greater enjoyment, stronger mental imagery, heightened emotional arousal, improved comprehension, and superior recall (Rodero and Lucas 2021).

These findings suggest that the expressive qualities of narration are not merely aesthetic features, but meaningful contributors to cognitive processing. Narrator performance relies heavily on prosodic variation, such as pitch modulation, rhythmic pacing, stress patterns, and intonational contour, which guide both structural interpretation and emotional understanding. By signaling emphasis, emotional shifts, and narrative tension, prosodic delivery may facilitate deeper immersion and more vivid mental representation of story events. As spoken narratives unfold temporally, listeners do not simply decode linguistic content; they actively construct internal simulations of scenes, characters, and actions. Expressive narration therefore influences not only comprehension accuracy, but also the richness of imagery and experiential engagement.

Despite growing evidence that narrator expressiveness enhances comprehension and memory, limited research has examined how these performance features interact with playback speed or individual differences in imagery ability. Understanding how prosodic delivery shapes both engagement and cognitive processing is essential for clarifying the mechanisms underlying audiobook comprehension.

Role of Visual Imagery and Impact of Aphantasia on Listening Comprehension

Mental imagery plays a vital role in language comprehension by supporting the construction of mental representations, often referred to as situation models, that integrate characters, actions, and spatial relationships within a narrative. Research has shown that both

reading and listening to language can evoke visual imagery, activating brain regions associated with perceptual simulation and multimodal processing (Zwaan and Radvansky 1998; Kosslyn et al. 2001). During auditory narrative comprehension, listeners frequently report forming mental images that mirror the events being described, suggesting that imagery contributes to engagement and understanding even in the absence of visual input.

However, the ability to generate mental imagery varies substantially across individuals. While many people experience vivid visual imagery, others report weak or inconsistent imagery, and a small subset experience little to no visual imagery at all, a phenomenon known as aphantasia (Zeman et al. 2015). Aphantasia is characterized by an absence of voluntary visual imagery despite intact perception, language abilities, and general cognition. Importantly, individuals with aphantasia are still able to comprehend spoken and written language, indicating that visual imagery is not a prerequisite for understanding narrative content.

Research examining comprehension in individuals with aphantasia suggests that they may rely more heavily on non-visual strategies, such as semantic processing, verbal reasoning, or abstract conceptual representations, rather than perceptual simulation (Zeman et al. 2015; Keogh and Pearson 2018). While these alternative strategies appear sufficient for accurate comprehension, some studies suggest that imagery vividness may influence aspects of narrative engagement, memory richness, or emotional resonance rather than basic understanding (Watkins 2018; Bainbridge et al. 2021).

In the context of audiobook listening, these individual differences in imagery ability may have important implications. Audiobooks often encourage immersive listening experiences through descriptive language, character voices, and emotional prosody, which may support mental imagery formation for some listeners while providing alternative cues for those with

limited imagery. For individuals with aphantasia, elements of SSF may play a compensatory role by enhancing comprehension and engagement through auditory rather than visual channels. Together, this body of research suggests that while visual imagery can enrich listening comprehension for many individuals, it is not essential for understanding spoken narratives. Instead, listeners appear to flexibly adapt their cognitive strategies based on their imagery abilities, highlighting the importance of considering individual differences when evaluating audiobook comprehension and engagement. These findings further underscore the potential role of narrators and SSF in shaping the listening experience, particularly for individuals who do not rely on visual imagery during comprehension.

Although playback speed research has identified threshold effects in listening comprehension, and prosody research has established the importance of suprasegmental cues for speech processing, few studies have examined how accelerated playback may alter access to prosodic information in narrative audiobook contexts. Understanding this interaction may clarify the mechanisms underlying variability in listener engagement and comprehension.

Gap in Existing Literature

Given the growing prevalence of accelerated audiobook consumption and the established importance of prosody in spoken language comprehension, this gap warrants systematic investigation. The proposed study addresses this need by examining how playback speed influences comprehension that depends on prosodic cues compared to comprehension that relies less heavily on these features. By comparing listener performance across varying playback speeds and question types, this research aims to clarify whether - and to what extent - accelerated playback alters the perception and cognitive processing of suprasegmental information.

A growing body of research demonstrates that increasing playback speed can reduce

comprehension as listeners approach cognitive processing limits, particularly in audio-only contexts where auditory resources are heavily taxed (Chen et al. 2024). Faster playback speed places strain on working memory and information processing, often resulting in decreased recall and understanding. While these findings establish a general effect of playback speed on comprehension, far less is known about the specific linguistic mechanisms underlying this decline. Limited attention has been given to the perception and interpretation of SSF, which conveys essential information for comprehension beyond lexical content. Prosodic cues enable listeners to disambiguate syntactic structure, signal emphasis, and decode emotional intent (Applebaum et al. 2014). Disruptions to these features may therefore produce comprehension deficits that are distinct from generalized cognitive overload. Yet existing playback speed research rarely isolates these mechanisms. Moreover, most studies have examined playback speed in lecture or multimedia contexts, where visual information such as slides or text may partially compensate for auditory distortions, potentially masking the isolated effect of accelerated speech on prosodic processing. Audiobook listening, in contrast, depends exclusively on auditory input and relies heavily on prosodic variation for narrative structure and engagement.

Despite the growing prevalence of accelerated audiobook consumption, few studies have examined how faster playback affects comprehension in this context, or how listener performance may vary according to the demands of prosodic versus lexical processing. Finally, individual differences, including imagery ability, attention, and working memory capacity, may further moderate how listeners process and interpret prosodic information. Existing research does not clarify whether comprehension loss under accelerated playback stems primarily from working memory limitations, disruption of SSF, or an interaction of both factors. Addressing this

gap is essential to understand the mechanisms of audiobook comprehension, particularly as listening speeds continue to increase in both educational and recreational contexts.

Research Questions

The present study aimed to determine if playback speed influences comprehension when answers to questions depend on prosodic cues and if such comprehension differed when answering questions that are less reliant on SSF. To test the following research questions, separate groups had different playback speeds and then answered the same sets of comprehension questions, one set that relied on prosodic cues, and another set that did not depend on these cues. By comparing listener performance across varying playback speeds and question types, the current study sought to clarify whether, and to what extent, accelerated playback alters the perception and cognitive processing of suprasegmental information.

The following research questions guide the present investigation:

1. Does accuracy on comprehension questions that depend on the perception of SSF decrease as audio playback speed increases?
2. Are participant scores on the VVIQ (vividness of visual imagery questionnaire) associated with accuracy on the comprehension questions that do not depend on the perception of SSF (i.e., factual questions)?
3. Are participant scores on the VVIQ (vividness of visual imagery questionnaire) associated with accuracy on the comprehension questions that depend on the perception of SSF?

Methods and Materials

Study Design

This study employed a between-subjects experimental design to examine the effect of audiobook playback speed on listening comprehension. The independent variable was playback speed, with three conditions, 1.0x, 1.4x, and 1.8x. Participants were randomly assigned to one of the three playback conditions. Question type served as a within-subjects factor and included two categories: questions requiring interpretation of suprasegmental features of speech (e.g., prosody, intonation, emphasis) and questions that could be answered based on lexical-semantic content alone. Accuracy for each question type served as the primary dependent measure. Visualization ability was included as a covariate to account for individual differences that may influence comprehension of auditory narrative material. This design allowed us to determine whether accelerated playback speed differentially affected comprehension for information conveyed through suprasegmental versus lexical-semantic cues.

Participants

Participants include 24 adult native English speakers between the ages of 18 and 35. The study included participants who self-reported a range of cognitive and perceptual profiles, including learning disabilities, attention-deficit/hyperactivity disorder (ADHD), dyslexia, mild traumatic brain injury (e.g., concussion), and visualization abilities ranging from aphantasia to high imagery. The researchers recruited participants from the Boulder area and the University of Colorado Boulder campus using a convenience sampling approach. The study did not target a specific ethnic distribution. As compensation for their participation, participants received a $15 electronic gift card upon completion of their first visit.

Materials

The study materials include an audiobook excerpt from chapter 7 of the novel, *Yours Truly,* by Abby Jimenez (2023), through Apple books platform, standardized assessments for

verbal IQ, verbal working memory, and visual imagery ability. To evaluate comprehension, two sets of comprehension questions were developed specifically for this book except used in this study. The comprehension questions include one set in which suprasegmental information (i.e., including pitch, intonation, stress, and rhythm) must be perceived and interpreted to answer correctly, and a second set that only required perception of speech (i.e., factual questions based on language alone). These two sets of questions allowed for analysis that could differentiate the effect of playback speed given different types of questions. Accuracy on these general comprehension questions also served as a control so that general comprehension could be assessed between groups.

Institutional Review

Approval from the Institutional Review Board (IRB) was received. All procedures complied with ethical standards for research involving human participants and followed IRB approved procedures.

Procedures

Each participant attended one in-person session at a dedicated testing laboratory in the Speech, Language, and Hearing Sciences building on the University of Colorado Boulder main campus. The testing room contained a conference table, comfortable seating, and a laptop equipped with high-quality noise-canceling headphones. Only the participants and a trained research assistant were present in the room during testing.

After completing the informed consent process, participants completed several standardized assessments. These included the Vividness of Visual Imagery Questionnaire (VVIQ), a self-report measure assessing the subjective clarity and vividness of mental visual

imagery. The VVIQ does not assess visual accuracy or perception, but rather an individual's perceived ability to generate mental images.

Participants also completed several standardized cognitive assessments, including a measure of auditory-verbal working memory (Woodcock–Johnson-IV, Test of Cognitive Abilities: Memory for Words) and two subtests assessing verbal intelligence (Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II): Vocabulary and Similarities). These assessments ensured that participants did not present with intellectual disability, language impairment, or deficits in auditory-verbal working memory that could confound performance on study measures.

Following standardized testing, participants listened to the audiobook excerpt consisting of the final 15 minutes of chapter 7. This excerpt was selected due to its rich emotional content, descriptive language, and complex prosodic features, including frequent shifts in tone, stress, rhythm, and intonation. The chapter includes both narration and dialogue, making it well-suited for assessing comprehension dependent on SSF. The selected excerpt was reviewed to ensure that it contained no overly explicit themes and was of moderate length to minimize participant fatigue. All participants listened to the same audiobook excerpt, but playback speed varied by condition. The excerpt duration varied by condition, lasting approximately 15 minutes at 1.0x speed, 10.7 minutes at 1.4x speed, and 8.3 minutes at 1.8x speed.

Immediately following the listening task, participants completed a comprehension assessment consisting of 25 multiple-choice questions. The questions assessed comprehension across four SSF categories: (1) pitch/intonation, (2) stress, (3) rhythm, and (4) general comprehension not reliant on suprasegmental features. Pitch and intonation were combined into

a single category because comprehension depended on the perception of intonational patterns rather than pitch alone.

Assessment Scoring

For each participant, standardized scores were calculated for all standardized subtests. Total VVIQ scores were also calculated according to published scoring procedures.

Comprehension Question Accuracy

Comprehension performance was recorded as total percent correct for the two main sets of questions, SSF-dependent questions, and non SSF-dependent questions.

Statistical Analyses

Between group analyses, the General Linear Model Analysis of Variance (ANOVA) was utilized to compare comprehension accuracy between groups. Within-group comparisons were made comparing percent accuracy for each question set. Additional Pearson correlation analyses were conducted for each group separately to examine possible associations between VVIQ and percent accuracy on the SSF comprehension questions. All statistical analyses were completed using SPSS version 31.0, and values of $p < 0.05$ were considered significant.

Results

Separate ANOVAs were used to compare group for the following variables: age, VVIQ, Verbal IQ (WASI-II VCI), and verbal working memory (WJ-IV Memory for Words). There were no significant effects of group for any of these variables (see Table 1).

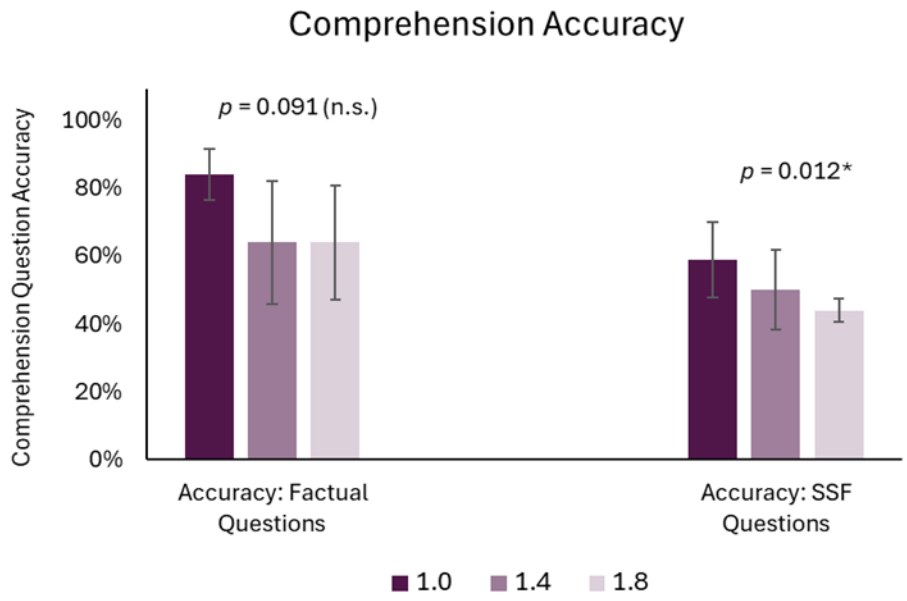
Table 1.

Participant Data

	Group 1.0x (n=5)	Group 1.4x (n=11)	Group 1.8x (n=7)	stats
Average Age (yrs)	20.4	20.6	21.4	$p = 0.65$ (n.s.)
Males/Females	2/3	3/8	1/6	$p = 0.64$ (n.s.)
Average VVIQ	48.0	56.1	57.9	$p = 0.75$ (n.s.)
Average WASI-II VCI	95.8	92.6	93.2	$p = 0.82$ (n.s.)
Average WJ-IV Memory for Words	106.4	114.2	106.3	$p = 0.13$ (n.s.)
Accuracy Factual Questions	84% (17)	64% (18)	64% (8)	$p = 0.91$ (n.s.)
Accuracy SSF Questions	59% (3)	50% (12)	44% (11)	$p = 0.012$*

To address the first research question, does accuracy on comprehension questions that depend on the perception of SSF decrease as audio playback speed increases, an ANOVA was used to compare accuracy on both types of questions between groups (see Figure 1). There was no significant effect of group for the factual questions, but there was a significant effect of group for the SSF questions ($F(2,23) = 5.659$, $p = 0.012$), with the 1.0x group having the highest accuracy and the 1.8x group having the lowest accuracy.

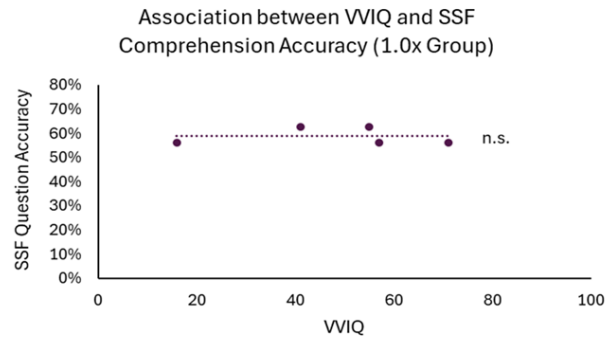
Figure 1. *Comprehension Accuracy by Group and Question Type*

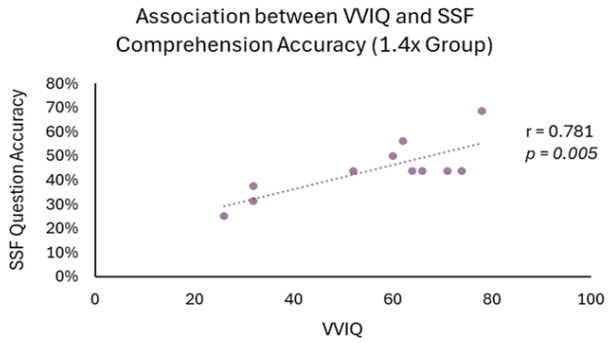


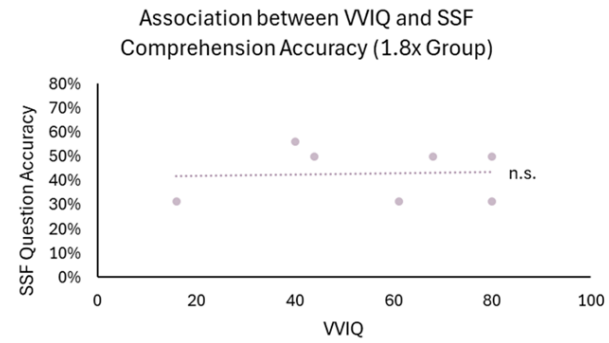
For the second research question, are participant scores on the VVIQ (vividness of visual imagery questionnaire) associated with accuracy on the comprehension questions that do not depend on the perception of SSF (i.e., factual questions), separate correlation analyses were completed for each group using VVIQ score and factual question accuracy. None of the correlations for any group were significant.

For the final research question, are participant scores on the VVIQ (vividness of visual imagery questionnaire) associated with accuracy on the comprehension questions that depend on the perception of SSF, separate correlation analyses were completed for each group using VVIQ score and SSF question accuracy. The correlations for the 1.0x and the 1.8x group were not significant, but the 1.4x group was ($r = 0.781$, $p = 0.005$) (see Figure 2).

Figure 2. *Correlation Results Between SSF Question Accuracy and VVIQ By Group*







Qualitative Observations

Across playback speed conditions, participants at 1.0x frequently reported difficulty sustaining attention, whereas those in the 1.8x condition often described uncertainty regarding their comprehension, consistent with their lower accuracy and confidence scores. Participants in the 1.4x condition appeared to exhibit greater engagement, offering more spontaneous commentary and reflections, suggesting that moderate playback speeds may balance attentional engagement with sufficient processing time.

Discussion

Accuracy on factual comprehension questions did not differ significantly between groups, although the group that had the 1.0x playback speed did have the highest accuracy. In contrast, average group accuracy for the SSF questions did show a significant effect of playback speed. The 1.0x group with the highest accuracy in SSF questions, while the group who had moderate acceleration (1.4x) had a lower accuracy, and the group who had the 1.8x playback speed demonstrated the lowest accuracy. The significantly lower accuracies observed of the 1.4x and 1.8x groups on the SSF questions, but not the factual questions, suggests the ability to answer general comprehension questions is less sensitive to increases in playback speed than questions that are dependent on SSF. The ability to perceive and access to prosodic cues such as stress, intonation, and emphasis, does facilitate comprehension of an oral narrative. While it remains unclear at exactly what speed this type of comprehension breaks down, the current study does reveal changes starting at the 1.4x speed. These results suggest that increased playback speed does appear to interfere with the perception and encoding of suprasegmental information critical for auditory comprehension.

Participant scores on the VVIQ were not found to be associated with accuracy on the factual comprehension questions. This may be because factual comprehension is more dependent on speech perception, not the perception of SSF. At the accelerated playback speeds used in this study (1.4x and 1.8x), speech perception may have remained sufficiently intact for listeners to extract information without relying on compensatory cognitive strategies such as visual imagery.

In contrast, VVIQ scores were found to be associated with accuracy on the SSF comprehension questions in the 1.4x group, but no significant associations were found in the 1.0x and the 1.8x groups. This may be because the 1.0x group did well on the SSF questions and did not need to utilize visual imagery to compensate like the 1.4x group did. Further, the 1.8x

group may have failed to perceive most of the SSF cues, and even those with strong visual imagery skills were unable to compensate for the lack of auditory cues available at that faster speed.

Several participants with lower VVIQ scores reported diagnoses of ADHD or ADD and described experiencing only mild visual imagery. Although this observation was not formally tested, it raises the possibility that attentional factors may influence self-reported imagery ability and in turn the processing of SSF information. These reports suggest that imagery vividness may support the processing, maintenance, or reconstruction of suprasegmental cues, when auditory input is temporally compressed, and perceptual signals become less distinct.

Overall, these findings indicate that audiobook playback speed differentially affects components of auditory comprehension. While factual comprehension appears relatively stable across speeds and individual imagery differences, suprasegmental comprehension is more sensitive to both playback speed and imagery vividness. These results support a component-based view of auditory comprehension in which prosodic processing relies on distinct cognitive resources, like VVIQ that are particularly vulnerable under increased speed and attentional demands.

The present findings further suggest that different cognitive mechanisms may underlie factual versus prosody-based comprehension. This highlights the importance of considering auditory processing constraints, presentation speed, and individual differences in visualization ability when evaluating audiobook comprehension. Observations about participant attention also motivate future research examining whether attentional resources interact with playback speed. Although not formally tested in this study, participants with lower imagery ability and attentional diagnoses reported greater difficulty maintaining focus at faster playback speeds. These

observations suggest that accelerated audiobook listening may not uniformly impair comprehension, but rather disproportionately affect processing that depends on fine-grained acoustic and expressive cues. Visual imagery ability may therefore moderate the impact of accelerated playback, reflecting a nuanced interaction between cognitive capacity and auditory perception.

Limitations and Future Directions

The findings of the present study should be interpreted considering several methodological limitations and considerations for future research. First, the participant sample was not evenly distributed by gender, with a greater number of female participants than male participants. Although gender effects were not a primary focus of this study, uneven representation may have introduced variability related to attentional strategies, listening habits, or self-reported imagery ability. Future research with more balanced or larger samples would allow for examination of potential gender-related influences on suprasegmental processing and imagery vividness.

Second, the group sizes were small and uneven. Future research should aim to have groups that are similar in size and larger. It is possible that non-significant effect may be found to be not significant with a larger sample size, or the opposite may be true with non-significant findings may reach significance with a larger sample size.

Additionally, the sample showed limited variability in imagery ability. Only two participants reported experiencing a complete absence of visual imagery, while the majority had a range that was reported mild to highly vivid imagery. This restricted range may have limited the strength of observed associations between VVIQ scores and suprasegmental comprehension

accuracy. Future research including a larger number of individuals with aphantasia or extremely low imagery may clarify the role of imagery vividness in auditory prosodic processing.

Another potential limitation is the nature of the auditory book sample, and the comprehension questions generated to assess comprehension of facts versus information conveyed via SSF. The materials created for this study have not been previously validated, and while careful attention to detail was used to create the two question sets, it is unknown if they really measure perception of SSF. Further, it is also unclear if similar results would be found for a different text or perhaps a different type of text (scientific book chapter, historical text, biography, etc.). Future research should aim to conduct similar studies using different types of texts and questions aligned to those book segments to determine if the current findings are replicated.

Finally, observations and comments from participants suggest that attentional factors may have interacted with both playback speed and self-reported imagery ability, even though this was not tested in this study. Several participants with lower VVIQ scores reported diagnoses of ADHD or ADD, raising the possibility that attentional differences influenced performance independently of imagery vividness. Although these observations were not formally analyzed, future research would benefit from the inclusion of standardized attention measures to disentangle the contributions of imagery, attention, and auditory processing speed.

Conclusion

This study demonstrates that audiobook comprehension is influenced by playback speed and individual differences in visualization vividness. While factual comprehension remained

largely stable, comprehension dependent on suprasegmental processing declined under accelerated conditions, highlighting the importance of prosodic cues for narrative understanding. Visualization ability may provide a compensatory advantage under moderate acceleration, though its effect diminishes at extreme speeds. These findings suggest that audio-only narratives can be accessible to a broad range of listeners, but speed and individual cognitive differences shape the quality of the listening experience. Future research should explore larger and more diverse samples, standardized attention measures, and a broader range of imagery abilities to fully elucidate the interactions between prosody, mental imagery, and comprehension.

Acknowledgements

Thank you to Dr. Arehart for growing my relationships on campus even once retired. She led me to form relationships on campus starting with my advisor Dr. Christine Brennan, who I thank so much for taking me under her wing. These two women alone pushed me to grow and learn as much as I can during my time in college. Not only do I have them to thank but my participants, thank you, for giving me your time and being so willing to be there for me with all your busy schedules, without you I wouldn't have been able to complete my thesis.

References

Applebaum L, Coppola M, Goldin-Meadow S. 2014. Prosody in a communication system developed without a language model. Sign Lang Linguist. 17(2):181–212. https://doi.org/10.1075/sll.17.2.02app

Bainbridge WA, Pounder Z, Eardley AF, Baker CI. 2021. Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. Cortex J Devoted Study Nerv Syst Behav. 135:159–172. https://doi.org/10.1016/j.cortex.2020.11.014

Bänziger T, Scherer K. 2005. The Role of Intonation in Emotional Expressions. Speech Commun. 46:252–267. https://doi.org/10.1016/j.specom.2005.02.016

Cain K, Bignell S. 2014. Reading and listening comprehension and their relation to inattention and hyperactivity. Br J Educ Psychol. 84(Pt 1):108–124. https://doi.org/10.1111/bjep.12009

Chen A, Kumar SE, Varkhedi R, Murphy DH. 2024. The Effect of Playback Speed and Distractions on the Comprehension of Audio and Audio-Visual Materials. Educ Psychol Rev. 36(3):79. https://doi.org/10.1007/s10648-024-09917-7

Christophe A, Gout A, Peperkamp S, Morgan J. 2003. Discovering words in the continuous speech stream: The role of prosody. J Phon. 31:585–598. https://doi.org/10.1016/S0095-4470(03)00040-8

Chung S, Cheon J, Lee K-W. 2015. [accessed 2026 Jan 27]. https://www.jstor.org/stable/43575306?utm_source=chatgpt.com&seq=1

Cutler A, Dahan D, van Donselaar W. 1997. Prosody in the comprehension of spoken language: a literature review. Lang Speech. 40 (Pt 2):141–201. https://doi.org/10.1177/002383099704000203

Foulke E, Sticht TG. 1969. Review of research on the intelligibility and comprehension of accelerated speech. Psychol Bull. 72(1):50–62. https://doi.org/10.1037/h0027575

Goldhaber GM, Weaver CH. 1968. Listener comprehension of compressed speech when the difficulty, rate of presentation, and sex of the listener are varied. Speech Monogr. 35(1):20–25. https://doi.org/10.1080/03637756809375562

Ji D, Liu B, Xu J, Gong J. 2024. [accessed 2026 Jan 26]. https://www.researchgate.net/publication/381635798_Why_Do_We_Listen_to_Audiobooks_The_Role_of_Narrator_Performance_BGM_Telepresence_and_Emotional_Connectedness

Karimi F. 2022. [accessed 2026 Jan 26]. https://www.cnn.com/2022/06/09/us/podfasters-audio-accelerated-speeds-cec/index.html

Keogh R, Pearson J. 2018. The blind mind: No sensory visual imagery in aphantasia. Cortex J Devoted Study Nerv Syst Behav. 105:53–60. https://doi.org/10.1016/j.cortex.2017.10.012

Kıyak YS, Budakoğlu İİ, Masters K, Coşkun Ö. 2023. The effect of watching lecture videos at 2× speed on memory retention performance of medical students: An experimental study. Med Teach. 45(8):913–917. https://doi.org/10.1080/0142159X.2023.2189537

Kosslyn SM, Ganis G, Thompson WL. 2001. Neural foundations of imagery. Nat Rev Neurosci. 2(9):635–642. https://doi.org/10.1038/35090055

Kruez R. 2023. [accessed 2026 Jan 26]. https://www.psychologytoday.com/us/blog/small-talk-and-big-questions/202307/is-speed-listening-right-for-you

Liu xiaohui et al. 2020. Modulating Your Voice: The Role of Paralinguistic Cues in Improving Podcasters' Competitiveness. PACIS 2020 Proc. [published online ahead of print] https://aisel.aisnet.org/pacis2020/145

Murphy D, Hoover K, Agadzhanyan K, Kuehn J. 2021. Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. ResearchGate. [published online ahead of print] https://www.researchgate.net/publication/356194270_Learning_in_double_time_The_effect_of_lecture_video_speed_on_immediate_and_delayed_comprehension

Ness I, Opdal K, Sandnes FE. 2021. On the Convenience of Speeding Up Lecture Recordings: Increased Playback Speed Reduces Learning. In: Huang Y-M, Lai C-F, Rocha T, editors. Innovative Technologies and Learning. Springer International Publishing; p 461–469. https://doi.org/10.1007/978-3-030-91540-7_47

Paivio A. 1991. Dual coding theory: Retrospect and current status. Can J Psychol Rev Can Psychol. 45(3):255–287. https://doi.org/10.1037/h0084295

Pastore R. 2012. The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. Comput Educ. 58(1):641–651. https://doi.org/10.1016/j.compedu.2011.09.018

Pierrehumbert J. 1990. Phonological and phonetic representation. J Phon. 18(3):375–394 (Phonetic Representation). https://doi.org/10.1016/S0095-4470(19)30380-8

Ritzhaupt A, Gomes N, Barron A. 2008. [accessed 2026 Jan 26]. https://www.researchgate.net/publication/223801608_The_effects_of_time-compressed_audio_and_verbal_redundancy_on_learner_performance_and_satisfaction

Rodero E, Lucas I. 2021. Sage Journals; [accessed 2026 Jan 24]. https://journals.sagepub.com/doi/epub/10.1177/14614448211024142. https://doi.org/10.1177/14614448211024142

Snelling M. 2021. The Audiobook Market and Its Adaptation to Cultural Changes. Publ Res Q. 37(4):642–656. https://doi.org/10.1007/s12109-021-09838-1

Song K et al. 2018. Does the Podcast Video Playback Speed Affect Comprehension for Novel Curriculum Delivery? A Randomized Trial. West J Emerg Med. 19(1):101–105. https://doi.org/10.5811/westjem.2017.10.36027

Streeter LA. 1978. Acoustic determinants of phrase boundary perception. J Acoust Soc Am. 64(6):1582–1592. https://doi.org/10.1121/1.382142

Wassenberg R et al. 2010. Speed of language comprehension is impaired in ADHD. J Atten Disord. 13(4):374–385. https://doi.org/10.1177/1087054708326111

Watkins NW. 2018. (A)phantasia and severely deficient autobiographical memory: Scientific and personal perspectives. Cortex J Devoted Study Nerv Syst Behav. 105:41–52. https://doi.org/10.1016/j.cortex.2017.10.010

Wimmer H, Mayringer H. 2002. Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. J Educ Psychol. 94(2):272–277. https://doi.org/10.1037/0022-0663.94.2.272

Wingfield A, Tun PA, McCoy SL. 2005. Hearing Loss in Older Adulthood: What It Is and How It Interacts With Cognitive Performance. Curr Dir Psychol Sci. 14(3):144–148. https://doi.org/10.1111/j.0963-7214.2005.00356.x

Wolf MC, Muijselaar MML, Boonstra AM, de Bree EH. 2019. The relationship between reading and listening comprehension: shared and modality-specific components. Read Writ. 32(7):1747–1767. https://doi.org/10.1007/s11145-018-9924-8

Yorio K. 2019. School Library Journal; [accessed 2026 Feb 8]. https://www.slj.com/story/audiobooks-increase-in-popularity-as-science-supports-their-value

Zeman A, Dewar M, Della Sala S. 2015. Lives without imagery—Congenital aphantasia. Cortex J Devoted Study Nerv Syst Behav. 73:378–380. https://doi.org/10.1016/j.cortex.2015.05.019

Zwaan RA, Radvansky GA. 1998. Situation models in language comprehension and memory. Psychol Bull. 123(2):162–185. https://doi.org/10.1037/0033-2909.123.2.162