Artificial Intelligence and Cognitive Constraints: A Working Memory Approach to Adult L2 Speech Perception

Siti Hajar binti Shariffudin, Muhammad Ridha bin Ali Huddin

Faculty of Language and Linguistics, University Malaya, Kuala Lumpur, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.100500830

Received: 28 May 2026; Accepted: 03 June 2026; Published: 15 June 2026

ABSTRACT

The vast growth and advancement of artificial intelligence (AI) have significantly reshaped most of our daily life activities including education among the adults. Suggesting that AI can overcome the age-related limitations among adult learners in language learning. AI-based systems offer enhanced input and instant outcome leading to the assumption that limitations linked to Critical Period Hypothesis may be reduced or overcome. However, they might overlook the role of cognitive constraints that fundamentally shape human language processing. This paper discusses that adult second languages (L2) remain inherently cognitively constrained with working memory functioning as a central mechanism governing the other components. Humans are subject to cognitive and attention constraints which affect the perceptual performance unlike AI systems. As a result, advancement in technological input do not necessarily translate into additional gains in perceptual accuracy or efficiency. Building on this idea, the paper proposes a theoretical framework in which AI is conceptualized as an adaptive while working memory acts as an internal processing constraint. By foregrounding working memory as a core component in adult L2 speech perception, this paper contributes to a more cognitively grounded account of language processing in the digital age, with the perspectives of linguistic and AI among adult learners of second language.

Keywords: Artificial intelligence (AI), working memory, speech perception, adult, second language (L2)

INTRODUCTION

The rapid advancement of artificial intelligence (AI) has significantly transformed the language learning perspective. AI-driven systems, particularly in speech processing demonstrates high level of accuracy, adaptability and stability and offer learners enhance exposure and real-time feedback. These developments have led to the assumption that technological advancements can compensate for limitations traditionally including adult language learning, especially second language (L2) acquisition. However, this assumption raises a fundamental theoretical issue, 'does improving input can necessarily improve the perception?' While AI systems optimize input conditions, they overlook the aspect of human speech perception that operates within a cognitively constrained system. Speech perception is not solely a function of acoustic input, but rather a cognitively mediated process shaped by limitations in attention, processing capacity, and, most critically, working memory. This paper discusses that current AI-driven approaches risk overemphasizing the input while underestimating the role of cognitive constraints. While AI systems operate with unconstrained computational resources, human learners process linguistic input within tightly bounded cognitive systems. This paper argues that AI should be understood as a form of external support, rather than a complete substitute for human cognitive process. By foregrounding working memory as the central component in speech perception, this paper proposes a reconceptualization from input-driven models toward cognitively grounded models in language learning in the digital age.

THEORETICAL FOUNDATIONS AND BACKGROUNDS

This section will discuss on the foundations and current models of speech perception, working memory and AI-based speech processing and their limitations.

Speech Perception Theories and Models

In the area of language and linguistics, speech perception has long been the centre topic with early theories since 1967. One of them is Motor Theory founded by Alvin Liberman and others which posits that listeners perceive speech by mapping the acoustic signals into articulatory signals. (Liberman et al. 1967). This is a complementary of the research before on categorical perception demonstrated that listeners interpret speech sounds in continuous acoustic variation into discrete phonological categories. (Liberman et al., 1957). Although these models provide insightful structure of speech perception, they still offer limited explanation into variability across listeners and contexts. This limitation has brought to later introduced cross-linguistics perspectives, particularly in second language (L2) acquisition.

The most influential second language (L2) phonology models adopted are Speech Learning Model, (Flege 1995; Flege & Bohn, 2020) and Perceptual Assimilation Model (Best 1995; Best & Tyler, 2007). First is the Speech Learning Model (SLM), developed by James Flege (1995), focuses on predicting the difficulties in the second language (L2) sounds, both perception and production in relation to existing first language (L1) categories, which may facilitate of hinder new category of language. The second one, the Perceptual Assimilation Model (PAM), associated with Catherine Best, predicts the challenges and possibilities in cross-language categorization of L1 and L2 sounds, in which the unfamiliar L2 sounds to L1, leading to perceptual difficulty (Best 1995; Best & Tyler, 2007).

While these models have significantly advanced in L2 speech perception which leads to revised version of SLM to SLM-r (Flege & Bohn, 2021), and PAM-L2 (Best & Tyler, 2007), they remain focused on phonological categorization and still offer limited information on cognitive mechanisms underlying the real processing of human perception. They do not explain how learners manage the temporal and information demands of inputs under cognitive constraints. This gap points to the need of a more cognitively grounded account of speech perception.

Traditional accounts of adult second language (L2) speech perception difficulties have often been framed in terms of age-related constraints, most prominently through the Critical Period Hypothesis (CPH) proposed by Eric Lenneberg (1967). The CPH posits that achieving a high level of language proficiency is only possible during a specific developmental period biologically. This has long been widely discussed on why adult L2 learners struggle to achieve native-like pronunciation. However, this topic has been discussed widely and is still a matter of debate (Siahaan, 2022).

Empirical data shows there are several adults who able to achieve high-level proficiency (Birdsong, 1999; DeKeyser, 2012). Learners can still achieve a high level of proficiency after the period, but they need to heavily rely on effort as the capacity for implicit phonological and syntactic learning declines (Birdsong, 2018). This variability shows that age alone cannot fully explain differences in L2 speech perception. The CPH is mainly biologically oriented and does not specify the cognitive mechanisms underlying age-related differences. These limitations highlight the need for a more mechanism-based account in which cognitive factors play as the central role.

Working Memory and Speech Processing

Working Memory (WM) is an important construct for understanding language learning. There are several models explored on this. The original model of working memory was introduced by Alan Baddeley and Hitch in 1974. After years, the original model was improvised into a multicomponent model, which consists of the phonological loop, visuospatial sketchpad, central executive, and episodic buffer (Baddeley, 2000, 2017, 2021). The phonological loop component is particularly relevant for speech processing as it supports the retention auditory-verbal information.

Miller's law (1956) and Cowan's model (2001) reveal that cognitive capacity limits are not arbitrary constraints but evolved mechanism in processing. Cowan's embedded-processes model (1999) shows that working memory represents activated portions of long-term memory (LTM) with certain attention focus. This provided as structure and blueprint for artificial cognitive systems.

The concept of working memory provides the main understanding of cognitive constraints involved in speech perception. It supports temporary storage and information manipulation and integration of information across time. The phonological loop is important for temporarily storing and rehearsing auditory-verbal information, thereby supporting tasks such as phoneme discrimination, vocabulary, and sentence processing and construction. It comprises a phonological store together with an articulatory rehearsal mechanism and plays an instrumental role in the storage and processing of novel phonological forms (Service et al., 2022). These cognitive functions are vital as it impose constraints on how speech is perceived, especially under conditions of increased complexity, noise, or unfamiliar phonological structures. For adult L2 learners, these constraints are further amplified, as processing often requires greater conscious effort and reliance on controlled cognitive resources.

Despite its relevancy, working memory has often been treated as a supporting variable rather than a main component in speech perception models. Thus, it will then lead to theoretical imbalance, where perceptual processes are explained mainly on input and representation with limited attention on the constraints from the learners themselves. Although there some language-based researchers who have used this framework in their studies (Martin & Ellis, 2012), there is still a need to bring forward that speech perception should be understood not only as a perceptual process, but a cognitively constrained activity shaped by the individual and processing resources. Many models do not integrate cognitive difficulties into their core assumptions, resulting in an incomplete account of perceptual processing.

Cognitive Load Theory and Cognitive Bottleneck

John Sweller (1988) developed Cognitive Load Theory (CTL) which proposes that learning performance is influenced by the working memory capacity and that the task designs or educational instructions should support the cognitive processing with relevant information. This theory has been refined and distinguishes three types of cognitive load; intrinsic load from material complexity, extraneous load from others and germane load from activities. When the learner's working memory is overloaded, he may not be able to process the information well, which leads to poor performance.

Based on this perspective, adaptive AI systems do not increase the working memory capacity, but they attempt to optimize instructional conditions like adjusting task difficulty, modifying processing speed, reducing unnecessary information and others based on the latest theory by Sweller and others (2019).

Figure 1 below shows a fundamental of human cognition. According to Baddeley's model (2000;2012), information must be temporarily maintained within working memory before it will be encoded to long-term memory (LTM). Sweller's Cognitive Load Theory (CLT) further argues that excessive information may overload the system and reduce effective learning and output. As information passes through this cognitive bottleneck, only information that receive sufficient attention, processing is likely to be transferred into long-term memory (LTM). This perspective aligns with CLT, which emphasizes that instructional systems should be designed to manage cognitive demands and transfer information to long-term memory rather than increasing the input presented (De Jong, 2010; Sweller, 2023). Thus, the current framework in Figure 2 suggests that the effectiveness of AI-assisted learning is not based on the volume of the input, but to which extend the input can be successfully processed within the learner's limited capacity.

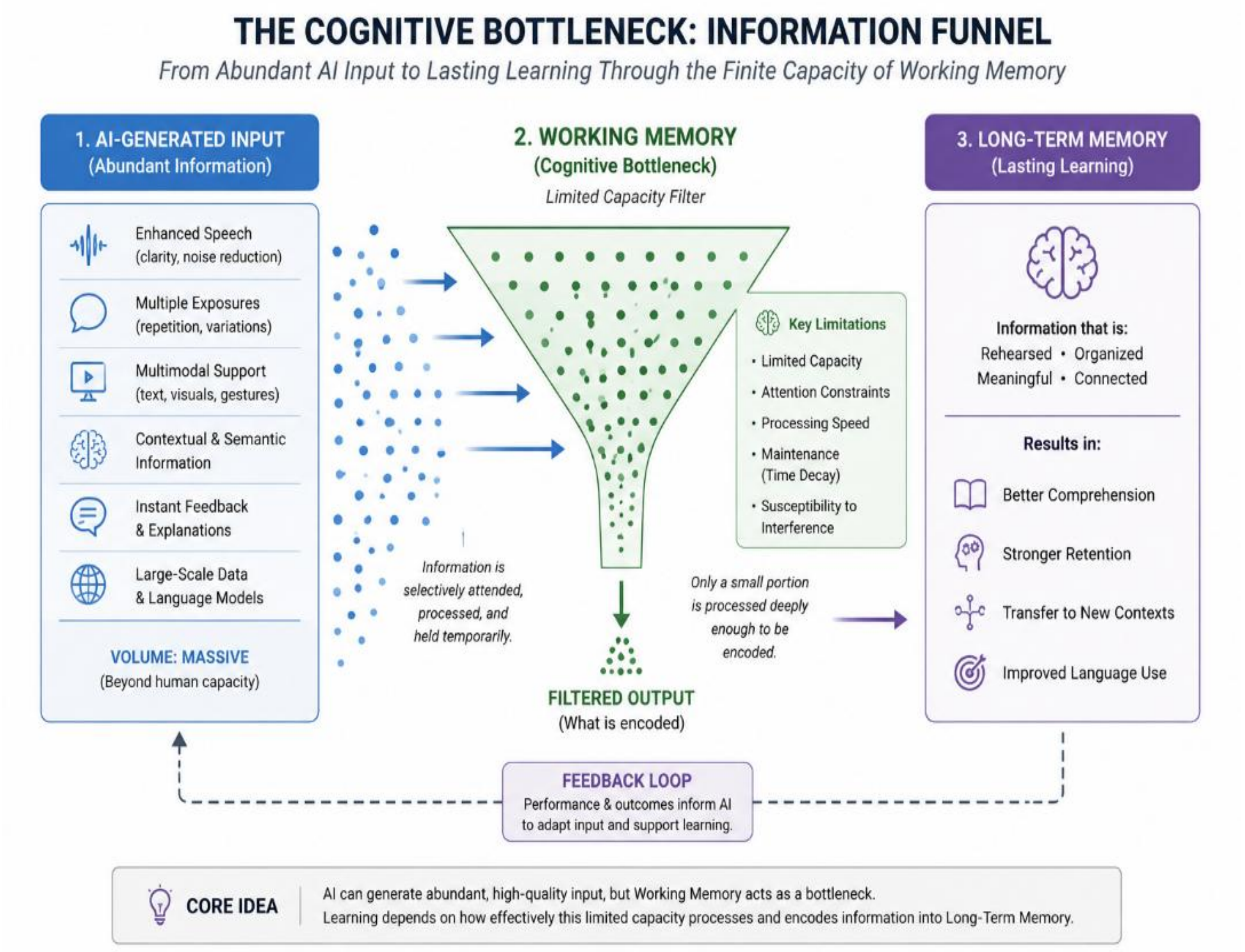

Figure 1

Artificial Intelligence and Speech Processing

The growth of development in artificial intelligence (AI) has significantly advanced the field of speech processing, especially through the evolution of automatic speech recognition (ASR) systems. The dominant approach for speech recognition has been the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM), which is based on context-dependent generative models of GMM and HMM (Zhang, 2025).

The introduction of deep learning around 2010, in speech processing has marked a major shift. Deep Neural Networks (DNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LTSM) models enabled a more effective modelling of sequential data. Recently, models like wav2ve 2.0 (Alexei Baevski et al., 2020) have advanced speech processing by learning speech representations directly from raw data. These models have improved to achieve high levels of performance across diverse speech tasks.

Despite these improvisation and advancements, AI-based speech systems are different from human speech perception. The differences between computer-generated speech and human speech are still hard to interpret because speech materials that may be unspecific to the human versus AI-driven speech contrast but could also be present for different human voices (Herrmann, 2023). AI systems work under conditions of high-capacity memory, parallel processing and computational optimization, allowing to process large volumes of input while maintaining the consistency and efficiency. However, human listeners process speech under a cognitively constrained system, which can be limited working memory capacity, attentional variability, and sequential processing demands. From Herrmann’s study (2023) suggesting that research on speech perception in older adults in modern AI-based synthesized speech is a need.

This limitation highlights an important theoretical issue. AI systems are often evaluated as ‘perfect’ either in terms of accuracy or efficiency, which lead to improved computational performance same goes to improved perceptual capability. In contrasts, such comparisons overlook the absence of cognitive constraints in AI systems, particularly working memory. They are not subject to the working memory constraints that shape how humans perceive, retain and integrate phonological information over time. They provide tools for enhancing inputs but do not model the internal processing limitations that define human speech perception.

AI cannot distinguish the working memory constraints, but it can shift, scaffold, or partially bypass them depending on the tasks and the interface. Some of the benefits in externalizing memory are decreasing cognitive load, improvisation of performance and increase in performance. However, there are some cons, include poor later recall, potential for altered metacognitive calibration, attentional costs, less effort, and dependence on device (Gilbert, 2024).

In response to the theoretical gaps before, the current study proposes a comprehensive and integrated approach in which the speech perception is conceptualized as an outcome of the interaction between external input conditions (AI) and internal cognitive constraints (working memory). By foregrounding working memory as a central component of speech perception, the study presents the foundation for a more cognitively grounded while maintaining the theoretically coherent account of L2 speech processing.

THE WAISP FRAMEWORK

To address the gap, this paper proposes *WAISP*, a cognitive-based approach to adult L2 speech perception while having AI as the input and moderator.

The core components can be divided into four, which are AI-input enhancement, adaptive AI, working memory constraints and speech perception performance. Speech perception emerges from the interaction between input (AI) and cognitive capacity (working memory), not from input alone. Thus, this framework reconceptualizes Adaptive AI as a regulatory mechanism that attempts to align input demands with learner’s cognitive abilities. The learner remains constrained but adaptive AI may improve the fit between the instructional demands and cognitive resources availability.

The current framework proposes that adult L2 speech perception does not determined only by the quality of the input, but from the interaction between AI-input enhancement, adaptive AI, and working memory constraints. AI-input enhancement acts as an external support system by improving the accessibility and clarity of speech through features. While adaptive AI plays as an intermediary mechanism that adjusts instructional conditions in response to learner performance and cognitive capabilities. The speech perception outcomes are based on the result of the dynamic interaction between enhanced input, adaptive support and learner’s cognitive capacity. These adaptations aim to optimize between the quality of the input and the learner’s cognitive capacity without removing or replacing the cognitive constraints but help by managing and operating within those constraints more efficiently.

Adult L2 learners still experience phonological difficulties in learning a new language and these processes draw upon working memory resources. Thus, adaptive AI systems could respond by adapting the tasks based on the learner’s cognitive capabilities. While the learner still relies on working memory to retain or process the information, adaptive AI may improve the processing efficiency without fundamentally altering cognitive capacity.

It does not replace the phonological loop or central executive, instead, it is creating and building conditions under which these components can operate better. Thus, speech perception performance should be understood as the outcome of the interaction between AI-input enhancement, adaptive AI instructional support and working memory constraints. The overview of the WAISP Framework is as the Figure 2 below.

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue V May 2026

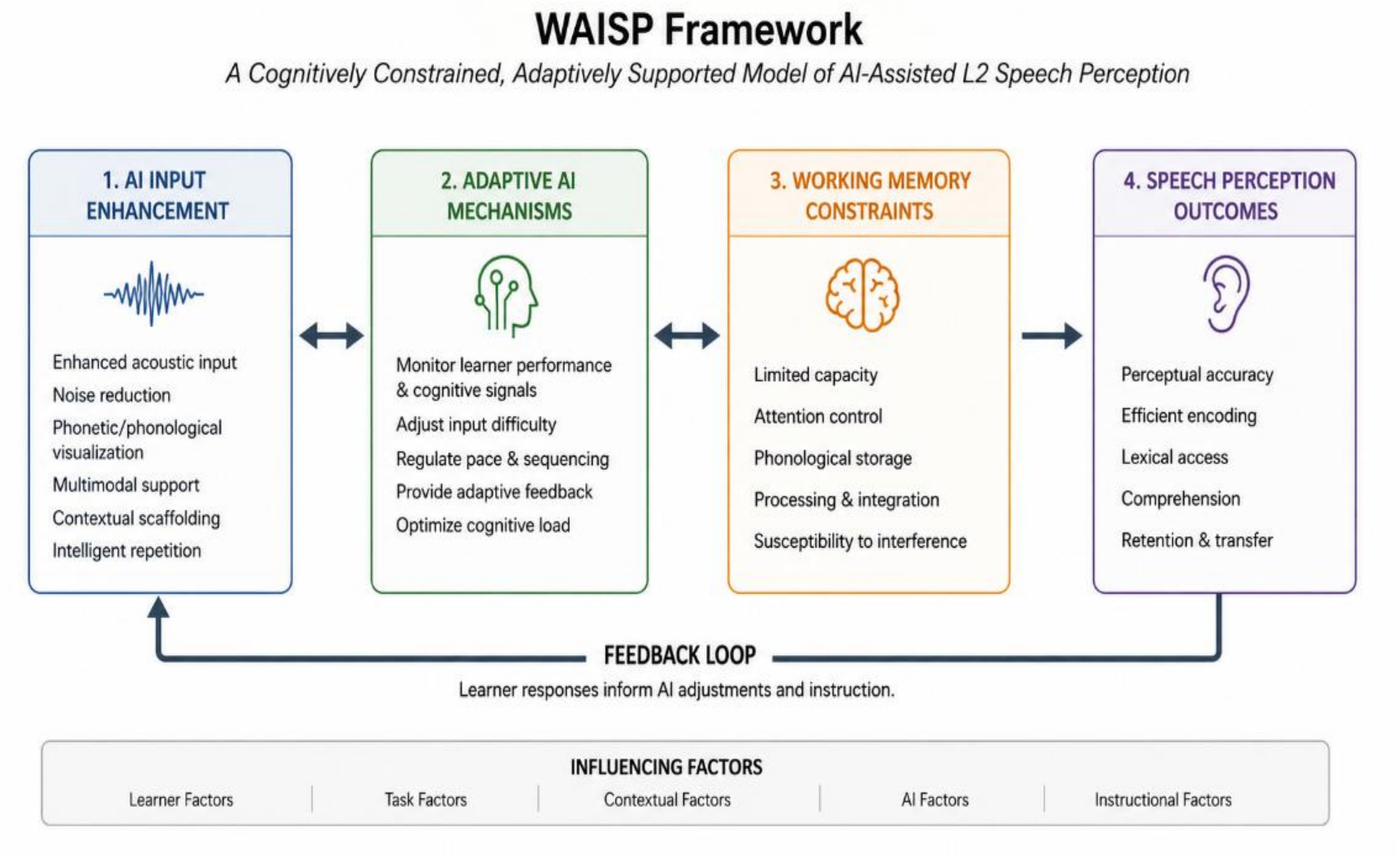

Figure 2

From the theoretical view, the framework is grounded with working memory as the fundamental bottleneck in speech processing which is consistent with Sweller et al (2019) and in line with Baddeley's multicomponent model (2000;2021) while adaptive AI functions as a cognitive support mechanism rather than a replacement with human cognition. Thus, this will shift the focus from input-centred of AI-assisted toward a cognitively constrained model in which AI are mediated by the learner's cognitive capacity.

IMPLICATIONS

There are three main implications regarding the proposed framework which have been divided into theory, language learning, social perspectives and framework implementation.

Theoretical Implications

This perspective calls for a reconceptualization of speech perception within linguistic theory integrated with working memory models. Rather than treating perception as a stable and uniform process, it should be understood as cognitively mediated, resource-dependent and individual differences. By foregrounding working memory as the main component, this paper suggests that current models focus on internal cognitive constraints rather than prioritize input and accuracy.

AI and Language Learning

The output also helps in designing and implementing AI-driven language tools. Before this, AI works as input, this system that assume uniform processing capabilities may fail to account for differences in working memory, leading to unbalanced outcomes. Thus, this proposed framework can help to reduce the concerns on cognitive inequality among individuals. There is a need for more cognitively informed AI systems that align with human speech perception.

Social Impact

AI technologies are often developed based on dominant languages and idealized user profiles in the context of linguistic diversity. Without consideration of cognitive constraints, such systems may not effectively support diverse learner populations, particularly adult L2 learners. A cognitively grounded approach to AI in language learning can contribute to more inclusive and equitable systems, ensuring that technological innovation aligns with the realities of human cognition.

Framework Implementation

The current framework suggests that AI systems should incorporate mechanisms in line with the Cognitive Load Theory (1988;2019). It demonstrates that excessive cognitive demands may affect negatively to the learning performance. Consequently, AI system should monitor the individual differences in cognitive capabilities. Rather than provide complex information, adaptive system should regulate task demands to maintain quality output. AI can enhance the input quality, but it cannot replace the fundamental constraints imposed by human working memory. Thus, AI developers should avoid assumptions that technology alone can automatically improve learning outcomes. Instead, AI should be informed as complement to human cognitive processes.

CONCLUSION

The paper discusses on a need to have a comprehensive framework on adult L2 speech perception in the context of rapid advancement in AI. While AI itself offers vast inputs in speech processing and language learning, it does not eliminate the cognitive constraints that define human speech perception. Speech perception is not solely determined by external input, but is fundamentally shaped by cognitive constraints, particularly those associated with working memory.

Based on established theories and models of speech perception and working memory, this paper has highlighted a key limitation in existing approaches which is the insufficient integration between cognitive constraints with perceptual performance. At the same time, despite the high levels of computational efficiency of AI-based speech processing, it still operates differently from human real cognition. In response to this gap, the paper proposes WAISP framework which conceptualized speech perception as the outcome of integration between AI-mediated input enhancement and working memory constraints. By foregrounding working memory, it has proposed a more realistic model of L2 speech processing that integrates technological and cognitive perspectives.

As a conclusion, this paper has argued that speech perception in adult L2 learners is best understood as a cognitively constrained and technologically mediated process. Recognizing the interaction between external input and internal cognitive constraints is therefore essential for developing more accurate theories of speech perception and more effective approaches to language learning in the digital era. By building AI systems that respect that human real processing structure, we move to a closer future where artificial intelligence becomes the partner of humanity's intellectual endeavor.

FUTURE RESEARCH DIRECTIONS

Despite considerable research, many questions remain, and new ones constantly emerge. There are some key areas for future investigation, one is focusing on the social perspective side. Language learning is not solely a cognitive process but also a fundamentally social, whereby social presence influences the attentional resources within working memory based on Vygotsky's Sociocultural Theory (Murphy, 2008).

Second, is exploring relationship between the variables in detail. As the current framework on evaluated using outcome measures alone, future research could explore the development of a hybrid efficiency metric that quantifies the relationship between AI-input and the learner's cognitive capacity. Thus, the effectiveness of AI enhancement may be conceptualized not simply by number of input but how efficiently that input can be processed within the limitations based on the relationship's metrics.

REFERENCES

1. Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production. *Journal of Memory and Language*, 61(4), 641–665.
2. Alexei Baevski, Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. *Neural Information Processing Systems*, *33*, 12449–12460.
3. András, I., & Balázs, G. (2025). Representation, Virtuality, and Cognition. Rethinking Language in the Digital Attention Economy. *Magyar Nyelvőr*, *149*(5), 637–648. https://doi.org/10.38143/nyr.2025.5.637
4. Baddeley, A. D., & Hitch, G. (1974). Working memory. In *The Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2
5. Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
6. Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839.
7. Baddeley Alan D. (2017). Modularity, working memory and language acquisition, *Second Language Research*,Vol. 33(3) 299–311, doi: https://doi.org/10.1177/0267658317709852
8. Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2021). A multicomponent model of working memory. In R.H. Logie, V. Camos & N. Cowan (Eds.), Working Memory: State of the Science. Oxford University Press.
9. Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). York Press.
10. Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Lawrence Erlbaum.
11. Birdsong, D. and Molis, M. (2001). On the Evidence for maturational constraints in second language acquisition. *Journal of Memory and Language.* 44, 235-249.
12. Birdsong, D. (2018). Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. *Frontiers in Psychology*, *9*(81). https://doi.org/10.3389/fpsyg.2018.00081
13. Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. *Language Experience in Second Language Speech Learning*, 13–34. https://doi.org/10.1075/lllt.17.07bes
14. De Jong, T. (2010). Cognitive Load theory, Educational research, and Instructional design: Some Food for Thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0
15. Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). York Press.
16. Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press. https://doi.org/10.1017/9781108886901
17. G. A. Miller. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychological review, vol. 63, no. 2, pp. 81–97
18. Herrmann, B. (2023). *The perception of artificial intelligence (AI) based synthesized speech in younger and older adults*. https://doi.org/10.1007/s10772-023-10027-y
19. Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.
20. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.
21. Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
22. Martin, K. I.; Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. Studies in Second Language Acquisition, v. 34, n. 3, p. 379-413. DOI 10.1017/S027226311 2000125
23. Murphy, R. (2008). Dynamic assessment precursors: Soviet ideology and Vygotsky. *The Irish Journal of Psychology*, *29*(3-4), 193-233.
24. N. Cowan. (1999). An embedded-processes model of working memory, Models of working memory: Mechanisms of active maintenance and executive control, vol. 20, p. 506

25. N. Cowan. (2001). "The magical number 4 in short term memory: A reconsideration of mental storage capacity," *Behavioral and brain sciences*, vol. 24, no. 1, pp. 87–114
26. Siahaan, F. (2022). The Critical Period Hypothesis of SLA Eric Lenneberg's. *Journal of Applied Linguistics*, *2*(2), 40–45. https://doi.org/10.52622/joal.v2i2.77
27. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, vol. 12, no. 2, pp. 257–285
28. Sweller, J., J. van Merriënboer, and F. Paas. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, vol. 31, pp. 261–292
29. Sweller, J. (2023). The Development of Cognitive Load Theory: Replication Crises and Incorporation of Other Theories Can Lead to Theory Expansion. *Educational Psychology Review*, 35, 95 https://doi.org/10.1007/s10648-023-09817-2
30. Zhang, Z. (2025). AI-Powered Intelligent Speech Processing: Evolution, Applications and Future Directions. *International Journal of Advanced Computer Science and Applications*, *16*(2). https://doi.org/10.14569/ijacsa.2025.0160291