DOI: 10.17951/j.2026.39.2.119-146

ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA

VOL. XXXIX, 2 SECTIO J 2026

Kitasato University. College of Liberal Arts and Sciences

YOSHIAKI ISHII
ORCID: 0009-0003-1171-2089
ishii.yoshiaki.nc@st.kitasato-u.ac.jp

Reconsidering the Mechanism of Listening Comprehension: Integrating the Two-Phase Model, Three-Phase Model, and Working Memory

Rekonceptualizacja mechanizmu rozumienia ze słuchu: integracja modelu dwufazowego, modelu trójfazowego oraz pamięci roboczej

HOW TO QUOTE THIS PAPER: Ishii, Y. (2026). Reconsidering the Mechanism of Listening Comprehension: Integrating the Two-Phase Model, Three-Phase Model, and Working Memory. *Annales Universitatis Mariae Curie-Skłodowska. Sectio J, Paedagogia-Psychologia*, *39*(2), 119–146. DOI: 10.17951/j.2026.39.2.119-146.

ABSTRACT

This study addresses the inherent difficulty of second language (L2) listening, particularly its unobservable cognitive nature, by proposing an integrated model of listening comprehension. Previous research has developed two-phase models (perception and comprehension) and three-phase models (perception, parsing, and utilisation), but their fragmented development has led to conceptual inconsistency. To resolve this, the present study integrates these models using working memory as a unifying framework, incorporating key components such as the phonological loop, central executive, and episodic buffer, along with the concept of perceptual sense units (PSUs). The proposed model clarifies how auditory input is processed from initial perception through linguistic analysis to higher-level meaning construction, demonstrating that these phases operate interactively rather than linearly. Furthermore, the model extends to include visual aids, highlighting the role of the visuospatial sketchpad in audio-visual processing. This integrated framework offers a more coherent and comprehensive explanation of listening comprehension mechanisms and provides practical implications for language teaching, such as improving instructional strategies and diagnosing learner difficulties. The study also identifies directions for future research, including the effects of working memory capacity and multimodal input on listening processes.

Keywords: listening comprehension; two-phase model; three-phase model; integrated model; working memory

INTRODUCTION

Listening is considered one of the most difficult skills to acquire by second (L2) or foreign language learners (Buck, 2001; Darti, Asmawati, 2017; Field, 2009; Ghaderpanahi, 2012; Graham, 2003, 2006; Hasan, 2000; Ishii, 2022, 2025; Kurita, 2012; Li, 2014; O'Malley et al., 1989; Rost, 1994; Rubin, 1995; Vandergrift, 1997; Xu, 2011). The primary reason for this is the unobservability of this skill. Every processing mechanism is performed inside the listener's brain; therefore, if the listener successfully catches up and comprehends spoken messages, it is externally unobservable (Ishii, 2025; Rost, 1994; Rubin, 1995). For example, not asking questions during a psychology class does not mean that students are not listening to the lecture. This unobservability of listening makes this skill arduous for language learners. Thus, addressing this unobservability is indispensable for improving the listening skills of language learners.

MECHANISM OF LISTENING COMPREHENSION

To address the unobservability of listening comprehension, several theoretical models have been proposed over the past few decades (Ishii, 2022). Those models are broadly divided into two categories: two-phase models (e.g. Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Mayer, 1997, 2001; Tamai, 2005) and three-phase models (e.g. Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986). These processing models are delineated by focusing on the scope and nature of the hypotheses following the core concepts shared by the two parties: working memory (Baddeley, 2000; Baddeley, Hitch, 1974), mental lexicon (Kadota, 2007, 2015, 2019; Kormos, 2006; Levelt, 1989, 1993, 1999a, 1999b; Levelt et al., 1999), and the accompanying two key elements (Kohno, 1993, 2001).

Working memory

Working memory is at the core of listening comprehension (Baddeley, 2000; Baddeley, Hitch, 1974; Churchill, 1999; Field, 2009; Kadota, 2007, 2015, 2019; Kohno, 1993, 2001). Among the models that have been devised, the working memory model devised by Baddeley and Hitch (Figure 1) is the most impactful and widely used (Baddeley, 2000; Baddeley, Hitch, 1974).

This model consists of one core component and three subcomponents: central executive, phonological loop, visuospatial sketchpad, and episodic buffer (Baddeley, 2000). The central executive functions as the managing system, similar to the CPU of the computer; it provides directions to retrieve

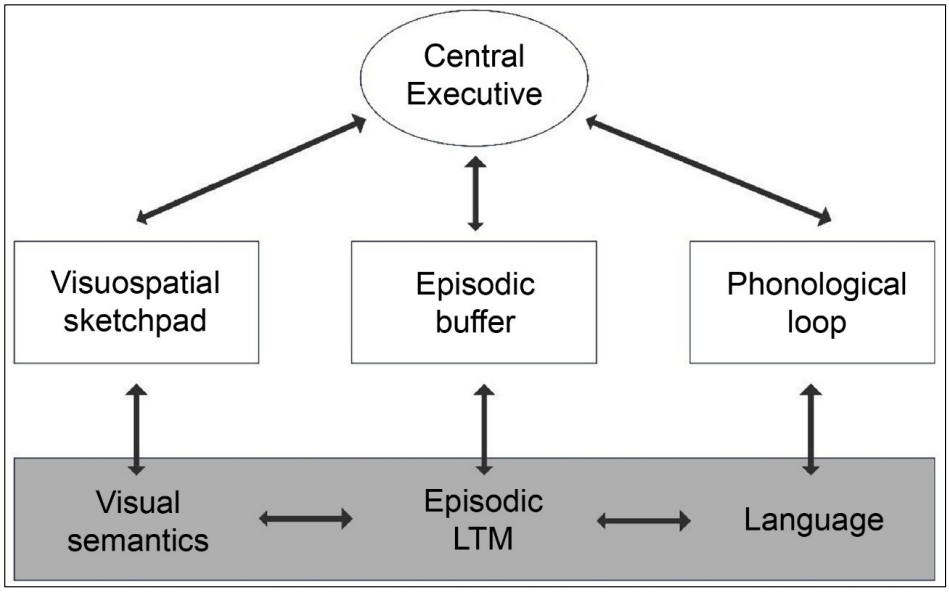

Figure 1. Alan Baddeley and Graham Hitch's Working Memory Model (Disclaimer: This figure was adapted with permission of Elsevier Science & Technology Journals, from Alan Baddeley (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. Permission conveyed through Copyright Clearance Center, Inc.)

information from long-term memory, sends it back to it, and allocates attention to relevant information as well as the consumption of cognitive resources (Baddeley, 2000; Baddeley, Hitch, 1974). The phonological loop is responsible for auditory inputs; it maintains phonological information tentatively through its frequent repeating that is called subvocal rehearsal and analyse it using various related knowledge stored in the long-term memory (Baddeley, 2000; Baddeley, Hitch, 1974). It also deals with language inputs in the form of phonological representation while analysing them using linguistic knowledge such as grammar, word classes, and meanings (Baddeley, 2000; Baddeley, Hitch, 1974). The visuospatial sketchpad maintains and analyses the visually presented information from optical inputs and visual imagery (Baddeley, 2000; Baddeley, Hitch, 1974). The episodic buffer manipulates information from episodic memory and integrates information from the phonological loop and visuospatial sketchpad; integrated information is considered to be sent to long-term memory (Baddeley, 2000). This subcomponent was not included in the initial model devised by Baddeley and Hitch (1976) but was later added by Baddeley (2000). This is a brief overview of the working memory structure. This system is considered the core of language comprehension mechanisms (Baddeley, 2000; Baddeley, Hitch, 1974; Field, 2009; Mayer, 1997, 2001).

Mental lexicon

The mental lexicon is a knowledge source for perceptual processing in the perception phase and analytical processing in the phonological loop of working memory, which is located in the long-term memory and is utilised when processing auditory inputs (Anderson, 1995; Baddeley, 2000; Goh, 2000; Kadota, 2007, 2015, 2019; Kormos, 2006; Levelt, 1989, 1993, 1999a, 1999b; Levelt et al., 1999; O'Malley et al., 1989; Vandergrift, Goh, 2021). It consists of lemma, which is syntactic information; lexeme, which is phonological and orthographic information; and concept, which is semantic information (Kadota, 2007, 2015, 2019; Kormos, 2006; Levelt, 1989, 1993, 1999a, 1999b; Levelt et al., 1999).

TWO-PHASE MODELS

Two-phase models seem to have been derived from studies of information processing (e.g. Atkinson, Shiffrin, 1968, 1971; Schneider, Shiffrin, 1977). The origin of the two-phase model is the model of information processing and human memory proposed by Atkinson and Shiffrin (1968, 1971). This model is called the "multi-store model" and is considered one of the most influential hypothetical mechanisms of information processing (Egan, 2018; Loftus, Loftus, 1976). Schneider and Shiffrin (1977) proposed a more sophisticated hypothesis to dissect the processing mechanism into two phases according to qualitative differences: controlled and automatic processing. Controlled processing is the deliberate and attention-dependent activation of controlled cognitive operations that are flexible but capacity-limited and are typically executed sequentially (Schneider, Shiffrin, 1977). By contrast, automatic processing is the effortless automatic activation of well-learnt sequences in long-term memory that proceeds without conscious control, attentional demand, or capacity limitations (Schneider, Shiffrin, 1977). These hypotheses have been incorporated into various models of listening comprehension (e.g. Eto, 2023; Kadota, 2007, 2015, 2019; Kohno, 1992).

Although a variety exists, two-phase models consist of perception and comprehension phases. The former corresponds to Schneider and Shiffrin's controlled processing or the first phase in Anderson's model, and the latter corresponds to Schneider and Shiffrin's automatic processing or the second and third phases in Anderson's model (Anderson, 1995; Frazier, Fodor, 1978; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; Kohno, 1992; Schneider, Shiffrin, 1977).

An early model of these two-phase assumptions entailing language processing mechanisms is seen in Frazier and Fodor's (1978) sausage machine model, which is also considered a type of two-phase model. The sausage machine model consists of the preliminary phrase packager and sentence structure supervisor as the first and second phases, respectively. In the preliminary phrase packager, sentences

are placed in a "narrow window" and dissected into packages in which several words are packaged together. In the sentence structure supervisor, these words are merged. In this model, Frazier and Fodor bore in mind that the minimal unit segmented from a stream of sentences in the first phase should be regulated by the length of time, not the absolute values of words or idioms. Moreover, they surmised that the length of time should be within "seven plus/minus two" (Frazier, Fodor, 1978, p. 293). This span is the same as that by Miller (1956). Thus, this model is considered one of the earliest to reflect the findings on information processing in the field of cognitive psychology.

Kohno's (1992) model[1] is also classified as a two-phase model (Kadota, 2007, 2015, 2019). This model is based mainly on Pimsleur's (1971) theoretical considerations and several information-processing hypotheses (Atkinson, Shiffrin, 1968; Bondarko et al., 1968/1970; Pisoni, Sawusch, 1975). In his model, the preliminary auditory processing and primary information analysis stages are set; the former covers the entire processing for perception, and the latter covers the remaining processing for comprehension (Kohno, 1992). According to Kohno, the preliminary auditory processing stage processes spoken inputs holistically in the Gestaltic way. By contrast, the primary information analysis stage processes inputs through detailed analysis using linguistic knowledge. This stage is intertwined with working memory, which indicates that the input analysis is performed mainly in the phonological loop according to the direction of the central executive (Baddeley, 2000; Baddeley, Hitch, 1974; Kohno, 1992). The detailed processing procedures in Kohno's (1992) model are as follows:

The spoken input that the listener initially captures enters the preliminary auditory processing stage. In this stage, spoken input is received by the ears, which are the sensory devices that process auditory signals, and is sent to a type of filter where the input is filtered and dissected into chunks by attention (Anderson, 1995; Broadbent, 1956, 1958; Cutler, Clifton, 1999; Field, 2009; Goh, 2000; Kadota, 2007, 2015, 2019; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). In this context, attention is considered selective; however, depending on the input type (e.g. audio-visual), this attention may be directed by feature-based or object-based mechanisms or may instead be distributed as divided attention (Kadota, 2007, 2015, 2019; Liu, 2019; Yantis, 2000). The unit of these channels is dissected according to "perceptual sense unit" (PSU) (Kohno, 1993, 2001). These dissected units are copied entirely to the echoic memory and are then converted into a phonetic representation, which is brought into the phonological loop of the working memory (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler,

[1] Unfortunately, Professor Kohno passed away in 2020 and this work became an orphan work. The author of the current study was unable to adapt Kohno's (1992) two-phase model. The English-translated version of this model is available in Kadota (2019, p. 14).

Clifton, 1999; Darwin et al., 1972; Field, 2009; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992, 2001; Mareschal, 2007; Neisser, 1967; O'Malley et al., 1989; Vandergrift, Goh, 2021). Echoic memory is a sensory memory storage system with an almost unlimited input capacity; however, it can keep the input for only a brief period, which is assumed to constitute a stage that precedes working memory in the information processing sequence (Baddeley, 2000; Baddeley, Hitch, 1974; Darwin et al., 1972; Kadota, 2007, 2015, 2019; Kohno, 1992, 2001; Nagle, Sanders, 1986; Neisser, 1967; Vandergrift, Goh, 2021). For the filtering of input until the creation of phonetic representation, listener's knowledge about phonological information (e.g. summer vacation = /ˈsʌmə veɪˈkeɪʃən/) is integral so that this knowledge is directly brought from the lexeme in the mental lexicon to the filter automatically (Anderson, 1995; Buck, 2001; Cutler, Clifton, 1999; Goh, 2000; Kadota, 2007, 2015, 2019; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). At this point, the processing of the input moves to the next stage (primary information analysis; Kohno, 1992).

In the primary information analysis stage, phonological representations kept in the working memory are carefully analysed (Kohno, 1992). Central executive retrieves various linguistic knowledge from mental lexicon to analyse its embedded information about linguistic knowledges, such as grammar, vocabularies, meanings (e.g. summer vacation = [noun], it is relatively long-term days off given to students between academic terms, usually between July and September), and other related knowledge such as contexts, prior knowledge (e.g. summer vacation comes after the first academic term ends), or pragmatic knowledge (e.g. summer vacation is coming = first academic term ends and homework will be imposed accordingly) or discourse knowledge (e.g. academic lecture, oral announcement, or informal talk; Anderson, 1995; Buck, 2001; Field, 2009; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). After the analysis of phonological representations is completed, the entire input is restored and sent to the long-term memory (Anderson, 1995; Field, 2009; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). These processes were surmised in Kohno's (1992) two-phase model.

Tamai (2005) also devised a theoretical model of listening comprehension based on Baddeley and Hitch's working memory model (1974), so that each explanation was well linked with information processing. However, Tamai's (2005) model was based on a previous working memory (Baddeley, Hitch, 1974) model, which did not include an episodic buffer. Moreover, Tamai's model showed that the auditory input entered the central executive and not the phonological loop (Baddeley, 2000; Baddeley, Hitch, 1974; Tamai, 2005). This lowered the validity of the model (Baddeley, 2000).

Kadota (2007, 2015, 2019) also considered a two-phase model of listening comprehension consisting of perception and comprehension. This was a simplified

version of Kohno's (1992) model. In Kadota's (2007, 2015, 2019) model, the hypothesised function of the perception phase was almost the same as that in Anderson's (1995) and Kohno's (1992) models. However, Kadota (2007, 2015, 2019) considered that processing in the parsing and utilisation phases should be conducted in the same phase. Eto (2023) updated Kadota's (2007, 2015, 2019) model by adding directional arrows indicating top-down and bottom-up processing. Top-down processing involves processing auditory inputs through analysis, mainly by relying on the listener's general social knowledge, such as prior knowledge or contextual information (Anderson, 1995; Eto, 2023; Field, 2009; O'Malley et al., 1989; Vandergrift, Goh, 2021). In contrast, bottom-up processing involves processing inputs through perception, mainly by relying on information available from perceived auditory inputs, such as phonemes, pronunciation, or words (Anderson, 1995; Eto, 2023; O'Malley et al., 1989; Vandergrift, Goh, 2021). In listening comprehension, both processing directions are considered to function in the mechanisms (Anderson, 1995; Eto, 2023; Field, 2009; Kadota, 2007, 2015, 2019; O'Malley et al., 1989; Vandergrift, Goh, 2021).

Aryadoust (2019) also developed a two-phase model, which consisted of two stages, each with two sub-stages, in which the first part focused on perceptual processing and the latter on the processing for comprehension (Aryadoust, 2019). At first glance, Aryadoust's (2019) model seemed highly accurate, but it crammed almost all the reading and listening processing essence centrally based on cognitive processing, making the overall mechanism extremely difficult to understand and significantly reducing its practical usefulness. Moreover, optical input and auditory input are different modes of input, so there should have been corresponding differences, such as the visuo-spatial sketchpad and phonological loop in working memory (Baddeley, 2000; Baddeley, Hitch, 1974), but this aspect was not appropriately reflected in Aryadoust's (2019) model, resulting in significant inconsistency with previous research.

Mayer (1997, 2001) considered a two-phase model created to consider audio-visual inputs so that both auditory and visual inputs were stipulated in the model (Mayer, 1997, 2001). Although Mayer considered working memory as the core for processing audio-visual input, the visuo-spatial sketchpad, phonological loop, and episodic buffer were not presented in the working memory; therefore, there were corresponding differences between this model and the system of working memory (Baddeley, 2000; Baddeley, Hitch, 1974; Mayer, 1997, 2001).

Minimal unit for perceptual processing as PSU

How listeners perceive sentences has been enthusiastically studied using two-phase models. Frazier and Fodor's (1978) model surmised that Miller's (1956) "seven plus/minus two" hypothesis was a minimal unit for perceptual processing

of spoken messages. This was eventually substantiated by Kohno (1993, 2001) through a series of experimental studies conducted in later years. Kohno is professor emeritus at Kobe City University of Foreign Studies in Japan. Through three experiments, Kohno (1993, 2001) identified the PSU, which is "a semantic unit consisting of approximately 7±2 or less syllables connected closely to one another by less than 330 ms intervals" (Kohno, 1993, p. 26). In the first experiment, junior high school students listened to English and Japanese stories with pauses inserted per word, phrase, clause, or sentence. Comprehension was highest when pauses occurred at the phrase level but declined when inserted after every word or removed entirely (Kohno, 1993, 2001). In the second experiment, advanced learners such as English teachers and graduate students showed the same pattern, with phrase-level pauses yielding the best results, although their overall comprehension scores were higher, suggesting that greater proficiency allowed longer processing units without losing efficiency (Kohno, 1993, 2001). In the third experiment, Japanese essays were manipulated either by inserting pauses at PSU boundaries or by artificially making syllable intervals of 200, 250, and 500 ms. The results showed that when participants listened to the audio recordings of these essays, comprehension declined when the essay's syllable intervals were 500 ms but was not influenced by long pauses between PSUs, highlighting the distinct roles of holistic and analytic processing (Kohno, 1993, 2001). These studies demonstrated that PSU is a semantically meaningful group of roughly seven or fewer syllables processed holistically and that listening comprehension is optimised when speech is segmented at these natural units (Kohno, 1993, 2001). This greatly reinforces Miller's (1956) "seven plus/minus two" hypothesis and indicates that perceived messages are dissected into PSUs, which are then converted into phonological representations (Anderson, 1995; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992, 2001; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021).

Dissociation between perceptual and analytical processing phases

In both the two- and three-phase models, the phases for perceptual and analytical processing are clearly separated but strongly connected (Anderson, 1995; Cutler, Clifton, 1999; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; O'Malley et al., 1989; Tamai, 2005; Vandergrift, Goh, 2021). Kohno (1993) conducted a series of experiments on rhythm perception and found that this dissociation is theoretically appropriate. In this research, he distinguished two mechanisms: the stage of holistic or Gestaltic processing for rapid sequences under 330 ms and an analytic processing phase for slower rhythms over 450 ms (Kohno, 1993, 2001), which corresponds well with the dissociation between the phase for perception and that for processing(s) for comprehension (Anderson, 1995; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; O'Malley

et al., 1989; Tamai, 2005; Vandergrift, Goh, 2021). In the first experiment, Kohno (1993, 2001) compared a patient with a cerebrovascular disorder to a healthy adult in a tapping task and showed that although the patient's right hand could process both fast and slow rhythms, their left hand could only follow rapid sequences, indicating a dissociation between holistic and analytic processing. In the second experiment, participants listened to nonsense words spoken at different syllable intervals and then performed mental calculations before recalling them; sequences with short (250 ms) intervals were recalled significantly better, demonstrating that holistically processed rapid rhythms serve as echoic memory resistant to interference (Darwin et al., 1972; Kohno, 1993, 2001; Neisser, 1967). In the third experiment, participants reproduced rhythms after calculation, and the results revealed that memory for fast rhythms was more accurate than that for slower rhythms, further confirming the robustness of holistic processing under a cognitive load (Kohno, 1993, 2001). Taken together, these three experiments established that rhythm perception operates through two distinct mechanisms: holistic processing for rapid sequences and analytic processing for slower sequences, and that these two phases support the structural phases of listening comprehension (Kohno, 1993, 2001).

THREE-PHASE MODELS

The three-phase models originate not only from information processing research but also from educational and pedagogical research (e.g. Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986). Among these, Anderson's (1995) three-phase model is one of the most notable and widely accepted models of listening comprehension (Jia, Hew, 2022). This model was devised by Anderson of Carnegie Mellon University (Anderson, 1995). As the name suggests, this model is classified as a three-phase model, consisting of the following phases: perception, parsing, and utilisation.

In "perception," the first phase of this model, each process is almost the same as that surmised by Kohno (1992) in his two-phase model. It includes dissecting inputs into PSUs and sending them to the echoic memory to create phonological representation and then sending the representation to the phonological loop in the working memory (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton, 1999; Darwin et al., 1972; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; Neisser, 1967; O'Malley et al., 1989; Vandergrift, Goh, 2021).

In the second phase, "parsing," dissected units are retained in the phonological loop through subvocal rehearsal to analyse their forms and meanings (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton, 1999; Field, 2009; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021).

While subvocal rehearsal maintains these channels, the central executive send directions to search grammatical rules, word classes, and meanings from the mental lexicon in the long-term memory (Anderson, 1995; Baddeley, 2000; Cutler, Clifton, 1999; Field, 2009; Goh, 2000; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). This knowledge is brought back to the phonological loop for analysis (Anderson, 1995; Field, 2009; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). Parsing is completed at this point, and the phase is moved to the next phase: utilisation (Anderson, 1995; Field, 2009; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021).

In "utilisation," dissected units whose grammatical rules, word classes, and meanings have already been analysed are further analysed using the listener's prior, pragmatic, or discourse knowledge in long-term memory (Anderson, 1995; Buck, 2001; Field, 2009; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). After these further analyses are performed, all dissected units are connected and sent to the long-term memory (Anderson, 1995; Field, 2009; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). Although this model is described as comprising three phases, they do not occur in a strictly linear sequence; they partially overlap, and the auditory input is processed continuously as it is received (Goh, 2000, 2023; Vandergrift, Goh, 2021). Consequently, the parsing phase does not commence only after the perception phase has been completed; rather, all three phases operate concurrently throughout the listening process.

Originally, this theoretical model focused on L1 listeners' cognitive processes, but it is also applicable to L2 listeners (Ishii, 2025; Goh, 2000). Goh (2000) confirmed the applicability of this theoretical model to L2 listeners. Drawing on learner diaries, group interviews, and immediate retrospective verbalisations from Chinese students in Singapore, his study identified real-time processing challenges (such as limited attentional resources, inferencing errors, and memory overload), which impeded comprehension and were not contradictory to the processing hypothesised in Anderson's (1995) model. Ishii (2025) investigated whether Anderson's (1995) model could be applied to Japanese EFL learners' listening skills. In this study (Ishii, 2025), data were collected from 107 Japanese undergraduates using the TOEFL-Junior listening section and a self-report questionnaire based on Richards' (1983) taxonomy and Anderson's (1995) model. Structural equation modelling and correlation analyses were conducted. The results showed that the three-phase framework statistically fits learners' cognitive processing of listening.

Vandergrift and Goh (2021) added metacognition to Anderson's (1995) model. Metacognition is the ability to monitor, evaluate, and regulate one's cognitive processes involved in perceiving, processing, and responding to information or stimuli (Vandergrift, Goh, 2021). Vandergrift and Goh (2021) explained that

metacognition aids in improving listening skills. According to research based on self-administered questionnaires, this ability seems to manifest across all three phases (Anderson, 1995; Goh, 2000; Ishii, 2025; Vandergrift, Goh, 2021).

Cutler and Clifton (1999) created a model similar to that of Anderson (1995). In this model, four phases were conceptualised, but the final phase showed only the processing of the integration of dissected PSUs whose embedded information had been analysed completely (Anderson, 1995; Cutler, Clifton, 1999; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). Thus, this model can be interpreted as an exceptional variation of the three-phase model (Cutler, Clifton, 1999).

Jia and Hew (2022) also created processing models for listening comprehension. Their model was based on Anderson's (1995) and Field's (2009, 2013) dissociation of processing assumptions into higher and lower levels. According to Field (2009, 2013), lower-level processes involve the automatic and rapid decoding of linguistic input, such as recognising words and parsing syntax. In contrast, higher-level processes involve a more controlled and slower construction of meaning, such as making inferences and integrating ideas into a coherent representation. However, this dissociation contradicts Kohno's (1993) findings from a series of experiments on processing mechanisms. Moreover, Jia and Hew's (2022) model does not reflect the structure of working memory (Baddeley, 2000; Baddeley, Hitch, 1974). Thus, although Jia and Hew's (2022) attempt to integrate Anderson's (1995) model with other theoretical models to provide a more sophisticated account of listening skills is commendable, their model unfortunately exhibits a significant lack of theoretical consistency.

Nagle and Sanders (1986) created a listening comprehension model to explain this skill. Their model consisted of three theoretical phases: short-term memory, long-term memory, and synthesis. Although Nagle and Sanders' model may be regarded as an early attempt to adapt the mechanism of listening comprehension to the field of second language comprehension, emphasising the significance of analysing this skill in terms of cognitive processing, their model reveals significant discrepancies with previous research, most prominently in its positioning of information integration after long-term memory, which raises questions regarding its theoretical validity (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Kadota, 2007, 2015, 2019; Kohno, 1992, 1993, 2001; Nagle, Sanders, 1986; O'Malley et al., 1989; Vandergrift, Goh, 2021).

Necessity to see the flow of processing through the integrated model

Although various models exist in a fragmented manner, and certain commonalities can be discerned upon close examination, this disjointedness tends to render the overall structure obscure and overly complex (Kadota, 2007, 2015,

2019). Integrating these models is necessary to better understand how humans comprehend spoken information.

PURPOSE OF THE STUDY

This study attempted to create a model of listening comprehension to visualise its inter- and intra-relationships by creating a new model based on the two-phase model (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Tamai, 2005), three-phase model (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986), and working memory (Baddeley, 2000; Baddeley, Hitch, 1974) integration, and peripheral important theories and hypothesis (Broadbent, 1956, 1958; Churchill, 1999; Field, 2009, 2013; Goh, 2000; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; O'Malley et al., 1989; Kohno, 1993, 2001; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Liu, 2019; Mareschal, 2007; Mayer, 1997, 2001; Neisser, 1967; Pimsleur, 1971; Pisoni, Sawusch, 1975; Schneider, Shiffrin, 1977; Vandergrift, Goh, 2021; Yantis, 2000). More specifically, this study aimed to construct a model that could adequately explain the process of listening comprehension by integrating the two-phase model (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Mayer, 1997, 2001; Tamai, 2005) into the internationally influential three-phase model (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986), employing a working memory (Baddeley, 2000; Baddeley, Hitch, 1974), which serves as a common foundation for both, as a bridge to ensure conceptual coherence and prevent inconsistencies.

In addition, several components have been uniquely developed and advanced within the framework of the two-phase model (e.g. PSU) (Kohno, 1993). These elements are also compatible with the processing mechanisms proposed in the three-phase model (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986) and can offer compelling explanations for aspects that have remained ambiguous in the latter. However, there is a lack of research that systematically integrates and discusses these elements in relation to the three-phase model (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986), resulting in their continued treatment as exclusive to the two-phase model (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Mayer, 1997, 2001; Tamai, 2005). Our study also addresses this gap. Therefore, following two research questions would arise from these above-mentioned theoretical backgrounds and previous research.

RQ1. How are perceptual, parsing, and utilisation processes functionally organised within the proposed integrated model of listening comprehension?

RQ2. To what extent does the proposed integrated model provide a more coherent explanation of listening comprehension processes compared to existing models?

INTEGRATED MODEL OF LISTENING COMPREHENSION

The above-mentioned elements and hypotheses reveal theoretical models focusing on two phases (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Tamai, 2005) and three phases (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986). Although it may be acknowledged that neither standpoint is inherently erroneous and that both maintain considerable validity, current polarisation engenders confusion and renders the overall picture substantially more difficult to discern, thereby constituting a matter of concern. Therefore, it is necessary to integrate these models.

Based on the above-mentioned elements and hypotheses (Atkinson, Shiffrin, 1968, 1971; Anderson, 1995; Baddeley, Hitch, 1974; Baddeley, 2000; Bondarko et al., 1968/1970; Buck, 2001; Broadbent, 1956, 1958; Churchill, 1999; Cutler, Clifton, 1999; Eto, 2023; Field, 2009, 2013; Goh, 2000, 2023; Ishii, 2022, 2025; Jia, Hew, 2022; Kadota, 2007, 2015, 2019; Kohno, 1992, 1993, 2001; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Liu, 2019; Mareschal, 2007; Mayer, 1997, 2001; Neisser, 1967; Nagle, Sanders, 1986; O'Malley et al., 1989; Pimsleur, 1971; Pisoni, Sawusch, 1975; Schneider, Shiffrin, 1977; Vandergrift, Goh 2021; Yantis, 2000), the model illustrated in Figure 2 can be conceptualised.

To avoid excessive complexity, preliminary auditory processing and primary information analysis stages (Kohno, 1992) were omitted from Figure 2. A model including this classification is shown in Figure 3.

Based on these two integrated models, the mechanism of listening information processing would be elucidated as follows. Spoken message as auditory input enters into the ears, which should be auditory sensory devices (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton 1999; Darwin et al., 1972; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; Neisser, 1967; O'Malley et al., 1989; Pimsleur, 1971; Vandergrift, Goh, 2021). This input then proceeds to the *filter device*, to which attention is paid according to the direction of the central executive of the working memory (Baddeley, 2000; Baddeley, Hitch, 1974). As noted in the "Two-Phase Models" section, the attention being paid to this process is basically selective attention as long as specific clues are given to listeners beforehand or while they are listening, or they already have background knowledge (which is also called "schema") about the topic of the listening information (Bartlett, 1932; Kadota, 2007, 2015,

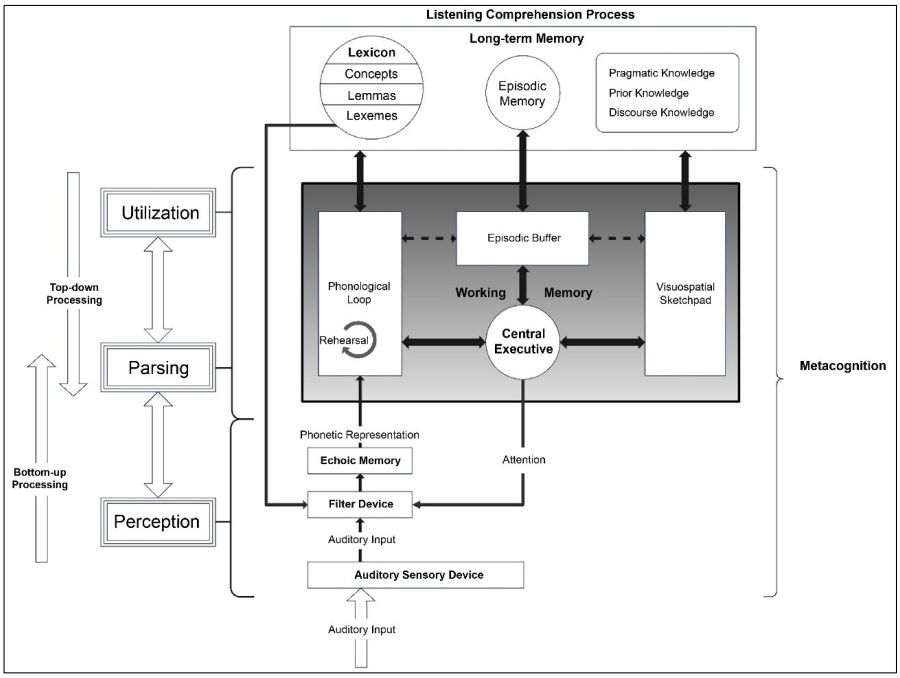

Figure 2. Integrated model of listening comprehension

Source: Author's own study.

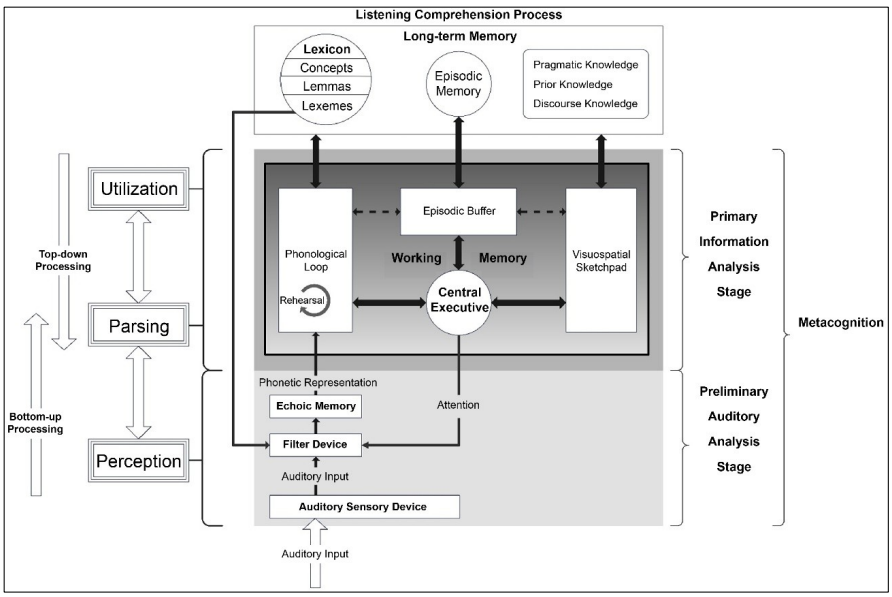

Figure 3. Integrated model of listening comprehension (with preliminary auditory processing and primary information analysis stages)

Source: Author's own study.

2019; Liu, 2019; Yantis, 2000). This is because the characteristics of these three phases of processing overlap considerably (Goh, 2023; Vandergrift, Goh, 2021). If clues or background knowledge are available when listeners listen to the input, they can listen more easily. For example, Madani and Kheirzadeh (2018) found that doing activities to activate schema before listening to auditory input better affects the processing. Moreover, based on Khono's findings (1993, 2001), it could be assumed that when a spoken message is segmented at the filter device through attention control, PSU constitutes the minimal unit. This is the chunk of words or phrases based on their meanings, spoken within 330 ms and whose syllables are in the 7±2 range (Kohno, 1993, 2001). Segmented PSUs are then sent to echoic memory, where phonetic representations for them are created; as it is a sensory device that can keep them only instantaneously, the representations created here are soon sent to the working memory (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton, 1999; Darwin et al., 1972; Goh, 2000; Kadota, 2007, 2015, 2019; Kohno, 1992; Mareschal, 2007; Neisser, 1967; O'Malley et al., 1989; Vandergrift, Goh, 2021). These belong to the processing carried out at the *perception* phase (Anderson, 1995). Moreover, this phase is entirely included in the *preliminary auditory processing stage*, as those processes need to be carried out for further complex analysis (Kohno 1992). Then, the representation transmitted is retained in the phonological loop of working memory (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Goh, 2000; O'Malley et al., 1989; Vandergrift, Goh, 2021). At the phonological loop, various components of information included in the representation are carefully analysed (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Goh, 2000; Kadota, 2007, 2015, 2019; O'Malley et al., 1989; Vandergrift, Goh, 2021). While the phonological loop keeps the representations through rehearsals, lexical, syntactical, and semantic information is retrieved by the direction of the central executive from the mental lexicon in the long-term memory into the phonological loop, and then it processes the representations using this retrieved information (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton, 1999; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). This belongs to the processing carried out at the *parsing* phase (Anderson, 1995). Analyses of vocabulary and grammatical information are conducted, and once these components of the representation have been processed, the analysis of other relevant information is promptly initiated (Anderson, 1995; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). In the subsequent phase, information that the central executive retrieves from long-term memory includes schemata and other components of knowledge related to the representation stored in the phonological loop (Anderson, 1995; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). These belong to the processing carried out during the *utilisation* phase (Anderson, 1995). Moreover, suppose mental imagery

is created in the visuospatial sketchpad based on the meaning of the representation or the listener's schemata during the parsing and utilisation phases. In that case, both the analysed representation and mental imagery are sent to the episodic buffer in the working memory and then integrated at the end of the utilisation phase (Anderson, 1995; Baddeley, Hitch, 1974; Baddeley, 2000; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). After all the analysis is done, the (integrated) representation is finally sent to the long-term memory (Anderson, 1995; Goh, 2000; Mareschal, 2007; O'Malley et al., 1989; Vandergrift, Goh, 2021). The parsing and utilisation phases are entirely included in the *primary information analysis stage*, as it processes representations that have passed through the preceding perceptual processing (Kohno, 1992).

Within these three phases, processed components are continuously routed to subsequent analyses, resulting in processes that are not entirely independent and exhibit substantial overlap (Vandergrift, Goh, 2021; Goh, 2000, 2023). However, the differences among these three phases are still evident, owing to the distinct information and knowledge processed in each phase. According to Goh's (2000) study involving a listening diary, students seem to "not recognise words they know" if the perception phase does not function properly (p. 61). Similarly, students seem to "quickly forget what is heard" (p. 60) or be "unable to form a mental representation from words heard" (p. 64) if the parsing phase does not function properly. Adding to these, students seem to "understand words but not the intended message" (Goh, 2000, p. 62) if the utilisation phase does not function properly. Based on these points, the two integrated models (Figures 2 and 3) are considered to account for the factors underlying listening information processing.

DIFFERENCES BETWEEN THE TWO-PHASE AND THREE-PHASE MODELS

In light of the points mentioned above, the differences between the two-phase and three-phase models should be reconsidered. The largest difference between the two models is in the parsing and utilisation phases (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Cutler, Clifton, 1999; Goh, 2000; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; Kohno, 1992, 1993, 2001; Mareschal, 2007; O'Malley et al., 1989; Pimsleur, 1971; Vandergrift, Goh, 2021). Although processing the input using linguistic knowledge (i.e. grammar, vocabulary, and meaning) and other related knowledge (i.e. prior, pragmatic, or discourse knowledge) are summarised together in two-phase models, they are separated according to knowledge type (Anderson, 1995; Baddeley, 2000; Baddeley, Hitch, 1974; Buck, 2001; Field, 2009; Goh, 2000; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; Kohno, 1992, 1993, 2001; Mareschal, 2007; O'Malley et al., 1989; Pimsleur, 1971; Vandergrift, Goh, 2021). As mentioned before, neither the

two-nor the three-phase models are incorrect, and both have a high degree of validity. However, the bifurcation of perspectives gives rise to uncertainty and substantially impairs the comprehensibility of the overall processing structure. Building upon the above, the potential integration of visual aids in the processing of listening comprehension was also considered.

VISUAL AID FOR LISTENING COMPREHENSION

The above theoretical models examined the processes of listening comprehension using only oral input. However, visual aid is also an important factor in listening comprehension; therefore, its effectiveness and drawbacks regarding comprehending spoken input and its mechanisms have also been studied (Aldera, Mohsen, 2013; Becker, Sturm, 2017; Garza, 1991; Gruba, 1997; Hayati, Mohmedi, 2009; Kashani et al., 2011; Kellerman, 1992; Mohsen, 2016; Mueller, 1980; Tanenhaus et al., 1995). The visual aid for listening comprehension is an optical image (Mueller, 1980; Tanenhaus et al., 1995), and the caption is a written transcription of spoken information adapted to audio-visual input and shown to listeners as an aid (Danan, 2004; Mohsen, 2016). Although optical images are considered to affect listeners' skills to extract context from auditory input (Mueller, 1980; Tanenhaus et al., 1995), captions can also influence listeners to perceive the messages (Garza, 1991; Hayati, Mohmedi, 2009). For instance, Mueller (1980) revealed a positive influence of optical images on listening comprehension. Through experiments using illustrations and video interviews, Mueller found that having students see visual aids affected students' understanding of auditory inputs positively (Mueller, 1980). Through eye-tracking experiments, Tanenhaus et al. (1995) found that optical images could support listeners in comprehending spoken messages. When an appropriate optical image is presented as a visual aid, listeners extract context from the aid and use the extraction to mediate the complexity or ambiguity embedded in the spoken messages (Tanenhaus et al., 1995). The captions also helped listeners in comprehending audio-visual input (Garza, 1991; Hayati, Mohmedi, 2009; Winke et al., 2010).

These findings have led to theoretical considerations of the mechanisms underlying the role of visual aids in listening comprehension (Baddeley, 2000; Gass et al., 2019; Mayer, Sims, 1994; Mayer, 1997, 2001, 2014). One of the core thoughts is that the processing of optical images and captions included in audio-visual input, such as video clips, is deeply intertwined with processing in working memory (Baddeley, 2000; Gass et al., 2019; Mayer, Sims 1994; Mayer, 1997, 2001, 2014). Mayer and Sims (1994) assumed that this process was based on Paivio's dual-coding theory (Clark, Paivio, 1991; Paivio, 1971, 1986). In their model, Mayer and Sims (1994) suggested that working memory (Baddeley, 2000) plays a crucial role in processing audio-visual inputs consisting of spoken

language and visual aids. Subsequently, Mayer (1997, 2001) devised a model that focused on multimedia learning. This updated model also employed working memory as its core.

As the visuospatial sketchpad processes visually presented information from optical input and visual imagery, captions are sent to the phonological loop after iconic memory (an optical-sensation-specific form of sensory storage that momentarily keeps optical input, functioning analogously to the above-mentioned echoic memory for the auditory input), conducting only the essential preliminary casting to create a visual representation (Baddeley, 2000; Baddeley, Hitch, 1974; Darwin et al., 1972; Neisser, 1967; Roland, Gulyás, 1994; Sperling, 1960). This path was omitted from the adapted model to avoid excessive complexity.

INTEGRATED MODEL OF LISTENING COMPREHENSION WITH VISUAL AID

Audio-visual input, which is an auditory input entailing a visual aid, can also be considered using the above model. As in several similar models for processing audio-visual inputs that have input direction to the system that processes visual inputs, the mechanism of listening comprehension with visual aids could also be covered by adding the path to the visuospatial sketchpad after the eyes, which are the optical sensory device that processed optical signals, and following iconic memory to the integrated model based on the above-mentioned elements and hypotheses about audio-visual inputs (Baddeley, 2000; Baddeley, Hitch, 1974; Clark, Paivio, 1991; Mayer, 1997, 2001; Mayer, Sims, 1994; Roland, Gulyás, 1994; Neisser, 1967; Paivio, 1971, 1986; Sperling, 1960), which is illustrated in Figure 4 and can be conceptualised as follows:

DISCUSSION AND IMPLICATIONS FOR FURTHER RESEARCH

This study aimed to integrate two-phase models (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Tamai, 2005), three-phase models (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew 2022; Nagle, Sanders, 1986), working memory (Baddeley, 2000; Baddeley, Hitch, 1974), and important peripheral theories and hypotheses (Bondarko et al., 1968/1970; Buck, 2001; Broadbent, 1956, 1958; Churchill, 1999; Field, 2009, 2013; Goh, 2000; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; O'Malley et al., 1989; Kohno, 1993, 2001; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Liu, 2019; Mareschal, 2007; Mayer, 1997, 2001; Neisser, 1967; Pimsleur, 1971; Pisoni, Sawusch, 1975; Schneider, Shiffrin, 1977; Vandergrift, Goh, 2021; Yantis, 2000) to make the theoretical mechanism of listening comprehension more observable and understandable.

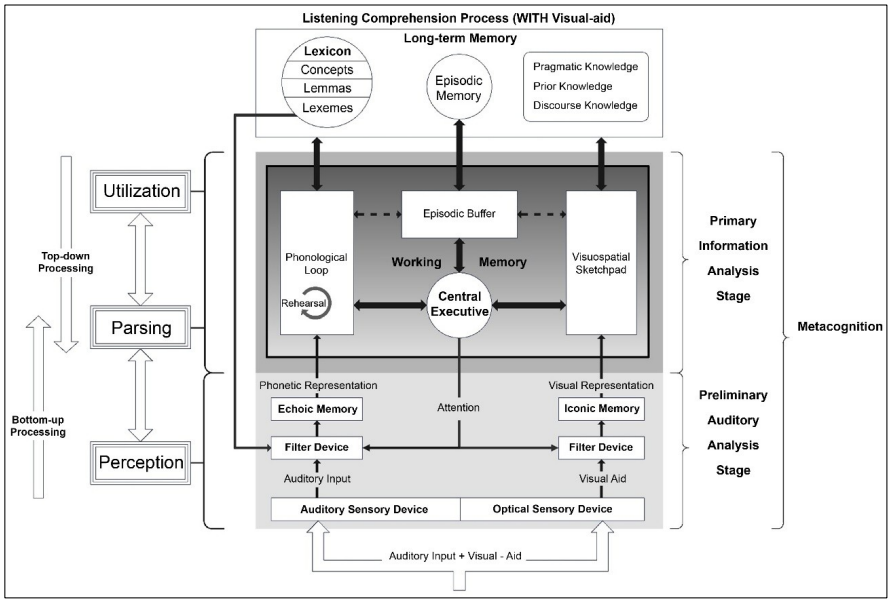

Figure 4. Integrated model of listening comprehension with visual aid

Source: Author's own study.

The research questions were: How are perceptual, parsing, and utilisation processes functionally organised within the proposed integrated model of listening comprehension? (RQ1) and: To what extent does the proposed integrated model provide a more coherent explanation of listening comprehension processes compared to existing models? (RQ2) Through the integration of theoretical models, the model proposed in the current study showed an obvious theoretical advantage to previous models, which would successfully offer answers to the abovementioned research questions.

First and the foremost, primary important cognitive processes, working memory (Baddeley, 2000; Baddeley, Hitch, 1974) and phasal structures of listening comprehension (Anderson, 1995; Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Cutler, Clifton, 1999; Eto, 2023; Frazier, Fodor, 1978; Jia, Hew, 2022; Kadota, 2007, 2015, 2019; Kohno, 1992; Nagle, Sanders, 1986; Tamai, 2005) which, had been discussed apart individually, were showed in a single model. It would offer a more detailed comprehensive view of how listening comprehension is carried out by language learners. As primary important theories of language comprehension, which are working memory and the mechanism of three phases (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986) and two-phases (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Tamai, 2005), are linked

to each other theoretically with referring to peripheral mechanisms (Bondarko et al., 1968/1970; Buck, 2001; Broadbent, 1956, 1958; Churchill, 1999; Field, 2009; 2013; Goh, 2000; Ishii, 2022, 2025; Kadota, 2007, 2015, 2019; O'Malley et al., 1989; Kohno, 1993, 2001; Kormos, 2006; Levelt, 1989, 1993, 1999a; 1999b; Levelt et al., 1999; Liu, 2019; Mareschal, 2007; Mayer, 1997, 2001; Neisser, 1967; Pimsleur, 1971; Pisoni, Sawusch, 1975; Schneider, Shiffrin, 1977; Vandergrift, Goh, 2021; Yantis, 2000), the overview of listening comprehension was made highly clearer. This would also lead to answer the first research question.

Second, this proposed model could be used to guide language teachers how to address learner's listening comprehension problems in more integrated ways. The clear link between processing phases and working memory would let teachers how specific tasks might affect language learner's processing phases of listening comprehension in more comprehensive ways (Field, 2009; Vandergrift, Goh, 2021). For example, Kadota (2007, 2012, 2015, 2019) stated that shadowing training (a training method of listening comprehension to follow and repeat incoming heard spoken inputs as precisely as possible) let language learners automatise their perceptual processing of linguistic inputs. This could be confirmed more precisely through the proposed model. In educational contexts, Anderson's three-phase model (Anderson, 1995) has been more predominant and widely utilized than other theoretical models in previous research (i.e. O'Malley et al., 1989; Vandergrift, Goh, 2021), the model proposed in the current study would offer more lucid and holistic overview of how specific language learning methods such as shadowing would have an affect on each three-phase as well as working memory of language learners. This would contribute to the answer to the second research question.

However, certain limitations and implication to the future research can be seen in this model. As for the limitation, unlike captions in videoclips, which serve as an aid for listeners, listening while reading requires them to pay equal attention to both written and listening inputs, but selectively (Kadota, 2007, 2015, 2019). If the written input is not the "aid," it is the input that listeners have to process enthusiastically while listening to spoken inputs through switching their attention selectively to each input modality (Kadota, 2007, 2015, 2019). Thus, dealing with these two main inputs (not as a cocktail of the main input and "aids," but from these distinct modalities) needs to be considered. Moreover, how the processing of each three phases would be compatible with biological structure of human brain has not been shed light on in this research. In previous studies, Goh (2000) and Ishii (2025) confirmed the possible compatibility between processing mechanisms in each three phases and auditory inputs. However, what was carried out in these studies were dialog and interview (Goh, 2000) or proficiency exam and questionnaire (Ishii, 2025). It means that the processing by each three phases were confirmed only on behavioral responses so that the neuropsychological study

using physiological tools such as fMRI or ERP would be expected. Next should be the implication for the future research. First, it is unclear how spoken messages that are difficult to segment into PSUs would be affected. In Kohno's (1993, 2001) studies, comprehension was negatively influenced when syllable intervals in the materials were 500 ms, even though the pauses by PSUs were available. If the quantity of PSUs constituting the phonological representations that are sent to working memory varies based on syllable interval length, it would influence the accuracy of both the parsing and utilisation phases involved in processing these representations. Second, how information or knowledge in the long-term memory would affect each phase also needs to be investigated further. Madani and Kheirzadeh (2018) found that doing activities before listening influences listeners' comprehension; however, what type of activities affects which phases more or less? The activities used by Madani and Kheirzadeh (2018) before having participants listen to the audio materials included teaching specific words, discussing the topic of the audio material, the teacher's extrinsic control over students' attention, and confirming listening questions. The influence of these four types of activities on processing in each of the three phases would be distinct, but in what way? Third, how the quality of the visual aid would affect each phase is still unclear. In Tanenhaus et al. (1995), an optical image was used as a visual aid. How would this change if the material were presented as captions in a video? Factors such as the font of the captions, the visual clarity of the video, and the overall perceptual salience could plausibly influence listening processes as well.

Therefore, although creating an integrated model of listening comprehension processes has been successful, the manner in which these individual processes influence each of the three phases, as well as how each phase is affected in turn, needs further investigation. More specifically, some hypotheses could be emanated from the proposed model of the current research. For example, (1) reading-while-listening may impose additional demands on working memory, resulting in reduced performance in each phase, specifically decrease available inputs due to their cognitive demands, or (2) the effectiveness of pre-listening activities may vary according to learners' working memory capacity, with higher-capacity learners benefiting more by functions of each phase and working memory in their processing.

CONCLUSIONS

Finally, the significance of the current study is summarised. As a result of the independent evolution and progress of the two-phase (Aryadoust, 2019; Atkinson, Shiffrin, 1968, 1971; Eto, 2023; Frazier, Fodor, 1978; Kadota, 2007, 2015, 2019; Kohno, 1992; Tamai, 2005) and three-phase (Anderson, 1995; Cutler, Clifton, 1999; Jia, Hew, 2022; Nagle, Sanders, 1986) models, the hypotheses

became polarised, making the overall picture of listening comprehension extremely difficult to grasp. Neither of these theories is incorrect, but the issue lies in how processing for comprehension is conceptualised. However, leaving this unresolved means that progress in the field largely depends on the research capabilities of individual researchers. In addition, it is ironic, but the absence or lack of essential components or internal coherence in a theoretical model serves to demonstrate the severe impairment in its adequacy. It is possible to regard these problems as among the fundamental underlying reasons why research on listening comprehension has received relatively little attention (Feyten, 1991; Nazarieh et al., 2022; Osada, 2004). Although the diversity of hypotheses plays a vital role in enabling theoretical explanations of diverse phenomena, it should be considered equally essential to strive for coherence and pursue an integrated, comprehensive perspective, particularly when multiple lines of scientific progress have emerged independently. This study focused on this issue.

DISCLAIMER

Figure 1. Alan Baddeley and Graham Hitch's Working Memory Model (Disclaimer: This figure was adapted with permission of Elsevier Science & Technology Journals, from Alan Baddeley (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. Permission conveyed through Copyright Clearance Center, Inc.)

DECLARATIONS

Funding: The author did not receive support from any organization for the submitted work.

Conflicts of interests: The author has no relevant financial or non-financial interests to disclose.

Ethics approval: This study was conducted solely based on a theoretical approach through extensive review of previously published literature and did not involve any experiments with human participants or animals, in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Therefore, no ethical approval was required so that it was waived.

Consent to participate: This study did not involve human participants or experiments and was conducted by theoretical approach through solely based on a review of existing literature; therefore, informed consent was not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The author does not analyse or generate any datasets, because this work proceeds within a theoretical approach. As for

Figure 2 to 4, high-resolution images for Figures 2–4 are available through appropriate citation (see the Open Practice Statement section).

Open Practice Statement: Figure 2 to 4, high-resolution images for Figures 2–4 are available on the following link: https://doi.org/10.5281/zenodo.19365686. Note that Figure 1, Alan Baddeley and Graham Hitch's Working Memory Model (Baddeley, 2000), was not available in this link as it was used through obtaining permission. As for this model, see the Disclaimer section above and References section below.

Code availability: Not applicable.

Authors' contributions: The author confirms the sole responsibility for the conception of the study, presented results and manuscript preparation.

REFERENCES

Aldera, A.S., Mohsen, M.A. (2013). Annotations in Captioned Animation: Effects on Vocabulary Learning and Listening Skills. *Computers & Education*, *68*, 60–75. DOI: 10.1016/j.compedu.2013.04.018.

Anderson, J.R. (1995). *Cognitive Psychology and Its Implications*. New York: W.H. Freeman and Company.

Aryadoust, V. (2019). An Integrated Cognitive Theory of Comprehension. *International Journal of Listening*, *33*(2), 71–100. DOI: 10.1080/10904018.2017.1397519.

Atkinson, R.C., Shiffrin, R.M. (1968). Human Memory: A Proposed System and Its Control Processes. *Psychology of Learning and Motivation*, *2*, 89–195. DOI: 10.1016/s0079-7421(08)60422-3.

Atkinson, R.C., Shiffrin, R.M. (1971). The Control of Short-Term Memory. *Scientific American*, *225*(2), 82–90. DOI: 10.1038/scientificamerican0871-82.

Baddeley, A. (2000). The Episodic Buffer: A New Component of Working Memory? *Trends in Cognitive Science*, *4*(11), 417–423. DOI: 10.1016/S1364-6613(00)01538-2.

Baddeley, A.D., Hitch, G.J. (1974). Working memory. In: G.A. Bower (Ed.), *Recent Advances in Learning and Motivation* (vol. 8). Academic Press. DOI: 10.1016/s0079-7421(08)60452-1.

Bartlett, F.C. (1932). *Remembering.* Cambridge: Cambridge University Press.

Becker, S.R., Sturm, J.L. (2017). Effects of Audiovisual Media on L2 Listening Comprehension: A Preliminary Study in French. *CALICO Journal*, *34*(2), 147–177. DOI: 10.1558/cj.26754.

Bondarko, L.V., Zagorujko, N.G., Koževnikov, V.A., Molčanov, A.P., Čistovič, L.A. (1970). A Model of Speech Perception in Humans. *Working Papers in Linguistics*, *6* (Trans.). Ohio State University (Original work published 1968), 88–132. Retrieved from: https://linguistics.osu.edu/sites/default/files/workingpapers/osu_wpl_06.pdf (access: 18.05.2025).

Broadbent, D.E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, *8*(4),145–152. DOI: 10.1080/17470215608416814.

Broadbent, D.E. (1958). *Perception and Communication.* London: Pergamon. DOI: 10.1037/10037-000.

Buck, G. (2001). *Assessing Listening.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511732959.

Churchill, Jr., E.F. (1999). The Role of Working Memory in SLA. *TUJ Working Papers in Applied Linguistics*, *14.* Retrieved from: https://www.tuj.ac.jp/grad-ed/publications/working-papers/vol-14/churchill (access: 12.03.2026).

Clark, J.M., Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. DOI: 10.1007/BF01320076.

Cutler, A., Clifton, Jr., C. (1999). Comprehending Spoken Language: A Blueprint of the Listener. In: C.M. Brown, P. Hagoort (Eds.), *The Neurocognition of Language.* Oxford: Oxford University Press.

Danan, M. (2004). Captioning and Subtitling: Undervalued Language Learning Strategies. *Meta*, *49*(1), 67–77. DOI: 10.7202/009021ar.

Darti, Asmawati, A. (2017). Analyzing Students' Difficulties Toward Listening Comprehension. *ETERNAL*, *3*(2), 206–220. DOI: 10.24252/Eternal.V32.2017.A9.

Darwin, C.J., Turvey, M.T., Crowder, R.G. (1972). An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage. *Cognitive Psychology*, *3*(2), 255–267. DOI: 10.1016/0010-0285(72)90007-2.

Egan, M.S. (2018). Memory and Learning: How We Learn by Remembering What Happened and Imagining What Could Have Happened. In: Ó.C. Siochrú (Ed.), *Psychology and the Study of Education; Critical Perspectives on Developing Theories*. Routledge.

Eto, H. (2023). Cognitive Process of Listening for Japanese EFL Learners; Pedagogical Implications for Listening Instruction. *Seinan Gakuin University Graduate School, 16*, 49–62. Retrieved from: http://repository.seinan-gu.ac.jp/bitstream/handle/123456789/2346/gs-n16-p49-62-eto.pdf?sequence=1&isAllowed=y (access: 24.03.2026).

Feyten, C.M. (1991). The Power of Listening Ability: An Overlooked Dimension in Language Acquisition. *The Modern Language Journal*, *75*(2), 173–180. DOI: 10.1111/j.1540-4781.1991.tb05348.x.

Field, J. (2009). *Listening in the Language Classroom.* Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511575945.

Field, J. (2013). Cognitive Validity. In: A. Geranpayeh, L. Taylor (Eds.), *Studies in Language Testing*, 35; *Examining Listening: Research and Practice in Assessing Second Language* (pp. 77–151). Cambridge: Cambridge University Press.

Frazier, L., Fodor, J.D. (1978). The Sausage Machine: A New Two-Stage Parsing Model. *Cognition*, *6*(4)*,* 291–325. DOI: 10.1016/0010-0277(78)90002-1.

Garza, T.J. (1991). Evaluating the Use of Captioned Video Materials in Advanced Foreign Language Learning. *Foreign Language Annals*, *24*(3), 239–258. DOI: 10.1111/j.1944-9720.1991.tb00469.x.

Gass, S., Winke, P., Isbell, D.R., Ahn, J. (2019). How Captions Help People Learn Languages: A Working-Memory, Eye-Tracking Study. *Language Learning & Technology*, *23*(2), 84–104. DOI: 10.64152/10125/44684.

Ghaderpanahi, L. (2012). Using Authentic Aural Materials to Develop Listening Comprehension in the EFL Classroom. *English Language Teaching*, *5*(6), 146–153. DOI: 10.5539/elt.v5n6p146.

Goh, C.C.M. (2000). A Cognitive Perspective on Language Learners' Listening Comprehension Problems. *System*, *28*(1), 55–75. DOI: 10.1016/S0346-251X(99)00060-3.

Goh, C.C.M. (2023). Learners' Cognitive Processing Problems During Comprehension as a Basis for L2 Listening Research. *System*, *119*, 103–164. DOI: 10.1016/j.system.2023.103164.

Graham, S. (2003). Learner Strategies and Advanced Level Listening Comprehension. *The Language Learning Journal*, *28*(1), 64–69. DOI: 10.1080/09571730385200221.

Graham, S. (2006). Listening Comprehension: The Learners' Perspective. *System*, *34*(2), 165–182. DOI: 10.1016/j.system.2005.11.001.

Gruba, P. (1997). The Role of Video Media in Listening Assessment. *System*, *25*(3), 335–345. DOI: 10.1016/S0346-251X(97)00026-2.

Hasan, A.S. (2000). Learners' Perceptions of Listening Comprehension Problems. *Language, Culture and Curriculum*, *13*(2), 137–153. DOI: 10.1080/07908310008666595.

Hayati, A., Mohmedi, F. (2009). The Effect of Film with and Without Subtitles on Listening Comprehension of EFL Learners. *British Journal of Educational Technology*, *42*(1), 181–192. DOI: 10.1111/j.1467-8535.2009.01004.x.

Ishii, Y. (2022). *Using Shadowing With a Communicative Task: An Exploratory Study.* [Unpublished master's thesis]. Sophia University, Japan.

Ishii, Y. (2025). Applicability and Reflection of Three-Phase Model of Language Comprehension on the Listening Skill of Japanese EFL (English as Foreign Language) Learners. *Linguistica Online Misc XVI. 28*, 1–61. Retrieved from: https://www.phil.muni.cz/linguistica/art/ishi/ish-001.pdf (access: 19.01.2026).

Jia, C., Hew, K.F.T. (2022). Supporting Lower-Level Processes in EFL Listening: The Effect on Learners' Listening Proficiency of a Dictation Program Supported by A Mobile Instant Messaging Application. *Computer Assisted Language Learning*, *35*(1–2), 141–168. DOI: 10.1080/09588221.2019.1671462.

Kadota, S. (2007). *Shadowing to ondoku no kagaku* [Science of shadowing and reading aloud]. Tokyo: CosmoPier Publishing Company.

Kadota, S. (2012). *Shadowing, ondoku to eigoshuutoku no kagaku* [Science of shadowing, reading aloud and English acquisition]. Tokyo: CosmoPier Publishing Company.

Kadota, S. (2015). *Shadowing, ondoku to eigo communication no kagaku* [Science of Shadowing, reading aloud and English communication]. Tokyo: CosmoPier Publishing Company.

Kadota, S. (2019). *Shadowing as a Practice in Second Language Acquisition: Connecting Inputs and Outputs*. Routledge. DOI: 10.4324/9781351049108.

Kashani, A.S., Sajjadi, S., Sohrabi, M.R., Younespour, S. (2011). Optimizing Visually-Assisted Listening Comprehension. *The Language Learning Journal*, *39*(1), 75–84. DOI: 10.1080/09571730903545236

Kellerman, S. (1992). "I See What You Mean": The Role of Kinesic Behaviour in Listening and Implications for Foreign and Second Language Learning. *Applied Linguistics*, *13*(3), 239–258. DOI: 10.1093/applin/13.3.239.

Kohno, M. (1992). *Eigojugyou no kaizou, 2nd edn.* [Redevelopment of English language class]. Tokyo: Tokyo Shoseki.

Kohno, M. (1993). Perceptual sense unit and echoic memory. *International Journal of Psycholinguistics*, *9*(1)[25], 13–31.

Kohno, M. (2001). *Onseigengi no ninshiki to seisei no mechanism* [Mechanisms for the recognition and production of spoken language]. Tokyo: Kinseido Publishing.

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah: Lawrence Erlbaum Associates Publishers.

Kurita, T. (2012). Issues in Second Language Listening Comprehension and the Pedagogical Implications. *Accents Asia*, *5*(1), 30–44. Retrieved from: https://accentsasia.org/issues/5-1/kurita.pdf (access: 21.12.2025).

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation.* Cambridge: The MIT Press. DOI: 10.7551/mitpress/6393.001.0001.

Levelt, W.J.M. (1993). The Architecture of Normal Spoken Language Use. In: G. Blanken, J., Dittman, H., Grimm, J.C. Marshall, C.W. Wallesch (Eds.)., *Linguistic Disorders and Pathologies: An International Handbook* (pp. 1–15). Berlin: de Gruyter.

Levelt, W.J.M. (1999a). Language Production: A Blueprint of the Speaker. In: C. Brown, P. Hagoort (Eds.), *Neurocognition of Language* (pp. 83–122). Oxford: Oxford University Press.

Levelt, W.J.M.(1999b). Models of Word Production. *Trends in Cognitive Science*, *3*(6), 223–232. DOI: https://doi.org/10.1016/s1364-6613(99)01319-4.

Levelt, W.J.M., Roelofs, A., Meyer, A.S. (1999). A Theory of Lexical Access in Speech Production. *Behavioral and Brain Sciences*, *22*(1), 1–38. DOI: 10.1017/s0140525x99001776.

Li, C. (2014). Application of Information Processing Strategies to Discourse Listening Some Cognitive Approaches. In: *Proceedings of the International Conference on Education, Language, Art and Intercultural Communication (ICELAIC-14)*. Atlantis Press. DOI: 10.2991/icelaic-14.2014.71.

Liu, T. (2019). Feature-Based Attention: Effects And Control. *Current Opinion in Psychology*, *29*, 187–192. DOI: 10.1016/j.copsyc.2019.03.013.

Loftus, G.R., Loftus, E.F. (1976). *Human Memory: The Processing of Information.* Mahwah: Lawrence Erlbaum. Retrieved from: http://dp20072271.lolipop.jp/dialogue/no03_2004/2004dialogue03_k4.pdf (access: 12.01.2026).

Madani, B.S., Kheirzadeh, S. (2018). The Impact of Pre-Listening Activities on EFL Learners' Listening Comprehension. *International Journal of Listening*, *36*(1), 53–67. DOI: 10.1080/10904018.2018.1523679.

Mareschal, C. (2007). *Student Perceptions of a Self-Regulatory Approach to Second Language Listening Comprehension Development.* [Unpublished doctoral dissertation]. University of Ottawa.

Mayer, R.E. (1997). Multimedia Learning: Are We Asking the Right Questions? *Educational Psychologist*, *32*(1), 1–19. DOI: 10.1207/s15326985ep3201_1.

Mayer, R.E. (2001). *Multimedia Learning.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139164603.

Mayer, R.E. (2014). Cognitive Theory of Multimedia Learning. In: R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43–71). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139547369.005.

Mayer, R.E., Sims, V.K. (1994). For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning. *Journal of Educational Psychology*, *86*(3), 389–401. 10.1037/0022-0663.86.3.389.

Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, *63*(2), 81–97. DOI: 10.1037/h0043158.

Mohsen, M.A. (2016). The Use of Help Options in Multimedia Listening Environments to Aid Language Learning: A Review. *British Journal of Educational Technology*, *47*(6), 1232–1242. DOI: 10.1111/bjet.12305.

Mueller, G.A. (1980). Visual Contextual Cues and Listening Comprehension: An Experiment. *The Modern Language Journal*, *64*(3), 335–340. DOI: 10.1111/j.1540-4781.1980.tb05202.x.

Nagle, S.J., Sanders, S.L. (1986). Comprehension Theory and Second Language Pedagogy. *TESOL Quarterly*, *20*(1), 9–26. DOI: 10.2307/3586386.

Nazarieh, M., Razmi, M.H., Azizian, M., Ghani, M.U. (2022). A Brief History of Listening Comprehension in Second Language Teaching and Learning. *Education Research International*, *1*, 1–5. DOI: 10.1155/2022/3894457.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

O'Malley, J.M., Chamot, A.U., Küpper, L. (1989). Listening Comprehension Strategies in Second Language Acquisition. *Applied Linguistics*, *10*(4), 418–437. DOI: 10.1093/applin/10.4.418.

Osada, N. (2004). Listening Comprehension Research: A Brief Review of the Past Thirty Years. *Dialogue*, *3*, 53–66. Retrieved from: https://linguistics.osu.edu/sites/default/files/workingpapers/osu_wpl_06.pdf (access: 12.12.2025).

Paivio, A. (1971). *Imagery and Verbal Processes*. Austin: Holt, Rinehart & Winston.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press.

Pimsleur, P. (1971). Some Aspects of Listening Comprehension. In: L.J. Chatagnier, G. Taggart (Eds.), *Laboratoires de langues: Orientations nouvelles.* Montréal: Editions Aquila Limitée.

Pisoni, D.B., Sawusch, J.R. (1975). Some Stages of Processing in Speech Perception. In: A. Cohen, S.G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 16–35). New York: Springer-Verlag. DOI: 10.1007/978-3-642-81000-8_2.

Richards, J.C. (1983). Listening Comprehension: Approach, Design, Procedure. *TESOL Quarterly*, *17*(2), 219–240. DOI: 10.2307/3586651.

Roland, P.E., Gulyás, B. (1994). Visual Imagery and Visual Representation. *Trends in Neuroscience, 17*(7), 281–297. DOI: 10.1016/0166-2236(94)90057-4.

Rost, M. (1994). *Introducing Listening.* London: Penguin Books.

Rubin, J. (1995). An Overview of a Guide for the Teaching of Second Language Listening. In: D.J. Mendelsohn, J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 7–11). Singapore: Dominie Press.

Schneider, W., Shiffrin, R.M. (1977). Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention. *Psychological Review*, *84*(1), 1–66. DOI: 10.1037/0033-295X.84.1.1.

Sperling, G. (1960). The Information Available in Brief Visual Presentations. *Psychological Monographs: General and Applied*, *74*(11),1–29. DOI: 10.1037/h0093759.

Tamai, K. (2005). *Listening shidouhou to shite no shadowing no koukanikansuru kenkyu* [Study of the effect of shadowing as a method for teaching listening comprehension]. Tokyo: Kazama Shobo.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. DOI: 10.1126/science.7777863.

Vandergrift, L. (1997). The Comprehension Strategies of Second Language (French) Listeners: A Descriptive Study. *Foreign Language Annals*, *30*(3), 387–409. DOI: 10.1111/j.1944-9720.1997.tb02362.x.

Vandergrift, L., Goh, C.C.M. (2021). *Teaching and Learning Second Language Listening: Metacognition in Action*. Routledge.

Winke, P., Gass, S., Sydorenko, T. (2010). The Effects of Captioning Video Used for Foreign Language Listening Activities. *Language Learning & Technology*, *14*(1), 65–86. DOI: 10.64152/10125/44203.

Xu, F. (2011). Anxiety in EFL Listening Comprehension. *Theory and Practice in Language Studies*, *1*(12), 1709–1717. DOI: 10.4304/tpls.1.12.1709-1717.

Yantis, S. (2000). Goal-Directed and Stimulus-Driven Determinants of Attentional Control. In: S. Monsell, J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII.* Cambridge: MIT Press. DOI: 10.7551/mitpress/1481.003.0010.

ABSTRAKT

Artykuł podejmuje problem inherentnej trudności rozumienia ze słuchu w języku drugim (L2), w szczególności jego nieobserwowalnego charakteru poznawczego, poprzez zaproponowanie zintegrowanego modelu rozumienia ze słuchu. Dotychczasowe badania wypracowały modele dwufazowe (percepcja i rozumienie) oraz modele trójfazowe (percepcja, parsowanie i wykorzystanie), jednak ich fragmentaryczny rozwój doprowadził do niespójności pojęciowej. Aby temu zaradzić, niniejsze badanie integruje te modele, wykorzystując pamięć roboczą jako ramę jednoczącą i uwzględniając kluczowe komponenty, takie jak pętla fonologiczna, centralny system wykonawczy oraz bufor epizodyczny, a także koncepcję percepcyjnych jednostek sensu (PSU). Zaproponowany model wyjaśnia, w jaki sposób bodźce słuchowe są przetwarzane od początkowej percepcji, przez analizę językową, aż do konstrukcji znaczenia na wyższym poziomie, pokazując, że fazy te działają w sposób interaktywny, a nie linearny. Ponadto model zostaje rozszerzony o uwzględnienie wsparcia wizualnego, podkreślając rolę szkicownika wzrokowo-przestrzennego w przetwarzaniu audiowizualnym. Zintegrowane ujęcie dostarcza bardziej spójnego i kompleksowego wyjaśnienia mechanizmów rozumienia ze słuchu oraz oferuje praktyczne implikacje dla nauczania języków obcych, takie jak doskonalenie strategii dydaktycznych i diagnozowanie trudności uczących się. W badaniu wskazano również kierunki

dalszych analiz, w tym wpływ pojemności pamięci roboczej oraz bodźców multimodalnych na procesy rozumienia ze słuchu.

Słowa kluczowe: rozumienie ze słuchu; model dwufazowy; model trójfazowy; model zintegrowany; pamięć robocza