English Language Teaching Educational Journal
Vol. 9, No. 1, April 2026, pp. 156-172
ISSN 2621-6485

The mondegreen effect in L2 speech perception: An investigation of phonological ambiguity and cognitive expectation in the speech perception of Indonesian university students

Adi Sutrisno [a,1*], Misnadin [b,2]

[a] Universitas Gadjah Mada, Bulaksumur, Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281
[b] Universitas Trunojoyo Madura, Jl. Raya Telang, Kamal-Bangkalan, Madura, East Java, Indonesia
[1] adisutrisno@ugm.ac.id, [2] misnadin@trunojoyo.ac.id

* Corresponding author

ARTICLE INFO

Article history
Received January 28, 2026
Revised February 23, 2026
Accepted April 27, 2026
Available Online April 30, 2026

Keywords
Indonesian EFL learners
L2 speech perception
Mondegreen effect
Phonological ambiguity
Top-down processing

ABSTRACT

This study investigates the Mondegreen effect in second language (L2) listening by examining how phonological ambiguity and cognitive expectation interact to shape auditory misperception among Indonesian university-level EFL learners. While prior research has often treated L2 listening difficulties as either phonological decoding problems or failures of top-down processing, this study offers an integrated account by analyzing misperceptions across multiple linguistic levels. Data were collected from 165 Indonesian undergraduate students through a listening transcription task involving 12 naturally produced English utterances characterized by connected speech, prosodic variation, and phonological reduction. A total of 1,980 responses were analyzed and categorized into word, phrase, clause, and sentence-level misperceptions. The findings show that 39.9% of responses involved misperception, with sentence-level reinterpretations emerging as the most frequent pattern, followed by word-level substitutions and lower rates at clause and phrase levels. The analysis demonstrates that auditory misperception is systematically triggered by reduced and ambiguous speech signals, which obscure segmentation and activate competing lexical candidates. Crucially, listeners do not merely fail to decode input but actively reconstruct meaning, often relying on top-down expectations that override bottom-up acoustic cues. This study contributes to L2 speech perception research by reframing Mondegreens as evidence of dynamic meaning construction rather than perceptual error. It also highlights the interaction between phonological processing and cognitive expectation in real-time listening. Pedagogically, the findings suggest the need for greater emphasis on prosodic training, segmentation awareness, and metacognitive strategies to enhance learners' perceptual resilience in authentic listening contexts.

How to Cite: Sutrisno, A., & Misnadin (2026). The mondegreen effect in L2 speech perception: An investigation of phonological ambiguity and cognitive expectation in the speeches of Indonesian university students. English *Language Teaching Educational Journal, 9*(1), 165-181. https://doi.org/10.12928/eltej.v9i1.15721

1. Introduction

Listening comprehension is widely recognized as one of the most demanding skills in second language acquisition. Unlike reading or writing, listening requires learners to process spoken input in real time, often under conditions of phonetic reduction, prosodic variation, and connected speech (Sutrisno, 2015; Sutrisno, 2018; West, 2023; Graham, 2017; Siegel, 2018; Vandergrift & Goh, 2012; Zhang, 2023; Aryadoust & Luo, 2023; Liu et al., 2025). EFL learners frequently misinterpret spoken utterances, sometimes believing they have understood correctly, only to later discover discrepancies between what was said and what was understood (Sutrisno, 2018; Vitevitch & Lachs, 2024). A striking illustration of

such misperception is the Mondegreen effect, where a listener replaces a misheard word or phrase with one that sounds similar but carries a different meaning.

The term *Mondegreen* was first coined by Wright (1954), who misheard a line from a Scottish ballad, "laid him on the green," as "Lady Mondegreen." Originally associated with misheard lyrics, the term has since gained broader recognition in studies of spoken language perception and is now widely used to describe how listeners reinterpret ambiguous auditory input when acoustic signals are degraded or unclear.

In Indonesia, a well-known example of a Mondegreen is found in children's rendition of the patriotic song *Garuda Pancasila*. The phrase *pribadi bangsaku* was often misheard as *pribang-pribangsaku*, a sequence with no lexical meaning but widely reproduced across generations. This distortion likely emerged from a combination of limited lexical knowledge, reduced auditory clarity, and unfamiliarity with abstract vocabulary. Interestingly, such misperceptions did not diminish the intended meaning, suggesting that listeners actively construct meaning even when perception is imperfect (Hartono, 2020).

These examples, though seemingly anecdotal, point to a deeper issue in EFL listening research: how learners interpret ambiguous speech input and what cognitive and phonological processes shape their misperceptions. Recent syntheses of L2 listening research emphasize that listening is a multidimensional construct involving decoding, inferencing, and strategic processing under real-time constraints (Aryadoust & Luo, 2023; Zhang & Shen, 2023; Liu et al., 2025). Moreover, listening performance is influenced not only by input characteristics but also by learner-related factors such as vocabulary knowledge, metacognitive awareness, and self-efficacy (Du & Man, 2022; Wang & Treffers-Daller, 2017). In addition, research on extensive listening demonstrates that sustained exposure to spoken input can enhance listening fluency and processing efficiency, highlighting the importance of input quantity and continuity in L2 listening development (Chang & Millett, 2016).

The Mondegreen effect has attracted attention across multiple disciplines. In applied linguistics, research has examined how phonological complexity and reduced speech forms interfere with lexical access. Psycholinguistic studies show that listeners rely heavily on predictive processing and prior expectations when interpreting ambiguous input, often minimizing prediction error to facilitate comprehension (Sohoglu & Davis, 2016). More recent evidence also indicates that perceptual learning allows listeners to adapt to degraded speech over time, although such adaptation does not eliminate ambiguity entirely (Scharenborg & Janse, 2019). These findings suggest that speech perception is not purely bottom-up but involves continuous interaction between incoming signals and internal expectations (Castro & Vitevich, 2023; Beck et al., 2014; Yasoda-Mohan et al., 2025)

One key challenge in second language listening lies in decoding connected speech, where processes such as vowel reduction, assimilation, and stress shifts obscure word boundaries (Cauldwell, 2013; Roach, 2009; Sutrisno, 2015; Tucker & Warner, 2020). Recent research further demonstrates that reduced forms in natural speech significantly affect L2 listeners' ability to recognize words, especially when canonical forms differ from their surface realizations (Reinisch et al., 2020). Additionally, aural decoding ability has been shown to play a central role in determining listening comprehension success (Ke & Wang, 2022; Leonard, 2019), while limitations in vocabulary knowledge constrain recognition of even high-frequency words in continuous speech (Matthews, 2018; Matthews & Cheng, 2015).

Such misperceptions are not random but reflect a systematic interaction between bottom-up and top-down processing. According to the Ganong effect (Ganong, 1980), listeners tend to resolve ambiguous input in favour of familiar lexical items. This aligns with ongoing debates in L2 listening pedagogy regarding whether instruction should emphasize bottom-up decoding or top-down inferencing. Rather than privileging one over the other, recent perspectives suggest that both processes are complementary and must be integrated for effective listening comprehension (Yeldham, 2018). Furthermore, challenging listening conditions, such as degraded input or noise, increase reliance on predictive mechanisms and contextual inference (Fujita, 2022).

Cross-linguistic influence further shapes this process. Learners often transfer phonological and rhythmic patterns from their first language, affecting how they segment and interpret second language input (Darcy et al., 2012; Saito & Plonsky, 2019). This is particularly relevant for Indonesian learners, whose syllable-timed language background contrasts with the stress-timed rhythm of English, making it difficult to detect reduced syllables and weak forms in continuous speech.

Another contributing factor is the gap between classroom input and real-world listening conditions. Many EFL materials rely on slow and clearly articulated speech, which differs significantly from the reduced, variable, and prosodically complex nature of authentic spoken language. This mismatch may leave learners underprepared for real-time listening, increasing their susceptibility to misperception. At the same time, increased exposure to naturalistic input has been shown to improve listening fluency, although it does not fully eliminate perceptual ambiguity (Chang & Millett, 2016).

Taken together, these studies suggest that auditory misperception emerges from the dynamic interaction of phonological variability, cognitive expectation, and cross-linguistic influence. However, relatively few studies have examined how these factors operate simultaneously in EFL listening contexts, particularly among Indonesian learners. More importantly, the Mondegreen effect remains underexplored as a lens for understanding how learners reconstruct meaning when speech input is ambiguous or degraded.

Given this gap, and in line with recent calls for more integrated models of L2 listening that account for both perceptual and cognitive dimensions (Aryadoust & Luo, 2023), the present study aims to investigate how phonological ambiguity and cognitive expectation interact in shaping auditory misperception. Specifically, this study addresses the following research questions:

1. What phonological and prosodic features contribute to Mondegreen occurrences in Indonesian EFL listening?
2. How do top-down cognitive expectations interact with bottom-up speech features in shaping auditory misperception?

2. Method

2.1. Participants

Participants were 165 undergraduate students drawn from multiple academic disciplines across five major Indonesian cities: Jakarta, Yogyakarta, Semarang, Surabaya, and Madura. The sample included students from both English-related majors (e.g., English Language Education and English Literature) and non-English majors (e.g., Economics, Engineering, and Social Sciences), allowing for a broader representation of EFL listening backgrounds. All participants had received formal English instruction as part of the Indonesian education system for approximately 6 to 12 years, beginning in primary or secondary school. However, their exposure to authentic spoken English varied, as English is primarily learned as a foreign language in classroom settings.

Eligibility criteria included active enrollment, age between 18 and 25, and no extended stay (over six months) in English-speaking countries. From an initial pool of 180 students, one submission was excluded due to incompleteness, and 14 were removed after data screening revealed highly similar transcriptions, suggesting potential collusion. The final dataset comprised 165 valid and independent responses.

Due to logistical constraints, standardized proficiency tests (e.g., TOEFL, IELTS) were not administered. While this limited detailed proficiency analysis, it allowed for a more ecologically valid sampling of diverse EFL listening abilities. All participants gave informed consent, and the study received ethical clearance.

2.2. Instrument

The main instrument was a listening misperception task consisting of 12 short English sentences, read naturally by a native American English speaker. These sentences were designed to include features that often trigger auditory misperception, such as elision, assimilation, stress shifts, and reduced forms. Participants were instructed to transcribe each sentence exactly as they heard it. Prior to implementation, the task was piloted with a separate group of participants to ensure clarity and appropriate difficulty. In addition, the instrument was reviewed by two experts in applied linguistics and L2 listening to establish content validity, particularly in terms of phonological features, prosodic variation, and potential for eliciting misperception. Feedback from both the pilot group and expert reviewers was used to refine the stimuli, ensuring that they consistently represented natural spoken English while maintaining an appropriate level of perceptual challenge.

2.3. Stimulus Sentences

To investigate auditory misperception in EFL listening, a set of twelve carefully designed stimulus sentences was employed. These sentences were selected to represent a range of phonological, prosodic, and lexical features commonly associated with perceptual ambiguity in natural spoken English. In particular, the stimuli were constructed to elicit potential Mondegreen effects by incorporating elements such as reduced forms, connected speech processes (e.g., elision and assimilation), and resegmentable sequences that may challenge listeners' parsing strategies.

In addition, the sentences vary in lexical frequency and semantic predictability, allowing for the observation of the interplay between bottom-up acoustic processing and top-down cognitive expectations. This design enables a more nuanced analysis of how Indonesian EFL learners reconstruct meaning under conditions of phonological uncertainty.

The stimulus sentences are presented as follows:

a) If Linda calls, tell her I am at the bank.
b) They're meant to help us.
c) I knew a guy once who worked in finance.
d) A toddler walks into the world with curiosity.
e) We met her in the hallway.
f) I scream for ice cream.
g) That's the show I'd like to see.
h) The tenancy has been terminated early.
i) He didn't mean to upset you, I guess.
j) I need a nice T-shirt.
k) Our new coffee table is made of copper.
l) I saw Phil and Tim at the scene of the accident.

These sentences were selected to present a range of phonological and prosodic challenges, including reduced forms, resegmentable sequences, and rhythm patterns that contrast with syllable-timed first-language backgrounds. Both frequent and less frequent lexical items were included to capture the interaction between top-down expectations and bottom-up perceptual processing.

2.4.Procedure

The task was administered in a group setting in quiet classroom environments across the participating institutions. All participants completed the task simultaneously under the supervision of the researcher or a trained assistant. The audio stimuli were played centrally using classroom audio (Listening) equipment (speakers connected to a laptop), ensuring that all participants received identical auditory input. Each sentence was presented once without repetition to simulate real-time listening conditions. Participants were given one minute to transcribe each sentence after it was played. The total duration of the task ranged from approximately 25 to 30 minutes. Participants were instructed to write down exactly what they believed they heard, even if they were uncertain, and were explicitly told not to use any external aids such as dictionaries, mobile phones, or peer discussion. The instructions emphasized reliance on auditory perception rather than guessing or correction.

2.5.Data Analysis

All transcriptions were collected in orthographic form, as participants were instructed to write down what they heard using standard English spelling rather than phonetic notation. This approach was selected to reflect natural listening conditions and to capture perceptual interpretation rather than phonetic precision. The dataset was first screened for completeness and originality. Completeness was defined as the presence of a written response for each stimulus item, excluding blank or unintelligible entries that did not allow meaningful interpretation. Originality was assessed by identifying highly similar or identical response patterns across participants; responses showing extensive overlap across multiple items were considered indicative of potential collusion and were excluded from analysis. Following this screening process, the final dataset included 165 participants × 12 items, resulting in 1,980 analyzable utterances.

All responses were then analyzed using a coding framework that categorized misperceptions into four linguistic levels: word, phrase, clause, and sentence. Mondegreens were defined as phonetically plausible

yet semantically altered reconstructions of the original utterance. Mishearing involving multiple words were classified at the highest relevant level, and components were not double-counted unless independently misperceived.

Two trained raters independently coded all transcriptions based on this framework. Interrater reliability was high (Cohen's Kappa = 0.89), indicating strong agreement. Any discrepancies were resolved through discussion and re-examination of the original audio stimuli to ensure consistency and accuracy.

The analysis combined quantitative and qualitative approaches. Quantitatively, frequencies of misperception types were calculated across linguistic levels. Qualitatively, patterns of misperception were examined in relation to (a) phoneme substitution or elision, (b) word boundary missegmentation, (c) lexical expectation and top-down inference, and (d) prosodic interference arising from stress and rhythm mismatches. This combined approach allowed for a comprehensive account of how and why Mondegreens emerge in real-time EFL listening.

3. Result

This section presents the quantitative results from a listening transcription task completed by 165 Indonesian EFL learners, each responding to 12 spoken English utterances. A total of 1,980 utterances were analyzed. Each response was classified into one of six categories: Word-level Mondegreen, Phrase-level Mondegreen, Clause-level Mondegreen, Sentence-level Mondegreen, Unintelligible response, or Verbatim transcription.

Mondegreens were observed in 790 of the 1,980 utterances, accounting for **39.9%** of all responses. These were further categorized by linguistic complexity as shown below:

Table 1. Linguistic level of mondegreen and each number of occurrences

Mondegreen	Number of occurrences
Word level	313 (15.81%)
Phrase level	38 (1.92%)
Clause level	51 (2.58%)
Sentence level	388 (19.60%)

In addition to Mondegreens, 196 utterances (9.90%) were unintelligible, and 994 utterances (50.20%) were correctly transcribed verbatim.

Table 2. Distribution of Mondegreen Types Across Stimuli

Stimulus	Word Level	Phrase Level	Clause Level	Sentence Level	Unintelligible	Verbatim
1	27	0	18	24	41	55
2	36	0	11	31	19	68
3	10	1	4	56	21	73
4	3	0	13	44	29	76
5	30	0	1	24	11	99
6	58	0	0	0	6	101
7	14	0	2	16	12	121
8	74	0	0	30	15	46
9	28	0	1	21	8	107
10	4	2	0	9	5	145
11	26	26	1	43	10	59
12	3	9	0	90	19	44
Total	**313**	**38**	**51**	**388**	**196**	**994**

The findings reveal that 39.9% of the utterances were misperceived in some form, representing a substantial proportion of perceptual deviation even among university-level learners. This result highlights the inherent difficulty of processing naturally connected speech in real-time listening.

The classification process was conducted by two trained raters using a standardized coding protocol. Interrater reliability was high (Cohen's Kappa = 0.89), indicating strong agreement and supporting the consistency of the data. Although a stimulus-by-stimulus breakdown is not the primary focus of this section, certain items consistently elicited higher rates of sentence-level misperceptions. To illustrate this pattern, Table X presents the distribution of sentence-level misperceptions across all stimuli. These patterns are further examined in the Discussion section.

Table 3. Sentence-Level Misperceptions Across Stimuli

Stimulus	Sentence-Level Misperceptions
1	24
2	31
3	56
4	44
5	24
6	0
7	16
8	30
9	21
10	9
11	43
12	90
Total	**388**

As shown in Table 3, sentence-level misperceptions are not evenly distributed across stimuli. Certain items (e.g., Stimuli 3, 4, 11, and 12) show notably higher frequencies, suggesting that specific phonological reduction patterns or prosodic configurations may increase susceptibility to reinterpretation at the sentence level.

4. Discussion

This section discusses the findings of the study by analyzing the types and causes of auditory misperception, specifically Mondegreens, observed among Indonesian EFL learners. As demonstrated in the Results section, misperceptions occurred across four linguistic levels: word, phrase, clause, and sentence. Among these, sentence-level Mondegreens were the most frequent, but all four levels revealed distinct patterns of listener error that reflect the complex interplay between bottom-up phonological input and top-down cognitive expectation, in line with the study's title and guiding research questions.

More precisely, RQ1 explored which phonological and prosodic features contribute to Mondegreen occurrences, while RQ2 examined how learners' internal expectations shape auditory interpretations under ambiguous conditions. The following analysis traces the distribution and nature of these misperceptions at each level, beginning with the word level, where isolated lexical items were replaced with phonetically similar alternatives. These instances serve as empirical evidence of how reduced syllables, unclear stress patterns, and segmental ambiguity can disrupt real-time decoding.

4.1. Word-level Mondegreens

Word-level Mondegreens, which accounted for 15.81% of all responses (313 out of 1980 utterances), typically involved misperceptions of individual lexical items due to segmental similarity and stress-induced phonological blending. Three salient examples from the dataset are analyzed below: "tenancy," "scream," and "meant." This level of misperception is particularly revealing, as it isolates the earliest stage of lexical access, where listeners must rapidly map acoustic input onto stored phonological representations under time pressure (Vitevitch & Lachs, 2024). This finding is also consistent with recent research highlighting the central role of aural decoding in L2 listening, where learners must efficiently map acoustic signals onto lexical representations under time constraints (Ke & Wang, 2022; Leonard, 2019). In addition, limitations in vocabulary size have been shown to constrain the recognition of reduced or less frequent lexical items in continuous speech (Matthews, 2018; Matthews & Cheng, 2015).

First, the word "tenancy" was frequently misheard as "tendency." This substitution exemplifies what Ganong (1980) described as lexical bias, where ambiguous phonemes are resolved in favor of more familiar lexical entries. More recent studies have reaffirmed that such bias is not merely frequency-driven but also shaped by probabilistic expectations and lexical competition within the mental lexicon

(Scharenborg & Janse, 2019; Vitevitch & Lachs, 2024). While "tenancy" is a relatively low-frequency English word, "tendency" exists as a high-frequency loanword in Indonesian (*tendensi*), making it a more accessible and cognitively primed alternative. This supports earlier findings by Roach (2009) and Scherling et al. (2022), who noted that L2 listeners tend to substitute unfamiliar phonemes or lexical items with those that align more closely with their L1 phonological and semantic inventories. Phonologically, the nasal-stop sequence in "tenancy" (/ˈtɛnənsi/) may be simplified or misperceived due to reduction of the medial unstressed syllables, resulting in a perceptual form closer to /ˈtɛndənsi/. This kind of reduction-induced ambiguity is also widely reported in studies of natural speech processing, where unstressed syllables are particularly vulnerable to perceptual loss (Erb & Obleser, 2019).

Second, the verb "scream" in the phrase "I scream for ice cream" was often interpreted simply as "cream." When preceded by the pronoun "I" and spoken with natural stress and linking in connected speech (e.g., /aɪ ˈskriːm fə ˈaɪsˌkriːm/), the initial phrase "I scream" merges perceptually into the noun "ice cream." This misperception reflects classic phonological ambiguity, where blending and elision erase clear word boundaries. Contemporary research on connected speech confirms that coarticulation, reduction, and resyllabification significantly weaken lexical segmentation cues, especially for L2 listeners (Cauldwell, 2013; Kang et al., 2019). Learners appear to activate the more familiar and meaningful collocation "ice cream," which is pragmatically plausible and commonly encountered, rather than decoding the verb "scream" in isolation. This example illustrates how top-down expectation (RQ2), informed by lexical familiarity and collocational salience, overrides ambiguous bottom-up input (RQ1). Such interaction between perceptual ambiguity and expectation-driven inference aligns with predictive processing accounts of speech perception, in which listeners continuously generate and update lexical hypotheses in real time (Pickering & Gambi, 2018).

Third, the modal verb "meant" in "They're meant to help us" was interpreted variously as "may," "might," or even "mind." This diversity of reconstructions suggests both phoneme-level confusion and expectation-driven reinterpretation. The original verb /mɛnt/ may be weakened by prosodic reduction or glottal masking, leading to mishearing as /meɪ/, /maɪt/, or /maɪnd/, all plausible English modals or verbs that would fit syntactically in the same environment. Learners' brains seem to favor more frequent or expected auxiliaries ("may," "might") over the less common "meant," especially when the auditory signal is compressed or lacks prominence. This supports Castro and Vitevitch's (2023) claim that dense phonological neighborhoods, in which many similar-sounding words compete for recognition, create conditions ripe for misperception. Recent work further suggests that lexical neighbors with higher frequency and stronger semantic associations are more likely to dominate perceptual outcomes under uncertainty (Vitevitch & Lachs, 2024).

Collectively, these word-level Mondegreens demonstrate how even single-word input can be distorted through the interaction of segmental similarity, syllabic reduction, lexical familiarity, frequency effects, and native language phonotactics. These findings resonate with probabilistic and noisy-channel models of language comprehension, which posit that listeners integrate imperfect acoustic input with prior expectations to arrive at the most likely interpretation (Gibson et al., 2019). They also reinforce the idea that EFL listening comprehension is not simply a matter of decoding speech sounds but involves active hypothesis testing under conditions of ambiguity. When acoustic cues are degraded or unfamiliar, Indonesian EFL learners fall back on probabilistic lexical selection strategies, reconstructing the most likely meaning based on what could be heard rather than what was said. This supports recent findings that listening performance is jointly shaped by decoding ability, vocabulary knowledge, and strategic processing, all of which interact dynamically during real-time comprehension (Du & Man, 2022; Wang & Treffers-Daller, 2017).

4.2. Phrase-level Mondegreens

Although less frequent than word or sentence-level errors, phrase-level Mondegreens accounted for 1.92% of the total responses (38 out of 1980 utterances). These misperceptions reflect how prosodic features, especially in stress-timed speech, can lead EFL learners to resegment lexical chunks into unintended phrases that are semantically and grammatically plausible. Recent research in L2 speech perception highlights that segmentation in continuous speech is particularly vulnerable to prosodic reduction and coarticulation, which obscure word boundaries and increase reliance on inferential processing (Erb & Obleser, 2019; Kang et al., 2019). Recent studies further indicate that learners engaged in self-regulated listening tasks tend to rely more heavily on inferencing and restructuring strategies when segmentation cues are unclear (Ozcelik et al., 2023). Unlike single-word mishearings, phrase-level

Mondegreens involve a restructuring of adjacent words based on perceived boundaries, often shaped by the listener's internalized expectations about likely word combinations. This process aligns with usage-based models of language, where frequently encountered lexical bundles and collocations strongly influence perceptual parsing (Ellis, 2016).

A representative case involves the phrase "our new" (from "Our new coffee table is made of copper"), which was frequently reinterpreted as "a new." In connected speech, "our new" is often pronounced with reduced articulation (/aʊə njuː/ or /ɑː njuː/), making it acoustically similar to "a new." For learners accustomed to syllable-timed rhythm and lacking familiarity with diphthongs or linking, the distinction between possessive pronoun and indefinite article becomes blurred. This aligns with RQ1, showing how stress alternation and vowel reduction in English contribute to misperception. From a cognitive standpoint (RQ2), "a new coffee table" is more frequent and semantically neutral, which increases its likelihood of selection under uncertainty. Such frequency-driven selection reflects probabilistic processing mechanisms, where listeners favor more common constructions when acoustic input is ambiguous (Gibson et al., 2019; Ellis, 2016).

Another striking example is the mishearing of "Phil and Tim at the scene of the accident" as "villain's attempts." This error reflects a complex case of rebracketing and phonological substitution. The original string contains proper nouns ("Phil and Tim") that may be unfamiliar to Indonesian learners, especially when spoken with natural coarticulation. As the boundary between "Phil and Tim" and "at the" is flattened, learners are left to reinterpret the auditory stream using available lexical schemas. The emergence of "villain's attempts," a semantically coherent phrase with thematic ties to accidents or conflict, suggests schema-driven reanalysis (Cutler, 2012), where top-down narrative expectations shape perceptual decoding. This phenomenon is also consistent with predictive processing accounts, in which listeners actively generate contextually plausible interpretations when bottom-up input is degraded (Pickering & Gambi, 2018; Huettig, 2015). This supports Scherling et al.'s (2022) observation that when proper names are unfamiliar, listeners substitute them with familiar collocations, particularly those associated with culturally salient scripts.

Similarly, the phrase "a nice T-shirt" was often transcribed as "night T-shirt" or "an iced tea, sir." These errors demonstrate the perceptual vulnerability of weak syllables in stress-timed languages, where unstressed vowels are frequently reduced or elided. The transition from "a nice" to "an ice" is phonetically plausible in rapid speech, especially when nasal assimilation and linking blur word boundaries (e.g., /ə naɪs/ → /ənaɪs/ → /ənaɪs/). Empirical studies on connected speech confirm that such reductions significantly impair L2 listeners' ability to identify word boundaries, particularly when function words are involved (Cauldwell, 2013; Field, 2019; Sutrisno, 2015). The interpretation "an iced tea, sir" reflects a fully plausible utterance in service settings, pointing again to pragmatic inference and cognitive schema activation (Castro & Vitevitch, 2023; Cutler, 2012). These reconstructions illustrate how learners apply top-down scenario templates to fill in gaps caused by unclear acoustic input. This aligns with findings that listeners rely heavily on situational schemas and discourse expectations when processing ambiguous speech (Van Engen & Peelle, 2014; Siegel, 2021).

Together, these phrase-level Mondegreens reveal a characteristic pattern of prosodic misalignment (RQ1) and expectation-driven repair (RQ2). Learners do not simply guess; they reconstruct what is most probable within their mental lexicon and pragmatic knowledge. This reconstruction process reflects an interaction between bottom-up acoustic cues and top-down probabilistic inference, as emphasized in contemporary models of speech comprehension (Pickering & Gambi, 2018; Gibson et al., 2019). These findings resonate with Field's (2004) argument that phrase-level processing is especially sensitive to prosodic cues and highly susceptible to disruption in L2 contexts. Moreover, for learners from syllable-timed language backgrounds like Indonesian, the stress-timed rhythm of English imposes a persistent decoding burden, particularly when weak function words precede strong lexical items. Recent pedagogical research further suggests that explicit training in connected speech features can help mitigate such segmentation difficulties in L2 listening (Kang et al., 2019; Siegel, 2021; Yenkimaleki et al., 2023).

4.3. Clause-level Mondegreens

Clause-level Mondegreens, which constituted 2.58% of total responses (51 out of 1980 utterances), represent a deeper layer of misperception where not just isolated words or phrases, but entire subject–predicate structures are misheard and reinterpreted. These errors demonstrate how EFL learners attempt to maintain syntactic coherence when confronted with degraded or ambiguous input. In such cases, misperception arises not only from phonological ambiguity (RQ1) but also from learners' effort to restore

clause-level meaning based on stored patterns and expectations (RQ2). This tendency is further reinforced by evidence that learners' metacognitive awareness and strategic processing play a crucial role in shaping how they reconstruct meaning from incomplete input (Maftoon & Alamdari, 2020; Rahimi & Abedi, 2014). This phenomenon aligns with contemporary models of predictive processing, which posit that listeners actively generate syntactic and semantic expectations during comprehension, especially under conditions of uncertainty (Pickering & Gambi, 2018; Huettig, 2015).

One prominent example emerged from the sentence "If Linda calls, tell her I am at the bank," which was frequently reconstructed as "Evelyn Dako is a teller at the bank." This mishearing involves both resegmentation and lexical substitution, transforming a conditional clause into a declarative clause with an identifiable subject and profession. The transformation from "If Linda calls" to "Evelyn Dako" illustrates the vulnerability of unstressed function words ("if," "her," "I am") in fast, connected speech, where they are easily elided or masked. Empirical research has shown that function words are particularly susceptible to perceptual loss in reduced speech, leading to major disruptions in syntactic parsing for L2 listeners (Erb & Obleser, 2019; Field, 2019). Learners, lacking clear bottom-up cues, rely on familiar proper name patterns as discussed by Scherling et al., (2022) and occupational schemas (e.g., "teller at the bank") to restore coherence. This supports Cutler's (2012) model of expectation-driven clause reconstruction, where listeners impose grammatical templates to fill gaps in auditory perception. More recent studies further suggest that such reconstruction is guided by probabilistic inference, where listeners select the most plausible syntactic structure given incomplete input (Gibson et al., 2019).

Similarly, in the sentence "A toddler walks into the world with curiosity," many learners reported hearing variants such as "I told her work…" or "Thought them walked…". These reanalyses reflect a loss of the low-frequency noun "toddler," which may not be firmly established in learners' mental lexicon, especially under time pressure. The substitution of "toddler" with "told her" or "thought them" reveals the semantic drift that occurs when input fails to activate a reliable lexical target. Beck et al. (2014) note that once an alternative interpretation is activated, even if incorrect, it can become stabilized through repetition or thematic fit, a process clearly visible here. This stabilization effect is consistent with findings in lexical activation research, where early candidate interpretations can persist and bias subsequent processing, even when they diverge from the intended input (Vitevitch & Lachs, 2024). Moreover, the insertion of pronouns ("her," "them") in place of the determiner+noun construction suggests a tendency to favor pronoun-subject clause templates, which are more commonly encountered in instructional contexts. This tendency also reflects usage-based learning, where frequent syntactic patterns become default templates in comprehension (Ellis, 2016).

A third case comes from the sentence "They're meant to help us," which was occasionally reinterpreted as "The man that helps us." Although this version preserves the helping intention, it shifts both grammatical subject and verb structure. Phonologically, the reduction of "they're meant to" (e.g., /ðɛə ˈmɛntə/) can obscure the contracted subject and auxiliary, making "the man" a more acoustically available and semantically plausible reconstruction. This aligns with the Ganong effect (Ganong, 1980) and lexical familiarity bias, where ambiguous auditory signals are resolved in favor of more concrete and frequent noun phrases. The phrase "that helps us" also reflects a familiar relative clause structure, commonly taught and reinforced in EFL classrooms. Thus, learners unconsciously apply known clause-building templates to fill in distorted input, precisely the type of top-down repair mechanism discussed by Castro and Vitevitch (2023). Such restructuring also aligns with evidence that listeners prefer syntactically complete and semantically interpretable structures when faced with ambiguous input, even if this involves substantial deviation from the original utterance (Pickering & Gambi, 2018).

Finally, the appearance of "If we are clever" as a possible reinterpretation of "If Linda calls…" illustrates how listeners gravitate toward semantically complete conditional clauses when the original conditional cue is phonetically weakened. The repetition of familiar if-clauses such as "If we are…" suggests syntactic projection based on internalized patterns, especially when the input signal offers only partial or masked cues. This reflects anticipatory processing mechanisms, in which listeners pre-activate likely syntactic frames based on prior linguistic experience (Huettig, 2015).

Across these examples, clause-level Mondegreens illustrate the intersection of ambiguous prosody and stored syntactic knowledge. These errors are not random but patterned: when prosodic boundaries are blurred due to connected speech, weak function words, or reduced syllables, learners reconstruct the clause using structures they find cognitively and pragmatically acceptable. Such findings align with research demonstrating that listening comprehension is influenced not only by linguistic input but also

by learners' self-efficacy and strategic engagement during processing (Du & Man, 2022). As shown in the Literature Review, this reflects the dual influence of phonological constraint (RQ1) and anticipatory processing (RQ2), the very mechanism at the heart of the Mondegreen effect in EFL listening. Taken together, these findings support a view of L2 listening as an active, predictive process in which listeners continuously integrate incomplete acoustic input with stored linguistic knowledge to construct coherent interpretations (Pickering & Gambi, 2018; Gibson et al., 2019).

4.4. Sentence-level Mondegreens

Sentence-level Mondegreens, which constituted 19.60% of total responses (51 out of 1980 utterances), represent a deeper layer of misperception where not just isolated words or phrases, but entire subject–predicate structures are misheard and reinterpreted. These errors demonstrate how EFL learners attempt to maintain syntactic coherence when confronted with degraded or ambiguous input. In such cases, misperception arises not only from phonological ambiguity (RQ1) but also from learners' effort to restore clause-level meaning based on stored patterns and expectations (RQ2). This tendency is further reinforced by evidence that learners' metacognitive awareness and strategic processing play a crucial role in shaping how they reconstruct meaning from incomplete input (Maftoon & Alamdari, 2020; Rahimi & Abedi, 2014). This phenomenon aligns with contemporary models of predictive processing, which posit that listeners actively generate syntactic and semantic expectations during comprehension, especially under conditions of uncertainty (Pickering & Gambi, 2018; Huettig, 2015).

Sentence-level Mondegreens represented the most frequent type of misperception in this study, accounting for 388 out of 1980 utterances or 19.60% of all responses. These instances are particularly significant because they involve a global reanalysis of entire utterances, not just lexical or syntactic fragments. In such cases, learners construct sentences that are grammatically and semantically coherent but entirely divergent from the original stimulus. These reconstructions offer compelling evidence for the dual role of phonological ambiguity (RQ1) and cognitive expectation (RQ2) in shaping real-time EFL listening comprehension. This type of large-scale reinterpretation is consistent with contemporary models of language processing that view comprehension as a probabilistic and reconstructive process, rather than a purely bottom-up decoding mechanism (Gibson et al., 2019; Pickering & Gambi, 2018). Recent studies further suggest that comprehension under degraded or complex input conditions relies heavily on predictive processing and contextual inference, particularly when listeners are exposed to variability in speech signals (Fujita, 2022).

One notable stimulus was "I saw Phil and Tim at the scene of the accident," which elicited the highest number of sentence-level Mondegreens (90 instances). However, the misperceptions of this utterance were highly varied, reflecting learners' different attempts to resolve the phonological ambiguity using top-down processing. Among the reconstructed versions, one illustrative example was "I so feel intimate the scene of the accident." This reinterpretation reflects heavy resegmentation, where the coarticulated proper nouns "Phil and Tim" are reanalysed as the more frequent emotional phrase "feel intimate." Meanwhile, the rest of the sentence remains partly intact, suggesting that learners retained fragments of the original while filling gaps with familiar structures. This case exemplifies how listeners apply pragmatic plausibility and emotional salience when reconstructing entire utterances under uncertainty. Such behaviour aligns with findings that listeners prioritize semantically coherent and affectively meaningful interpretations when acoustic input is degraded, even at the expense of fidelity to the original signal (Van Engen & Peelle, 2014; Huettig, 2015).

Another highly misperceived sentence was "I knew a guy once who worked in finance," which resulted in numerous variations. One recurrent form was "A new guy who wants to work in finance." This shift illustrates multiple reconfigurations: "knew" becomes "new," "once who worked" becomes "wants to work," and the temporal frame shifts from past to present. These changes reflect the impact of vowel reduction and prosodic smoothing, as well as the dominance of canonical sentence schemas in learners' minds. Such reinterpretations indicate that learners default to syntactic constructions that are more accessible and frequently encountered in input materials, an example of top-down schema imposition described by Castro and Vitevitch (2023). This tendency is also consistent with usage-based accounts of language processing, where high-frequency constructions serve as default templates in both production and comprehension (Ellis, 2016). In addition, recent research highlights that learners' ability to manage such variability depends on their capacity to integrate decoding, inferencing, and strategic processing during real-time listening (Liu et al., 2025).

A third example comes from the stimulus "Our new coffee table is made of copper," which was occasionally reconstructed as "I knew a comfortable table." Unlike unintelligible responses, this version remains semantically plausible while significantly diverging from the original input. The phrase "our new" is reinterpreted as "I knew," and "coffee table" becomes "comfortable table," reflecting both phonological similarity and lexical familiarity. This reconstruction illustrates how listeners actively reorganize ambiguous input into meaningful structures, rather than producing random or fragmented output. It supports the argument that when phonological and prosodic cues are weakened, learners rely more heavily on top-down cognitive processes, drawing on familiar lexical patterns and syntactic templates to restore coherence. Recent research in noisy-channel models further supports this view, suggesting that listeners integrate uncertain input with prior expectations to infer the most probable meaning (Gibson et al., 2019; Crisol, 2024). Moreover, learner-related factors such as vocabulary knowledge and processing efficiency may further influence how successfully such reconstructions are carried out in real time (Du & Man, 2022; Wang & Treffers-Daller, 2017).

Together, these sentence-level Mondegreens provide clear evidence for the study's central claim: that auditory misperceptions in EFL listening are not random but arise from a dynamic interplay of ambiguous bottom-up input and top-down inferencing. Learners reconstruct entire utterances using mental templates, pragmatic expectations, and lexical familiarity, especially when the original input is masked by stress-timed rhythm, coarticulation, and connected speech. These findings align with psycholinguistic theories of comprehension under noise (Cutler, 2012; Beck et al., 2014), as well as with Roach's (2009) work on the perceptual challenges imposed by English prosody for L2 learners. They also resonate with contemporary perspectives that conceptualize listening as an active, predictive, and inference-driven process, in which meaning is continuously negotiated between incoming acoustic signals and stored linguistic knowledge (Pickering & Gambi, 2018; Huettig, 2015). Sentence-level misperceptions are therefore not merely errors; they are evidence of learners' adaptive strategies to manage real-time meaning construction under conditions of auditory uncertainty. In this sense, Mondegreens can be understood as windows into the cognitive mechanisms that underlie L2 listening, revealing how perception, prediction, and interpretation operate together as an integrated system in real-time language use.

4.5. Cross-level Synthesis and Pedagogical Implications

Across all levels of analysis: word, phrase, clause, and sentence, a consistent pattern emerges: the less robust the phonological and prosodic signal, the more learners rely on top-down cognitive processes to construct meaning. This interplay between ambiguous bottom-up input and expectation-driven inference lies at the heart of the Mondegreen effect in EFL listening. The data affirm that misperceptions are not isolated or erratic; rather, they reflect systematic listener responses to prosodic uncertainty, lexical unfamiliarity, and segmental overlap, validating the central claim posed in the article's title. This pattern is consistent with contemporary accounts of speech perception that conceptualize comprehension as a probabilistic and inference-driven process, where listeners continuously integrate uncertain input with prior knowledge to arrive at the most plausible interpretation (Gibson et al., 2019; Pickering & Gambi, 2018). Recent research further emphasizes that this integration process is mediated by learner-specific factors such as vocabulary knowledge, metacognitive awareness, and self-efficacy, which jointly influence listening success (Du & Man, 2022; Wang & Treffers-Daller, 2017).

In relation to RQ1, the findings confirm that several phonological and prosodic features—such as stress shifts, vowel reduction, assimilation, linking, and coarticulation—contribute significantly to misperception. These features blur word boundaries, reduce the perceptibility of function words, and heighten the difficulty of lexical segmentation, especially for learners whose first language lacks stress-timed rhythm (Roach, 2009; Scherling et al., 2022). As Cauldwell (2013) emphasizes, the "sound substance" of spontaneous speech often departs drastically from dictionary forms, leading to a "sight–sound mismatch" that disrupts decoding for learners unaccustomed to reduced forms and connected speech. Moreover, Vitevitch and Luce (1999) explain how words that reside in dense phonological neighbourhoods are more prone to competition and misrecognition, further compounding learners' difficulty in isolating the correct lexical target during rapid speech input. More recent research further demonstrates that such segmentation difficulties are amplified in naturalistic listening conditions, where acoustic reduction and variability increase perceptual load (Erb & Obleser, 2019; Field, 2019).

Concerning RQ2, the study illustrates how learners activate stored syntactic templates, semantic frames, and discourse schemas to reconstruct plausible meanings when bottom-up cues are insufficient.

This aligns with psycholinguistic theories emphasizing anticipatory processing, lexical probability, and pragmatic fit (Vitevitch & Lachs, 2024). Field (2004) likewise emphasizes that L2 listeners engage in "on-line reconstruction," whereby they test interpretive hypotheses against incoming signals to achieve coherence. Our data show that such top-down inferencing is not merely reactive but constitutes an active listening strategy, especially in sentence-level misperceptions, where learners restructure the utterance to fit familiar grammatical templates. Goh (2000) adds that L2 listeners frequently struggle with lexical segmentation, storing partial word forms and filling in perceived gaps with mental approximations shaped by context and expectation. This behavior aligns with predictive processing models, which propose that listeners continuously generate and update expectations about upcoming linguistic input during comprehension (Pickering & Gambi, 2018; Huettig, 2015).

Vandergrift (2007) notes that effective L2 listening involves a recursive process of prediction, monitoring, and revision, precisely the pattern evidenced in this study's Mondegreen responses. The data also align with Chang and Read's (2006) observation that when learners are not supported by clear contextual cues, they resort more heavily to schematic knowledge and inferencing. These tendencies were especially visible in the clause and sentence-level misperceptions, where learners projected familiar event structures (e.g., workplace scenarios, emotional expressions, transactional settings) to resolve ambiguity. Recent studies further confirm that such reliance on schema-based processing becomes more pronounced under conditions of increased perceptual difficulty, highlighting the adaptive nature of top-down strategies in L2 listening (Siegel, 2021; Liu et al., 2025).

Beyond the pedagogical domain, these findings also resonate with major models of L2 speech perception. Flege's Speech Learning Model (1995) predicts that L2 learners assimilate new sounds into their closest L1 categories, a mechanism visible in the systematic substitutions and resegmentations observed here. Best's Perceptual Assimilation Model (Flege,1995; Best & Tyler, 2007) similarly accounts for why ambiguous or reduced contrasts are perceived as familiar L1-like categories, reinforcing the patterned nature of these mishearing. Moreover, the high incidence of sentence-level Mondegreens aligns with Munro and Derwing's (1995, 2011) research on intelligibility, showing that misperceptions reflect reduced comprehensibility rather than random error. By situating Mondegreens within these theoretical frameworks, this study extends pronunciation research by demonstrating how perceptual models and intelligibility constructs play out in real-time L2 listening. Importantly, these findings also connect with emerging perspectives that view L2 listening as an adaptive system shaped by continuous interaction between perceptual constraints and cognitive expectations (Van Engen & Peelle, 2014).

These insights carry important pedagogical implications. First, they call for a reorientation in listening instruction, away from scripted, slow-paced input toward exposure to authentic, prosodically rich spoken English. Learners must be trained not only in lexical recognition but also in prosodic sensitivity—the ability to detect and interpret stress patterns, reductions, and rhythm. As Field (2009) and Cauldwell (2013) advocate, instruction must move beyond "what is said" to "how it is said," using spontaneous speech models and focused practice on decoding acoustic features of connected speech. Recent pedagogical research supports this shift, demonstrating that explicit instruction in connected speech features and prosodic awareness significantly improves L2 listening performance (Kang et al., 2019; Siegel, 2021). This is further supported by studies showing that both segmental and suprasegmental training contribute to measurable gains in listening comprehension among EFL learners (Yenkimaleki et al., 2023).

Second, teachers should integrate ambiguity tolerance and strategic inferencing into the curriculum. Rather than treating misperceptions as mere comprehension failures, educators can use Mondegreens as diagnostic tools to uncover where and how learners reconstruct meaning. Classroom activities that involve identifying, reflecting on, and correcting misheard phrases can raise learners' metacognitive awareness of their own listening strategies. As Vandergrift (2007) suggests, such reflective strategies enhance learners' ability to monitor and evaluate their comprehension in real time. This approach aligns with recent calls for strategy-based listening instruction that emphasizes metacognitive awareness and adaptive processing in L2 learners (Siegel, 2021). Moreover, process-based metacognitive instruction has been shown to significantly enhance learners' listening awareness and performance, particularly in managing ambiguity and uncertainty during comprehension (Maftoon & Alamdari, 2020; Milliner & Dimoski, 2021).

Finally, this study invites educators and researchers alike to treat listening not as a passive decoding process but as a dynamic, interpretive act—one in which language users negotiate meaning through

constant interaction between what is heard and what is expected. The Mondegreen effect, rather than a humorous footnote in language learning, should be recognized as a powerful indicator of how learners navigate the complexity of real-time spoken input in a second language. In this sense, Mondegreens offer valuable insights into the cognitive architecture of L2 listening, revealing how perception, prediction, and interpretation operate together as an integrated system in real-time language use.

5. Conclusion

This study examined how phonological ambiguity and cognitive expectation interact to produce auditory misperceptions, specifically the Mondegreen effect, among Indonesian EFL learners. Based on transcription data from 165 participants and 12 spoken English utterances, the study identified misperceptions across four linguistic levels: word, phrase, clause, and sentence.

Findings show that EFL listening comprehension is shaped not merely by perceptual decoding, but by the dynamic interplay between ambiguous bottom-up input and top-down inferential processing. In response to RQ1, several phonological and prosodic features, such as stress shifts, vowel reduction, assimilation, and connected speech, were found to disrupt accurate segmentation and lead to lexical misinterpretation. Learners, particularly those from syllable-timed language backgrounds, struggled to perceive weak grammatical elements, making function words especially prone to mishearing.

Addressing RQ2, the study demonstrated that learners rely heavily on cognitive expectations, including familiar syntactic templates, lexical associations, and pragmatic schemas, to reconstruct plausible meanings. This was most evident in sentence-level Mondegreens, where learners generated grammatically coherent but unintended interpretations based on probabilistic reasoning rather than faithful perception.

These findings affirm that auditory misperceptions are not mere signs of inattention or low proficiency. Rather, they reveal the learner's active role in meaning-making under uncertain input conditions. Far from being random, these misperceptions are strategic reconstructions informed by linguistic experience and contextual reasoning.

Theoretically, this study contributes a nuanced, level-based account of misperception in EFL listening. By situating Mondegreens within established frameworks such as Flege's Speech Learning Model, Best's Perceptual Assimilation Model, and Munro & Derwing's work on intelligibility, the study demonstrates that mishearing represent predictable outcomes of cross-linguistic perception and provide a window into the mechanisms of L2 pronunciation processing. Pedagogically, it underscores the importance of training learners not only in decoding but in navigating prosodic variation, managing ambiguity, and deploying inference strategies. Incorporating Mondegreens into classroom practice may foster greater metacognitive awareness and better prepare learners for real-world listening.

Ultimately, the Mondegreen effect offers a compelling lens through which to understand both the vulnerability and adaptability of second language listeners. It highlights not only how learners negotiate meaning amid the inherent fluidity of spoken language, but also how their perceptual strategies extend, and in some cases challenge, current models of L2 speech perception and pronunciation research.

Acknowledgment

This research was financially supported by the Department of Language and Literature, Faculty of Cultural Sciences, Universitas Gadjah Mada, under Research Contract No. 2226/UN1/FIB.1.3/PT/PT.01.03/2025. The authors gratefully acknowledge this institutional support, which enabled the successful completion of this study. The authors also acknowledge the limited use of AI-assisted language tools, specifically ChatGPT, solely for minor editorial refinement and language polishing. All conceptualization, research design, data analysis, and interpretation of findings were carried out independently by the authors, in full adherence to academic integrity and scholarly standards.

Declarations

Author contribution : The authors contributed equally to all stages of the research process, including the conceptualization of the study, design development,

data collection, analysis and interpretation of findings, and the drafting and revision of the manuscript.

Funding statement : This study was supported by institutional research funding from the Department of Language and Literature, Faculty of Cultural Sciences, Universitas Gadjah Mada, under Research Contract No. 2226/UN1/FIB.1.3/PT/PT.01.03/2025.

Conflict of interest : The author(s) declare that there are no competing interests related to this study.

Ethical declaration : This study involved non-invasive procedures and did not address sensitive or potentially harmful issues; therefore, formal ethical approval from an institutional review board was not required. All participants were informed about the purpose of the study and participated voluntarily, with informed consent obtained prior to data collection. Participants' anonymity and confidentiality were strictly maintained throughout the research process. The study was conducted in accordance with established ethical principles for research involving human participants and adhered to relevant institutional academic guidelines.

We support ELTEJ in maintaining high standards of personal conduct, practicing honesty in all our professional practices and endeavors.

Additional information : No additional information is available for this paper.

REFERENCES

Aryadoust, V., & Luo, L. (2023). The typology of second language listening constructs: A systematic review. *Language Testing, 40*(2), 375–409. https://doi.org/10.1177/02655322221126604

Beck, J. M., Miyamoto, R. T., & Heinz, M. G. (2014). Perceptual inference in the brain: Neural correlates of listening under uncertainty. *Neuron, 83*(1), 14–28. https://doi.org/10.1016/j.neuron.2014.05.026

Best, C. T., Tyler, M., Bohn, O., & Munro, M. (2007). Nonnative and second-language speech perception. *Language experience in second language speech learning*, *17*, 13-34.

Castro, N., & Vitevitch, M. S. (2023). The influence of lexical and phonological competition on spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001212

Cauldwell, R. (2013). Phonology for listening. *Birmingham: Speech in Action*.

Chang, A. C.-S., & Millett, S. (2016). The effects of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal, 68*, 31–40. https://doi.org/10.1093/elt/cct052

Chang, A. C.-S., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly, 40*(2), 375–397. https://doi.org/10.2307/40264527

Crisol, L. G. D. (2024). A semantic analysis of cross-linguistic mondegreens: Implications on how Filipinos interpret meanings. *Southeastern Philippines Journal of Research and Development, 29*(1), 19–41. https://doi.org/10.53899/spjrd.v29i1.285

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research*, *28*(1), 5-40.

Du, G., & Man, D. (2022). Person factors and strategic processing in L2 listening comprehension: Examining the role of vocabulary size, metacognitive knowledge, self-efficacy, and strategy use. *System*, *107*, https://doi.org/10.1016/j.system.2022.102801

Ellis, N. C. (2016). Usage-based approaches to language acquisition. *Language Learning, 66*(S1), 3–28. https://doi.org/10.1111/lang.12177

Erb, J., & Obleser, J. (2019). Upregulation of cognitive control networks in challenging listening conditions. *Cerebral Cortex, 29*(5), 2054–2065. https://doi.org/10.1093/cercor/bhy090

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System, 32*(3), 363–377. https://doi.org/10.1016/j.system.2004.05.002

Field, J. (2009). *Listening in the language classroom*. Cambridge University Press.

Field, J. (2019). *Rethinking the second language listening test*. Cambridge University Press.

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.

Fujita, R. (2022). The role of speech-in-noise in Japanese EFL learners' listening comprehension. *International Journal of Listening, 36*(2), 118–137. https://doi.org/10.1080/10904018.2021.1963252

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gibson, E., Bergen, L., & Piantadosi, S. T. (2019). Rational integration of noisy evidence and prior semantic expectations. *Proceedings of the National Academy of Sciences, 116*(16), 8051–8056. https://doi.org/10.1073/pnas.1815715116

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55–75. https://doi.org/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*, 103164. https://doi.org/10.1016/j.system.2023.103164

Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting. *Language teaching*, *50*(1), 107-119.

Hartono, D. (2022). Fenomena kesadaran bela negara di era digital dalam perspektif ketahanan nasional. *Jurnal Lemhannas RI, 8*(1), 14–33. https://doi.org/10.55960/jlri.v8i1.301

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Kang, O., Rubin, D., & Pickering, L. (2019). Suprasegmental measures of accentedness. *TESOL Quarterly, 53*(2)

Ke, Z., & Wang, Y. (2022). Exploring the relationship between aural decoding and listening comprehension. *System, 104*, 102688. https://doi.org/10.1016/j.system.2021.102688

Leonard, K. R. (2019). Decoding and comprehension in L2 listening. *System, 87*, https://doi.org/10.1016/j.system.2019.102149

Liu, L., Darmi, R. H., & Wan Mohtar, W. I. (2025). Improvement of listening performance among EFL learners: A systematic review. *World Journal of English Language, 15*(7), 278–289. https://doi.org/10.5430/wjel.v15n7p278

Maftoon, P., & Alamdari, E. F. (2020). Metacognitive strategy instruction and listening performance. *International Journal of Listening, 34*(1), 1–20. https://doi.org/10.1080/10904018.2018.1462739

Matthews, J. (2018). Vocabulary for listening. *System, 72*, 23–36. https://doi.org/10.1016/j.system.2017.11.005

Matthews, J., & Cheng, J. (2015). Recognition of high-frequency words in speech. *System, 52*, 1–13. https://doi.org/10.1016/j.system.2015.03.004

Milliner, B., & Dimoski, B. (2021). Metacognitive intervention and listening self-efficacy. *Language Teaching Research*. https://doi.org/10.1177/13621688211004604

Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility. *Language Teaching, 44*(3), 316–327. https://doi.org/10.1017/S0261444811000103

Newton, J. M., & Nation, I. S. P. (2020). *Teaching ESL/EFL listening and speaking*. Routledge.

Ozcelik, H. N., Van den Branden, K., & Van Steendam, E. (2023). Listening comprehension problems of FL learners in a peer interactive, self-regulated listening task. *International Journal of Listening*, *37*(2), 142-155.

Pickering, M. J., & Gambi, C. (2018). Predictive processing in language comprehension. *Psychological Bulletin, 144*(10), 1002–1044. https://doi.org/10.1037/bul0000153

Rahimi, M., & Abedi, S. (2014). The relationship between listening self-efficacy and metacognitive awareness of listening strategies. *Procedia-Social and Behavioral Sciences*, *98*, 1454-1460.

Reinisch, E., & Llompart, M. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(8), 1590–1610. https://doi.org/10.1037/xlm0000832

Roach, P. (2009). *English phonetics and phonology* (4th ed.). Cambridge University Press.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning, 69*(3), 652–708. https://doi.org/10.1111/lang.12315

Scharenborg, O., & Janse, E. (2019). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01675-x

Scherling, J., Kornder, L., & Kelly, N. (2022). Perception and reinterpretation of English song lyrics by native speakers of Japanese: A case study of samples from the TV-show Soramimi-Hour. *Frontiers in Communication, 7*, 780279. https://doi.org/10.3389/fcomm.2022.780279

Siegel, J. (2018). *Second language listening instruction: Principles and practice*. Cambridge University Press.

Siegel, J. (2021). *Exploring L2 listening instruction*. Routledge.

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences, 113*(12), E1747–E1756. https://doi.org/10.1073/pnas.1523266113

Sutrisno, A. (2015). *Kesulitan mempersesi bunyi ujaran bahasa Inggris sebagai bahasa asing di Indonesia* [Unpublished doctoral dissertation]. Universitas Gadjah Mada.

Sutrisno, A. (2018). Problems of speech perception experienced by the EFL learners. *Theory and Practice in Language Studies, 8*(1), 143–149. https://doi.org/10.17507/tpls.0801.18

Tucker, B. V., & Warner, N. (2020). What it means to be phonologically reduced. *Laboratory Phonology*. https://doi.org/10.5334/labphon.213

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and speech comprehension. *Frontiers in Human Neuroscience, 8*, 577. https://doi.org/10.3389/fnhum.2014.00577

Vandergrift, L. (2007). Recent developments in second language listening comprehension research. *Language Teaching, 40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Vitevitch, M. S., & Lachs, L. (2024). Using network science to examine audio-visual speech perception with a multi-layer graph. *PLOS ONE, 19*(3). https://doi.org/10.1371/journal.pone.0300926

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation. *Journal of Memory and Language, 40*(3), 374–408. https://doi.org/10.1006/jmla.1998.2618

Wang, Y., & Treffers-Daller, J. (2017). Listening comprehension and vocabulary knowledge. *System, 65*, 139–150. https://doi.org/10.1016/j.system.2017.01.007

West, D. (2023). Strangers in the night, exchanging glasses. *English Text Construction, 16*(2), 197–213. https://doi.org/10.1075/etc.00060.wes

Wright, S. (1954). The Death of Lady Mondegreen. *Harper's Magazine*, *209*(1254), 48-51.

Yeldham, M. (2018). L2 listening instruction: More bottom-up or more top-down? *The Journal of Asia TEFL, 15*(3), 805–810.

Yenkimaleki, M., van Heuven, V. J., & Afshar, H. S. (2023). Pronunciation instruction and listening comprehension. *Language Learning Journal, 51*(6), 734–748. https://doi.org/10.1080/09571736.2021.1954683

Yasoda-Mohan, A., Chen, F., & Vanneste, S. (2025). Unveiling the mind's ear: Understanding auditory processing using illusions. *Hearing Research, 459*, 109227. https://doi.org/10.1016/j.heares.2025.109227

Zhang, X., & Chen, Y. (2023). Chinese EFL learners' perception of English prosodic focus. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2023)* (pp. 92–96). ISCA. https://doi.org/10.21437/Interspeech.2023-1781