Studies in Second Language Acquisition (2025), 1–22
doi:10.1017/S0272263125101447

METHODS FORUM

Optimizing L2 phonetic learning

Spaced vs. massed training schedule

Kazuya Saito[1,2] and Siying Chen[1]

[1]University College London, London, United Kingdom and [2]Tohoku University, Sendai, Japan
Corresponding author: Kazuya Saito; Email: k.saito@ucl.ac.uk

(Received 12 June 2025; Revised 24 October 2025; Accepted 30 October 2025)

Abstract

This methodological study investigated how the distribution of training sessions—massed, equal spacing, and expanding spacing—affects L2 phonetic learning, focusing on Mandarin-speaking learners' perception of the English /ɛ/–/æ/ contrast. While most previous phonetic training studies have used massed schedules, the current quasi-experimental design revealed that both types of spaced practice significantly outperformed massed practice in terms of immediate gains and long-term retention. Effect sizes in the spaced groups were approximately double those of the massed group. No significant differences emerged between equal and expanding spacing. These findings suggest that distributed practice—regardless of spacing type—can enhance both the magnitude and durability of L2 phonetic learning. Crucially, this study makes it possible to revisit past findings based on massed training paradigms and to consider whether adopting alternative timing schedules could unlock greater learning potential—for instance, by doubling the size and durability of training effects through the use of spaced conditions.

Keywords: distributed practice; L2 phonetic learning; methodological renovation; perceptual training; spacing effect

Attaining accurate L2 phonetic perception is fundamental to successful communication, and explicit training has been shown to improve learners' phonetic competence. However, most existing studies have relied on massed practice schedules, where training sessions are delivered in quick succession with minimal intervals. Recent work in cognitive psychology and L2 vocabulary learning has highlighted the benefits of distributed practice, where learning sessions are spaced over time. Despite its potential importance, this issue has rarely been investigated in L2 phonetic training. Addressing this gap, the present study examines how different training schedules—massed, equal spacing, and expanding spacing—affect Mandarin learners' acquisition of the English /ɛ/–/æ/ contrast.

L2 phonetic training research

Over the past 50 years, much attention has been given to the effects of training in L2 phonetic learning. A range of empirical studies has demonstrated that provision of explicit phonetic training promotes L2 phonetic perception on both segmental (e.g., Bradlow, 2008 for Japanese speakers' English [r] and [l] acquisition) and suprasegmental levels (e.g., Wang et al., 2003 for Americans' acquisition of Mandarin lexical tones). Such training effectiveness can be amplified through the use of multiple talkers (e.g., High Variability Phonetic Training; Uchihara et al., 2025), and key features are acoustically enhanced (e.g., F3 enhancement in English [r] and [l] contrast; Iverson et al., 2005).

Recent meta-analyses offer further support for these findings, revealing that phonetic training has a medium to large positive effect on L2 phonetic competence (e.g., Saito & Plonsky, 2019; Sakai & Moorman, 2018; Yao et al., 2025). In a recent comprehensive meta-analysis, Yao et al. (2025) synthesized findings from 65 studies involving 2,793 learners and reported a large average effect size of $d = 0.762$ for phonetic training. More specifically, analyzing 25 years of research, Sakai and Moorman's (2018) meta-analysis found that perceptual training led to medium-sized improvements in perception ($d = 0.92$) and small but significant improvements in production ($d = 0.54$). This suggests that perceptual training can indeed have a positive effect on production abilities in L2 phonetic learning (Flege & Bohn, 2021). Finally, Saito and Plonsky (2019) conducted a meta-analysis of 77 studies and found that L2 pronunciation instruction is generally effective with a moderate overall effect size ($d = 0.61$). Greater improvements were observed when targeting specific features and using objective, controlled assessments rather than global, subjective evaluations.

In a typical phonetic training study, participants take a pre-test, complete a series of short training sessions (typically 10–30 minutes), and then undergo immediate and delayed post-tests. When it comes to the *perception* domain aspects of L2 phonetic training (i.e., the main focus of the study), recent review articles demonstrated that a few hours of training typically result in approximately 5–10% of gains (e.g., Saito et al., 2022). One methodological issue that has become standard practice—without substantial scholarly scrutiny—is the distribution schedule of these training sessions. Most existing studies have employed massed or evenly spaced training input (e.g., using a one-day inter-session interval). More broadly, L2 phonetic learning research has long emphasized the importance of input. There is ample evidence that both the quantity of input (e.g., length of immersion; Trofimovich & Baker, 2006) and the quality of input (e.g., the frequency and intensity of L2 interaction; Derwing & Munro, 2013) shape the rate and ultimate success of speech development. However, the question of how best to schedule such input—particularly in short-term instructional contexts—remains largely underexplored.

Massed practice is a learning strategy in which study or skill rehearsal is concentrated into a single, continuous, and intensive period with little to no rest between repetitions. Often referred to as "cramming," it contrasts with distributed practice, which spaces learning sessions over time (for a meta-analysis of verbal recall in cognitive psychology, see Cepeda et al., 2006). In research in L2 learning outside phonetics, it has been shown that training outcomes can vary—and sometimes improve—depending on the distribution

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

schedule (e.g., massed vs. equal vs. expanding spacing; Suzuki, Nakata, & DeKeyser, 2019). Despite its potential importance, this issue has received little attention in L2 phonetics, with Alfotais et al. (2025) being a rare exception (see below).

In the case of L2 phonetic training, previous studies have employed a wide range of schedules, varying in duration, number of sessions, and overall intensity. As shown in Uchihara et al.'s (2025) a methodological synthesis, programs have ranged from highly condensed formats delivered within a few days to more extended schedules lasting several weeks, with total training time varying from as little as 60 minutes (i.e., massed learning) to more than 20 hours (i.e., spaced learning; see Table 2 on p. 810). However, these spacing decisions were typically made on pragmatic rather than theoretical grounds, and no consistent pattern has been established linking timing/spacing parameters to learning outcomes. While perception gains tend to increase with overall training time up to a certain threshold (approximately six hours), neither the number of sessions nor their temporal distribution systematically predicted outcomes. Thus, although extant L2 phonetic training research provides abundant evidence for the efficacy of L2 phonetic training, the specific role of training distribution has not yet been directly addressed. The current study was designed to correspond to these concerns.

Distributed practice in L2 vocabulary research

Turning to L2 vocabulary literature, researchers have begun to investigate how the *distribution* of input—not just its amount or quality—affects learning outcomes (Suzuki et al., 2019). This line of inquiry draws from the well-established *spacing effect* in cognitive psychology, which posits that learning is more robust when practice is spaced out over time rather than massed into a single session (Baddeley, 1997; Cepeda et al., 2006). One influential origin of this research is Ebbinghaus's (1913) discovery that forgetting follows a predictable curve, but can be mitigated by repeated retrieval spaced over time.

In language learning, *spaced practice* has consistently been shown to yield better retention outcomes than *massed practice* (e.g., Bloom & Shuell, 1981; Hamouda, 2021; Namaziandost et al., 2020). Studies show that spaced learning (either equal or expanding) can lead to 20–40% higher retention compared to massed practice, and this effect is especially clearly observed in long-term retention (e.g., after a week or more).

Within the broader category of spaced practice, scholars have further distinguished between *absolute spacing* (i.e., total time between repetitions) and *relative spacing* (i.e., the pattern of intervals between repetitions), both of which have been shown to influence learning. Relative spacing has received particular attention in L2 vocabulary acquisition. It includes *equal spacing*—where intervals between study sessions are fixed —and *expanding spacing*—where intervals gradually increase across repetitions. Expanding spacing is often argued to be more effective because it increases retrieval difficulty over time, which can promote stronger memory consolidation (Karpicke & Bauernschmidt, 2011; Landauer & Bjork, 1978; Pyc & Rawson, 2009).

However, empirical evidence comparing equal and expanding spacing has yielded mixed results. Nakata (2015), in a carefully controlled study of 128 Japanese learners, reported that expanding spacing resulted in significantly higher receptive vocabulary scores than equal spacing, across short, medium, and long absolute spacing schedules. Similarly, Schuetze (2015) found advantages for expanding spacing in short-term retention, though equal spacing outperformed in the long-term. Karpicke and Roediger (2007)

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

reported comparable findings, suggesting that the benefits of expanding spacing may be more pronounced in the early stages of learning.

Conversely, other studies have failed to find consistent advantages. Kang et al. (2014), for example, compared expanding and equal spacing schedules in a four-week vocabulary learning study and found no significant differences between the two groups on delayed post-tests. This echoes earlier work by Pyc and Rawson (2007), who also found no long-term advantage of expanding over equal spacing in their study of Swahili-English word pair retention. Overall, while the spacing effect is well established, the specific efficacy of *expanding* vs. *equal* spacing remains unsettled.

It is important to note that much of the existing research has focused on vocabulary learning. Moreover, outcomes may depend on moderating factors such as the total number of exposures, the timing of retrieval, the nature of the linguistic targets, and whether feedback is provided during practice. As stressed earlier, surprisingly little is known about the optimal timing of input in the context of L2 *phonetic* learning and training.

The mechanisms underlying L2 vocabulary and phonetic learning may differ substantially. On the one hand, vocabulary learning (as tested in the existing literature) involves the conscious cognitive operation of higher-order information, such as facts, conceptual knowledge, vocabulary, and morphosyntactic rules. In this case, learning typically refers to the explicit acquisition of form-meaning mappings using declarative memory. In such contexts, introducing different types of input variability through distributed practice (massed vs. spaced; equal vs. expanding spacing) can strengthen memory traces and retrieval (Landauer & Bjork, 1978). On the other hand, phonetic learning takes place on relatively lower-order and perceptual levels. Such learning processes entail the readjustment of abstract, acoustic, and articulatory details of new sounds relative to L1 phonetic categories. Here, learning relies on procedural rather than declarative memory, and the process tends to be implicit, subconscious, and gradual (Best & Tyler, 2007).

Given these fundamentally different learning mechanisms, researchers in L2 acquisition and teaching have increasingly shown interest in testing the generalizability (or specificity) of findings across different domains of L2 learning (e.g., extending task-based language teaching from vocabulary and grammar to phonology; Mora-Plaza et al., 2024; Xu et al., 2024). Following this line of thought, we argue that it is important to explore whether, to what degree, and how insights on distributed practice (derived mainly from L2 vocabulary and grammar research) can be generalized to L2 phonetic learning. At the very least, it is possible that the effects of distributed practice differ across vocabulary/grammar learning and phonetic training, and such discrepancies should be examined to further our understanding of the role of distributed practice in language learning more broadly.

Motivation for current study

To our knowledge, Alfotais et al.'s (2025) recent empirical study represents the very first systematic attempt to compare the effects of massed and spaced repetition schedules. In their quasi-experimental investigation, 49 Saudi EFL learners were trained on the acquisition of the English /b/–/p/ contrast. Using a within-subjects design, the study assessed L2 phonetic development both immediately after training and following a four-week delay. The results revealed that spaced repetition consistently outperformed

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

massed repetition for both short- and long-term retention, enhancing effectiveness by approximately 55–80% compared with massed repetition—a pattern strikingly similar to the figures reported in L2 vocabulary research (Nakata, 2015).

While Alfotais et al. (2025) represent an important first step in applying distributed practice frameworks to L2 phonetic learning, their work raises two important questions that merit further investigation. First, can the benefits of spaced over massed repetition be replicated in a different L2 phonetic learning context? Second, might the effects of spaced repetition be further enhanced by considering *types* of spacing—specifically, by distinguishing between equal and expanding spacing formats, two theoretically significant subtypes of distributed practice?

The present study built on this foundation by not only replicating the massed vs. spaced comparison but also extending it to examine the relative efficacy of equal vs. expanding spacing schedules. Given that most existing phonetic training studies have employed massed practice, our ultimate aim was to provide principled methodological guidance on the extent to which the effectiveness of the same phonetic training can change or increase when researchers adopt different timing regimes—massed, equal, or expanding. These findings will ultimately inform key theoretical questions concerning the temporal distribution of phonetic input in L2 speech training and how differently distributed input may affect L2 phonetic learning.

In the context of Mandarin learners' acquisition of the English vowels /ɛ/ and /æ/, the current study was designed to address the following three research questions, each accompanied by a corresponding prediction:

1. Does a spaced practice schedule facilitate L2 phonetic learning more effectively than a massed practice schedule? If so, how much?
2. How do different types of spaced practice schedules (equal vs. expanding) influence L2 phonetic learning? If so, how much?
3. Do the relative benefits of input distribution (massed, equal, expanding) vary as a function of learners' perceptual–cognitive abilities?

Prediction 1: Massed vs. spaced practice schedules

Following previous research in L2 vocabulary learning (e.g., Nakata, 2015), we first aimed to test the generalizability of the spacing effect—specifically, whether the relative advantage of spaced over massed practice would extend to L2 phonetic learning. Drawing on prior phonetic training studies (e.g., Alfotais et al., 2025), we predicted that spaced practice would lead to greater learning gains than massed practice. This effect was expected to be substantial, as Alfotais et al. reported significantly larger effect sizes for spaced conditions ($d > 2$) than for massed conditions ($d = 1–2$) at both immediate and delayed post-tests. Such findings suggest that spaced practice can result in nearly *twice* the retention gains of massed practice (Alfotais et al., 2025; Nakata, 2015).

Prediction 2: Equal vs. expanding spacing

Although spaced repetition is a well-established method for enhancing memory, debate *remains* as to whether expanding spacing—gradually increasing the intervals between reviews—yields greater learning gains than equal spacing, where intervals remain constant. Proponents of expanding spacing argue that increasing retrieval difficulty

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

strengthens memory by engaging effortful retrieval and reconsolidation processes (Karpicke & Bauernschmidt, 2011; Landauer & Bjork, 1978; Pyc & Rawson, 2009). In contrast, equal spacing provides a simpler and more consistent schedule, which may be more effective when the initial learning task is particularly demanding (e.g., Kang et al., 2014). In the present study, given the moderate difficulty Mandarin learners experience in acquiring the English /ɛ/–/æ/ contrast, we predicted that expanding spacing would lead to greater improvement than equal spacing. Previous research suggests that the gains from expanding spacing may yield moderately higher retention —approximately 10–20% more—than equal spacing (Nakata, 2015).

Prediction 3: Aptitude–treatment interaction

Recent research has shown that individual differences in perceptual-cognitive abilities can strongly influence the outcomes of L2 speech instruction. While well-structured, explicit training tends to benefit learners broadly, those with lower aptitude profiles may struggle more under less scaffolded (e.g., Suzukida & Saito, 2023 for the null aptitude effects in classroom L2 speech learning), more naturalistic conditions—such as study abroad settings (Saito & Tierney, 2025), incidental rather than intentional phonetic training (Correia et al., 2025; Saito et al., 2022), or meaning-oriented over form-focused instruction (Ruan & Saito, 2023; Xu et al., 2024). In the current study, we hypothesized that learners with lower aptitude would make limited gains in the more demanding massed condition, where cognitive load is concentrated and consolidation opportunities are minimal. In contrast, spaced practice—particularly expanding spacing—was expected to mitigate these individual differences, leading to robust gains across aptitude levels.

Methodological implications

In response to R1-R3, the paper aims to provide critical methodological implications regarding whether, to what degree, and how introducing different practice schedules (equal and expanding spacing) can influence and enhance the effectiveness of the same phonetic training compared to traditional massed practice. Given that many existing studies traditionally rely on massed practice schedules, the findings may allow us to reinterpret their potential effectiveness if the distribution of training sessions is adjusted.

Method

This study employed a quasi-experimental design with a pre-test, an immediate post-test, and a delayed post-test. A total of 60 participants took part and were randomly assigned to either one of two experimental groups ($n = 39$) or a control group ($n = 20$). All participants completed the pre-test before undergoing six training sessions. Each experimental group received the same phonetic training—focusing on the identification and discrimination of the English vowels /e/ and /æ/ with feedback—delivered under different temporal schedules (see Figure 1 for a visual summary).

Following the well-established HVPT paradigm, first introduced by Logan, Lively, and Pisoni (1991), the training stimuli consisted of naturally produced tokens of real English words, recorded by multiple native speakers. This methodological framework has been extensively validated in more than 50 empirical studies over the past three

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

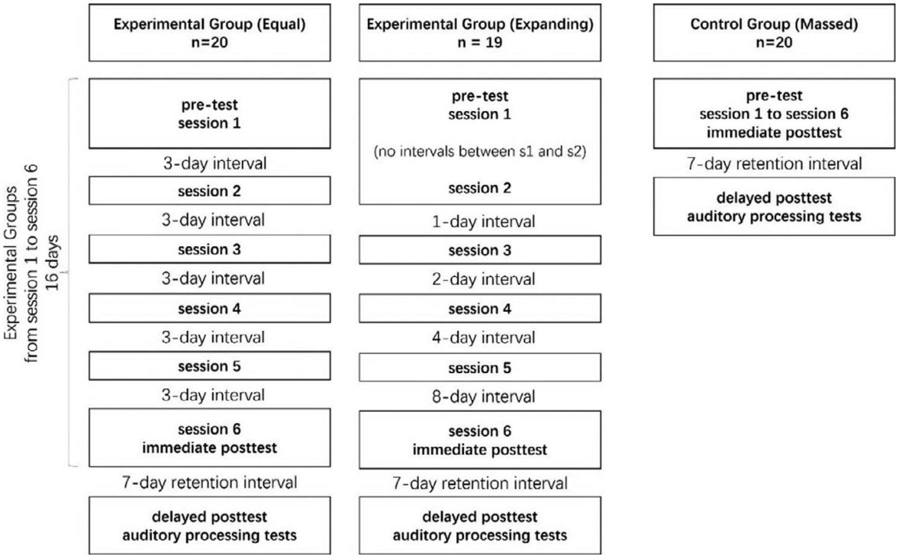


Figure 1. Overview of the experimental procedure across three conditions: Equal spacing (three-day fixed intervals), Expanding spacing (intervals increasing from 0 to 8 days), and Massed practice (six sessions completed in one day). All participants completed a pre-test, and the two experimental groups received immediate post-tests after session 6, followed by a seven-day retention interval and delayed post-tests. Auditory processing aptitude tests were conducted after the delayed post-tests.

decades and was recently synthesized in a comprehensive meta-analysis (Uchihara et al., 2025), which confirmed its robustness in promoting durable gains in L2 speech perception and production.

Upon completion of the six sessions, participants took an immediate post-test, a delayed post-test, and a battery of auditory processing tests to assess aptitude. The design incorporated two of the two methodological criteria proposed by Nakata (2015) for evaluating the effects of spacing: (1) provision of trial-by-trial feedback during training and (2) inclusion of multiple retention intervals via both immediate and delayed post-tests. These conditions were deemed particularly important for contrasting the effects of equal and expanding spacing schedules.

The three-day spacing interval in the Equal Spacing group was selected to align with principles of optimal distributed practice, where intersession intervals are proportional to the desired retention interval (Cepeda et al., 2006). Given the seven-day delay before the final post-test, a three-day interval was expected to strike an appropriate balance between retrieval difficulty and memory consolidation for the moderately difficult /ɛ/–/æ/ contrast. This spacing schedule also allowed for consistent learner engagement and operational feasibility across the 16-day training window.

In the Expanding Spacing group, the intersession intervals gradually increased from 1 to 8 days (1–2–3–4–5–8). This schedule was grounded in the expanding retrieval hypothesis, which posits that learning is enhanced when retrieval becomes increasingly effortful but remains successful over time (Karpicke & Bauernschmidt, 2011; Landauer & Bjork, 1978). The initial short intervals supported early encoding of the novel phonetic contrast (/ɛ/–/æ/), while later, more widely spaced sessions fostered durable consolidation in anticipation of the seven-day delayed post-test. This structure was intended to optimize both short- and long-term gains by balancing encoding support and memory challenge across the training timeline.

Participants

The participants were 60 adult Chinese ESL learners recruited through social media platforms in a specific region of China. Participation was incentivized by advertising the project as a free English phonetic training program, thereby attracting only individuals genuinely interested in the project. This approach was intended to avoid recruiting participants who might otherwise participate solely for monetary incentives and thus complete the training without sufficient concentration. All were over 18 years old, self-reported no hearing impairments, and had received at least 7 years of formal English education as part of China's compulsory schooling system. According to national curriculum guidelines, this educational background corresponds to CEFR A2–B1 proficiency levels. None of the participants reported extensive study abroad experience. Interested individuals contacted the researcher to enroll. Prior to participation, they received an information sheet and signed a consent form. One participant withdrew before completing the study due to personal reasons, leaving a final sample of 59 participants. The average age was 29.6 (Range = 20–54). On average, they began learning English at age 7.6 (Range = 1–13 years). Participants were randomly assigned to one of three groups: Expanding spacing (n = 19), Equal spacing (n = 20), or Control (Massed practice) (n = 20).

Power analysis of sample size

Importantly, for robust statistics, participants with extremely high scores (>90% accuracy) were excluded at the outset of the project to avoid ceiling effects. This resulted in a final sample of 53 participants for analysis. To estimate the required sample size, a power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Although Alfotais et al. (2025) reported relatively large effects (d = 2) in their precursor research, we adopted a more conservative approach and set the predicted effect size to "medium," in line with Yao et al.'s (2025) meta-analysis, which suggested that L2 phonetic training typically yields medium effects (d = 0.762). For three groups (massed, equal spacing, and expanding spacing) assessed across three testing points (pre, immediate post, and delayed post), achieving medium effects (f = 0.25) with sufficient power (0.95) required a sample size of 54. Thus, our final sample size (n = 53) was considered adequate. Notably, this was comparable to existing research (n = 49; Alfotais et al., 2025).

Setup

The experiment was conducted online via the Gorilla platform (Anwyl-Irvine et al., 2020), with participants completing all tests and training individually on their own computers and headsets under the guidance of a researcher (a native Mandarin speaker). Several precautionary measures were implemented to ensure that participants fully understood and engaged with the tasks, remained focused, and completed the procedures without distraction. Any participants who showed evidence of deviation were excluded to maintain data quality. First, we explicitly advertised the project as an opportunity to learn English pronunciation for free, ensuring that participants were genuinely motivated by the project's objective (i.e., improving their L2 English speech proficiency). Second, the investigator arranged an individual videoconferencing meeting with each participant, explained the project's purpose and procedures in their L1 (Mandarin), and confirmed comprehension. During this meeting, participants also completed an online survey about their EFL experience on Gorilla as a test run and checked the sound level of their headsets. Third, the investigator monitored participants' progress in each training session via Gorilla. Pilot data collection confirmed that each training session lasted about 10 minutes. A threshold of 15 minutes per session was set for potential elimination; however, no participants exceeded this limit. In sum, we are confident that all participants included in the final analyses (*n* = 53) completed the tasks with adequate levels of understanding, engagement, and concentration.

Speech stimuli

Target of training

The focus of this study was on the perception of the English vowels /ɛ/ and /æ/, a high-functional-load contrast in English that is particularly difficult for Mandarin speakers. Prior research (e.g., Xiao & Chen, 2021) has shown that Mandarin speakers often assimilate both /ɛ/ and /æ/ to their native /a/ (or /e/) category, resulting in poor perceptual discrimination and inaccurate production.

Flege's Speech Learning Model (SLM; Flege & Bohn, 2021) attributes this perceptual difficulty to category assimilation. The likelihood of establishing a new L2 phonetic category depends on the perceived dissimilarity between the L2 sound and its closest L1 counterpart. Since Mandarin lacks the /æ/ vowel, learners typically assimilate it to /a/ or /e/, resulting in persistent misperception. The SLM also emphasizes the reciprocal relationship between perception and production, whereby inaccurate perception reinforces inaccurate production and vice versa. Best and Tyler's (2007) Perceptual Assimilation Model offers a similar account: both English vowels are assimilated to a single Mandarin category (/a/ or /e/). This Single-Category assimilation prevents learners from forming distinct representations of /ɛ/ and /æ/, leading to consistently poor discrimination performance in empirical studies.

Indeed, research has repeatedly shown that this contrast is among the most problematic for Mandarin learners, often proving more difficult to acquire than other challenging pairs (e.g., /i/-/ɪ/, /u/-/ʊ/; Thomson, Nearey, & Derwing, 2009). This perceptual difficulty is further compounded by Mandarin speakers' reliance on acoustic cues that differ from those used by native English speakers. While native English speakers employ both spectral (vowel height, F1) and durational cues to distinguish /ɛ/ and /æ/, Mandarin learners often struggle to use these cues in a native-like manner, which hinders the establishment of distinct phonological categories for the two vowels (Kachlicka, Symons, Saito, Dick, & Tierney, 2024).

Lexical items

L2 phonetic training in the current study employed words that contrasted /ɛ/ and /æ/ in a consonant-vowel-consonant structure. A total of 20 lexical items were selected—10 trained items, which appeared in the training and both tests, and 10 untrained items, which appeared only in the tests. This allowed for an assessment of both trained learning and generalization to novel items. The majority of items (19 out of 20) belonged to the first 2,000 word families in the British National Corpus (BNC), based

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

Table 1. Target Word pairs used during training and testing (trained items) vs. testing only (untrained items)

Trained items	Untrained items
1. pen-pan	1. said-sad
2. beg-bag	2. guess-gas
3. head-had	3. leg-lag
4. bed-bad	4. pet-pat
5. men-man	5. bet-bat

on frequency profiles available via Tom Cobb's Vocabulary Profiler. Thirteen items were classified as K-1, and six as K-2. The remaining word (*lag*) was from the K-5 level but followed regular phonological patterns and was therefore deemed appropriate. According to Milton (2009), a vocabulary size of 1,500–2,500 words corresponds to CEFR A2 level, suggesting that the selected items were appropriate for the study's target participants, all of whom had a minimum estimated proficiency of CEFR A2. A list of the target items was displayed in Table 1.

Talkers

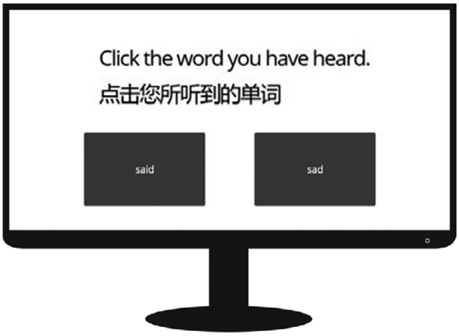
Two talkers (one male, one female) were recruited to produce the auditory stimuli. Both were native speakers of British English who spoke with Received Pronunciation and had experience as professional voice actors. Prior to recording, the talkers were informed of the nature of the study and instructed to clearly distinguish between the English vowels /ɛ/ and /æ/ in their pronunciations. In total, they produced 40 stimuli (20 per speaker). All recordings were completed in their own home studios using professional-grade equipment to ensure high audio quality.

It is important to acknowledge that the training involved only two talkers. According to Uchihara et al.'s (2025) methodological synthesis, this number falls at the lower end of HVPT research, which typically ranges between 2 and 30 talkers. While their meta-analysis suggested that the number of talkers may influence outcomes for highly proficient L2 learners, our participants were relatively lower proficiency (CEFR A2–B1). We therefore assumed that two talkers would be sufficient to trigger learning in this context. Supporting this assumption, Nagle, Bruun, and Zarate-Sandez (2025) found no advantage of using six over two talkers when training English speakers on relatively simple Spanish stop consonants.

Pre/post-test measures

To assess the effects of training on learners' perception of the English vowels /ɛ/ and /æ/ , a forced-choice identification test was administered at pre-test, immediate post-test, and delayed post-test. To accommodate participants' availability and geographical distribution, the tests were delivered online using the Gorilla Experiment Builder platform. During the test, participants completed a two-alternative forced-choice identification task. Using their own devices in a quiet environment, they listened to pairs of minimally contrasting words (e.g., *pen* vs. *pan*) that differed only in the target vowel sound. The stimuli included both trained and untrained items (see Table 1), and each item was recorded by 3 native speakers, resulting in 60 audio tokens in total. For each trial, participants heard a single word and were asked to identify what they heard by selecting one of two written options displayed on the screen (see Figure 2).

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press



Click the word you have heard.

点击您所听到的单词

Figure 2. Sample screenshot of the two-alternative forced-choice identification task interface used in the perception test. On each trial, participants listened to a spoken word and selected the corresponding written form from two options (e.g., *said* vs. *sad*). Instructions were presented bilingually in English and Chinese ("Click the word you heard." / 点击您所听到的单词) to ensure clarity for all participants. Responses were submitted by clicking one of the two options, after which the task automatically advanced to the next item.

Phonetic training

Training setup

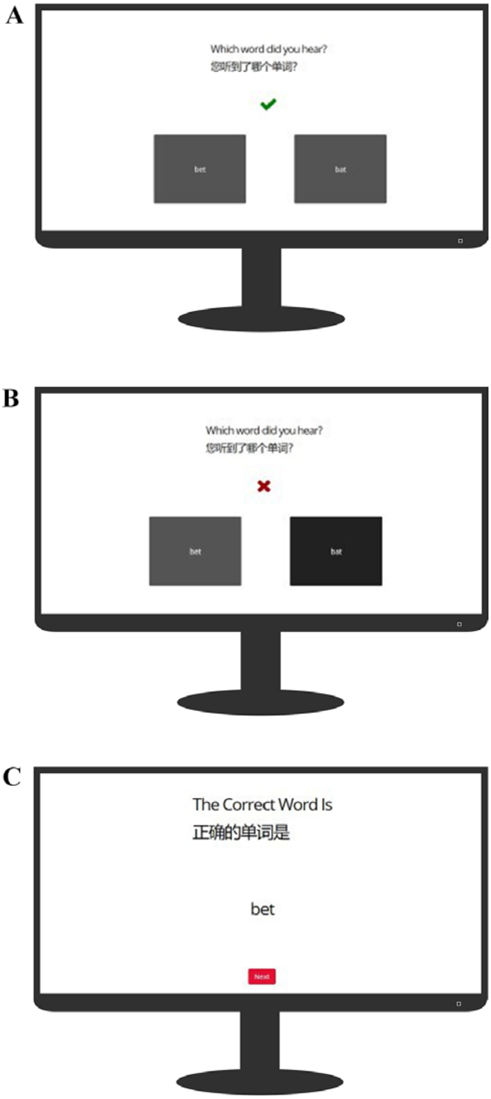
Following the pre-test, participants completed six sessions of perceptual training using a high-variability paradigm with trial-by-trial feedback. Each session lasted approximately 10 minutes, amounting to about 1 hour of training in total. In each session, participants identified 20 audio tokens of the 10 trained words (see Table 1), spoken by both a male and a female talker. After each response, participants received immediate visual feedback, followed by a replay of the correct word accompanied by a bilingual message (see Figure 3). This feedback format was informed by Nakata (2015), who emphasize the importance of corrective feedback and repeated exposure for consolidating difficult phonetic contrasts. To minimize reliance on rote memorization, the order of word presentation was randomized in each session. The six-session structure was based on research suggesting that 2–20 exposures are typically required to establish lexical knowledge (Saragi et al., 1978; Uchihara et al., 2022; Webb, 2007). In the present design, each word was presented at least 12 times across sessions (6 sessions × 2 talkers), providing sufficient repetition to support perceptual learning.

Group conditions

Upon enrolment, participants were randomly assigned to one of three groups: the Massed Practice group (control), the Equal Spacing group (experimental), and the Expanding Spacing group (experimental).

All participants followed the same overall sequence: a pre-test, six training sessions, an immediate post-test administered directly after the final training session (Retention Interval = 0 days), and a delayed post-test administered seven days later. Auditory processing aptitude tests were completed within two days after the delayed post-test. Notably, the first training session began immediately after the pre-test.

Each group differed in how their training sessions were spaced. Drawing from Landauer et al. (1978), which originally employed three repetitions following initial exposure (e.g., 5–5–5), the current study implemented five intersession intervals,



Which word did you hear?
您听到了哪个单词?

The Correct Word Is
正确的单词是

Figure 3. Example screenshots from the phonetic training session. After selecting one of the two word options, participants received immediate visual feedback indicating whether their response was correct (green tick; 3A) or incorrect (red cross; 3B). This was followed by a feedback screen presenting the correct word along with its pronunciation and a bilingual prompt (“The correct word is” / 正确的单词是; 3C). The two-step feedback procedure was designed to enhance phonetic category learning by encouraging accurate form-meaning mapping and reinforcing correct auditory representations.

allowing each target item to be reviewed six times across the six sessions. The intersession interval was measured in time (hours/days), and each group followed a distinct schedule:

- **Massed Group (0–0–0–0–0):** All six sessions were completed consecutively without delay between them (e.g., session 2 followed immediately after session 1).

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

- **Expanding Group (0–1–2–4–8):** ISIs increased across sessions—session 2 followed immediately after session 1, session 3 was held one day later, session 4 after two days, session 5 after four days, and session 6 after eight days.
- **Equal Group (3–3–3–3–3):** All sessions were spaced evenly, with three days (72 hours) between each.

To control for absolute spacing, which may influence the effectiveness of relative spacing (Dobson, 2012; Maddox, Balota, Coane, & Duchek, 2011; Nakata, 2015), both the expanding and equal groups were matched on total time span (15 days) across the 5 intersession intervals (Expanding: 0 + 1 + 2 + 4 + 8 = 15 days; Equal: 3 × 5 = 15 days). This controlled design ensured that any observed differences could be attributed to relative rather than absolute spacing.

Although intersession intervals were planned in days, the implementation was based on exact hours. For example, if a participant in the Equal group completed session 1 at 9:00 a.m. on Day 1, session 2 would become available at 9:00 a.m. on Day 4. Given the practical constraints of participant schedules, some flexibility was introduced: each training session could be completed within a six-hour window after the scheduled start time (e.g., between 9:00 a.m. and 3:00 p.m.).

Auditory processing measures

To assess participants' individual differences in auditory processing abilities, two types of tests were administered: (a) auditory acuity tasks targeting spectral and temporal discrimination, and (b) auditory–motor integration tasks assessing rhythm and melody reproduction. All tasks were delivered via the Gorilla platform. Written instructions were provided in both English and Mandarin, and participants completed a brief practice trial before each task to ensure comprehension. All test materials are available through *SLA Speech Tools* (Mora-Plaza, Saito, Suzukida, Dewaele, & Tierney, 2022), with full methodological details and validation reported in Saito and Tierney (2025).

Auditory acuity test

Three adaptive three-alternative forced-choice discrimination tests were used to assess participants' ability to detect subtle differences in duration, formant, and pitch. Each trial presented three non-verbal sounds (with a fixed 0.5s inter-stimulus interval); participants were asked to identify which of the first or third sound differed from the standard second sound by clicking "1" or "3" on the screen. The stimuli were 500-ms four-harmonic complex tones with a fundamental frequency of 330 Hz. The specific test parameters were as follows. In the duration discrimination task, the standard stimulus was set at 250 ms, and target durations varied from 252.5 to 500 ms in 2.5 ms increments. In the pitch discrimination task, the standard tone was 330 Hz, while the target frequencies ranged from 330.3 Hz to 360 Hz in 0.3 Hz steps. For the formant discrimination task, the stimuli were complex tones with a fundamental frequency of 100 Hz, F1 at 500 Hz, and F3 at 2,500 Hz. The second formant was set at 1,500 Hz in the standard stimulus and varied from 1,502 to 1,700 Hz in 2 Hz increments in the target stimuli.

Following Levitt's (1971) adaptive threshold procedure, task difficulty increased after three consecutive correct responses and decreased after each incorrect response. The task ended after either 8 reversals or 70 trials. Final scores ranged from 0 to 100, with lower scores indicating finer discrimination ability.

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

Audio-motor integration test

To evaluate participants' ability to reproduce temporal and spectral auditory information, two reproduction tasks were administered: rhythm reproduction and melody reproduction. In the rhythm reproduction task, participants listened to rhythmic patterns that were 3.2 seconds in length and composed of drum hits, adapted from Povel and Essens (1985). Each pattern was presented three times per trial, after which participants were instructed to reproduce the rhythm by tapping the spacebar on their keyboard. In the melody reproduction task, participants heard short melodies consisting of seven notes, with each note lasting 300 milliseconds. The melodies were constructed using six-harmonic complex tones with fundamental frequencies of 220, 246.9, 277.2, 311.1, and 329.6 Hz. After listening to each melody three times, participants reproduced it by pressing number keys from 1 to 5, where "1" represented the lowest pitch and "5" the highest. Prior to the actual test, participants were given a brief familiarization period to ensure they understood the correspondence between the number keys and the pitch scale.

In contrast to the acuity tasks, higher scores in the reproduction tasks indicated better performance. All tasks have been validated and used extensively in recent studies examining auditory processing in adult L2 acquisition (Kachlicka, Saito, & Tierney, 2019; Saito et al., 2022; Shao et al., 2023).

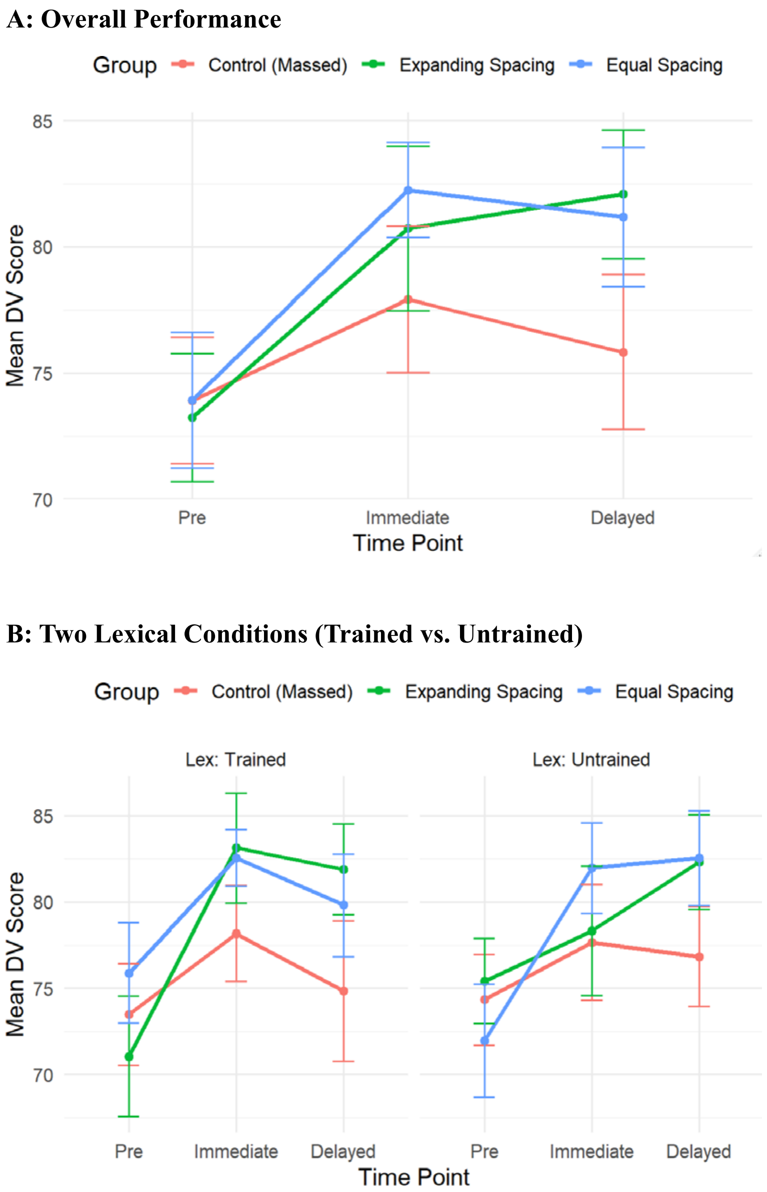
Results

The data and R script are publicly available on OSF: https://osf.io/9x48y. As observed in existing phonetic training studies of this kind, participants' test scores were negatively skewed. Following established procedures, six participants who scored above 90% accuracy at the onset of the project were removed from the dataset, as these participants had little room for improvement (*n* = 3 from Expanding Spacing, *n* = 3 from Equal Spacing; see Iverson, Pinet, & Evans, 2012, for a similar decision). According to Kolmogorov–Smirnov tests, the final dataset (*n* = 53) did not show significant deviations from normality for the Control, Expanding, and Equal groups (*D* = .129, .089, and .122, respectively; p > .05). Descriptive statistics for participants' vowel perception scores by group (Control, Expanding, Equal) and lexis condition (trained, untrained) across the three test points (pre-, immediate post-, and delayed post-test) are summarized in Table 2 and visually presented in Figure 4. Participants' average accuracy ranged between 70% and 80%, which is comparable to other studies that have targeted and reported significant improvements in the context of challenging L2 speech acquisition cases (see Saito et al.'s 2022 synthesis of L2 phonetic training studies). This suggests that the Mandarin participants in the present study did experience difficulty with the target contrasts (English /ɛ/ and /æ/), leaving substantial room for improvement.

Table 2. Descriptive statistics of participants' vowel perception scores at different time points as per group conditions

		Pre		Immediate post		Delayed post	
Group	Lexical conditions	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control/massed	Trained	73.5	13.1	78.1	12.4	74.8	18.1
(*n* = 20)	Untrained	74.3	11.7	77.6	14.9	76.8	12.8
Expanding spacing	Trained	73.1	14.4	83.5	12.1	82.9	10.4
(*n* = 16)	Untrained	77.7	11.5	79.8	14.8	83.8	11.3
Equal spacing	Trained	75.8	11.9	82.5	6.7	79.8	12.2
(*n* = 17)	Untrained	71.9	13.5	81.9	10.8	82.5	11.3

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press



A: Overall Performance

B: Two Lexical Conditions (Trained vs. Untrained)

Figure 4. Vowel perception performance across groups and time. A visual summary of the three groups' (Control, Expanding, Equal) vowel perception performance (%) at three different time points (Pre, Immediate, Delayed) overall (4A) and across two lexical conditions (4B). While the control/massed group showed improvement at the immediate post-tests, their performance declined at the delayed post-tests. Both expanding and equal groups maintained their performance up to the delayed post-tests.

A linear mixed-effects regression analysis was performed using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in the R statistical environment (R Core Team, 2025). The following model (DV ~ Group * Time * Lex + (1 | ID)) was constructed to examine the extent to which participants in three different groups

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

(Group: Controlled, Expanding, Equal) improved their L2 vowel perception scores (%) over time (Time: Pre, Immediate Post, Delayed Post) and whether these improvements extended to untrained lexical contexts (Lex: Trained, Untrained).

As shown in Table 3 (and visually plotted in Figure 4A), the analysis revealed significant main effects of Time, as well as significant Group × Time interaction effects, indicating that the trajectory of improvement differed across the three groups. However, the Group × Time × Lex interaction was not statistically significant (p = .132), suggesting that the pattern of improvement was consistent across both trained and untrained lexical conditions. This implies that the observed learning gains were generalizable beyond the specific lexical items used during training.

To unpack the significant interaction effects of Group and Time, we performed a set of post-hoc multiple comparisons. These comparisons were analyzed using the *emmeans* package (Lenth & Piaskowski, 2025) in R and are summarized in Table 4. The results indicated that the Control/Massed group showed rapid improvement from pre-test to immediate post-test, with a medium effect size ($p = .040$, $d = 0.55$, 95% CI [0.10, 0.99]). However, these gains were not statistically significant at the delayed post-test ($p = .474$). In contrast, both the Expanding and Equal groups demonstrated significant improvements at both the immediate and delayed post-tests (p < .001 for all comparisons with pre-test scores). Effect sizes varied slightly between these two groups. The Expanding group exhibited large effects at both the immediate ($d = 1.02$) and delayed ($d = 1.21$) post-tests. Similarly, the Equal group showed a large effect at the immediate post-test ($d = 1.14$), which remained substantial at the delayed post-test ($d = 0.99$). When comparing improvements from pre-test to immediate post-test, the Expanding ($d = 1.02$) and Equal ($d = 1.14$) groups demonstrated effect sizes nearly twice as large as that of the Control group ($d = 0.55$). Furthermore, at the delayed post-test, the Control group's gains were not statistically significant and reflected only a small effect size ($d = 0.26$), whereas the Expanding ($d = 1.21$) and Equal ($d = 0.99$) groups maintained large effect sizes, clearly outperforming the Control group in terms of sustained improvement.

The objective of the final statistical analyses was to examine the extent to which participants' improvement over time was linked to their individual differences in

Table 3. Results of mixed-effects model analysis for L2 vowel perception

Fixed effect	*SS*	*F*	p
Group	53.03	0.49	.614
Time	2807.77	26.07	<.001*
Lex	0.31	0.01	.939
Group: Time	528.47	2.45	.046*
Group: Lex	26.10	0.24	.785
Group: Time: Lex	385.14	1.79	.132

Note: * indicates $p < .05$

Table 4. Results of multiple comparison analyses

	Pre to immediate post			Pre to delayed post		
Group	*t*	p	*d* [95% CI]	*t*	p	*d* [95% CI]
Control	2.413	.040	0.545 [0.096, 0.994]	1.168	.474	0.261 [0.185, 0.708]
Expanding	4.088	<.001	1.022 [0.516, 1.528]	4.826	<.001	1.207 [0.697, 1.715]
Equal	4.682	<.001	1.136 [0.642, 1.629]	4.076	<.001	0.989 [0.497, 1.480]

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

auditory processing abilities. To this end, participants' relative gain scores were calculated for two different contexts (pre to immediate post, pre to delayed post) using the following formula: (pre-test minus post-test scores)/pre-test scores. Using their gain scores as dependent variables (one for pre-to-immediate, the other for pre-to-delayed), the following linear regression models were constructed—i.e., DV ~ Group * formant + Group * duration + Group * melody + Group * rhythm. As summarized in Table 5, not surprisingly, the main effects of Group reached statistical significance as both Expanding and Equal outperformed Control/Massed at both immediate and delayed post-tests (p < .05). However, none of the main and interaction effects of auditory processing variables reached statistical significance (p > .05). This suggests that the participants' improvement patterns were unrelated to their perceptual-cognitive individual differences.

Discussion

Over the past 30 years, a substantial body of research has investigated various phenomena underlying instructed L2 phonetic learning. As shown in recent meta-analyses, a few hours of targeted phonetic training typically yield medium-sized learning gains (e.g., d = 0.762 as reported in Yao et al., 2025). According to Saito et al. (2022), such improvements generally fall within the range of 5–10%. However, most existing studies have exclusively employed massed training conditions, where sessions are delivered in close succession with minimal intervals. In contrast, psycholinguistic research increasingly suggests that the distribution of practice—how learning sessions are spaced over time—can significantly enhance L2 learning outcomes in domains such as vocabulary and grammar (e.g., Suzuki et al., 2019). The primary aim of the present study was to explore how alternative practice schedules (massed, equal spacing, and expanding spacing) affect the efficacy of L2 phonetic training. In doing so, we aim to offer methodological insights into how the findings of past studies might be reinterpreted through the lens of practice distribution.

After just one hour of training, participants in the Control/Massed group showed approximately a 5% improvement, with a medium effect size at the immediate post-test (d = 0.545). Since neither the main nor interaction effects of Lex (trained vs. untrained lexical contexts) were observed for the Control/Massed group (or any other spacing group), the training gains were not influenced by whether performance was tested on trained or untrained lexical contexts. This suggests that the training gains generalized from trained to untrained lexical contexts. When it comes to the durability of such

Table 5. Results of multiple regression analyses

	Pre to immediate post			Pre to delayed post		
Fixed effect	*SS*	*F*	p	*SS*	*F*	p
Group	0.040	5.706	.021*	0.104	4.114	.024*
Formant	0.001	0.156	.694	0.003	0.259	.613
Duration	< 0.001	<0.001	.999	0.001	0.001	.978
Melody	0.008	0.763	.387	0.002	0.227	.636
Rhythm	0.001	0.113	.738	0.001	0.017	.895
Group: Formant	0.005	0.255	.775	0.010	0.398	.674
Group: Duration	0.017	0.783	.463	0.003	0.137	.872
Group: Melody	0.027	1.238	.301	0.008	0.344	.710
Group: Rhythm	0.013	0.629	.538	0.008	0.335	.717

Note: * indicates p < .05

training effects, however, they were unable to retain these gains, with nearly half of the improvement lost by the delayed post-test ($d = 0.261$). The magnitude of the initial gain is comparable to previous findings (e.g., Yao et al., 2025, $d = 0.762$), and the developmental pattern—rapid improvement followed by decay—is consistent with trends observed in explicit L2 training research (e.g., Li, 2010). Crucially, when training sessions were distributed either with expanding intervals (0–1–2–4–8) or equal intervals (3–3–3–3–3), learning gains increased by approximately 10%. The effect sizes at the immediate post-test were large ($d = 1.002$ for Expanding and $d = 1.136$ for Equal), roughly double that of the Massed group (5% gain; $d = 0.545$). Most importantly, both spacing groups retained their gains at the delayed post-test, with effect sizes remaining large ($d = 1.207$ for Expanding and $d = 0.989$ for Equal). Here, we did not observe clear group differences between the Expanding and Equal groups.

Building on Alfotais et al. (2025), the current study provided another empirical support that introducing spaced practice schedule may not only double the effectiveness of traditional phonetic training paradigm with massed practice schedule at immediate post-tests but also boost the durability of such large gains over time. The findings here could be comparable to other dimensions of L2 learning (e.g., Nakata, 2015 for vocabulary; Suzuki, 2018 for grammar). Given that most of the existing L2 phonetic training studies have exclusively relied on massed conditions or short intersession intervals (one day), the methodological implications of the study and Alfotais et al. (2025) could be substantial. Whereas they typically note medium effects (e.g., Yao et al., 2025, $d = 0.762$), it is logically possible that replicating these studies with spaced practice sessions could help amplify the size and durability of their initial effectiveness. Whereas scholars have paid much attention to a range of methodological factors such as variability in input (e.g., Sakai & Moorman, 2018) and nature of training (e.g., task-based language; Mora-Plaza et al., 2024) as a way to enhance the existing L2 phonetic methods, the current study proposes a crucial methodological and pedagogical message that manipulating practice schedule could be another way with a view of acquisitionally rich instructed L2 phonetic learning.

An unpredicted finding of the current study was that the introduction of two different spacing schedules—Expanding (0–1–2–4–8) and Equal (3–3–3–3–3)—did not result in clear differences in training effectiveness or retention. The null finding itself has important pedagogical implications. For classroom practice, it suggests that what matters most is implementing some form of distributed practice rather than massing sessions together. Teachers therefore need not be concerned with optimizing the exact spacing formula (e.g., equal vs. expanding intervals); instead, ensuring that practice is spaced out—even moderately—appears sufficient to yield robust improvements in phonetic learning.

From theoretical perspectives, this outcome contrasts with findings in L2 vocabulary research, where expanding spacing has often shown superior benefits (e.g., Nakata, 2015). The current results suggest that, in the context of instructed L2 phonetic learning, what may matter most is that input is distributed over time, allowing learners sufficient opportunity to encode, proceduralize, and restructure their phonetic systems. However, this raises the question: Why might the specific type of spacing (equal vs. expanding) not significantly affect outcomes in L2 phonetic training?

One possible reason is the potentially different learning mechanisms underlying L2 vocabulary and phonetic learning. Much of the empirical support for expanding spacing comes from research on vocabulary, which involves form-meaning mapping using declarative memory. In such contexts, expanding intervals are thought to introduce “desirable difficulties” that strengthen memory traces through effortful yet

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

successful retrieval and encoding variability (Landauer & Bjork, 1978). In contrast, L2 phonetic learning—particularly the acquisition of speech perception and production skills—primarily involves procedural memory. This domain relies on the implicit formation of perceptual and motor routines, such as tuning the auditory system to distinguish unfamiliar phonetic contrasts or developing accurate articulatory patterns.

Procedural learning may respond differently to practice schedules. For instance, motor learning literature suggests that consistent and relatively immediate repetitions early in training can help stabilize new sensorimotor mappings, minimize the risk of fossilizing errors, and promote fluent automaticity (Best & Tyler, 2007). From this perspective, equal spacing may offer a more reliable and sustainable rhythm of practice, especially during the early stages of phonetic training.

Moreover, acquiring L2 phonetic categories involves detecting and re-weighting fine-grained acoustic cues—an inherently difficult and subtle process that differs from the more explicit mappings involved in vocabulary or grammar learning (see Kachlicka et al., 2024, for a cue-weighting account of L2 speech learning). If expanding intervals increase too quickly, learners may face retrieval failure rather than productive difficulty, particularly when forming new phonetic categories that require high precision. This risk may diminish the potential advantages of expanding spacing in this specific domain.

Finally, it is notable that the relative efficacy of distributed practice was not influenced by participants' perceptual–cognitive individual differences. The null aptitude effects observed here may seem surprising, given our initial predictions. In this study, although the type of input distribution varied (massed vs. spaced), the training itself was explicit: participants were fully aware of what they were learning and consciously focused on improving in each session. Under such conditions, the role of aptitude may be less fundamental. Indeed, previous research has shown that aptitude effects in L2 speech learning are more clearly observed when training conditions are more demanding—communicatively authentic, meaning-oriented, incidental, and/or implicit—thus posing greater challenges for individuals with lower perceptual-cognitive abilities (Correia et al., 2025; Xu et al., 2024). Taken together, our findings provide *tentative* suggestion that spaced practice is a broadly effective instructional strategy, likely to benefit the majority of L2 learners regardless of their perceptual–cognitive profiles.

Conclusion

In the domains of L2 vocabulary and grammar learning, there is growing empirical support for the benefits of distributed practice, with studies showing that spacing practice sessions can substantially enhance the effectiveness and durability of instruction (e.g., Suzuki et al., 2019). Although prior research on L2 phonetic training has reported moderate learning gains (e.g., Yao et al., 2025), these studies have almost exclusively relied on massed training schedules. Some have explicitly called for investigations into alternative distribution methods (cf. Alfotais et al., 2025). Addressing this gap, the present study examined the effects of two types of spaced training—expanding and equal—relative to a traditional massed training condition, focusing on Chinese learners' acquisition of the English /ɛ/–/æ/ contrast.

Results showed that both spaced conditions led to significantly larger improvements —approximately double the gains of the massed condition at the immediate post-test— and these gains were successfully retained over time. From a methodological standpoint, the findings suggest that previous studies using massed practice may underestimate the full potential of L2 phonetic training. Had spaced practice been implemented, the reported effects could have been considerably larger and more durable. Notably, the present study indicated that effect sizes can be approximately twice as large for spaced conditions relative to massed conditions.

While distributed practice clearly outperformed massed practice—echoing patterns found in other areas of instructed L2 learning (e.g., Suzuki et al., 2019)—no significant differences were observed between the expanding and equal spacing conditions. Although these two approaches may differ in their effectiveness for declarative learning (e.g., vocabulary or grammar), their impact on procedural learning, such as L2 phonetic acquisition, may be more limited. In this context, subtle distinctions between spacing types might yield only marginal gains, which are difficult to detect given the inherent variability in L2 learning outcomes.

These results highlight the potential for distributed practice—regardless of spacing type—to serve as a robust method for enhancing both the magnitude and durability of L2 phonetic learning. This underscores the need to revisit past findings based on massed training paradigms and to consider whether adopting alternative timing schedules could unlock greater learning potential—for instance, by doubling the size and durability of training effects through the use of spaced conditions.

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at: https://osf.io/9x48y.

Acknowledgments. This project originated from Siying Chen's Master's dissertation submitted to University College London. We thank Tatsuya Nakata for his valuable feedback on the experimental design and acknowledge anonymous reviewers and Editor Luke Plonsky for their insightful comments on an earlier version of the manuscript. This research was supported by a Leverhulme Trust Grant (RPG-2024-391) and a UK-ISPF Grant (1185702223).

References

Alfotais, A., Mahdi, H. S., & Alkhammash, R. (2025). The effect of spaced vs massed repetition on variability phonetic training among Saudi English as foreign language learners. *The Journal of the Acoustical Society of America*, *157*(1), 265–274.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407.

Baddeley, A. (1997). *Human memory: Theory and practice* (Rev. ed.). Psychology Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, *74*(4), 245–248.

Bradlow, A. R. (2008). Training non-native language sound patterns. In J. Hansen & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). John Benjamins.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380.

Correia, S., Rato, A., Ge, Y., Fernandes, J. D., Kachlicka, M., Saito, K., & Rebuschat, P. (2025). Effects of phonetic training and cognitive aptitude on the perception and production of non-native speech contrasts. *Studies in Second Language Acquisition*, *47*(1), 440–457.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 fluency and comprehensibility over time: Impacts of L1 background, instruction, and experience. *Applied Linguistics*, *34*(5), 554–577.

Dobson, J. L. (2012). Effect of uniform versus expanding retrieval practice on the recall of physiology information. *Advances in Physiology Education*, *36*, 6–12.

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Hamouda, A. (2021). The effect of massed vs. distributed practice on EFL vocabulary learning and retention. *Theory and Practice in Language Studies*, *11*(5), 482–489.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r–l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*(5), 3267–3278.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145–160.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, *192*, 15–24.

Kachlicka, M., Symons, A. E., Saito, K., Dick, F., & Tierney, A. T. (2024). Tone language experience enhances dimension-selective attention and subcortical encoding but not cortical entrainment to pitch. *Imaging Neuroscience*, *2*, 1–19.

Kang, S. H., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval?. *Psychonomic Bulletin & Review*, *21*, 1544–1550.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250–1257.

Karpicke, J. D., & Roediger III, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 704.

Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Lenth, R., & Piaskowski, J. (2025). emmeans: Estimated marginal means, aka least-squares means (R package version 2.0.0). Retrieved from https://CRAN.R-project.org/package=emmeans

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309–365.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Maddox, G. B., Balota, D. A., Coane, J. H., & Duchek, J. M. (2011). The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. *Psychology and Aging*, *26*(3), 661.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

Mora-Plaza, I., Mora, J. C., Ortega, M., & Aliaga-Garcia, C. (2024). Is L2 pronunciation affected by increased task complexity in pronunciation-unfocused speaking tasks? *Studies in Second Language Acquisition*, *46* (4), 1117–1149.

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com. http://doi.org/10.17616/R31NJNAX

Nagle, C., Bruun, S., & Zarate-Sandez, G. (2025). Comparing lower and higher variability multi-talker perceptual training. *Applied Psycholinguistics*, *46*, e14.

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning?. *Studies in Second Language Acquisition*, *37*(4), 677–711.

Namaziandost, E., Anwar, C., & Neisi, L. (2020). Comparing the impact of spaced instruction and massed instruction in learning collocations among Iranian EFL learners. *EduLite: Journal of English Education, Literature and Culture*, *5*(1), 55–65.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*(8), 1917–1927.

https://doi.org/10.1017/S0272263125101447 Published online by Cambridge University Press

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?. *Journal of Memory and Language*, *60*(4), 437–447.

R Core Team. (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Ruan, Y., & Saito, K. (2023). Less precise auditory processing limits instructed L2 speech learning: Communicative focus on phonetic form revisited. *System*, *114*, 103020.

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and multimodal high variability phonetic training: Potential, limits, and future directions. *Language Learning*, *72*(4), 1049–1091.

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.

Saito, K., & Tierney, A. (2025). Roles of domain-general auditory processing in second language speech learning revisited: What degree of precision makes a difference? *Language Learning*. https://doi.org/10.1111/lang.12722

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224.

Saragi, T. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78.

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*(1), 28–42.

Shao, Y., Saito, K., & Tierney, A. (2023). How does having a good ear promote instructed second language pronunciation development? Roles of domain-general auditory processing in choral repetition training. *TESOL Quarterly*, *57*(1), 33–63.

Suzuki, Y. (2018). The role of procedural learning ability in automatization of L2 morphology under different learning schedules: An exploratory study. *Studies in Second Language Acquisition*, *40*(4), 923–937.

Suzuki, Y., Nakata, T., & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal*, *103*(3), 713–720.

Suzukida, Y., & Saito, K. (2023). Detangling experiential, cognitive, and sociopsychological individual differences in second language speech learning: Cross-sectional and longitudinal investigations. *Bilingualism: Language and Cognition*, *26*(4), 762–775.

Thomson, R. I., Nearey, T. M., & Derwing, T. M. (2009). A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels. *The Journal of the Acoustical Society of America*, *126*(3), 1447–1460.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30.

Uchihara, T., Karas, M., & Thomson, R. I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*.

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, *44*(2), 357–380.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65.

Xiao, Q., & Chen, M. (2021). On Chinese students' English pronunciation problems and countermeasures. *Learning & Education*, *10*(5), 17–20.

Xu, J., Saito, K., & Mora-Plaza, I. (2024). Task-based pronunciation teaching: Lack of auditory precision but not memory hinders learning. *System*, *127*, 103532.

Yao, Y., He, M., Chen, F., & Zhu, J. (2025). A meta-analysis of second language phonetic training: Exploring overall effect and moderating factors. *Journal of Speech, Language, and Hearing Research*, *68*(4), 1784–1802.

Cite this article: Saito, K., & Chen, S. (2025). Optimizing L2 phonetic learning: Spaced vs. massed training schedule. *Studies in Second Language Acquisition,* 1–22. https://doi.org/10.1017/S0272263125101447