System 127 (2024) 103532

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Task-based pronunciation teaching: Lack of auditory precision but not memory hinders learning

Jiying Xu [a], Kazuya Saito [a,*], Ingrid Mora-Plaza [a,b]

[a] University College London, UK
[b] Universitat de Barcelona, Spain

ARTICLE INFO

Keywords:
Auditory processing
Individual differences
Phonetic instruction
Second language speech
Task-based language teaching
Working memory

ABSTRACT

This study examined the impact of individual differences in perception and cognition on the effectiveness of task-based pronunciation teaching (TBPT). Seventy young adult Chinese speakers were randomly assigned to either an experimental or control group. The experimental group participated in a 30-min TBPT session, engaging in meaning-oriented tasks designed to help them notice and practice two English vocalic contrasts: [ɛ]-[æ] (e.g., “set” vs. “sat”) and [ɪ]-[i] (e.g., “sit” vs. “seat”). Participants' learning patterns were then assessed based on their aptitude on perceptual (auditory processing) and cognitive (working memory) levels. Analysis of pre- and post-test results, conducted through forced-choice identification tasks, revealed significant improvements in vowel performance (approximately 10% gain), particularly among participants with normative auditory precision. However, the effectiveness of TBPT was limited among those with lower auditory precision. These findings highlight the negative impact of lower-order aptitude, specifically auditory processing, on the efficacy of TBPT.

1. Introduction

Over the past few decades, research in second language acquisition (SLA) has extensively explored the benefits of task-based language teaching and learning (TBLT; Ellis, 2003; Long, 2015; Robinson, 2011; Skehan, 1998). This pedagogical approach has proved successful in directing learners' attention to desired linguistic forms within meaningful real-life communication, with the goal of enhancing cognitive processes that lead to second language (L2) acquisition. While most SLA research has mainly concentrated on the benefits of task design and manipulation for L2 grammatical (Gurzynski-Weiss & Baralt, 2014), lexical (Révész & Han, 2006) or pragmatic development (Barón et al., 2020), less attention has been paid to the impact of TBLT on L2 *phonology* (Gurzynski-Weiss et al., 2017; Mora & Levkina, 2017; Mora-Plaza et al., 2024).

While L2 pronunciation teaching and TBLT research acknowledge the importance of pronunciation for the development of intelligibility, comprehensibility (Levis, 2022) and overall L2 oral communication (Long, 2015), and some studies have recognized the occurrence of phonology-related episodes during task-based interactions (Ellis et al., 2001; Gurzynski-Weiss & Baralt, 2014), a growing amount of attention has been given to investigating task-based pronunciation teaching—TBPT (e.g., Gordon, 2021; Mora-Plaza et al., 2024).

This study sets out to contribute to this emerging field of research by exploring the effectiveness of TBPT, targeting two instances of English vocalic contrasts—i.e., [ɛ]-[æ] (e.g., “set” vs. “sat”) and [ɪ]-[i] (e.g., “sit” vs. “seat”). Further, we explore the roles of perceptual

* Corresponding author. 20 Bedford Way, London, WC1H0AL, UK.
E-mail address: k.saito@ucl.ac.uk (K. Saito).

https://doi.org/10.1016/j.system.2024.103532
Received 12 April 2024; Received in revised form 29 October 2024; Accepted 29 October 2024
Available online 8 November 2024

and cognitive individual differences in the potential and limits for TBPT. It is important to acknowledge the substantial variability observed in L2 acquisition outcomes by adult learners. Even when individuals engage in similar amounts and types of language practice, they frequently exhibit varying levels of L2 learning. This variability extends beyond mere exposure to the L2 and is influenced by individual perceptual and cognitive factors associated with explicit language learning, commonly referred to as language aptitude (Li, 2016; Skehan, 2019). Here, we test our hypotheses on the aptitude-treatment interaction: Whereas those learners with relatively normative (or high) aptitude can make the most of TBPT, such instructional effectiveness could be limited among those with relatively low aptitude.

2. Background

2.1. Task-based language teaching and L2 pronunciation

L2 phonetic teaching research grounded on the Communicative Language Teaching (Celce-Murcia et al., 2010) and Automatization in Communicative Contexts of Essential Speech Segments (Gatbonton & Segalowitz, 2005) frameworks has underscored the potential of learning L2 phonetic and phonological forms while being involved in contextualized meaning-oriented activities. Communicative form-focused instruction has been found to enhance automaticity of L2 phonetic and phonological processing and generalization from classroom practice to actual behavior (e.g. Darcy & Rocca, 2023; Ruan & Saito, 2023).

The integration of focus-on-form approaches in communicative language teaching is a central tenet in TBLT (Long, 1985, 2015), where attention to language is not artificially imposed but rather driven by the context the task is placed in. Tasks are real-world communicative activities that orient learners' attention to meaning while using their own linguistic resources, prepare learners for task completion outside the L2 classroom and, most importantly, involve several cognitive processes that may promote L2 development (Ellis, 2003, 2009; Ellis et al., 2020; Skehan, 1998). During task performance, interaction provides crucial opportunities for learners to restructure their interlanguage by drawing their attention to linguistic code features during negotiation of form and meaning (Pica, 1994), which fosters SLA processes of input noticing and processing, intake, L2 knowledge processing and output (Gilabert et al., 2016; Swain, 1985). While several proactive (e.g. input enhancement, input elaboration, task essentialness) and reactive (e.g. corrective feedback) form-focused instructional techniques have been found to enhance the probability to orient attention to particular lexico-grammatical aspects of the language code in purely meaningful situations (see Saito & Lyster, 2012 as an early example of form-focused instruction in L2 phonology), their benefits for L2 phonetic development are still under-researched. Therefore, TBPT stands as an optimal avenue to investigate the development of L2 speech during meaning-based interaction.

To date, there is one study examining L2 oral intervention gains on phonetic levels. De Ridder et al. (2007) included an independent measure of pronunciation (i.e., distinctness, intelligibility, naturalness) and intonation (i.e., clearness, intelligibility, naturalness, melodic phrasing), but these were only assessed subjectively by two raters. The minor presence of pronunciation in interaction studies has been mainly through references to phonology-related episodes arising from negotiation of meaning during spontaneous conversations (Ellis et al., 2001; Gurzynski-Weiss & Baralt, 2014). Despite instances of phonology-related episodes being relatively infrequent, they have been found to be implicated in numerous intelligibility breakdowns during task-based interaction (Loewen & Isbell, 2017).

More recently, several studies have investigated the role of task design and manipulation in raising learners' awareness of phonetic features during communicative task performance and have assessed improvement in L2 pronunciation accuracy (see Gurzynski-Weiss et al., 2017; Mora & Levkina, 2017). On the one hand, TBPT studies targeting L2 suprasegmental features have shown that interacting via face-to-face modality (Loewen & Isbell, 2017), hearing and producing the same words repetitively (Jung et al., 2017) and increasing task complexity (Gordon, 2021) has a positive impact on the development of L2 lexical stress while engaging in collaborative meaningful tasks. On the other hand, manipulating task design so that learners necessitate the accurate perception and production of the target linguistic features for successful task completion (i.e. *task-essentialness;* Loschky & Bley-Vroman, 1993) has been employed in several tasks to investigate the acquisition of specific L2 segmental features (Mora-Plaza et al., 2024; Sicola, 2008; Solon et al., 2017). Conducting a longitudinal TBPT intervention in a high school context, Mora-Plaza et al.'s (2024) investigation underscored the effectiveness of TBPT in the development of perceptual discrimination (via ABX discrimination), lexical encoding (via lexical decision) and controlled production (via delayed sentence repetition) of two challenging contrasts (/i/-/ɪ/,/æ/-/ʌ/) for Catalan/Spanish EFL learners. The aforementioned studies also showed generalization to untrained items and voices, indicating robustness of L2 speech learning. Last, TBPT studies exploring the role of task complexity on vowel performance (Solon et al., 2017) and learning (Mora-Plaza et al., 2024) showed that L2 segmental features generally became more target-like in cognitively complex than cognitively simple L2 oral tasks.

These studies provide solid evidence that task-based methodologies which promote noticing and attention to differences between L1 and L2 phonological systems while learners focus on communicating meaning may create opportunities for learners to adjust and acquire L2 phonological representations, which may eventually result in global development of L2 pronunciation proficiency (Saito & Plonsky, 2019). Besides Mora-Plaza et al. (2024), research evaluating the effects of TBPT on L2 vowel perceptual learning remains relatively scarce.

2.2. Perceptual-cognitive individual differences in L2 speech learning

This paper also investigates the influence of perceptual and cognitive individual differences on TBPT. Individuals differ in their perceptual ability to encode and reproduce basic sound characteristics (e.g., formants, pitch, and duration; Surprenant & Watson, 2001). As this is the first ability individuals rely on when receiving aural input, the auditory precision theory posits that auditory processing (AP) serves as a bottleneck for various learning behaviors, including emotion, music, and language.

In the context of L1 (Goswami, 2015) and L2 acquisition (Kachlicka et al., 2019), there is growing cross-sectional evidence that those with advanced, robust linguistic proficiency likely have precise AP, and a lack of such precision may impede the rate of language learning, even with similar learning experiences. While individuals may receive identical language instruction for the same duration, their L2 proficiency outcomes often diverge significantly (Saito et al., 2017). This variation is attributed to differences in individuals' abilities to notice, elaborate on, and effectively utilize input opportunities—collectively known as "aptitude."

Studies on the impact of language aptitude on L2 speech have revealed that L2 learners with higher phonemic coding assessment scores tend to exhibit superior pronunciation skills at both segmental and suprasegmental level (Saito, 2019). However, there is limited knowledge regarding the specific components of phonemic coding that contribute to learning L2 segments and the perceptual and cognitive abilities that influence the prosodic aspects of L2 pronunciation development (Saito et al., 2020). Evidence from cross-sectional (Kachlicka et al., 2019) and longitudinal investigations have revealed that learners with more precise AP demonstrate gains when they receive intensive phonetic training (e.g., Saito et al., 2021) or engage in form-focused communicative instruction (Ruan & Saito, 2023). Given the well-established significance of AP in L1 and L2 acquisition, the current study primarily centers on exploring how variations in AP may impact L2 segmental perception.

In addition to AP ability, research findings also highlight the crucial role of working memory (WM) in determining success in achieving high-level L2 proficiency (Linck et al., 2013). In comparison to lower-order auditory perception, WM is regarded as a higher-order cognitive ability that enables learners to retain, elaborate on, and refine linguistic information they receive (Li, 2016).

Interactionist SLA research has delved into the role of WM in facilitating awareness during feedback and interaction-driven L2 learning. For instance, Mackey et al. (2010) examined how WM capacity moderated task-based interaction learning. Their findings revealed that individuals with higher WM capacity were more proficient at recognizing recasts during conversational interaction compared to those with lower WM capacity. Drawing upon the established notion that WM significantly influences how individuals respond to corrective feedback in task-based language interactions (see Révész et al., 2011), it becomes imperative to investigate its role in mediating the effectiveness of TBLT approaches in improving learners' perception and production of L2 individual segments.

Concerning the study of WM in L2 speech studies, Darcy et al. (2015) found an L2 complex WM span measure to be correlated to individual L2-English phonological scores (i.e. a composite of segmental categorization, lexical stress and phonotactics). In the same vein, Christiner and Reiterer (2016) revealed that WM capacity was moderately correlated with L2 English accent imitation and English text reading, suggesting that WM skills contribute significantly to aptitude for imitating and repeating unfamiliar speech in musicians.

To our knowledge, only Mora-Plaza (2023) investigated the potential contributions of WM on L2 phonological performance and development as a result of TBPT. Their results shed light on the mediating role of complex WM on learners' accuracy in the perception of L2 vowel contrasts at pre-lexical and lexical levels. Therefore, this study aims at providing further empirical evidence on the relationship between IDs in AP and WM and L2 perceptual gains obtained through a short TBPT intervention.

3. Current study

The current study is motivated by the relatively limited attention given to L2 speech learning within the framework of TBLT. While previous studies have investigated the effectiveness of TBPT to enhance learners' pronunciation accuracy (Gordon, 2021; Mora-Plaza et al., 2024; Solon et al., 2017), less research has focused on the impact of TBPT on the development of L2 speech perception.

Despite recent theoretical and empirical advances in the intersection of IDs in AP and L2 speech development, the role of AP and WM in various learning conditions and with diverse participant profiles still remains underexplored, highlighting the need for further empirical investigations (Ruan & Saito, 2023; Saito et al., 2021; Shao et al., 2023). As a result, this study aims to address this gap by examining how students with varying AP and WM abilities might differently benefit from a TBPT intervention.

Specifically, the goal was to examine the effectiveness of task design on the perception of two difficult vowel contrasts for Chinese EFL learners. The selected phonological contrasts were/ɛ/-/æ/(e.g., *pen* vs. *pan*) and/ɪ/-/i/(e.g., *ship* vs. *sheep*) which are known to be challenging for Chinese speakers because the absence of the/æ/sound in the Chinese phonetic inventory often leads to nasalization and potential confusion with/ɛ/, and English vowels/i/-/ɪ/tend to be categorized as/i/in Chinese (e.g., Wang & Munro, 2004). In addition, these two vowel contrasts appear in a number of minimal pairs, hence, have a high functional load (Brown, 1988) and may likely compromise listeners' global understanding of L2 speech (Suzukida & Saito, 2022). Therefore, the current study set out to address the following research questions:

1. To what extent does task-based instruction help Chinese learners of English improve their L2 vowel perception (English/ɛ/-/æ/ and/ɪ/-/i/)?
2. To what extent do IDs in domain-general AP and WM mediate the L2 learning outcomes?

As for R1, replicating Mora-Plaza et al.'s (2024) prior work, we predicted that TBPT would help L2 learners significantly improve their vowel performance. Given the demanding nature of the treatment, which pushes L2 learners to practice new target sounds while maintaining their primary focus on task completion, we speculated that the outcomes could exhibit considerable individual variation, with some learners showing larger gains than others. As for R2, following the domain-general paradigm of L2 speech learning, such individual differences could be accounted for by participants' diverse profiles at lower-order perceptual levels (i.e., auditory processing;) and higher-order cognitive levels (Darcy et al., 2015).

4. Method

4.1. Design

This study employed a quasi-experimental pre- and post-test design, involving a total of 70 participants who were randomly assigned to either the experimental group ($n = 50$) or the control group ($n = 20$). A larger number of participants were recruited for the experimental group to ensure robust investigation into the role of individual differences in aptitude in L2 vowel acquisition. The ample sample size was deemed necessary to provide sufficient statistical power for the analyses conducted within the experimental group. Including the control group was essential to account for potential test-retest effects since identical materials were utilized for both pre- and post-tests.

The treatment session was administered as an extracurricular activity outside of regular classroom hours. The participants were volunteer young EFL students (as described in further detail below). During preliminary interviews with potential participants, many expressed a preference for extracurricular activities to be conducted online due to their busy schedules and geographical distance. As a result, we opted to deliver the sessions using a tutoring format via the Zoom platform, allowing for interactive engagement between two students and a teacher. All test and aptitude data were collected through an online psychology experiment platform known as Gorilla (Anwyl-Irvine et al., 2020).

In Week 1, all participants completed the aptitude tests and pre-tests. In Week 2, the experimental group engaged in a 30-min TBPT intervention, while the control group received 30-min lexicogrammar instruction (without any focus on pronunciation accuracy). Subsequently, in Week 3, all participants completed the post-tests (see Fig. 1).

During both the pre- and post-test phases, participants completed a forced-choice identification task. This task aimed to assess the influence of instruction on the participants' perception of English/ɛ/-/æ/and/ɪ/-/i/, embedded in both trained and untrained English minimal pairs.

4.2. Participants

The study involved young adult Chinese EFL learners hailing from Fujian province in southeastern China. These learners were native Mandarin Chinese speakers. To recruit participants, advertisements were circulated among students via online channels and email. The digital flyer specified that we were seeking participants without any perceptual, cognitive, or language disorders. Interested individuals reached out to the researcher, expressed their willingness to participate, and completed the necessary consent forms, also scheduling their pre-test appointments. The study comprised a total of 70 young Chinese EFL learners, consisting of 25 males and 45 females. The average age across the entire sample was 18 years old (*range* = 17–20). All participants had previously undertaken a vocabulary test on the VocabularySize.com website, revealing an average vocabulary size of 6158-word families, ranging from 3800 to 10500. The test development process was detailed in Schmitt et al. (2001).

Despite having substantial prior EFL learning experience ($M > 5$ years), none had experienced full immersion in an English-speaking environment. Given their location in China, opportunities for interaction with native and non-native English speakers were notably scarce. Following the pre-tests, participants were randomly allocated to either the experimental group ($n = 50$) or the control group ($n = 20$). None of the participants reported any hearing impairments.

4.3. Instructional treatment

4.3.1. Instructor

The teacher for both groups of students was a highly proficient English speaker who was a native Mandarin speaker. They had extensive experience teaching English in China for multiple years and held an MA degree in Teaching English to Speakers of Other Languages.

4.3.2. Target of instruction

The study focused on two phonological contrasts, namely, English/ɛ/-/æ/(e.g., *pen* vs. *pan*) and English/ɪ/-/i/(e.g., *ship* vs. *sheep*).

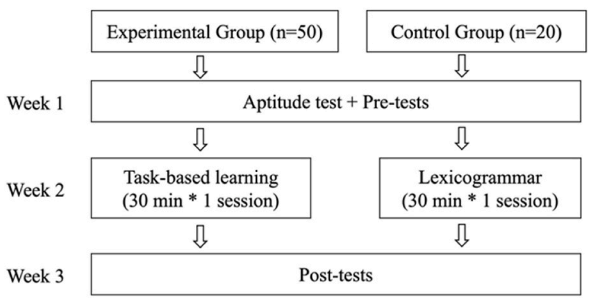

Fig. 1. Summary of research design.

For the purpose of consistency, the instruction referred to Received Pronunciation. According to the functional load principle, which ranks the relative communicative values of phonological contrasts in English,/ɛ/-/æ/ranked 10 and/ɪ/-/i/ranked 8 in a 10-point scale of importance which means these two pairs of phonemes should be prioritized in instructed SLA (Brown, 1988).

English vowels/ɪ/and/i/differ in articulation, duration, and spectral characteristics. Native speakers of English often rely on formant values (Hz) as the primary cue, particularly in General American English (Gottfried et al., 1990). In Received Pronunciation, there is a distinct long-short vowel length contrast (/ɪ/and/i/) (Ladefoged & Johnson, 2014) and they may also consider phonemic length (ms) to distinguish L2-contrastive vowels perceptually (Hillenbrand et al., 2000). L1 Mandarin learners of English often struggle to differentiate between/ɪ/and/i/, tending to categorize both as/i/(Wang & Munro, 2004).

According to the perceptual assimilation model (Best, 1995; Best & Tyler, 2007), L1 Mandarin learners of English tend to perceive English/ɪ/and/i/as equivalents to the Mandarin/i/because Mandarin's vowel inventory only includes one unique sound/i/. In addition, English/æ/may be perceived as highly similar to Chinese/ɛ/due to their shared front and unrounded vowel characteristics, primarily differing in tongue height. According to the speech learning model (Flege, 1995), this perceived similarity can hinder the establishment of a distinct phonetic category for English/æ/, resulting in L2 pronunciation difficulties. For other similar theoretical accounts, see the Linguistic Perception Second Language model (Escudero, 2009) and the Universal Perceptual Model (Georgiou & Dimitriou, 2023).

As observed in Chan's (2010) study, the participants exhibited significant inaccuracies in producing/æ/, with an accuracy rate of only 16.4%. Notably, 97.7% of participants mispronounced/æ/as/ɛ/, while 30.7% of them mispronounced/ɛ/as/æ/. This challenge could be attributed to the absence of the English/æ/sound in Chinese, often leading to nasalization and potential confusion with/-ɛ/(Swan & Smith, 2001).

Last, the mispronunciation of/ɛ/-/æ/and/ɪ/-/i/might affect learners' comprehensibility and intelligibility due to their high communicative value (Brown, 1988; Munro & Derwing, 2006; Suzukida & Saito, 2022), as these L2 vowel contrasts include many minimal pairs in frequently-used word contexts. Therefore, targeting the accurate pronunciation of these two pairs of L2 vowels is crucial for the development of their overall L2 pronunciation proficiency.

4.3.3. Content of treatment

Regarding the experimental group, participants underwent a 30-min task-oriented intervention, focusing on negotiating a summer holiday travel plan. The instructional materials were based on Mora-Plaza (2023), who included various information-gap problem--solving communicative tasks that aimed at enhancing learners' perception and production of English phonemes/i-ɪ/and/æ-ʌ/by Catalan/Spanish EFL learners. These materials were adapted to assess Chinese learners' perception of English/ɛ/-/æ/and/ɪ/-/i/phonemes.

A set of 10 consonant-vowel-consonant (CVC) minimal pairs, which only differed in terms of/ɛ/-/æ/and/ɪ/-/i/sounds, were incorporated into the task (see Table 1). All the target tokens belonged to the first 2,000 most common word families as per the BNC/COCA Word Frequency List (Nation, 2012). The word frequency of the target words was carefully controlled for in order to minimize the influence of word frequency and familiarity on learners' task and test performance.

Following Willis' (1996) task-based learning framework, during the 5-min pre-task phase, the teacher elicited some topic-related words and phrases to introduce the topic of travelling. Then, participants heard one of the words from the visually represented minimal pairs twice and had to select the corresponding image. They listened to the same target words in sentence contexts and were asked to verify their responses. Last, participants completed a listening comprehension task that mirrored the type of conversations learners would need to engage in the main task, and answered some comprehension questions.

Afterwards, they proceeded to the main task (25 min). Tasks were two-way, close and convergent (Pica, 1994) and involved cognitive processes such as reasoning in decision-making. Following Ellis' (2003, 2009) definition, this information-gap task provided

Table 1
Summary of 20 target words (10 trained; 10 untrained) from the perception test.

Trained items	Untrained items
English/ɛ/-/æ/	English/ɛ/-/æ/
• End – and	• Beg – bag
• Bend – band	• Dead – dad
• Guess – gas	• Head – had
• Merry - marry	• Lend – land
• Pen – pan	• Men – man
English/ɪ/-/i/	English/ɪ/-/i/
• Bean - bin	• Beat – bit
• Cheap – chip	• Feel – fill
• Meal – mill	• Seat – sit
• Sheep – ship	• Seek – sick
• Steel – still	• Sleep slip

Note. The lexical items are listed alphabetically in the table. They were presented in random order in the tests. They all fall among the first 2,000 most common word families according to BNC/COCA Word Frequency List (Nation, 2012).

opportunities for using specific L2 phonological features communicatively, mirrored a real-world activity and had a clearly-defined non-linguistic outcome.

In addition, this pronunciation-focused task was manipulated to enhance learners' attention to difficult L2 vowel contrasts (/ɛ/-/æ/ ,/ɪ/-/i/) in order to successfully complete the task while focusing primarily on meaning; in other words, L2 vowel contrasts were made *task-essential* (Loschky & Bley-Vroman, 1993). Target minimal pairs were also typographically enhanced in bold as a technique to enhance further noticing of L2 phonological contrasts. During the main task, participants interacted in pairs while the teacher passively observed their performance. The task required students to engage in a conversation with their partners, share information, and collectively choose three destination preferences based on several individual preferences and conditions. At the end, they were asked to plan the task outcome and orally present their travel plan. The task materials can be found and downloaded in the SLA Speech Tools open-science repository (Mora-Plaza et al., 2022: http://sla-speech-tools.com/).

As for the control group, participants spent a similar amount of time (30-min session) learning English vocabulary and grammar explicitly with the researcher. The materials largely consisted of supplementary exercises related to the content covered in their regular school EFL lessons. These sessions did not involve any oral communication practice or specific phonetic training. The inclusion of the control group aimed to discern any potential test-retest effects, as identical materials were employed for both the pre-test and post-test assessments. Since all participants resided in China and had access solely to grammar-focused classroom English instruction, none of them received any additional pronunciation training involving English/ɛ/-/æ/and/ɪ/-/i/between the pre-test and post-test stages.

4.4. Measure of L2 speech development

In order to assess the influence of task-based instruction on how participants perceived specific L2 phonological contrasts in English (/ɛ/-/æ/,/ɪ/-/i/), a forced-choice identification task was employed at pre- and post-tests.

4.4.1. Materials

The forced-lexical identification task comprised not only 10 trained minimal pairs but also 10 untrained CVC minimal pairs (Table 1). Similar to the 10 target minimal pairs, the 10 untrained minimal pairs also belonged to the first 2,000 word families as per the BNC/COCA Word Frequency List (Nation, 2012). The untrained words were employed to assess the degree to which instructional effects could be generalized to unfamiliar words. A female native speaker of British English was recruited. She was asked to read aloud the 20 trained and untrained minimal pairs and the recordings were used as stimuli in the identification task.

4.4.2. Procedure

After receiving instructions from the researcher, participants used their personal computers equipped with headphones to access the test materials available on the online platform Gorilla (Anwyl-Irvine et al., 2020). A total of 80 stimuli, consisting of 40 trained and untrained target words were presented to them in a randomized sequence. For each auditory stimulus, participants were required to determine the word they heard by selecting one of two options displayed orthographically on the computer screen (see Fig. 2). Their accuracy rate (out of 80 stimuli) was calculated to represent participants' gain scores before and after the treatment.

4.5. Measures of auditory processing

In alignment with previous studies (e.g., Kachlicka et al., 2019), two distinct AXB discrimination tasks were employed to assess participants' auditory sensitivities to particular acoustic dimensions. These cues were chosen as native speakers rely on both cues to perceive English/ɛ/-/æ/and/ɪ/-/i/(Ladefoged & Johnson, 2014). Synthesized, nonverbal stimuli were utilized with simplified acoustic characteristics, making them unrecognizable as speech by individuals with normal hearing.

In each trial, participants encountered three non-verbal sounds and were instructed to identify the one (either the first or third) that

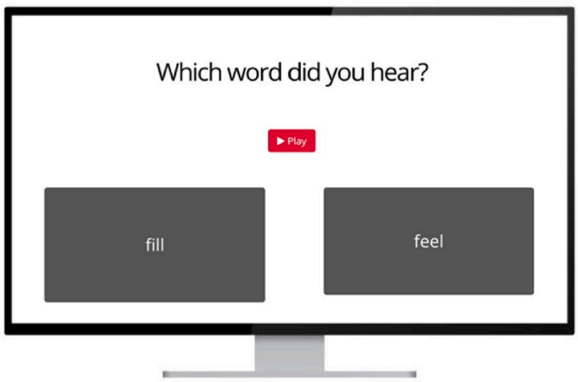

Fig. 2. Screenshot of the forced-choice identification task. Participants were asked to listen to a target stimulus and select which word they heard. No feedback was provided during the task performance.

differed from the other two by selecting the corresponding number ("1" or "3") using a mouse (refer to Fig. 3). Since each sound varied in one acoustic parameter (formants), each task aimed to measure the smallest discernible difference participants could perceive in the target dimension.

Many L2 learners often struggle with acquiring tense/lax vowels due to their tendency to rely heavily on duration cues rather than spectral cues. However, native speakers primarily rely on spectral cues for vowel perception (Hillenbrand et al., 2000). Given this discrepancy, we hypothesized that individuals with more precise AP abilities for spectral information, particularly formants, would exhibit greater improvements in English tense/lax vowel acquisition. Therefore, we utilized the formant discrimination task to assess participants' AP abilities relevant to the instructional target of tense/lax vowel acquisition.

Custom MATLAB scripts were employed to generate one reference stimulus and 100 target stimuli. The stimuli comprised 250-ms-long complex tones with a fundamental frequency (F0) of 330 Hz. These complex tones with an F0 of 100 Hz and consisting of three formants were used (F1 = 500 Hz, F2 = 1500–1700 Hz, F3 = 2500 Hz). While all other acoustic parameters remained constant, F2 was manipulated to vary from trial to trial. The reference stimulus had an F2 set at 1500 Hz, while the F2 of the target stimuli varied from 1502 Hz to 1700 Hz in 2 Hz increments.

Utilizing Levitt's (1971) adaptive up-down procedure, the test initially began at Level 50 (out of 100 levels) and adjusted the difficulty level based on the participants' performance. Difficulty increased by 10 steps after three consecutive correct responses and decreased by 10 steps after a single incorrect response. The step size reduced to five after the first reversal, then to two after the second, and finally to one after the third, continuing until task completion. Participants' final scores were recorded on a 100-point scale, with lower scores indicating more precise AP concerning formant.

As we explain in the results section, we used AP scores (out of 100 points) as both continuous values to index the experimental participants' individual differences and as a categorical factor to group them into Normative (n = 24; Range = 5.80 to 33.83) and Low (n = 26; Range = 36.00 to 70.83). For the latter, we developed an in-house benchmark specific to the formant discrimination abilities among adult Chinese learners of English in a precursor study (Saito and Tierney, forthcoming). We asked a total of 373 Mandarin Chinese learners of English in both China and the UK to take a range of AP tests, including the formant discrimination test. While all participants reported normal hearing, their AP scores for formant discrimination varied widely (M = 31.64, 95% CI [29.05–34.23]). In the context of the formant discrimination test, we classified participants whose raw scores were within the upper range of the 95% confidence interval ($x < 34.23$) as Normative and those beyond this upper level ($x > 34.23$) as Low.

4.6. Measures of working memory

Working memory was assessed using the visual and text-entry digital span tests, which were adapted and customized based on Dean (2020). It was originally designed for diagnosing children's learning difficulties related to verbal WM. During these tests, pre-recorded sequences of numbers were displayed on the screen at a rate of one digit per second, with a brief fixation point in between digits (lasting 100ms). Participants were instructed to recall the numbers by entering them into an input box once each sequence concluded (see Fig. 4). Responses were obligatory, and participants needed to press the "return" key on the keyboard to confirm their response and proceed to the next trial.

The digit span forward test was employed to evaluate participants' phonological short-term memory capacity. In this task, participants were required to recall the numbers in the exact order they were presented. Conversely, the digit span backward test assessed complex WM processing, where participants had to input numbers in the reverse order they were displayed. The digit span forward test included 9 spans, ranging from 2 digits to 10 digits, while the digit span backward test consisted of 8 spans (from 2 digits to 9 digits). Each span comprised two trials, with successful completion of at least one trial leading to the next level (one digit longer than the previous one) until participants were unable to correctly recall either of the two trials of the same length. The WM capacity was determined by the longest spans participants successfully entered without error. The two tests were administered sequentially, with the digit span forward test preceding the digit span backward test. An optional break between the tasks was provided, and this segment

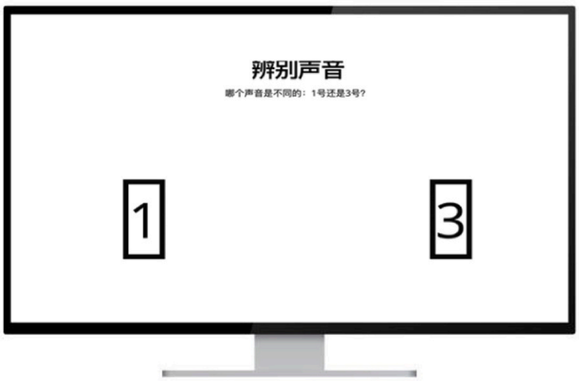

Fig. 3. Screenshot of auditory processing discrimination tasks. Participants listened to a sequence of three sounds and were asked to choose which sound (first or third) differed from the other two sounds. The second sound remained the same throughout the task.

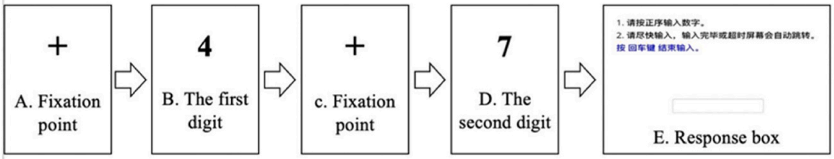

Fig. 4. Sequences of a two-digit span test.

typically took approximately 5–10 min to complete (Dean, 2020).

5. Results

5.1. Overall improvement (pre-to post-tests)

Table 2 and Fig. 5 display the pre-test and post-test correct identification scores (%) obtained from the forced-choice identification task across three conditions: overall (n = 40 items), trained (n = 20 items), and untrained (n = 20 items). The normality of pre-test scores for both the experimental and control groups was evaluated using a Kolmogorov-Smirnov test, which revealed no significant deviation from normal distribution ($p > .05$). To examine potential pre-existing differences in perceptual accuracy of target sounds (English/ɛ/-/æ/and/ɪ/-/i/), independent t-tests were conducted on total pre-test scores. The results indicated no significant between-group differences at the pre-test stage ($t = -.129$, $p = .898$, $d = -.034$), suggesting comparable vowel perception performance before the treatment.

To assess the impact of TBPT on participants' L2 vowel performance across two lexical conditions (trained and untrained), a mixed-effects modeling regression analysis was conducted using the R environment. The results of the initial model (DV ~ Group-*Time*Context + (1|ID)) are summarized in Table 3. The model identified significant interaction effects for Time and Group ($b = -8.9$, $t = -2.58$, $p = .012$), while neither the main effects nor the interaction effects of Context reached statistical significance. Further analysis of multiple comparisons revealed that the experimental group exhibited significant improvement over time ($b = -5.80$, $t = -3.43$, $p < .001$), whereas the control group's improvement did not reach statistical significance ($b = 1.25$, $t = .46$, $p = .648$). The findings suggest that participants who underwent TBPT demonstrated enhanced vowel performance in both trained and untrained lexical conditions, confirming the generalizable effectiveness of TBPT. Conversely, the control group did not exhibit significant improvement, indicating minimal influence from test-retest effects and lexical conditions in the study.

The subsequent objective analyses aimed to explore how the progress of participants in the experimental group might be influenced by their aptitude profiles in AP and WM. Another model was constructed to investigate the interaction effect of time (i.e., TBPT effectiveness) and the aptitude variables (AP, WM)—DV ~ Time*Context + Time*AP + Time*WM + (1|ID). The results of Spearman nonparametric correlation analyses revealed significant medium associations between AP and WM ($r = -.344$, $p = .015$), indicating that while these two scores somewhat overlap, they essentially measure distinct constructs of aptitude. As outlined in Table 4, the main effects of Time remained marginally significant ($b = 17.37$, $p = .055$). Notably, interaction effects of Time were significant for AP ($b = -.23$, $p = .04$) but not for WM ($b = -.20$, $p = .624$). The results suggest that AP mediated the effectiveness of TBPT in the current dataset.

To further explore the relationship between AP and participants' learning progress, those in the experimental group were divided into Normative and Low AP subgroups (as detailed in the Method section). In the precursor project (Saito and Tierney, forthcoming) involving 373 adult Chinese learners of English, we developed the benchmark to determine who could be identified as relatively low auditory precision in the context of the formant discrimination test (i.e., x > 34.23 out of 100 scores). In the current datasets, the Normative group (n = 24) had raw AP scores ranging from 5.80 to 33.83 (M = 19.20, SD = 8.95), while the Low group (n = 26) had raw AP scores ranging from 36.00 to 70.83 (M = 49.28, SD = 9.02). For a similar approach, see Chandrasekaran et al. (2010) and Perrachione et al. (2011).

To explore the significant interaction effects of Time and AP, the following model was constructed to examine how participants with different AP profiles (Normative vs. Low AP) developed their L2 vowel perception scores over time: DV ~ Group_Thresh*Time + (1|ID), where Group_Thresh represented the participants in the Normative AP and Low AP groups. As shown in Table 5, the interaction

Table 2
Summary of L2 vowel identification scores (%).

	Experimental			Control		
	M (%)	SD	CI	M (%)	SD	CI
Overall (pre)	70.7	16.2	[66.1–75.3]	71.2	16.3	[64.0–78.5]
Trained (pre)	72.0	16.1	[67.4–76.7]	70.5	21.5	[60.1–80.3]
Untrained (pre)	69.0	21.8	[63.1–74.9]	72.0	15.8	[65.1–78.9]
Overall (post)	77.1	14.4	[73.0–81.1]	70.3	17.8	[62.2–78.3]
Trained (pre)	78.2	14.3	[74.1–82.3]	71.5	20.3	[62.2–80.8]
Untrained (post)	74.4	19.5	[68.9–79.9]	68.5	18.9	[60.3–76.7]

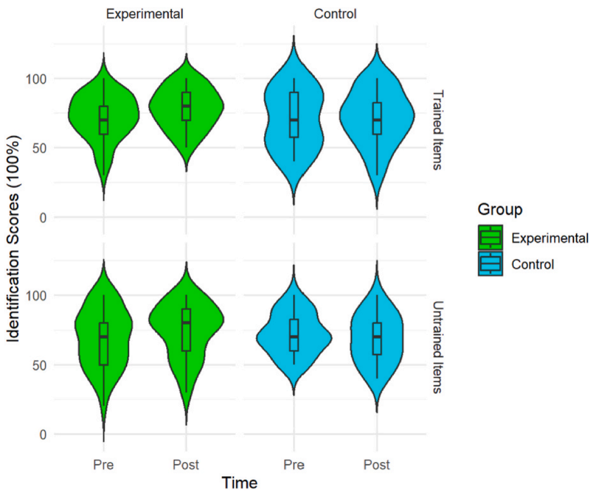

Fig. 5. Distribution of Identification Scores (100%) by Group, Time, and Context.
The Experimental Group showed a significant improvement in identification scores from pre-to post-tests for both trained and untrained lexical items (approximately 10%). In contrast, the Control Group's performance remained consistent throughout the study. The green and cyan violins represent the Experimental and Control groups, respectively, with boxplots indicating the median and interquartile range of the scores. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Summary of model 1 (Group, context, and time).

	b	*SE*	*t*	*p*	Conditional R^2	Marginal R^2
Intercept	68.52	4.20	16.27	<.001	.572	.023
Group	5.93	4.98	1.18	.240		
Context	3.01	3.89	.77	.443		
Time	3.56	2.90	1.20	.233		
Context: Group	.85	4.60	.17	.863		
Time: Group	−8.92	3.44	−2.58	.012[a]		
Context: Time	−.456	5.09	−.88	.380		
Context: Time: Group	3.78	6.02	.61	.541		

Note.
[a] For $p < .05$.

Table 4
Summary of model 2 (Roles of AP and WM).

	b	*SE*	*t*	*p*	Conditional R^2	Marginal R^2
Intercept	63.17	10.90	5.79	<.001	.583	.139
Time	17.37	8.99	1.93	.055		
Context	−3.00	2.44	−1.23	.221		
AP	−.07	.13	−.53	.598		
WM	.70	.49	1.43	.159		
Time: Context	−.80	3.45	−.23	.817		
Time: AP	−.23	.11	−2.13	.035[a]		
Time: WM	−.20	.40	−.49	.624		

Note.
[a] For $p < .05$.

effects reached statistical significance ($b = -4.725$, $p = .013$). Using the emmeans package, a set of follow-up pairwise comparison analyses were performed. Participants with normative AP significantly improved their performance to a greater degree ($b = -8.12$, $t = -3.622$, $p < .001$), while those with low AP showed improvement that did not reach statistical significance ($b = -3.400$, $t = -1.547$, $p = .128$). Additionally, the Normative AP group outperformed the Low AP group at the time of the post-test ($b = 11.07$, $t = 2.598$, $p = .011$), whereas this group difference was not significant prior to the project ($b = 6.340$, $t = 1.489$, $p = .141$). For a visual representation of group performance, please refer to Fig. 6.

Table 5
Summary of Model 3 (Normative vs. Low AP Groups).

	b	*SE*	*t*	*p*	Conditional R^2	Marginal R^2
Intercept	79.883	6.775	11.791	<.001[a]	.759	.115
Group_Thresh	−6.342	4.259	−1.489	.141		
Time	12.85	4.996	4.89	.001[a]		
Group_Thresh: Time	−4.725	3.141	−2.572	.013[a]		

Note.
[a] For $p < .05$.

6. Discussion

6.1. TBLT and L2 vowel learning

The first research question asked whether task-based instruction could help Chinese learners of English improve L2 vowel perception (English/ɛ/-/æ/and/ɪ/-/i/). The task-based intervention had a remarkably positive impact on L2 vowel perception, indicating that meaning-based interaction that induces learners to notice the differences between their interlocutors' productions and their own's may enhance awareness of vowel quality differences between L2 vowel contrasts, leading to perceptual learning. The experimental group exhibited significant improvements of 9.05% overall with medium-to-large effect sizes, comprising gains for trained and untrained words. These findings indicate that even a short 30-min TBPT session may have a significant positive impact on L2 vowel perception, with effects comparable and even superior to prior studies assessing L2 vowel identification and discrimination as a result of implicit high-variability perceptual training (Saito, Hanzawa, et al., 2022), communicative form-focused instruction (Ruan & Saito, 2023) and TBPT (Mora-Plaza et al., 2024).

Furthermore, results revealed that participants who received task-based instruction showed improvement not only in the perceptual identification of trained words (i.e. those included in the task-based intervention) but also in the perceptual identification of untrained words (those only included in the pre- and post-tests). These findings echo prior research showing the potential of pronunciation-focused tasks to promote robustness of L2 perceptual vowel learning (Mora-Plaza et al., 2024); hence, it is a reliable indication that L2 learners were not merely memorizing L2 phonological cues in short-term memory, but they generalized L2 learning to novel lexical contexts, likely enhancing the formation of long-term memory representations (Carlet & Cebrian, 2019).

Taken together, these findings carry important implications for language pedagogy, particularly regarding the use of authentic tasks, which orient learners' attention to challenging L2 phonological features, in meaning-based conversational classes (Gurzynski-Weiss et al., 2017; Mora & Levkina, 2017). However, it is important to emphasize that sustained, long-term progress necessitates the consistent integration of pronunciation-focused tasks in L2 classrooms over an extended period of time (Mora-Plaza, 2023).

6.2. Aptitude and L2 vowel learning

The second research question explored whether individual differences in language aptitude at perceptual (auditory processing) and

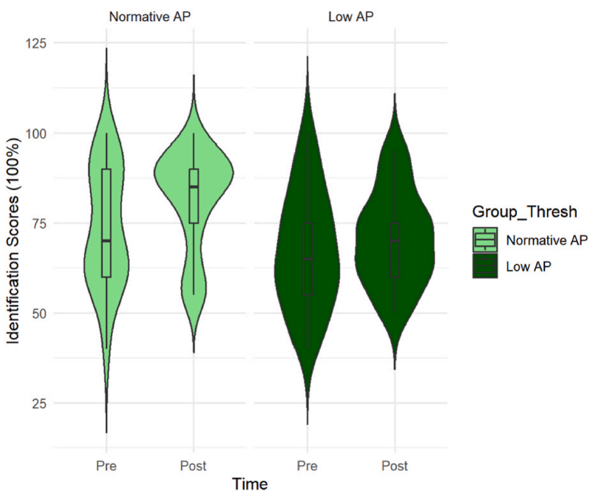

Fig. 6. Distribution of Identification Scores (100%) by Subcategories within the Experimental Group (Normative vs. Low AP). The Normative AP group showed significant improvement in identification scores. In contrast, the Low AP group exhibited less clear improvement. This figure highlights the differential effects of AP abilities on identification performance in the given tasks.

cognitive (working memory) levels mediated L2 learning outcomes. The results indicated a significant association between TBPT effectiveness and participants' AP scores but not their working memory (WM) scores. This finding is consistent with previous research highlighting the crucial role of AP in instructed L2 speech learning (e.g., Ruan & Saito, 2023), while also reflecting the less clear relationship between WM and L2 speech learning observed in other studies (e.g., Mora-Plaza, 2023).

One possible explanation for this asymmetric pattern of the aptitude-learning link is provided by the theory of auditory scaffolding (Conway et al., 2009), which posits a hierarchical relationship between perceptual abilities (AP), cognitive abilities (WM), and language learning abilities. According to this theory, AP serves as the foundational ability upon which language processing relies when learners receive auditory input. Consequently, even if individuals possess strong cognitive abilities at higher levels (WM), any individual differences at the lower perceptual level (AP) can significantly impact other aspects of L2 learning, acting as a bottleneck in speech learning processes. In the context of L2 speech learning, recent evidence suggests that learners' profiles of AP may be more predictive of learning outcomes than their WM profiles (Saito et al., 2022). In contrast, the effects of WM on L2 speech learning have been mixed (e.g., Huensch, 2024, vs. Darcy et al., 2015). However, we also acknowledge that in the field of L1 acquisition, there is evidence indicating that individuals with poor auditory precision can still achieve satisfactory linguistic outcomes, potentially by leveraging other cognitive resources, such as executive functions, to compensate for the lack of precise AP (for a summary, see Rosen, 2003).

Importantly, the follow-up analyses showed some individual differences in TBPT effectiveness between participants with normative/high and low AP abilities. This suggests that when engaging in real-life, task-based interventions, individuals with precise AP abilities may experience rich language acquisition opportunities. Conversely, those with lower AP abilities may struggle to encode, analyze, and proceduralize information from such meaning-oriented discourse without explicit pronunciation instruction. Indeed, for individuals with lower AP abilities, TBPT may pose challenges. These learners may prioritize task completion, focusing on meaning, over refining phonological accuracy, which is a secondary objective (corresponding to the definition of tasks in TBLT and TBPT, see Ellis, 2003). This dual focus could be taxing for those with limited AP abilities, as they may struggle to process and discriminate between subtle phonetic contrasts while engaging in meaning-oriented tasks. Therefore, adapting TBPT approaches to provide additional support and scaffolding for learners with lower AP abilities could be beneficial to enhance their learning outcomes. These findings support the threshold view of aptitude-treatment interaction, suggesting that learners above a certain threshold can benefit equally from intensive L2 speech training, while those below the threshold may not fully capitalize on TBPT. Similar findings (i.e., the disadvantage among lower aptitude learners) were observed especially when they engaged in more real-life L2 speech learning experiences, such as communicative focus on form (Ruan & Saito, 2023), high variability phonetic training (Perrachione et al., 2011), study abroad (Saito and Tierney, forthcoming), and incidental, multimodal learning (Correia et al., 2024; Saito et al., 2022).

The question has now become: How can we help all learners equally make the most of TBPT regardless of their diverse AP profiles? One solution could be the provision of explicit phonetic instruction prior to TBPT (i.e., task supported language learning; Li et al., 2016). Various explicit phonetic instructional methods have been proposed to help learners notice the perceptual and articulatory characteristics of L2 sounds relative to their L1 counterparts (see Saito & Plonsky, 2019). In fact, when learners receive such elaborate phonetic instruction, its effectiveness may be irrelevant to aptitude factors, suggesting that all L2 learners can equally improve their pronunciation proficiency regardless of their aptitude profiles (Suzukida & Saito, 2023). When learners are identified with low AP, engagement with real-life L2 speech learning via TBPT could be demanding. Before such treatment, they can be acquisitionally *ready* if they receive explicit training to help raise their awareness of target sounds via a range of strategies, such as exaggerated teacher talk (Saito & van Poeteren, 2012), acoustically enhanced speech (Iverson et al., 2005), and visualization of articulatory movements (Bryfonski, 2023).

Another solution is to provide focused training on AP, especially for those with relatively low auditory precision, given that even brief training has been proven effective in both L1 acquisition (Merzenich et al., 1996) and L2 acquisition (Saito et al., 2022). For example, if some L2 learners are diagnosed with relatively low AP, they can engage in focused auditory training before progressing to language training. This approach allows them to fully concentrate on developing domain-general abilities to perceive basic acoustic information (e.g., pitch, formant, duration, and amplitude) without the need to attend to domain-specific linguistic aspects (e.g., phonology, vocabulary, and grammar). Once they reach certain thresholds for normative AP (e.g., 34.23 out of 100 for formant discrimination test), they can be guided to proceed to real-life L2 speech training (e.g., TBPT) to maximize the benefits from every input opportunity.

7. Limitations and future directions

While this study has made significant progress in empirically examining aptitude effects on TBPT, it is essential to acknowledge its limitations and chart potential future directions.

One limitation of this study is the relatively small sample size, coupled with the relatively brief intervention period. Future research could consider expanding the participant pool and extending the intervention duration to investigate potential variations in outcomes longitudinally in face-to-face classroom environments.

The exclusive focus on L2 vowels within this study represents another limitation. To offer a more comprehensive view of the aptitude-acquisition link, future research should incorporate a wider array of L2 pronunciation targets, including segmental (i.e. consonants) and suprasegmentals (i.e. lexical stress, rhythm) features of L2 speech.

This research primarily examined the influence of TBPT on L2 speech perception. To obtain a more comprehensive understanding of the pedagogical advantages of TBPT in the development of L2 perception, forthcoming investigations should incorporate a wider range of tasks examining L2 perceptual improvement, namely, categorization, identification and discrimination tasks gauging accuracy and reaction times (Lively et al., 1994).

To provide a more holistic understanding of the effectiveness of TBPT, future research should expand its focus beyond L2 perception. This might entail investigating the effects of TBLT on L2 segmental and suprasegmental features in controlled and spontaneous production tasks. In addition, L2 pronunciation needs to be assessed in terms of acoustic measurements and global assessments of accentedness, comprehensibility and fluency (Saito & Plonsky, 2019) to be able to adequately characterize the impact of task design and task manipulations on L2 pronunciation development at fine-grained and perceptible levels.

To comprehensively gauge the long-term effectiveness of TBPT, future research should contemplate adding a delayed post-test, and collecting teachers and learners' qualitative perceptions of TBPT through questionnaires, semi-structured interviews or stimulated recalls.

Finally, a reviewer suggested that the orthography of words might have influenced the performance of Chinese learners of English. Georgiou (2021) highlighted that L2 speech perception could be affected by participants' reliance on graphene-phoneme correspondence. In the present study, it is likely that the orthography of words might have influenced the performance of Chinese learners of English. Therefore, it would be interesting if future studies could compare two groups of L2 learners with different degrees of graphene-phoneme correspondence in their L1 systems.

8. Conclusion

This study investigated the impact of task-based pronunciation teaching on L2 speech learning, focusing on the perception of English vowel contrasts/ɛ/-/æ/and/ɪ/-/i/by 70 Chinese EFL learners. Our findings revealed that a 30-min session of TBPT resulted in significant improvements in L2 vowel perception, extending beyond the trained lexical items. This highlights the potential of TBPT in enhancing the robustness of L2 vowel perception. Additionally, we found that individual differences at lower-order perception levels could act as a bottleneck for how much learners could benefit from TBPT: AP played a crucial role in the effectiveness of TBPT, with individuals possessing more precise perceptual abilities benefiting more from the training. However, those with less precise perceptual abilities did not show significant improvements. This underscores the importance of considering learners' AP profiles when implementing TBPT. To ensure that all learners can benefit equally from TBPT, it may be necessary to incorporate more explicit phonetic instruction or AP training before they proceed to TBPT.

CRediT authorship contribution statement

Jiying Xu: Writing – original draft, Project administration, Methodology, Investigation, Conceptualization. **Kazuya Saito:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Ingrid Mora-Plaza:** Writing – review & editing, Conceptualization.

References

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.

Barón, J., Celaya, M. L., & Levkina, M. (2020). Learning pragmatics through tasks. When interaction plays a role. *Applied Pragmatics, 21*, 1–25.

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in Infancy Research, 9*, 217, 217.

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 learning & Language teaching* (vol. 17, pp. 13–34). John Benjamins Publishing Company.

Brown, A. (1988). Functional load and the teaching of pronunciation. *Tesol Quarterly, 22*(4), 593.

Bryfonski, L. (2023). Is seeing believing?: The role of ultrasound tongue imaging and oral corrective feedback in L2 pronunciation development. *Journal of Second Language Pronunciation, 9*(1), 103–129.

Carlet, A., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters—in honor of Ocke-Schwen Bohn* (pp. 91–119). Department of English, School of Communication & Culture, Aarhus University.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages* (2nd ed.). Cambridge: Cambridge University Press.

Chan, A. Y. W. (2010). Advanced Cantonese ESL learners' production of English speech sounds: Problems and strategies. *System, 38*(2), 316–328.

Chandrasekaran, B., Sampath, P., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*(1), 456–465.

Christiner, M., & Reiterer, S. (2016). Music, song and speech: A closer look at the interfaces between musicality, singing and individual differences in phonetic language aptitude. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 131–156). John Benjamins Publishing Company.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science, 18*(5), 275–279.

Correia, S., Kachlicka, M., & Saito, K. 2024. Effects of phonetic training and cognitive aptitude on the perception and production of non-native speech contrasts. Studies in Second Language Acquisition..

Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences, 40*, 63–72.

Darcy, I., & Rocca, B. (2023). Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. *Journal of Second Language Pronunciation, 8*(3), 328–362.

De Ridder, I., Vangehuchten, L., & Gomez, M. S. (2007). Enhancing automaticity through task-based language learning. *Applied Linguistics, 28*(2), 309–315.

Dean, P. (2020). Digit span task (visual & text entry). *Gorilla*.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings: Task-based language teaching. *International Journal of Applied Linguistics, 19*(3), 221–246.

Ellis, R., Baskturkmen, H., & Loewen, S. (2001). Learning uptake in communicative ESL lessons. *Language Learning, 51*, 281–318.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
Escudero, P. (2009). *Linguistic perception of "similar" L2 sounds*. Mouton de Gruyter.
Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–277). Timonium, MD: York Press.
Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325–353.
Georgiou, G. P., & Dimitriou, D. (2023). Perception of Dutch vowels by Cypriot Greek listeners: To what extent can listeners' patterns be predicted by acoustic and perceptual similarity? *Attention, Perception, & Psychophysics, 85*(7), 2459–2474.
Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics, 36*, 117–135.
Gordon, J. (2021). Pronunciation and task-based instruction: Effects of a classroom intervention. *RELC Journal, 52*(1), 94–109.
Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience, 16*(1), 43–54.
Gottfried, T. L., Miller, J. L., & Payton, P. E. (1990). Effect of speaking rate on the perception of vowels. *Phonetica, 47*(3–4), 155–172.
Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition, 36*, 1–37.
Gurzynski-Weiss, L., Long, A., & Solon, M. (Eds.). (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? [Special issue]. *Studies in Second Language Acquisition, 39*(2).
Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013–3022.
Huensch, A. (2024). Clarifying the role of inhibitory control in L2 phonological processing: A preregistered, close replication of Darcy et al. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263124000238, 2016.
Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress patterns through auditory priming tasks. *Studies in Second Language Acquisition, 39*(2), 319–346.
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.
Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage learning.
Levis, J. M. (2022). Teaching pronunciation: Truths and lies. In C. Bardel, C. Hedman, K. Rejman, & E. Zetterholm (Eds.), *Exploring Language education: Global and local perspectives* (pp. 39–72). Stockholm: Stockholm University Press.
Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467–477.
Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(4), 801–842.
Li, S., Ellis, R., & Zhu, Y. (2016). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics, 36*, 205–229.
Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency: High-level aptitude. *Language Learning, 63*(3), 530–566.
Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 2076–2087.
Loewen, S., & Isbell, D. R. (2017). Pronunciation in face-to-face and audio-only synchronous computer-mediated learner interactions. *Studies in Second Language Acquisition, 39*(2), 225–256.
Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. Hyltenstam, & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Clevedon: Multilingual Matters.
Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley.
Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. *Tasks and Language Learning, 1*, 123–167.
Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning, 60*(3), 501–533.
Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271*(5245), 77–81.
Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition, 39*(2), 381–399.
Mora-Plaza, I. (2023). Task-based pronunciation teaching and learning of L2 vowels in EFL learners: Task complexity effects. *Doctoral Thesis)*. http://hdl.handle.net/10803/688956.
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*. https://doi.org/10.1017/S0272263124000470
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.
Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*(4), 520–531.
Nation, P. (2012). *The BNC/COCA word family lists*. Victoria University of Wellington.
Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America, 130*(1), 461–472.
Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning, 44*(3), 493–527.
Révész, A., & Han, Z. (2006). Task content familiarity, task type and efficacy of recasts. *Language Awareness, 15*, 160–179.
Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and second language development. In P. Robinson (Ed.), *Researching task complexity: Task demands, task-based language learning and performance* (pp. 203–238). Amsterdam, the Netherlands: Benjamins.
Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–37). John Benjamins.
Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics, 31*(3–4), 509–527.
Ruan, Y., & Saito, K. (2023). Less precise auditory processing limits instructed L2 speech learning: Communicative focus on phonetic form revisited. *System, 114*, Article 103020.
Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English/ɹ/pronunciation attainment in classroom settings. *Applied PsychoLinguistics, 40*(1), 183–204.
Saito, K., Cui, H., Suzukida, Y., Dardon, D. E., Suzuki, Y., Jeong, H., … Tierney, A. (2022). Does domain-general auditory processing uniquely explain the outcomes of second language speech acquisition, even once cognitive and demographic variables are accounted for? *Bilingualism: Language and Cognition, 25*(5), 856–868.
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., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning, 62*, 595–633.
Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., & Tierney, A. (2022). Training auditory processing promotes second language speech acquisition. *Journal of Experimental Psychology: Human Perception and Performance, 48*(12), 1410.
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., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied PsychoLinguistics, 41*(5), 1083–1112.
Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning, 71*(3), 669–715.
Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*(4), 439–462.
Saito, K., & Tierney, A. (forthcoming). Roles of domain-general auditory processing in L2 speech learning revisited: What degree of precision makes a difference?.
Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–88.
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.
Sicola, L. (2008). *"No, they won't 'just sound like each other'": NNS-NNS negotiated interaction and attention to phonological form on targeted L2 pronunciation tasks (unpublished doctoral dissertation)*. Philadelphia, PA: University of Pennsylvania.
Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics, 18*, 268–286.
Skehan, P. (2019). Language aptitude implicates Language and cognitive skills 1. In Z. Edward Wen, P. Skehan, A. Biedroń, S. Li, & R. L. Sparks (Eds.), *Language aptitude* (1st ed., pp. 56–77). Routledge.
Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition, 39*(2), 347–380.
Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System, 106*, Article 102754.
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.
Swain, M. (1985). Communicative competence Some roles of comprehensible input and comprehensible output in its development. In S. Gassand, & C. Madden (Eds.), *Input and Second Language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
Swan, M., & Smith, B. (Eds.). (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge University Press.
Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System, 32*(4), 539–552.
Willis, J. (1996). *A framework for task-based learning*. Harlow: Longman.