Brain & Language 265 (2025) 105568

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Daytime naps consolidate Cantonese tone learning through promoting cross-talker perception: The role of prior knowledge

Quentin Zhen Qin [a,*], Ruofan Wu [a,b], Caicai Zhang [b]

[a] Speech, Learning, and the Brain (SLaB) Lab, Division of Humanities, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
[b] Neurocognition of Language, Music and Learning (NLML) Lab, Research Centre for Language, Cognition and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Hong Kong

ARTICLE INFO

Keywords:
Lexical tones
Memory consolidation
Perceptual learning
Prior knowledge
Cross-talker perception
Daytime naps
EEG

ABSTRACT

This study investigates whether daytime naps facilitate perceptual learning of Cantonese tones and how prior knowledge mediates the consolidation effect. Ninety Mandarin native speakers were pseudo-randomly assigned to either a nap group, who napped for 1.5 h with brain activities recorded, or the non-nap group, who rested for 1.5 h. They were trained with Cantonese contour-level tonal contrasts and level-level tonal contrasts, followed by a tone identification task (trained talker) before the nap manipulation, and were re-tested (trained and novel talkers) after the nap. The results showed that naps facilitated Cantonese tone learning, with the nap group outperforming the non-nap group in the cross-talker perception. The cross-talker perception effect was specific to contour-level tonal contrasts (consistent with prior knowledge) and was associated with increased sleep spindles and slow oscillations. The findings suggest that prior knowledge plays an important role in prioritizing contour-level tonal contrasts for memory consolidation.

1. Introduction

Adult "late" second language (L2) learners often find it difficult to learn a new language (e.g., non-native speech) especially when the L2 is learned later in life (Myers, 2014; Myers et al., 2017). Learning (non-native) languages requires the integration of newly-learned linguistic knowledge into long-term memory networks, and sleep has been found to facilitate different aspects of language learning through *memory consolidation* (see a special issue by Rasch, 2017). Recent research has provided converging evidence that sleep (where memory consolidation takes place), immediately following training (where encoding takes place) induces an offline improvement and supports the learning of various linguistic aspects, such as speech sounds, words, and grammar by adult L2 learners (Batterink & Paller, 2017; Davis et al., 2008; Fenn et al., 2003; Weighall et al., 2017). Notably, a daytime nap (henceforth, nap) also offers consolidation advantages in stabilizing and reinforcing declarative memories of newly-acquired knowledge, such as novel words (see Farhadian, Khazaie, Nami, & Khazaie, 2021 for a review). The current study adopted a nap design (nap group vs. non-nap group) to investigate whether nap-mediated consolidation benefits perceptual learning of novel words contrasting in *lexical tones* (i.e., pitch variation signaling word identity), by focusing on the effect of *prior knowledge* (i.e., information that relates to existing knowledge) on the early formation of tonal representations.

1.1. Prior knowledge in memory consolidation

Current memory consolidation theory assumes that learning new information involves transforming from fragile episodic memories to stabilized long-term declarative memories (i.e., the information that can be explicitly recalled) so that newly-formed categories (e.g., robust linguistic representations) can emerge (see the complementary learning systems in McClelland, 2013). For example, the Active System Consolidation theory assumes a mechanism in which new (declarative) memories are spontaneously reactivated during non-rapid eye movement (NREM) sleep and the temporary representations (encoded/learned in hippocampal networks) are then redistributed into long-term knowledge (stored in neocortical networks) through sleep-mediated consolidation (Rasch, 2017; Schreiner & Rasch, 2017). Accumulating evidence also has demonstrated that a daytime nap, provided that it contains enough NREM sleep, offers a comparable advantage of declarative memory consolidation (henceforth, memory consolidation) as nocturnal sleep (e.g., Alger et al., 2010; Lahl et al., 2008). However, what remains less clear is why certain linguistic information is often selectively

* Corresponding author at: The Division of Humanities, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.
E-mail addresses: hmzqin@ust.hk (Q.Z. Qin), ruofan-ann.wu@connect.polyu.hk (R. Wu), caicai.zhang@polyu.edu.hk (C. Zhang).

https://doi.org/10.1016/j.bandl.2025.105568
Received 11 November 2024; Received in revised form 2 March 2025; Accepted 4 March 2025
Available online 13 March 2025

consolidated during naps (see Cordi & Rasch, 2021 for a review on memory consolidation of declarative knowledge). That is, some information may be prioritized for nap-mediated consolidation (e.g., Studte et al., 2017).

Numerous studies have shown an effect of prior knowledge (i.e., schema) on memory consolidation, suggesting that new information that is consistent with prior knowledge is consolidated more rapidly than that which is not (McClelland, 2013; Tse et al., 2007, 2011). For instance, Hennies et al. (2016) tested the effect of prior knowledge on sleep-mediated consolidation of new facts. Participants first established a schema of knowledge by learning a set of biological facts (e.g., information about cells) and then learned new biological facts, which were either related to prior knowledge or completely unrelated (e.g., information about arthropods). The results showed the benefits of prior knowledge, with schema-related memories recalled better than non-schema memories after sleep. Lewis and Durrant (2011) thus proposed a concept of "information overlap to abstract" (iota), which is aligned with the active system consolidation theory, to explain the (active) integration of newly-learned knowledge into the existing framework/structure of knowledge through memory consolidation and the role of prior knowledge (i.e., schema effect) during the integration process. The iota explanation is about selective memory strengthening by information overlap, specifically stating that new and old memories are replayed simultaneously during sleep (e.g., slow-wave sleep), with information shared across new and old memories more strongly activated than that from only a single source.

In line with the studies above, language learning studies have provided supportive evidence for the beneficial effect of prior linguistic knowledge (e.g., long-term knowledge of familiar sounds/words in L1) on nap-mediated memory consolidation in L2 word learning and learning of morphological-phonological regularities (Cordi et al., 2023; Zion et al., 2019). Recent studies provided further evidence of memory consolidation in phonological learning (novel spoken words; Havas et al., 2018) and speech perceptual learning (Hindi dental/retroflex consonants; Earle & Myers, 2015a). Compared with other linguistic domains (e.g., words), speech perceptual learning is particularly challenging because listeners need to abstract sound categories from speech signals that are often acoustically variable and multidimensional (Myers, 2014; Myers et al., 2017). This learning process is even more challenging for non-native learners whose prior linguistic knowledge may interfere with L2 speech perceptual learning (Best, 1995). Thus, it is theoretically important to investigate whether a nap plays a similar/different role in consolidating non-native speech (e.g., tone) learning and enabling listeners to generalize to new tokens (produced by novel talkers). To deepen our understanding of how robust tonal representations emerge in *early* stages considering the well-documented difficulty in learning L2 tones (see Pelzl, 2019 for a review), this study examines the role of prior knowledge in the nap-mediated consolidation of lexical tones (also see Chapelle et al., 2022).

1.2. Memory consolidation of lexical tones

Lexical tones, signaling the identity of each syllable and/or word in Chinese languages, reside at the interface between speech and lexical-semantic information (Yip, 2002). Lexical tones are contrastively used to distinguish words in Chinese languages. Segmentally identical words that contain different tones, for example, in Cantonese (/si 55/ 'silk' (Tone 1 (T1)), /si 25/ 'history' (T2), /si 33/ 'to try' (T3), /si 21/ 'time' (T4), /si 23/ 'city' (T5), and /si 22/ 'matter' (T6)) differ in meaning (Matthews & Yip, 2011). Given a high degree of homophony, lexical tones are weighed significantly in Cantonese word recognition. Understanding whether, and if so how, lexical tones are consolidated under the influence of prior knowledge (e.g., cues used to encode tonal contrasts in a native language, L1) is important from a learning perspective (Pelzl, 2019).

First, it is well established that an individual's prior tonal knowledge, that is, cues used to encode tonal contrasts in L1, has an impact on how they learn different types of lexical tones (Francis et al., 2008; Gandour, 1983; Jongman et al., 2017). While pitch is a primary cue that contrasts tones, different pitch cues are proposed to contrast different types of tonal contrasts; both need to be learned: pitch height (average height), which is used to distinguish higher- and lower- *level tones*, and pitch contour (tone shape), which is used to contrast falling and rising *contour tones*. Depending on L1 experience (e.g., contour-tone system), L2 learners may learn certain tones well but learn others poorly, at least at an early learning stage (Francis et al., 2008; Hao, 2012). While the Cantonese tonal system is balanced in contour tones (contrasting in pitch contour) and level tones (contrasting in pitch height), Mandarin listeners encode pitch (contour) cues only in their contour-tone system (/pa 55/ 'eight' T1, /pa 35/ 'to pull out' T2, /pa 214/ 'to hold' T3, /pa 51/ 'father' T4) (Yip, 2002). A previous work (Qin & Jongman, 2016) found that Mandarin-speaking participants showed a different pattern in distinguishing Cantonese contour vs. level tones. Mandarin listeners found it relatively easy to differentiate contour-level tones (a low-rising tone T5, /23/ vs. a low-level tone T6, /33/; see Fig. 2 below) that were primarily distinguished by pitch contour differences. However, they had difficulties distinguishing Cantonese level-level tones (a mid-level tone T3, /33/ vs. a low-level tone T6, /22/; see Fig. 2 below), which are mainly cued by pitch height differences, potentially due to the lack of level-level tonal contrast in their L1 tone inventory (also see Jongman et al., 2017). In contrast, the English-speaking participants, who did not consistently use pitch cues to encode lexical meanings, did not show such a bias.

Since memory consolidation is a necessary step for successful tone learning (i.e., forming robust tonal representations), learners' prior linguistic knowledge (i.e., specific pitch cues) may influence the consolidation of lexical tones (Cordi et al., 2023; Zion et al., 2019). Cantonese has a complex tonal system with contour tones and level tones, making it a good case for testing the effect of prior knowledge on the consolidation of different types of lexical tones. A pilot study of tone consolidation examined whether prior tonal knowledge (of Mandarin) influences the consolidation of non-native Cantonese tonal contrasts by Mandarin-speaking participants (Qin et al., preprint; also see Qin, Jin, et al., 2022). The study adopted an overnight design, with the participants trained/tested in the evening on two non-native contrasts of contour tones and level tones and re-tested after a 12-hour overnight interval. The results showed that the Mandarin-speaking participants consolidated non-native contour tones to a larger degree than non-native level tones after the overnight sleep, possibly attributed to their familiarity with using pitch contour to encode tones in their L1. However, the results might have been driven by circadian rhythm differences of tests (evening) and re-tests (morning), because the study did not have a control group that was tested at the same time of day but did not go through memory consolidation, that is, participants who stayed awake across the full night. Note that sleep deprivation would introduce other confounding influences (e.g., alertness differences) while controlling for circadian rhythm effects.

One way to address the limitation is to adopt a nap design (participants randomly assigned to a nap or non-nap condition) which benefits from control of time-of-day effects and being able to keep interference to an absolute minimum (e.g., Heim et al., 2017). Therefore, this study adopted a nap design to investigate the effect of prior knowledge (e.g., pitch cues) on nap-mediated memory consolidation of Cantonese contour and level tones.

In addition to different cues signaling tonal categories, lexical tones have multiple sources of variations that are related to different talkers, genders, and tonal contexts (e.g., C. Zhang & Chen, 2016). The variability of lexical tones poses persistent difficulty even for advanced L2 learners, especially when their L1 (e.g., English) is not contrastive in tones (Pelzl, 2019). Such variations place a demand on L2 learners to extract abstract representations from tonal exemplars across different talkers to accommodate talker-specific tonal variations (K. Zhang et al., 2018), to distinguish different tone categories successfully. Previous research has provided evidence supporting that (post-training) overnight sleep might be beneficial for the consolidation of newly-learned tonal contrasts by promoting the *generalization* of non-native tones to a novel talker (Earle & Myers, 2015a; Qin, Gong, et al., 2021; Qin & Zhang, 2019). For instance, Qin & Zhang (2019) tested the effect of overnight consolidation on talker-independent learning of non-native Cantonese tones. Young adult L1-Mandarin speakers were divided into two groups – a morning group and an evening group – and were trained on the identification of three Cantonese (level) tones. Training in tone identification was conducted on a trained talker, and assessment was conducted on both the trained talker and a novel talker. Results showed that the morning group, without the benefit of immediate sleep, showed declining performance in the tone identification tasks, while the evening group showed an increased identification performance that was maintained over the next day. Importantly, this group difference was found for both the trained and novel talkers, suggesting sleep-mediated (talker-independent) generalization of newly-learned tonal categories. Given the findings about the beneficial effect of overnight consolidation on talker generalization, the current study also investigates whether nap-mediated memory consolidation promotes the generalization of Cantonese tones to a novel talker (i.e., cross-talker perception).

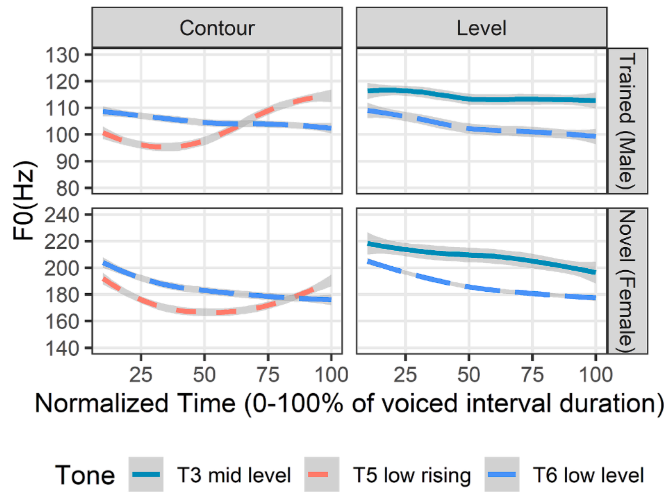

Fig. 2. Time-normalized F0 tracks of Cantonese contour tone and level tone contrasts by a male and a female speaker. Two tones in contour contrast differ in pitch contour, with T5 (red) being a rising tone while T6 (blue) being a level tone. Meanwhile, T3 (green) and T6 in the level tone contrast differ primarily in pitch height.

1.3. Sleep-EEG signaling memory consolidation of language learning

Sleep-EEG activities have been implicated to be associated with language learning (see a special issue by Rasch, 2017), that is, a process of integrating new memories (e.g., novel words) into an existing language structure (Tamminen et al., 2010, 2013). The specific components of sleep architecture, that is, sleep spindles and slow oscillations (see Fig. 1 for a canonical diagram illustrating these two components), are found to align with hippocampal ripples temporally, and closely implicate the hippocampal-neocortical dialogue for consolidation of declarative (semantic) memory (Rasch & Born, 2013). As illustrated in Fig. 1, sleep spindles are defined as sinusoidal and spindle-like brain waves ranging between 11–16 Hz frequency and 0.5–2 s duration, usually maximal in amplitude in the central part (Berry et al., 2017). Some studies further found that fast sleep spindles (13–15 Hz) preferentially play a key role in memory consolidation of language learning, possibly by enforcing slow oscillation-spindle cycles (Mölle et al., 2011; Friedrich et al., 2020; Xia et al., 2023). Slow oscillations, also referred to as slow-wave activity by previous studies that examined activity within a particular frequency band such as delta activity (Simon et al., 2017), were defined as brain waves that have a frequency between 0.5–2 Hz and peak-to-peak amplitude larger than 75 μV and are typically measured over the frontal regions (Berry et al., 2017).

Sleep spindles and slow oscillations were found to play an important role in integrating meaning of novel spoken words into existing semantic structures. For instance, Tamminen (2013) found that the sleep-EEG components mediated the integration of novel spoken words with different semantic neighborhoods into pre-existing semantic knowledge. Their findings showed an advantage of integrating a spare semantic neighborhood over a dense semantic neighborhood, which was more likely to include new knowledge inconsistent with prior knowledge. Increased slow-wave activity (left slow oscillations, 0.5–1 Hz) and sleep spindles (fast sleep spindles, 13–15 Hz) are associated with a rapid integration of words of sparse semantic neighborhood. Likewise, Hennies et al., (2016) showed the beneficial effect of prior knowledge in their behavioral results and conducted analyses on fast sleep spindles (13–15 Hz; stage N2). The results showed an association between spindle density (number of sleep spindles per minute) and augmented overnight consolidation of memories consistent with prior knowledge (but not those inconsistent with prior knowledge), with increases in spindle density for better learning of novel words (also see Studte et al., 2017). Prior studies have examined nap-mediated consolidation of lexical tones and found a beneficial effect of naps on tone learning (e.g., Chapelle et al., 2022). However, a fine-grained analysis of individual differences in sleep spindles and slow oscillations is necessary to clarify which components of sleep architecture are critical to memory consolidation of lexical tones, and specifically, whether they mediate the potential effect of prior tonal knowledge during the perceptual tone-learning process.

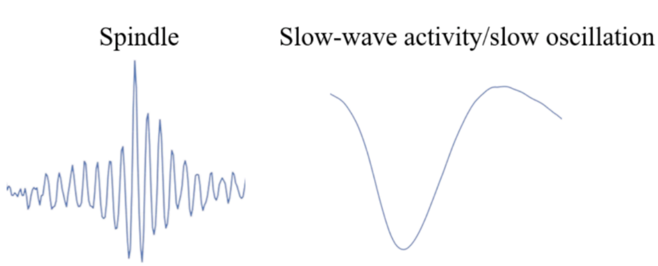

Fig. 1. An illustration of sleep spindle (left) and slow oscillation (right). They were plotted based on samples selected from our EEG data.

1.4. The present study

To examine the effect of prior knowledge on nap-mediated memory consolidation and cross-talker perception during tone learning, the present study adopted a nap design to investigate whether naps help Mandarin speakers consolidate Cantonese tones under the influence of their prior tonal knowledge and through the promotion of cross-talker perception, with specific research objectives and hypotheses stated below.

First, the present study primarily investigates whether Mandarin listeners' prior knowledge of using pitch contour to encode L1 tones influences their nap-mediated consolidation of Cantonese tones. The experiment adopted the nap design (nap vs. non-nap group) across multiple sessions (pretests, training, and posttests) and encompassed two sets of tonal contrasts (Cantonese contour tones or level tones) conducted on Mandarin-speaking L2 learners. Given previous findings showing the effect of prior knowledge on memory consolidation (e.g., Hennies et al., 2016) and the consolidation effects for tone learning (Chapelle et al., 2022), it was hypothesized that prior knowledge of lexical tones would influence Mandarin listeners' memory consolidation of Cantonese tones. Specifically, contour tones (that are consistent with prior knowledge) were hypothesized to be prioritized by showing a greater improvement (or maintenance) than level tones after naps by the nap participants, while the bias would be reduced or absent among the control participants without taking naps.

Given the variability of lexical tones across talkers, the current study also investigates whether naps help Mandarin speakers consolidate Cantonese tones by transferring the tonal knowledge from a learned talker to a novel talker (i.e., cross-talker perception). The participants were tested using tone stimuli produced by a familiar-trained (male) talker as well as those produced by a novel (female) talker after the nap manipulation. If there is an effect of nap-mediated consolidation on cross-talker differences in tone learning (see Qin & Zhang, 2019), the participants who napped were hypothesized to show better cross-talker perception (e.g., smaller differences in accuracy for tone stimuli produced by the trained and novel talkers) than the participants who did not nap. The effect would be more pronounced for contour tones than level tones due to the potential effect of prior knowledge on memory consolidation.

Lastly, the current study investigates whether sleep spindles and slow oscillations predict the effect of prior knowledge on memory consolidation or cross-talker differences in perceiving Cantonese tones. To measure sleep-related brain activities, the nap participants who napped for 1.5 h in the manipulation session had their individual brain EEG activities recorded and analyzed. Given previous findings demonstrating the positive effect of sleep spindles and slow oscillations on the integration of new linguistic information into the existing knowledge structure (Hennies et al., 2016; Tamminen et al., 2013), it was hypothesized that individual learners with higher density in sleep spindles and slow oscillations would show better nap-mediated consolidation or better cross-talker perception, preferentially for contour tones, than their counterparts with lower density in sleep spindles and slow oscillations.

2. Methods

2.1. Participants

Ninety participants (67 females; age = 23.32; age range: 18–30 years) were recruited from universities in Hong Kong via on-campus and social platform advertisements (see a power analysis for sample size justification in Supplementary Materials A). All the participants met the following four criteria: 1) they were all native to Mandarin and did not know any Southern Chinese dialect varieties (e.g., Southern Min, Shanghainese), which may contain level-level tonal contrasts; 2) they had minimal exposure to Cantonese (i.e., their residence in Hong Kong[1] was no more than one year); (3) they had less than four years of any kind of professional music training, such as vocal training and musical instrument training (following the criterion of Li & DeKeyser, 2017); (4) they did not exhibit any sleep disorders and maintained good sleep quality which was indicated by a Pittsburgh Sleep Quality Index (PSQI) score equal to or less than 7 (Buysse et al., 1989). None of the participants reported any diagnosed hearing impairment or neurocognitive disorders. Given the task design of the present study (i.e., recognizing colored pictures), only the participants without color blindness were recruited for the experiment.

The ninety participants were pseudo-randomly assigned into a nap group (N = 50, 38 females; age range: 18–30 years) and a non-nap group (N = 40, 29 females; age range: 18–28 years). The grouping was partly based on their pretest performance, such as pitch processing abilities, musical aptitude, and cognitive abilities given that previous studies have shown that pre-training performance may influence participants' learning and consolidation processes (Perrachione et al., 2011; Qin, Zhang, et al., 2021). Participants' demographic details and pretest performance in measuring pitch processing abilities, musical aptitude, and memory capacity are reported in Table 1. The Bayesian factor tests (BF_{10}[2]) provided evidence for the null hypothesis that the two groups were not different in their demographic variables or pretest performance. The participants' habitual nap routines were also considered in the grouping. While twenty-two participants self-reported being habitual nappers, the number of habitual nappers was roughly balanced between the two groups (12 habitual nappers in the nap group; 10 habitual nappers in the non-nap group). Most participants in each group (78 % for the nap group and 50 % for the non-nap group) self-reported as the neutral chronotype based on a Morningness-Eveningness Questionnaire (MEQ; Loureiro & Garcia-Marques, 2015).

Table 1
Demographics and pretest performance of the nap and non-nap groups.

	Nap	Non-nap	BF_{10}
No. of participants	50 (38F, 12 M)	40 (29F, 11 M)	/
Age (year)	23.36 year-old (2.55)	23.28 year-old (2.24)	0.22
Length of residence in HK (month)	6.18 months (3.04)	6.98 months (3.48)	0.40
Pretests			
Pitch threshold_Discrete (semitone)	0.71 semitones (1.36)	1.31 semitones (2.24)	0.63
Pitch threshold_Glide (semitone)	0.51 semitones (0.77)	0.62 semitones (0.76)	0.27
Discrimination_Contour (%)	95 % (9)	93 % (11)	0.35
Discrimination_Level (%)	74 % (26)	73 % (24)	0.23
MBEA pitch (%)	86 % (8)	86 % (9)	0.24
Pitch memory span (no. of tones)	6.10 tones (1.42)	6.44 tones (1.96)	0.33
Digit memory span (no. of digits)	10.06 digits (1.83)	9.91 digits (2.15)	0.23

Note. The mean scores (with the standard deviation in parentheses) of participants, with the Bayesian factor ($BF_{10} < 1$) showing no difference between the two groups, are reported for each pretest (see details of pretests in Supplementary Materials A).

The study protocol was approved by the Human Subjects Ethics Subcommittee of the Hong Kong University of Science and Technology and the Hong Kong Polytechnic University. Prior to joining the study, all participants provided written informed consent, and they received monetary compensation for their participation.

2.2. Stimuli

The auditory stimuli consisted of two types of tonal contrasts in Cantonese: a contour-level tone contrast (low-rising contour tone T5 /23/ vs. low-level tone T6 /22/), henceforth, a *contour tone contrast*, carried by two base syllables /ji/ and /fɐn/; and a level-level tone contrast (mid-level tone T3 /33/ vs. low-level tone T6 /22/), henceforth, a *level tone contrast*, carried by two base syllables /jɐu/ and /fu/. Syllable familiarity was balanced between the tonal contrasts. While Cantonese syllables of /ji/ and /fu/ are found in Mandarin, Cantonese syllables of /fɐn/ and /jɐu/ are novel to Mandarin speakers because of vowel differences. The other two Cantonese tones, T1 (/55/, a high-level tone) and T2 (/25/, a high-rising tone), were not included because they are acoustically similar to the Mandarin T1 (/55/, a high-level tone) and T2

[1] We used their self-reported residence length in Hong Kong as a rough index of Cantonese exposure. Future studies could use a questionnaire to collect more detailed information on the participants' exposure or familiarity with spoken Cantonese.

[2] The Bayes factor (BF_{10}) is a continuous, non-binary number that indicates the relative plausibility of H1 (i.e., two groups are different) over H0 (i.e., two groups are comparable). A $BF_{10} < 1$ generally indicates evidence towards H0, although some researchers also suggest more fine-grained subdivisions (Schönbrodt & Wagenmakers, 2018).

(/35/, a high-rising tone), respectively. The eight tone-words had a similar syllable frequency and pronounceability to Mandarin speakers based on a norming study (Francis et al., 2008).

For each word, audio recordings were recorded from a male and a female native speaker of Hong Kong Cantonese who did not merge tones in their production. Recordings were made in a soundproof room on a PC workstation with an Azden ECZ990 microphone (Azden, Mt. Arlington, NJ). All the recordings were made at a sampling rate of 44,100 Hz with 16 bits per sample. During the recording, target words were embedded in a carrier phrase "呢個係_lei55 kɔ33 hɐi22 [target word]" ('this is [target word]'), and the speakers were asked to repeat the carrier phrase (and the target words) six times in total. These tokens were then compared with each other, and five tokens were selected for each word and each gender by the first author based on pronunciation accuracy and intelligibility. All stimuli were normalized to 500 ms in duration and 70 dB in intensity using Praat (Boersma & Weenink, 2018). There were eight words (2 contrasts × 2 tones × 2 syllables), which were produced by one male talker and one female talker.

A visualization of the fundamental frequency (F0) tracks of the two tonal contrasts is shown in Fig. 2. Tonal contours were measured using ten measurement points in Hertz (Hz). Regarding the tone stimuli produced by the two talkers, on the one hand, the male talker's tokens (F0 mean of contour tone contrast: $M_{T5} = 103.43$, SD = 8.30, $M_{T6} = 105.01$, SD = 3.91; F0 mean of level tone contrast: $M_{T3} = 114.27$, SD = 5.53, $M_{T6} = 103.20$, SD = 5.94) were used in training tasks and post-training identification tasks to examine memory consolidation of the same tone stimuli before and after the nap. The tone stimuli produced by the male talker were chosen as the training stimuli because the contour and level contrasts differed in pitch contour and pitch height, respectively, and the differences in each contrast were acoustically salient to a similar extent. On the other hand, the tokens in female voice (F0 mean of contour tone contrast: $M_{T5} = 176.19$, SD = 11.82, $M_{T6} = 184.71$, SD = 11.07; F0 mean of level tone contrast: $M_{T3} = 207.99$, SD = 15.76, $M_{T6} = 187.35$, SD = 9.88) were used in a post-training identification task to assess generalization to new talkers because of different acoustic profiles of the tone stimuli (see below). The female talker's voice preserved the contour and level contrasts cued by pitch contour and height differences, while they were distinguished from the male talker's voice by mean F0 (i.e., the female talker = 189.06 Hz; the male talker = 106.48 Hz), F0 ranges (i.e., the female voice had a range of 160–220 Hz; the male voice had a range of 95–120 Hz) and the tonal distribution. Importantly, the stimuli produced by the (novel) female talker were not used in the tests before the nap manipulation (i.e., the first post-training test and discrimination pretest) because talker variability of tone stimuli introduced in the test session may destabilize learning and prevent consolidation-based performance improvements (Fuhrmeister & Myers, 2017, 2020).

In addition, visual stimuli were selected to pair with the tone stimuli and presented in training. These encompassed eight pictures of novel objects from the Novel Object and Unusual Name (NOUN) database (Horst & Hout, 2016). The novel objects were selected based on their shape and color to ensure that all these objects were visually distinct to facilitate the learning of tonal contrasts.

2.3. Procedure

Fig. 3 outlines the experimental procedures conducted on the nap and non-nap groups for contour tones and level tones across two days. Before the experiment, the participants were screened for their language and music backgrounds and sleep patterns, and only those who satisfied our requirements were invited to participate. On an experimental day (Day 2), both groups were first tested in a pretest session in the morning (10:30 am-12:00 pm), then trained after a break (e.g., a light lunch) and tested immediately after training (1:00–2:00 pm). They participated in either a nap or a rest session in the afternoon (2:00–3:00 pm), and then were re-tested after the manipulation session (after 4 pm).

To ensure that the participants' performance difference was not affected by their sleepiness state (or the related time-of-day effects), all the participants reported their sleepiness levels on a 1–7 scale through the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973) at three different points: (1) at the beginning of the pretest session (SSS-1), (2) at the beginning of the training session (i.e., before the nap manipulation) (SSS-2), and (3) before the second ID task (i.e., after the nap manipulation) (SSS-3).

2.3.1. Pretests

Participants were advised to keep a normal sleep-wake pattern several days before the experimental day. They were instructed to abstain from caffeine and alcohol consumption on the experimental day. They arrived at the laboratory at around 10:30 am and participated in a set of behavioral pretests for grouping purposes, that is, ensuring that the nap group and non-nap groups were comparable in their pitch aptitude, musical aptitude, and short-term memory span (see their comparable pretest performance in Table 1 and details of pretests in Supplementary Materials A).

2.3.2. Training and ID tasks

After the pretests and the lunch break, both the nap and non-nap groups were trained to identify two Cantonese contour-level (T5-T6) tonal contrasts and two Cantonese level-level (T3-T6) tonal contrasts and tested on their training outcome of the two kinds of tonal contrast respectively (see an illustration of training trials in Fig. A1 of Supplementary Materials).

The training task was designed for the participants to learn to associate novel objects with Cantonese pseudo-words. Given the arbitrary mapping between objects and pseudo-words, the training was necessary for completing the ID tests. As mentioned earlier, there were eight words and eight objects in total (four for the contour contrast and four for the level contrast), and the word-object associations were fixed across all participants. In each trial, participants first heard a spoken word and then immediately saw two novel objects on the left and right sides of the screen. The auditory stimuli were presented over headphones (Sony MDR-7506) at a comfortable volume (adjustable by individual participants). The participant's task was to press the left or right key to choose the corresponding object that represented the word in a self-paced manner. As soon as they responded, a feedback message ("Correct" or "Incorrect") was shown in the middle of the screen for 1000 ms, and participants would learn the correct word-object association through feedback. The training adopted a blocked design so that each block only contained one syllable, and two syllables (i.e., /ji/ and /fɐn/ for the contour tone contrast; /jɐu/ and /fu/ for the level tone contrast) alternated within the task. There were six blocks (2 syllables of each tonal contrast × 3 blocks), with 240 trials (2 tones × 5 tokens × 4 repetitions × 6 blocks) in the training, for each tonal contrast. The order of syllables (i.e., their corresponding block) for each contrast was counterbalanced within and across two groups of participants. The order of the tonal contrasts was also counterbalanced within and across groups (i.e., for each group, half of the participants trained on contour-level first and the other half trained on level-level first). The training session took around 40–50 min to learn the tonal contrasts.

The training of the two tonal contrasts was immediately followed with a first tone identification (ID-1) task before the nap manipulation, using the stimuli produced by the trained (male) talker. For each participant, the training task of one tonal contrast type was always followed at once by the tone identification task of the same tone type before the participant was trained and assessed for the other tonal contrast type. The ID task was to examine the learning outcome of the training task. Unlike training, there was no feedback in the ID task, and the participants were instructed to respond as fast and as accurately as possible when they saw two objects on the screen. The time-out limit was 1000 ms. Participants were carefully instructed and given 20 practice trials (excluded from analysis) to ensure they were sufficiently

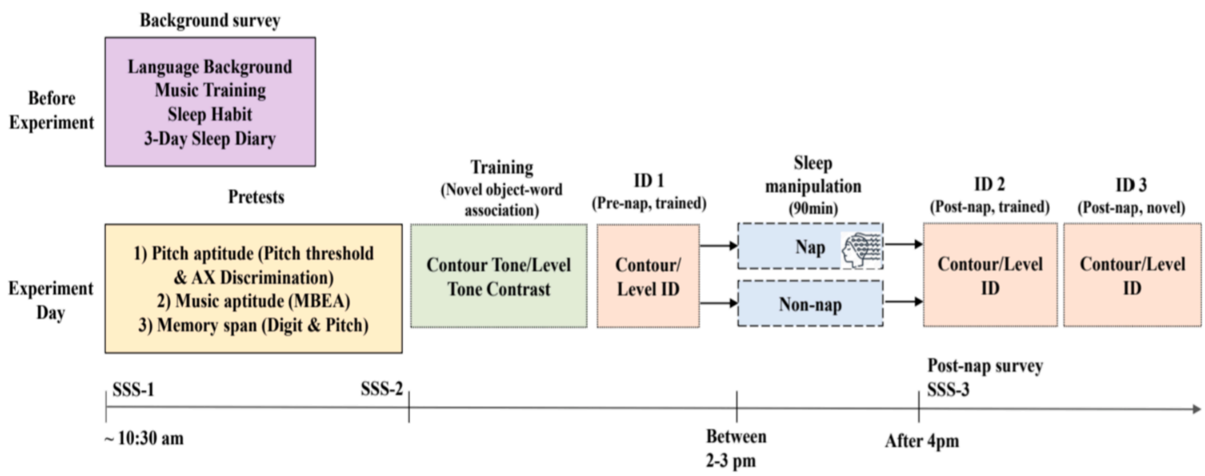

Fig. 3. An overview of the experimental procedure and the timing of each phase.

familiar with the new ID task. Each ID task contained 80 test trials (2 syllables × 2 tones × 5 tokens × 4 repetitions). Each ID session took around 10 min.

2.3.3. The nap manipulation

Upon completion of the ID-1 task for contour and level tone contrasts, participants were pseudo-randomly assigned to the nap group (i.e., the target group of nap manipulation) and the non-nap group (i.e., the control group). For the nap group, the participants were invited to a sleep lab and were given the opportunity to take a nap on a comfortable bed located in this room for a maximum of 90 min (one sleep cycle, which ends with the REM stage that usually lasts for 90 min, see Heim et al., 2017), with their brain EEG activity recorded during the 90-minute nap. The sleep lab was soundproof and dark so that the participants were maximally shielded from light and noise. The participants were instructed to sleep no more than 7 h on nights before the experimental day to facilitate afternoon napping (Chapelle et al., 2022), but to sleep as much as they could when they were given the opportunity of a 90-minute nap on the experimental day. An experimenter monitored the participants' brain waves to ensure that they fell asleep and, if necessary, woke them up after 90 min. For the non-nap group, the participants were invited into another soundproof lab room and asked to have a rest session for 90 min while staying awake. To achieve this, they were given the option of watching a silent documentary on cooking and/or playing simple computer games (e.g., Zuma). All these tasks were non-verbal and had minimal cognitive demands, and the purpose was to keep the participants awake with minimal interference. An experimenter actively monitored the alertness level of the participants via a webcam installed in the soundproof lab. If the participants were found to be sleepy, the experimenter would orally remind them to stay awake.

After the 90-minute session of nap manipulation, the participants were allowed to rest for up to 20 min to account for sleep inertia (the nap participants were additionally asked to fill out a post-nap survey asking about nap quality). After that, their performances of consolidation and cross-talker perception were assessed with the second ID task (ID-2) using stimuli produced by the trained talker, and then finally, a third ID (ID-3) task using the stimuli produced by the novel (female) talker.

2.4. Data acquisition and processing

2.4.1. EEG data acquisition

To investigate the relationship between neural activity during sleep (e.g., sleep spindles, slow oscillations, etc.) and differences across identification tasks, sleep EEG data were recorded on the participants of the nap group, using SynAmps 2 amplifier (NeuroScan, Charlotte, NC, U.S.) at a sampling rate of 1000 Hz. Scalp electrodes were arranged using the international 10–20 system at ten electrodes: FP1, FP2, F3, F4, C3, C4, P3, P4, O1, and O2 (Mölle et al., 2011; Hennis et al., 2016; Farthouat et al., 2017). Left and right electro-oculogram (HEOG, VEOG) and a ground electrode were also attached. An electrode located between the Cz and CPz served as the online reference. The impedance of all electrodes was checked to be stably no more than 5 kΩ (Mölle et al., 2011).

2.4.2. Sleep staging scoring

The sleep EEG data was preprocessed (resampled to 100 Hz; band-pass filtered between 0.1–40 Hz; re-referenced to grand average) and visualized with the EEGLAB toolbox (Delorme & Makeig, 2004) based in MATLAB (R2021a, The Mathworks, Inc.). Each 30 s epoch of sleep was visually scored into the wake, rapid-eye-movement (REM)-sleep, or NREM-sleep stages 1, 2, or 3 (i.e., N1, N2, N3 in which 1 is light sleep and 3 is slow-wave sleep), according to standard criteria of the American Academy of Sleep Medicine Manual (AASM V.2.4; Berry et al., 2017). Specifically, the scoring of sleep stages in this study was based on the AASM criteria applied to EEG and EOG channels. For each sleep stage, the duration in minutes and the percentage of time with reference to the total sleep time (TST) were then calculated. To ensure the accuracy of scoring, we compared the scoring results from the first 15 nap participants between two experienced sleep researchers, and the inter-rater reliability was 80%, indicating a high reliability in the scoring result. When instances of discrepancy arose, decisions were made by the primary scorer, who was an experienced researcher with expertise in sleep-memory research and thoroughly studied the AASM manual to ensure a comprehensive understanding of the scoring rules.

2.4.3. Sleep spindles and slow oscillations

[3] As mentioned in the introduction, the relevance of sleep spindles (henceforth, spindles) and slow oscillation (henceforth, SO) to memory consolidation is substantiated by empirical evidence: spindles (e.g., Tamminen et al., 2013) are a signature of N2 sleep, and SOs occur mostly during N3 sleep (see Fig. 1). In our study, spindles and SOs were extracted with an open-source, Python-based machine learning algorithm Yet Another Spindle Algorithm (YASA hereafter; Vallat & Walker,

[3] Recent studies suggested that SO-spindle temporal coupling may predict enhanced memory formation and index the development of memory consolidation (Hahn et al., 2020). Most such studies focus on full-night sleep, allowing for long N3 sleep and sufficient data of SO-nested spindles for analysis. One nap study examined SO-spindle coupling (Wüst et al., 2024) showed that SO-nested spindles in naps (i.e., N2/N3) decreased significantly from full-night sleep. Given the limited data on SO-nested spindles in a nap design (some nap participants did not have the coupling data), this study opted to focus on N2-spindles and SOs for analysis.

2021). As recommended by the YASA developer, EEG data were first down-sampled to 100 Hz and then bandpass filtered between 0.1–40 Hz. Then, with built-in functions in YASA, spindles were detected from N2 sleep (Hennies et al., 2016), and SOs were detected from N3 sleep (Massimini et al., 2004). Detected spindles were averaged across the abovementioned ten electrodes (i.e., the international 10–20 system) following common practice in previous research, while detected SOs were averaged over frontal electrodes (i.e., F3, F4) where SOs are expected to be prominent among young adults (Berry et al., 2017). The spindle detection method consisted mainly of three steps. First, a FIR filter bandpass filtered the EEG data between 13–15 Hz for (fast) spindles (Tamminen et al., 2013). Second, three thresholds (sigma power relative to broadband power; correlation between sigma and broadband; root mean square) were calculated, and only the data that met all three thresholds would be considered as a potential spindle. YASA then computed a soft threshold by smoothing the decision vector with a 100 ms window and found the proper start and end times by finding the spindle at which two out of the three thresholds were crossed. Finally, spindles close to each other (less than 500 ms) were merged, and spindles that were too short or too long were removed. Spindle density was calculated as the number of spindles per minute in N2 sleep.

The SO detection algorithm worked as follows. First, the EEG data was filtered between 0.5–2.0 Hz using an FIR filter with a transition band of 0.2 Hz (Xia et al., 2023). Then, all positive (between 10 to 150 μV) and negative peaks (between −40 to −200 μV) were detected. The nearest subsequent positive peak was identified for each negative peak, as these two peaks together preliminarily determine a potential SO. Then, a set of logical thresholds was applied to determine true SOs. An optional automatic outlier rejection was also applied to remove further abnormal SOs. The SO density was calculated as the number of SOs per minute in N3 sleep.

2.5. Data analysis

All data files and data analysis scripts of this study to conduct data analysis can be found at Open Science Framework repository (https://osf.io/9rcnm/). To investigate the effect of tonal contrasts and nap conditions on nap-based memory consolidation of lexical tones, mixed-effects logistic regression models were fitted on response accuracy (binary coded, 1 for correct and 0 for incorrect) in R (R Development Core Team, 2008) using the *lme4* package (Bates et al., 2015). All categorical predictors were deviation-coded (−0.5, 0.5) to test the main effect. For training performance, the first two and the last two training blocks were coded as initial and outcome blocks to track participants' performance improvement throughout the training: group (nap vs. non-nap; deviation coding: −0.5, 0.5), tone (contour vs. level; deviation coding: −0.5, 0.5) and block (initial vs. outcome; deviation coding: −0.5, 0.5) were entered as fixed effects.

For identification tasks, the consolidation effect (i.e., before and after a nap, for the trained talker) was examined with differences between ID 2 and ID 1, and the cross-talker effect (i.e., after a nap and between the trained and the novel talker) was examined with differences between ID 3 and ID 2. For the consolidation effect, group (nap vs. non-nap; deviation coding: −0.5, 0.5), tone (contour vs. level; deviation coding: −0.5, 0.5), and session (ID 1 vs. ID 2; deviation coding: −0.5, 0.5) were entered as fixed effects. For the cross-talker perception, group (nap vs. non-nap; deviation coding: −0.5, 0.5), tonal contrast (contour vs. level; deviation coding: −0.5, 0.5), and session (ID 2 vs. ID 3; deviation coding: −0.5, 0.5) were entered as fixed effects. Regarding the random effects, we started from the most complex by-subject random effect structure (i.e., session*tone*group slope and intercept) plus a by-trial intercept, since variation across the participants is the primary source of variation resulting from repeated measurement. When the model could not converge, we simplified the random structure by removing fixed effects from the by-subject random slope by adopting a theoretical approach (Barr et al., 2013).

Generalized linear models were constructed to investigate how sleep-EEG components, that is, neural activities indexing the hippocampal-neocortical cycle underlying memory consolidation of language learning (spindle density and SO density) contribute to performance differences across identification tasks (i.e., ID 2 compared to ID 1 for consolidation; ID 3 compared to ID 2 for talker-difference) of contour and level tones respectively. To account for the unique effect of each sleep-EEG measure on the consolidation/cross-talker perception, spindle density (continuous numeric variables), SO density (continuous numeric variables), and session (ID 2-ID 1 or ID 3-ID 2, depending on the analysis of ID performance) were entered as fixed effects for the model of each tone type. Regarding the random effects, we started from a by-subject random effect structure (i.e., session slope and intercept) plus a by-trial intercept. When the model could not converge, we simplified the random structure by removing fixed effects from the by-subject random slope.

3. Results

3.1. Behavioral data: Memory consolidation and cross-talker perception

To ensure that all the nap participants fell asleep for a reasonable duration, we excluded two participants in the nap group from the analysis of ID performance due to their limited nap length based on EEG sleep scoring (i.e., overall sleep length $<$ 10 mins with little Stage N2 and N3 sleep recorded). We assessed the ID accuracy of contour and level tone contrasts across the three identification tasks by the remaining forty-eight participants in the nap group and forty participants in the non-nap group.

We firstly confirmed that the participants were effectively trained with the two categories of non-native (Cantonese) tonal contrasts and that the two groups had a similar baseline performance (i.e., ID) for both tonal contrasts (see the results in Supplementary Materials B). We also conducted an exploratory analysis to rule out the possibility that sleepiness levels (i.e., SSS scores) accounted for the ID performance (see the results in Supplementary Materials C). This section examines the consolidation effect (i.e., before vs. after the nap manipulation for trained talker) by comparing ID 2 (post-nap) and ID 1 (pre-nap) between the two groups, and then turns to the cross-talker perception (i.e., the stimuli produced by the trained vs. the novel talker, after the nap manipulation) by comparing ID 3 (novel talker) and ID 2 (trained talker) between the two groups.

Fig. 4 displays the change in response accuracy across the three identification tasks by forty-eight participants in the nap group and forty participants in the non-nap group. To examine the nap-based memory consolidation effect of the trained talker's voice first, we ran a logistic mixed-effects model on the participants' identification accuracy in ID 1 (pre-nap; trained) and ID 2 (post-nap; trained). Fixed effects were session (ID 1 vs. ID 2), group (nap vs. non-nap), and tone (contour vs. level). The random effects were by-subject intercepts and slopes of session and group, as well as by-trial intercept. The results revealed a main effect of tonal contrast and a main effect of session ($\beta = 0.12$, $SE = 0.05$, $z = 2.27$, $p = 0.02$; see detailed statistics in Table D1 of Supplementary Materials). The results showed that the contour tone contrast yielded higher accuracy than the level tone contrast ($\beta = -0.64$, $SE = 0.03$, $z = -18.88$, $p < 0.001$) regardless of sessions and groups. The ID 2 yielded higher accuracy than the ID 1 regardless of tonal contrasts and groups. Crucially, no interactions were found, for instance, between the group and the session. The results showed that the nap and non-nap groups might have consolidated the contour and level tone contrasts, which were produced by the trained talker, to a similar degree.

We subsequently investigated the cross-talker perception with logistic mixed-effects models on the participants' identification accuracy in ID 2 (post-nap; trained talker) and ID 3 (post-nap; novel talker). Session (ID 2 vs. ID 3), group (nap vs. non-nap), and tone (contour vs. level) were included as fixed effects. The random effects are by-subject

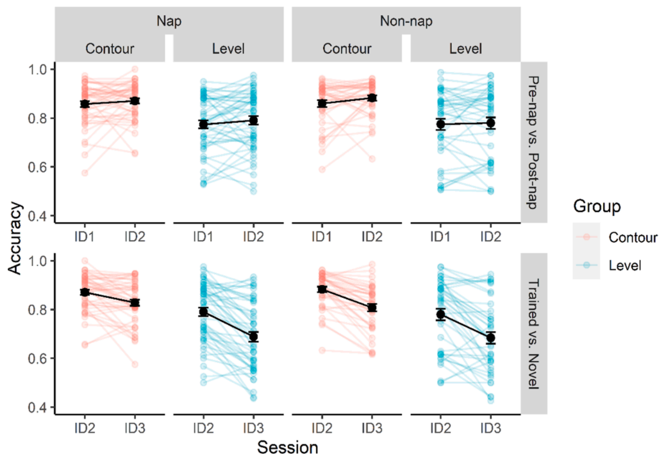

Fig. 4. The response accuracy of the nap group (left) and the non-nap group (right) in identifying contour (red) and level (green) tones. Three identification tasks: ID 1 (pre-nap; trained) and ID 2 (post-nap; trained) in the top panel for consolidation effect; ID 2 (post-nap; trained) and ID 3 (post-nap; novel) in the bottom panel for the cross-talker perception. Error bars indicate the standard error of the mean value.

intercepts and slopes of session and group and by-trial intercepts. The results revealed a significant main effect of tonal contrast and session ($\beta = -0.53$, $SE = 0.05$, $z = -10.58$, $p < 0.001$; see detailed statistics in Table D1 of Supplementary Materials). The contour tone contrast yielded higher accuracy than the level tone contrast ($\beta = -0.72$, $SE = 0.03$, $z = -22.17$, $p < 0.001$) regardless of sessions and groups. The results of ID 3 (i.e., novel talker) yielded a lower accuracy than ID 2 (i.e., trained talker) regardless of tonal contrasts and groups. Crucially, the results yielded a three-way interaction ($\beta = -0.31$, $SE = 0.13$, $z = 2.37$, $p = 0.02$) between group, session, and tone.[4]

To better understand the sources of the three-way interaction, follow-up models, with by-subject intercepts and slopes of the session and by-trial intercepts, were run on the ID accuracy for each tonal contrast, respectively. The results showed a significant interaction between session and group in the contour tone contrast ($\beta = -0.27$, $SE = 0.12$, $z = -2.13$, $p = 0.03$), but not in the level tone contrast ($\beta = -0.005$, $SE = 0.13$, $z = -0.04$, $p = 0.97$). The subsequent analysis showed that the nap group outperformed the non-nap group by demonstrating a smaller difference in ID accuracy between the trained talker and the novel talker; in other words, they demonstrated smaller ID 3- ID 2 differences that phonetically imply better cross-talker perception. The group effect on the cross-talker perception was found for the contour tone contrast but not for the level tone contrast. Likewise, follow-up models, with by-subject intercepts and slopes of the session and by-trial intercepts, were also run on the ID accuracy for each group, respectively. The results suggested that there was a significant interaction between session and tone in the nap group ($\beta = -0.21$, $SE = 0.09$, $z = -2.41$, $p = 0.02$), but not in the non-nap group ($\beta = 0.10$, $SE = 0.10$, $z = 1.02$, $p = 0.31$). The subsequent analysis for each group indicated that the nap group demonstrated a smaller difference in ID accuracy between the trained talker and the novel talker for contour tone contrast than for level tone contrast. In contrast, no such effect of tonal contrast on cross-talker perception was found for the non-nap group.

Overall, the results of tone ID performance suggested that compared with the non-nap group, a 90-minute nap might have fostered the nap participants' consolidation of newly-learned tonal contrasts by diminishing the differences between the novel and trained talkers (but not necessarily improving the tone ID accuracy within the same talker). Importantly, the plausible effect of nap-mediated consolidation on the cross-talker perception was contingent on the participants' prior knowledge, with the contour tone contrast prioritized over the level tone contrast. Given the significant findings about the cross-talker perception for the nap group, the next section focuses on the individual analysis of the nap participants and addresses the question of whether the cross-talker perception effect for each tonal contrast (e.g., ID 3-ID 2 of the contour tone) is accounted for by their differences in sleep spindles and SOs.

3.2. Sleep-EEG data: Spindle density and SO density predict the cross-talker perception

This section investigates the effect of sleep-EEG components on the consolidation effect by benefitting the cross-talker perception. Three participants were further excluded from the analysis of sleep-EEG variables due to EEG problems (e.g., difficulty in differentiating sleep stages and identifying sleep-related brain activities based on their EEG). The sleep data of the remaining forty-five participants of the nap group (see details of sleep data in Supplementary Materials E; see the analysis of sleep-EEG components on ID 2- ID 1 in Supplementary Materials F) were analyzed. (Fast) spindles and SOs[5]are crucial in predicting memory consolidation of language learning (Tamminen et al., 2013) and were thus analyzed in terms of the cross-talker effect found in the behavioral data.

Fig. 5 illustrates the relationships between sleep-EEG variables and cross-talker perception, indexed by the ID differences between the novel talker and the trained talker (i.e., ID 3-ID 2 accuracy). The larger the difference values are, the better the cross-talker perception is. Two participants had extreme values of spindle density or SO density (i.e., 2.5 S.D above the mean). They were kept for subsequent analysis, considering the availability of N3 sleep data (the pattern remained unchanged even after the extreme values were removed from data analysis).

We fitted a logistic mixed-effects model on contour tones and level tones, respectively, for spindle density and SO density. Before they were entered into the models, the multi-collinearity of the two sleep-EEG variables was first checked. The ordinary and Bayesian correlation test results showed no correlation between spindle density and SO density, $r(32) = -0.24$, $p = 0.18$, and $BF_{10} = 0.82$. To capture the unique effects of spindle density and SO density in contour and level tone contrasts, fixed effects were entered as spindle density and SO density (continuous variables), as well as session (ID 2 vs. ID 3). The random effects were by-

[4] To rule out the possibility that group differences in MEQ chronotype (evening/morning type vs. neutral type) and sleepiness (i.e., SSS-3; before ID 2 and 3) differentially impacted memory performances for the nap and non-nap groups, mixed-effects models with chronotype and SSS-3 as covariates were conducted on ID 2 and ID 3. The three-way interaction between group, session, and tone remained significant ($\beta = 0.31$, $SE = 0.13$, $z = 2.36$, $p = 0.02$), indicating that group differences in chorotype and SSS-3 did not explain their differences after the nap manipulation. Likewise, a mixed-effects model with training outcome as a covariate was also conducted on the same data. The three-way interaction between group, session, and tone remained significant ($\beta = 0.30$, $SE = 0.13$, $z = 2.34$, $p = 0.02$), indicating that group differences in the training outcome (e.g., level tones) did not explain their differences after the nap manipulation.

[5] Thirty-four out of forty-five participants successfully recorded the N3 sleep. Only these participants were included in the analyses of SOs, given that SOs were extracted from N3 sleep.

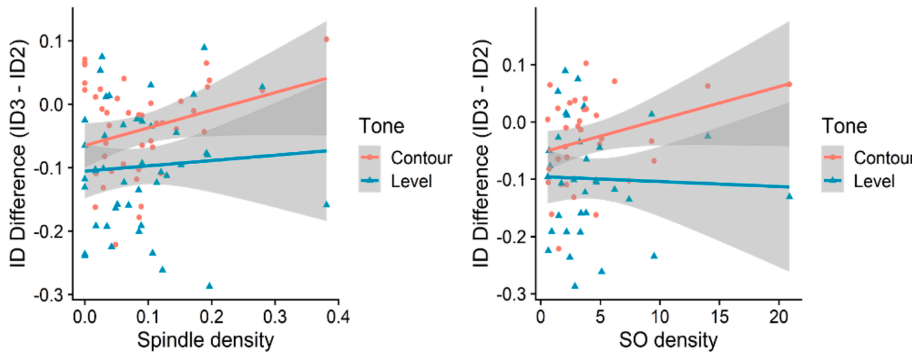

Fig. 5. Spindle density (left) and SO density (right) of the nap participants. They positively predicted the difference in identification accuracy between post-nap trained (ID 2) and post-nap novel (ID 3) talkers for the contour tone contrast (red), but not the level tone contrast (green).

subject session slopes and intercepts, and by-trial intercepts. The results[6] (see detailed statistics in Table D2 of Supplementary Materials) showed that only in the contour tone contrast, both the spindle density ($\beta = 2.11$, $SE = 1.09$, $z = 1.93$, $p = 0.05$) and SO density ($\beta = 0.05$, $SE = 0.02$, $z = 2.08$, $p = 0.04$) had significant interaction with the session, with higher spindle and SO densities associated with better performances of cross-talker perception. That is, smaller differences in ID accuracy were associated with better sleep if ID 3 was lower than ID 2 (i.e., participants with negative y-axis values in Fig. 5), and greater differences in ID accuracy were associated with better sleep if ID 3 was higher than ID 2 (i.e., participants with positive y-axis values in Fig. 5). For the level tone contrast, neither the spindle density nor SO density manifested a significant interaction with the session.

These results showed that spindle density and SO density positively predicted the effect of nap-mediated consolidation on cross-talker perception for the contour tone contrast but not the level tone contrast. The increased SOs, together with denser spindles, were associated with the better cross-talker performances. The association was specific to the manipulation of prior knowledge, that is, the contour tone contrast. This finding suggests that naps preferentially facilitate the consolidation of schema-related memories consistent with prior knowledge. The individual differences in N2 sleep (i.e., spindle density) and slow-wave sleep (i.e., SO density) may have mediated the integration of newly-learned tonal contrasts into the existing tonal structure (i.e., L1 tonal inventory) through facilitating cross-taker perception among the nap participants.

4. Discussion

This study adopted a nap design to investigate the effect of prior knowledge on nap-mediated memory consolidation (i.e., pre-nap ID 1 and post-nap ID 2) and cross-talker perception (i.e., trained ID 2 and novel ID 3) of Cantonese contour and level tones among Mandarin-speaking L2 learners. The results of tone identification performance between the nap and non-nap groups showed that they successfully gained improvement (ID 1 and ID 2) for both tonal contrasts upon completion of the training task, without a group difference. The behavioral results of ID 2 and ID 3 showed that the Mandarin-speaking nap participants exhibited a smaller difference in ID accuracy between the trained talker and the novel talker (i.e., a better cross-talker perception) than their non-nap counterparts through the nap manipulation. Importantly, the nap effect on the cross-talker perception was only found for the contour tone contrast, consistent with Mandarin speakers' prior tonal knowledge. The cross-talker effect of the contour tone contrast (but not the level tone contrast) was also found to be significantly predicted by sleep-related brain activities, with increased spindles and SOs positively associated with better cross-talker perception among the nap participants.

4.1. The role of prior knowledge in nap-mediated consolidation of lexical tones

Consistent with prior work on tone perception and training (Francis et al., 2008; Gandour, 1983; Jongman et al., 2017), the behavioral results of tonal contrasts showed higher accuracy of the contour tone contrast than the level tone contrasts, regardless of groups and blocks/sessions, for the training and tone identification assessments. The tonal contrast difference indicates that tonal contrasts distinguished by pitch contour differences (e.g., Cantonese T5-T6) were better trained and perceived than those distinguished by pitch height differences (e.g., Cantonese T3-T6) by Mandarin speakers. This is highly expected as prior knowledge also affected the encoding phase before sleep, with schema-related knowledge better encoded than schema-unrelated knowledge (see similar findings in Cordi et al., 2023; Hennies et al., 2016). Mandarin-speaking participants used pitch contour to encode L1 tonal contrasts and thus showed an advantage in perceiving the L2 contour tone contrast over the L2 level tone contrast (Jongman et al., 2017; Qin & Jongman, 2016).

Despite the perceptual advantage regarding tonal contrasts, the results showed that the nap group and non-nap group both had an improvement in tone identification (i.e., ID 2 > ID 1) regardless of the tonal contrasts, and thus might have similarly consolidated the tone stimuli produced by the trained talker. The lack of the tonal contrast effect did not support a significant role of prior knowledge in the nap-mediated consolidation of lexical tones. This may appear to stand at odds with the findings of a recent study that reported a possible role of prior knowledge in the overnight consolidation of lexical tones (Qin et al., preprint). The differences in the overnight and nap designs may account for the discrepancy in the results. The overall sleep length (i.e., the number of sleep cycles) and relative sleep architecture (e.g., spindle and SO characteristics) often differ between a full-night sleep and a daytime 90-minute nap (see van Schalkwijk et al., 2019 for a direct comparison). It is possible that the prioritized consolidation of contour tones was not evident for the trained tone stimuli because not every nap participant had a full (NREM-REM) sleep cycle (e.g., only nine participants recorded a REM stage) in this study. Given the possibility that 90 min might have been insufficient for average adults to complete the full sleep cycle, future studies may consider providing the participants with

[6] To explore the possibility that initial encoding strengths (i.e., ID1 accuracy) accounted for the cross-talker perception (e.g., Tucker & Fishbein, 2008), the participants' ID1 accuracy was put into new mixed-effects models of contour tones and level tones, respectively. The two-way interaction between ID1 and the session was not significant for contour tones ($\beta = 0.93$, $SE = 1.50$, $z = 0.63$, $p = 0.53$) and level tones ($\beta = -1.55$, $SE = 0.90$, $z = -1.74$, $p = 0.08$). Crucially, the results of spindles and SOs largely remained unchanged, indicating that initial encoding strength might not account for cross-talker perception performances of the nap participants.

a 2-hour rest (see Farhadian, Khazaie, Nami, & Khazaie, 2021 for a review of 2-hour nap studies) in examining the nap effect on tone learning. Also, the overnight study did not control for interfering activities (e.g., L1 interference, Earle & Myers, 2015b) between experiment offset (i.e., training and immediate posttests) and sleep onset. Given the methodological discrepancy, the findings between the current nap study and the overnight study might not be comparable.

While we hypothesized an advantage of the nap group over the non-nap group for the tone stimuli produced by the trained talker, the non-nap group, who took the 90-minute rest session, showed a similar ID change (i.e., an offline improvement) to the nap group after the nap manipulation. The ID accuracy, especially that of the contour tone contrast, was improved in both the nap and non-nap groups (see Fig. 4). This finding is not surprising as the pre-nap ID 1 and post-nap ID 2 were identical. A possible repetition effect (i.e., a better performance when retested) could not be ruled out. Alternatively, the beneficial effect of 90-minute naps on stabilizing memories is not universal and might be less reliable than previously assumed by the initial sleep studies (see Cordi & Rasch, 2021; Farhadian, Khazaie, Nami, & Khazaie, 2021). Recent evidence also suggests that (unoccupied) wakeful rest facilitates memory consolidation in a manner similar to that during sleep (see Wamsley, 2019 for reviews). One possibility is that, despite the non-verbal activities (e.g., video watching) employed in our manipulation, the wakeful rest session provided an opportunity for quiet (unoccupied) rest by protecting newly-learned tonal knowledge from interfering activities and may have allowed for offline consolidation, especially for the contour tone contrast that is consistent with the participants' prior knowledge.

4.2. The effect of prior knowledge on cross-talker perception and sleep-EEG variables

The behavioral results of cross-talker perception (i.e., ID 3- ID 2) showed a smaller difference between novel and trained talkers (i.e., better cross-talker perception) by the nap group than the non-nap group for the contour tone contrast. Specifically, the results of the nap participants (but not those of the non-nap participants) implied that the contour tone contrast might have been better applied across the trained and novel talkers than the level tone contrast. The effect related to tonal contrasts may tentatively suggest their enhanced cross-talker perception due to nap-mediated memory consolidation (Qin & Zhang, 2019). That is, the Cantonese contour tone contrast (primarily cued by pitch contour differences) was consistent with Mandarin speakers' prior knowledge and was thus prioritized over the level tone contrast for memory consolidation, with better cross-talker perception between trained and novel talkers. The current study seems to suggest the beneficial role of prior knowledge on memory consolidation in terms of using acoustic cues. For our case, pitch contour cueing contour tone differences are consistent between L1-Mandarin and L2-Cantonese. This may be aligned with a "cue-weighting" idea that acoustic cues for perceiving L2 sounds are perceptually weighted as a function of their informativeness for cueing L1 sound contrasts (Francis & Nusbaum, 2002; Holt & Lotto, 2006). That is, pitch contour was weighted more than pitch height by Mandarin listeners to perceive and consolidate Cantonese tones. As a result, generalizing the contour tone contrast (i.e., schema-related memories) across talkers might be prioritized for nap-mediated consolidation over the level tone contrast (i.e., non-schema memories) (Hennies et al., 2016; Tse et al., 2007, 2011). However, the beneficial effect of prior knowledge on memory consolidation needs to be cautiously interpreted as it may not necessarily be generalized to the cases where the cues used in L1 sound contrasts are not consistent with the target cues used in L2 speech perceptual learning (e.g., English listeners distinguishing Hindi dental-retroflex consonant contrasts; Earle & Myers, 2015a). The generalizability of the prior knowledge effect should be further examined in perceptual learning of multiple sound contrasts, which are cued by the same versus different acoustic information in L1 and L2 (Francis & Nusbaum, 2002; Holt & Lotto, 2006).

Importantly, the analyses of sleep-EEG data further supported the possibility that learning to identify the contour tone contrast across talkers was prioritized for nap-mediated consolidation. Specifically, spindle density and SO density, which have been shown to be associated with memory consolidation of language learning (see a special issue by Rasch, 2017), positively predicted the cross-talker perception of the contour tone contrast. These findings, consistent with previous studies on sleep architecture (Hennies et al., 2016; Tamminen et al., 2013), further suggest that sleep spindles and slow-wave activity might play a significant role in integrating non-native (Cantonese) tonal contrasts into the existing tonal structure (i.e., Mandarin tonal inventory). The Cantonese contour tone contrast seemed to be preferentially consolidated, and its integration may have proceeded rapidly during naps because of the beneficial role of prior knowledge. Hence, applying the learned contour tone contrast to a novel talker yielded an association with increased spindles and SOs. However, the Cantonese level tone contrast is more likely to include inconsistent knowledge and thus resulted in a relatively slow nap-mediated consolidation and/or integration to the existing tonal structures. Hence, learning the level tone contrast showed a lack of association with spindles and SOs. The findings have important theoretical implications for perceptual learning of lexical tones by advancing our understanding of neurophysiological sleep correlates (i.e., spindles and SOs) that are most critical to nap-mediated memory consolidation of lexical tones (Chapelle et al., 2022).

The current findings about the role of prior knowledge (i.e., tonal contrasts) align with the iota model proposed by Lewis and Durrant (2011), which provides a theoretical framework explaining the significant role of prior knowledge in memory consolidation. The memory consolidation model, which supports an active system consolidation process (Rasch & Born, 2013), states that new and old memories are replayed simultaneously during slow-wave sleep, with information shared across new and old memories (i.e., schema-related memories) more strongly activated than that from only a single source (i.e., non-schema memories). This study provided empirical evidence supporting the theoretical account, suggesting that sleep might not only passively protect memories from interfering influences but also actively shape the newly-learned declarative knowledge (e.g., lexical tones) and make qualitative changes to them *selectively* based on the relationship between incoming information and the existing knowledge framework (Lewis & Durrant, 2011; McClelland, 2013). Importantly, both the old and new memories might have been replayed simultaneously during slow-wave sleep, with supporting evidence from the finding that increased SOs lead to a better abstraction of tonal categories (e.g., across talkers) or insight into hidden regularities of tonal representations.

4.3. The implications of memory consolidation and language learning

A significant theoretical contribution of the current study is that an effect of prior knowledge on memory consolidation of L2 word learning (Cordi et al., 2023; Zion et al., 2019) was also found for L2 tone learning. On the one hand, this may suggest, although speech is acoustically variable and multidimensional (Myers et al., 2017), that speech (e.g., tone) learning is more comparable to other types of learning (e.g., word or vocabulary learning) than previously thought, at least for the (beneficial) role of prior knowledge on memory consolidation. This finding aligns with learning and memory theories (McClelland, 2013; Rasch, 2017), suggesting that sleep-mediated memory consolidation process in the brain can benefit language learning in general and may provide the basis for stabilizing and integrating newly-learned linguistic information. On the other hand, different from previous studies on the memory consolidation of L2 word learning (Cordi et al., 2023; Zion et al., 2019), the facilitative effect of naps was found to benefit L2 tone learning, specifically, by promoting cross-talker perception. The results related to cross-talker perception are not unexpected if learning to generalize across talkers is more essential to perceptual learning of lexical tones than language learning in other domains (e.g., L2 word learning in Cordi et al., 2023). The outcomes related to cross-talker perception might be explained by the properties of lexical tones, that is, the prevalence of tonal variations induced by different talkers (e.g., C. Zhang & Chen, 2016). Building talker-independent representations for tonal abstraction is a great challenge for L2 learners, but it is also a necessary step in identifying tones produced by different talkers to accommodate talker variations in real-life communication (e.g., K. Zhang et al., 2018). Learning to deal with talker-induced variability is often incorporated into tone training studies, and talker generalization (i.e., using novel talkers) has been used to assess the learning outcome in training studies (e.g., Perrachione et al., 2011; Wang et al., 1999).

Admittedly, linguistic representations of speech (e.g., lexical tones), words, and grammatical rules are not the same, and learning them may involve consolidation of declarative memories (Davis et al., 2008; Fenn et al., 2003; Weighall et al., 2017), procedural memories (Batterink & Paller, 2017; Zion et al., 2022), or both. How the aspects of languages might be (differentially) shaped by memory consolidation is still poorly understood. This study is one of the first nap studies that focus on perceptual learning of lexical tones (also see Chapelle et al., 2022) and examines nap-mediated consolidation of specific declarative knowledge (i.e., associative tonal memories and their association with sleep-EEG components). More studies on the consolidation of declarative and procedural memories during full-night sleep and daytime naps might be needed to understand the exact role of sleep-brain activities (e.g., SO-spindle coupling; NREM vs. REM sleep) in learning multiple language skills (see Rasch, 2017 for the discussion).

This study has addressed an important issue concerning adult language acquisition by focusing on perceptual learning of lexical tones, which often induce great talker-related variability and pose challenges for L2 learners (Pelzl, 2019); it has leveraged insights from theories of memory consolidation (Lewis & Durrant, 2011; McClelland, 2013). The findings imply that which types of tones are consolidated and generalized across talkers largely depends on learners' prior knowledge (e.g., using pitch cues in L1), at least at an early learning stage. These findings have advanced our understanding of the neural mechanisms of how sleep supports successful L2 tone learning through memory consolidation. This may enable researchers to use the technique of re-exposure during sleep (i.e., Targeted Memory Reactivation-TMR) to reconstruct neural representations during sleep, and directly examine whether contour tone contrasts are reactivated more (e.g., increased spindle density, greater spindle power etc.) than level tone contrasts during slow-wave sleep (e.g., Ngo & Staresina, 2022; Schreiner & Rasch, 2017).

In closing, it is necessary to acknowledge some limitations of this study. The current design did not allow us to rule out the possible effect of encoding memory strength on memory consolidation (but see our exploratory analyses including ID 1 accuracy in new models; footnote 6), as participants' prior knowledge (inevitably) benefited encoding (i.e., perceptual learning) of contour tonal contrasts over level tonal contrasts (Cordi et al., 2023; Hennies et al., 2016). A future direction could be to manipulate training strength and provide additional training opportunities for low prior knowledge conditions (see Cordi et al., 2023), for example, level tonal contrasts, to better disentangle the roles of prior knowledge and encoding strength in nap-mediated memory consolidation. In addition, a group with a comparable number of participants, who do not have prior knowledge of lexical tones in their L1 (e.g., English-speaking L2 learners in Qin et al., preprint), is also needed to investigate further the effect of prior knowledge on nap-mediated consolidation of lexical tones in a different scenario. Finally, the implication on cross-talker perception needs to be cautiously interpreted, as the participants' accuracy on the tone stimuli produced by the novel talker was not examined as a function of the nap manipulation and should be examined in future studies (but see Fuhrmeister & Myers, 2017, 2020 for the findings that test-stimuli variability may compromise memory consolidation). Given speaker idiosyncrasies in tone production, it would be appropriate to use (at least) two novel talkers for future studies and examine talker generalization conclusively by counterbalancing the trained and novel talkers within and across the manipulation groups.

5. Conclusions

Aligned with the idea that sleep facilitates language learning through memory consolidation (see a special issue by Rasch, 2017), the current study provided empirical evidence supporting the important role of prior knowledge in nap-mediated consolidation of lexical tones. The findings suggest that a short nap benefited participants' tone consolidation by facilitating the cross-talker perception between the trained and the novel talker. The effect of cross-talker perception was specific to the contour tone contrast that was consistent with prior knowledge. Its associations with increased sleep spindles and slow-wave activity, indexing the hippocampal-neocortical cycle underlying memory consolidation processes, further supported the role of prior knowledge in integrating newly-learned memories. Overall, novel contour tones (i.e., schema-related memories) were found to be prioritized over level tones for nap-mediated consolidation under the influence of participants' prior knowledge.

CRediT authorship contribution statement

Quentin Zhen Qin: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Ruofan Wu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Caicai Zhang:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the RGC-Hong Kong Research Grant (ECS #26600623) and Sustainable Smart Campus as a Living Lab Fund (FS #107) awarded to Quentin Zhen Qin. The authors would like to thank Yin-To Chui, Rui Jin, Susu Lai, and Kangdi Liu for their help in data collection and data analysis. Portions of this work have been presented at the Society for the Neurobiology of Language 16th Annual Meeting in Brisbane, Australia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2025.105568.

Data availability

All data files and data analysis scripts of this study to conduct data analysis can be found at Open Science Framework repository (https://osf.io/9rcnm/).

References

Alger, S. E., Lau, H., & Fishbein, W. (2010). Delayed onset of a daytime nap facilitates retention of declarative memory. *PLoS ONE, 5*(8), Article e12131. https://doi.org/10.1371/journal.pone.0012131

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., & Paller, K. A. (2017). Sleep-based memory processing facilitates grammatical generalization: Evidence from targeted memory reactivation. *Brain and Language, 167*, 83–93. https://doi.org/10.1016/j.bandl.2015.09.003

Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., Troester, M. T., & Vaughn, B. V. (2017). AASM Scoring Manual Updates for 2017 (Version 2.4). *Journal of Clinical Sleep Medicine, 13*(05), 665–666. https://doi.org/10.5664/jcsm.6576

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. [Computer program].

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28*(2), 193–213.

de la Chapelle, A., Savard, M. A., Restani, R., Ghaemmaghami, P., Thillou, N., Zardoui, K., Chandrasekaran, B., & Coffey, E. B. J. (2022). Sleep affects higher-level categorization of speech sounds, but not frequency encoding. *Cortex, 154*, 27–45. https://doi.org/10.1016/J.CORTEX.2022.04.018

Cordi, M. J., & Rasch, B. (2021). How robust are sleep-mediated memory benefits? *Current Opinion in Neurobiology, 67*, 1–7. https://doi.org/10.1016/j.conb.2020.06.002

Cordi, M. J., Schreiner, T., & Rasch, B. (2023). Is prior knowledge essential? Additional training opportunities restore sleep-associated memory benefits under conditions of low prior knowledge. *Journal of Sleep Research, 32*(4). https://doi.org/10.1111/jsr.13834

Davis, M. H., Di Betta, A. M., Macdonald, M., & Gaskell, G. (2008). Learning and consolollidation of novel spoken words: Behavioral and neural evidence. *Cerebral Cortex*, 803–820.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Earle, F. S., & Myers, E. B. (2015a). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America, 137*(1), EL91–EL97. https://doi.org/10.1121/1.4903918

Earle, F. S., & Myers, E. B. (2015b). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance, 41*(6), 1680–1695. https://doi.org/10.1037/xhp0000113

Farhadian, N., Khazaie, H., Nami, M., & Khazaie, S. (2021). The role of daytime napping in declarative memory performance: a systematic review. *Sleep Medicine, 84*, 134–141. https://doi.org/10.1016/j.sleep.2021.05.019

Farthouat, J., Gilson, M., & Peigneux, P. (2017). New evidence for the necessity of a silent plastic period during sleep for a memory benefit of targeted memory reactivation. *Sleep Spindles & Cortical Up States, 1*(1), 14–26. https://doi.org/10.1556/2053.1.2016.002

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425*(6958), 614–616. https://doi.org/10.1038/nature01951

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Friedrich, M., Mölle, M., Friederici, A. D., & Born, J. (2020). Sleep-dependent memory consolidation in infants protects new episodic memories from existing semantic memories. *Nature Communications, 11*(1), 1298. https://doi.org/10.1038/s41467-020-14850-8

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *The Journal of the Acoustical Society of America, 142*(5), EL448–EL454. https://doi.org/10.1121/1.5009688

Fuhrmeister, P., & Myers, E. B. (2020). Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, and Psychophysics, 82*, 2049–2065. https://doi.org/10.3758/s13414-019-01925-y

Gandour, J. T. (1983). Tone perception in far eastern-languages. *Journal of Phonetics, 11*(2), 149–175.

Hahn, M. A., Heib, D., Schabus, M., Hoedlmoser, K., & Helfrich, R. F. (2020). Slow oscillation-spindle coupling predicts enhanced memory formation from childhood to adolescence. *ELife, 9*, 1–21. https://doi.org/10.7554/eLife.53730

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Havas, V., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Quarterly Journal of Experimental Psychology, 71*(6), 1469–1481. https://doi.org/10.1080/17470218.2017.1329325

Heim, S., Klann, J., Schattka, K. I., Bauhoff, S., Borcherding, G., Nosbüsch, N., Struth, L., Binkofski, F. C., & Werner, C. J. (2017). A nap but not rest or activity consolidates language learning. *Frontiers in Psychology, 8*(MAY), 1–8. https://doi.org/10.3389/fpsyg.2017.00665

Hennies, N., Ralph, M. A. L., Kempkes, M., Cousins, J. N., & Lewis, P. A. (2016). Sleep spindle density predicts the effect of prior knowledge on memory consolidation. *Journal of Neuroscience, 36*(13), 3799–3810. https://doi.org/10.1523/JNEUROSCI.3162-15.2016

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America, 119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America, 142*(2), EL163–EL169. https://doi.org/10.1121/1.4995526

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research, 17*(1), 3–10. https://doi.org/10.1111/j.1365-2869.2008.00622.x

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition, 39*(4), 593–620.

Loureiro, F., & Garcia-Marques, T. (2015). Morning or Evening person? Which type are you? Self-assessment of chronotype. *Personality and Individual Differences, 86*, 168–171. https://doi.org/10.1016/j.paid.2015.06.022

Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *The Journal of Neuroscience, 24*(31), 6862–6870. https://doi.org/10.1523/JNEUROSCI.1318-04.2004

Matthews, S., & Yip, V. (2011). Cantonese: A comprehensive grammar. In *Cantonese: A Comprehensive Grammar* (2nd edition). Routledge, London.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*(4), 1190–1210. https://doi.org/10.1037/a0033812

Mölle, M., Bergmann, T. O., Marshall, L., & Born, J. (2011). Fast and slow spindles during the sleep slow oscillation: Disparate coalescence and engagement in memory processing. *Sleep, 34*(10), 1411–1421. https://doi.org/10.5665/SLEEP.1290

Myers, E. B. (2014). Emergence of category-level sensitivities in non-native speech sound learning. *Frontiers in Neuroscience, 8*, 238. https://doi.org/10.3389/fnins.2014.00238

Myers, E. B., Johns, A. R., Sayako, F., & Xie, X. (2017). The invariance problem in the acquisition of non-native phonetic contrasts: From instances to categories. In *The Speech Processing Lexicon* (pp. 52–84). De Gruyter. https://doi.org/10.1515/9783110422658-004.

Ngo, H. V. V., & Staresina, B. P. (2022). Shaping overnight consolidation via slow-oscillation closed-loop targeted memory reactivation. *Proceedings of the National Academy of Sciences of the United States of America, 119*(44). https://doi.org/10.1073/pnas.2123428119

Pelzl, E. (2019). What makes second language perception of Mandarin tones hard? *Chinese as a Second Language 汉语教学研究, 54*(1), 51–78. https://doi.org/10.1075/csl.18009.pel

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461–472. https://doi.org/10.1121/1.3593366

Qin, Q. Z., Jin, R., & Wu, R. (preprint). The role of prior knowledge in second-language learners' overnight consolidation of Cantonese tones. https://doi.org/10.2139/ssrn.4978553.

Qin, Z., Gong, M., & Zhang, C. (2021). Neural responses in novice learners' perceptual learning and generalization of lexical tones: The effect of training variability. *Brain and Language, 223*, Article 105029. https://doi.org/10.1016/j.bandl.2021.105029

Qin, Z., Jin, R., & Zhang, C. (2022). The effects of training variability and pitch aptitude on the overnight consolidation of lexical tones. *Journal of Speech, Language, and Hearing Research, 65*(9), 3377–3391. https://doi.org/10.1044/2022_JSLHR-22-00058

Qin, Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech, 59*(3), 318–338. https://doi.org/10.1177/0023830915590191

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLOS ONE, 14*(12), Article e0221498. https://doi.org/10.1371/journal.pone.0221498

Qin, Z., Zhang, C., & Wang, W. S. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America, 149*(1), 435–446. https://doi.org/10.1121/10.0003330

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rasch, B. (2017). Sleep and language learning. *Brain and Language, 167*, 1–2. https://doi.org/10.1016/j.bandl.2017.02.002

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schreiner, T., & Rasch, B. (2017). The beneficial role of memory reactivation for language learning during sleep: A review. *Brain and Language, 167*, 94–105. https://doi.org/10.1016/j.bandl.2016.02.005

Simon, K. N. S., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L., & Gómez, R. L. (2017). Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain and Language, 167*, 3–12. https://doi.org/10.1016/j.bandl.2016.05.002

Studte, S., Bridger, E., & Mecklinger, A. (2017). Sleep spindles during a nap correlate with post sleep memory performance for highly rewarded word-pairs. *Brain and Language, 167*, 28–35. https://doi.org/10.1016/j.bandl.2016.03.003

Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *Journal of Neuroscience, 33*(39), 15376–15381. https://doi.org/10.1523/JNEUROSCI.5093-12.2013

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience, 30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., Bito, H., & Morris, R. G. M. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333*(6044), 891–895. https://doi.org/10.1126/science.1205274

Vallat, R., & Walker, M. P. (2021). An open-source, high-performance tool for automated sleep staging. *ELife, 10*, Article e70092. https://doi.org/10.7554/eLife.70092

van Schalkwijk, F. J., Sauter, C., Hoedlmoser, K., Heib, D. P. J., Klösch, G., Moser, D., Gruber, G., Anderer, P., Zeitlhofer, J., & Schabus, M. (2019). The effect of daytime napping and full-night sleep on the consolidation of declarative and procedural information. *Journal of Sleep Research, 28*(1). https://doi.org/10.1111/jsr.12649

Wamsley, E. J. (2019). Memory Consolidation during Waking Rest. *Trends in Cognitive Sciences, 23*(3), 171–173. https://doi.org/10.1016/j.tics.2018.12.007

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649–3658. https://doi.org/10.1121/1.428217

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language, 167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Wüst, L. N., Antonenko, D., Malinowski, R., Khakimova, L., Grittner, U., Obermayer, K., Ladenbauer, J., & Flöel, A. (2024). Interrelations and functional roles of key oscillatory activities during daytime sleep in older adults. *Journal of Sleep Research, 33*(3). https://doi.org/10.1111/jsr.13981

Xia, T., Yao, Z., Guo, X., Liu, J., Chen, D., Liu, Q., Paller, K. A., & Hu, X. (2023). Updating memories of unwanted emotions during human sleep. *Current Biology, 33*(2), 309–320.e5. https://doi.org/10.1016/j.cub.2022.12.004

Yip, M. J. W. (2002). *Tone*. Cambridge.

Zhang, C., & Chen, S. (2016). Toward an integrative model of talker normalization. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1252–1268. https://doi.org/10.1037/xhp0000216

Zhang, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in Psychology, 9*(OCT), 1–13. https://doi.org/10.3389/fpsyg.2018.01982

Zion, D. B., Gabitov, E., Prior, A., & Bitan, T. (2022). Effects of sleep on language and motor consolidation: evidence of domain general and specific mechanisms. *Neurobiology of Language, 3*(2), 180–213. https://doi.org/10.1162/nol_a_00060

Zion, D. B., Nevat, M., Prior, A., & Bitan, T. (2019). Prior knowledge predicts early consolidation in second language learning. *Frontiers in Psychology, 10*, 1–15. https://doi.org/10.3389/fpsyg.2019.02312