Carnegie Mellon University

Study Protocol and Statistical Analysis Plan (SAP)
NCT04509024
Incidental Auditory Category Training for Language Learning
1R03HD099382-01

February 26, 2024

Table I

	Incidental, Nonspeech (nonspeech)	Explicit, Speech (speech)	No Training (none)
Relevant Classroom, Mandarin L2 (+instruction)	+ instruction / nonspeech *Preliminary Study 2*	+ instruction / speech *Preliminary Study 2*	+ instruction / none
Irrelevant Classroom, French/Spanish L2 (-instruction)	- instruction / nonspeech *Preliminary Study 1*	- instruction / speech *Prior literature*	- instruction / none

Design Overview. We propose an 8-week training study among university students with no prior experience with Mandarin (or any tonal language). We will track learning through two modalities: explicit classroom instruction and web-based incidental training. Half of the participants will come from first-year Mandarin classes at Carnegie Mellon University (CMU). These students (+instruction, see **Table I**) will undergo traditional L2 learning through explicit instruction, as is the typical pedagogical approach of university classrooms. This approach emphasizes the four specific tonal contours and the lexical role of tone in Mandarin word recognition through explicit instruction. On average, these students will spend 200 minutes/week in classroom instruction, and an additional 300 minutes expected preparation outside of class. The other half of the participants will be recruited from the same university-student population, but will be engaged in non-Mandarin, non-tonal foreign language instruction for Spanish or French at an identical 200 minutes/week. In this manner, these participants (–instruction, **Table I**) will be similarly motivated to learn a foreign language, yet will have no tonal training: they will not be taught the four tone categories or the lexical role tone carries. In this way, the –instruction group models naïve participants typical of laboratory L2 training studies, with the added control of being an equally motivated cohort. Participants will undergo either supplementary incidental nonspeech training or supplementary explicit speech training. A third group will serve as a control and undergo no additional training.

Participants and Power. We will test 150 participants (N=25/condition). The sample size was selected so that multiple regression analyses with all predictors have high power (1-β=0.8, p<0.05, two-tailed) to detect even a small (Cohen's $f2$=0.08) reliable deviation from 0 of a single linear regression coefficient (G*Power). This sample size is twice that of previous laboratory incidental L2 training studies (N=13, 80; N=10, 63), appreciating that we will need a somewhat larger sample to accommodate the potential for attrition across the retention interval. Even if we were to experience as much as 33% attrition, this sample size will assure a robust test (N=17; Cohen's $f2$=0.08; power=.6) in line with prior research.

All participants will be native-English young adults (18-28 years) with normal hearing, no history of speech or language disorder, no prior experience with a tonal language, and background in musical training equated across groups. Half of the participants will be enrolled in first-year Mandarin at CMU (+instruction). The other half will be enrolled in first-year Spanish or French at CMU (–instruction). First-year Mandarin, Spanish, and French language courses are offered in Fall and Spring terms with 100 students completing each of these courses in 2017-2018. Instructors of each language group make use of the same, shared lesson plans and undergo thorough training to ensure uniformity. Our plan is to recruit participants across four academic semesters, presenting an excellent participant pool with the majority of enrolled students meeting inclusion criteria in recent semesters.

Protocol Details Common to All Participants. ***Pre-test Measures.*** Participants from each of the six groups (**Table I**) will undergo language, sensory, and cognitive assessments to examine factors that may interact with learning. *Online Battery of Questionnaires*. We will document participants' language background, dominant hand, musical training, and self-reported history of language

disorders, speech impairments, videogame expertise, and self-reported hearing loss using batteries established by prior research. *Basic Hearing Screening.* We will measure participants' pure-tone frequency thresholds using the adaptive maximum likelihood procedure as implemented in the MLP MATLAB toolbox, which converges quickly (~30 trials). *Cognitive Tests*. Since auditory working memory and phonological ability are associated with language learning, we will use the Woodcock-Johnson Tests of Oral and Cognitive Abilities to evaluate phonological awareness, working memory and auditory working memory using the Sound Blending, Numbers Reversed and Auditory Working Memory subtests, as in prior research. These tests will be administered before training and/or classroom instruction commences to accommodate exclusion criteria and to match groups on music training. Pre-test measures will allow assessment of group differences and individual scores will be used as regressors in analyses to examine the relationship of individual differences to learning outcomes using linear mixed-effect regression models in R to capture both fixed effects of linguistic/cognitive background and training, as well as the random effects of participant.

Measures to Assess Learning. Whereas +instruction learners will have been explicitly instructed in Mandarin tone in the classroom, –instruction listeners will not. This marked difference impacts the methods available to assess learning, in part because naïve listeners cannot overtly label the tones. To meet this challenge, we employ three assessments that do not depend upon explicit knowledge. *(1) Perceptual Tasks: Pitch Pattern Discrimination (PPD) and Pitch Direction Identification (PDI) Tests.* Prior research has shown that participants who exhibit good Mandarin tone learning tend to exhibit better sensitivity to F0 contours in pitch contour discrimination and identification tests. These two measures will allow us to ascertain pre- to post- and delayed post-test change in Mandarin tone perception in a manner that can be readily applied across +/– instruction groups, and to assess perceptual weighting of pitch contour versus pitch height (and its change with learning). Additionally, these tests provide an assessment of individual differences in F0 sensitivity (pre-test). The PPD and PDI tests will follow the approach of our prior research. We will use the PPD data to calculate the sensitivity (d') and reaction time (offset of the second stimulus to response) of each participant at pre-test, post-test (after instruction/training), delayed post-test (3 months after training). We will use multidimensional scaling (MDS; Individual Differences Scaling, INDSCAL) with the PPD data to examine the underlying perceptual dimensions listeners use to discriminate tone contours. This MDS approach models successful prior research and provides us with a perceptual metric of the impact of learning across groups that do not depend on explicit awareness of tone. We will use the PDI data to calculate accuracy for each participant at pre-test, post-test, and delayed post-test.

(2) Artificial language-learning task. Previous research from our team has demonstrated the feasibility of using an artificial Mandarin-like tonal language to track and test learning across speakers with different levels of overt awareness of Mandarin tone, such as learners early in explicit classroom instruction (+instruction) compared to naïve learners (-instruction). PI Wiener designed an artificial language that mimicked Mandarin's linguistic characteristics, including identical tone contours and syllable structures, but remained unintelligible to both English and Mandarin speakers. Because the artificial language was modeled on Mandarin, the results revealed not only how native speakers represent new Mandarin-like words, but also how non-native L2 learners acquire Mandarin-like tones in the context of the language acquisition task. In this multi-day

training and testing paradigm, participants learn 130 consonant-vowel syllable+tone nonce words, each paired with a black and white nonce symbol. Daily learning is assessed through a 4-alternative forced choice (4AFC) task in which participants hear a target syllable+tone and click on a symbol. Feedback guides learning. This study also provides a way to compare online tonal sensitivity across Mandarin L2 learners and naïve learners with no prior experience with tone.

Protocol Details for Incidental Nonspeech Training. Training will take place in the videogame paradigm. Participants will train 60 min/week across 8 weeks for a total of 8 hours, online at their convenience. Our prior research demonstrates that overall time-on-task is closely related to the number of stimulus exposures in the videogame, even though participants differ in ultimate *gaming* achievements like Highest Level achieved. This is because at higher, faster levels "trials" are shorter with fewer stimulus repetitions, but there are more of them. Thus, our principal concern is with time-on-task and not with equating experience across listeners, *per se*. Protocol compliance will be monitored electronically and regular contact with our research team will be tailored to encourage high retention, as has been successful in our prior research. Learners will train on a set of nonspeech hums derived from a larger set of Mandarin talkers (N=20 instead of N=4 in Preliminary Study 2).

The training regimen we propose is substantially longer (8 hrs over 8 wks) than has been investigated in prior studies of incidental auditory category learning (~2.5 hrs in 5 days; 62). *This presents the opportunity to take a data-intensive approach*. We will use the continuous, fine-grained online data generated in the videogame to address the fundamental question of how categories are formed via incidental learning.

Protocol Details for Web-based Speech Training. Groups undertaking explicit speech training also will train 60 min/week across 8 weeks for a total of 8 hours. Training sessions will be broken into 20-minute modules, each of which takes place online. Protocol compliance will be monitored electronically and regular contact with our research team will be tailored to encourage high retention. Explicit speech training groups will train on standard L2 audio-lingual materials, which aim to develop learners' phonological awareness of tone. These tasks include same/different tone discrimination, AXB discrimination, 2-alternative-forced choice, 4-alternative-forced-choice, and tone labeling (roughly 4 minutes per task). Though the tasks are currently designed to be used by students enrolled in the university curriculum, all tasks are purely perceptual and do not require linguistic knowledge. For example, a participant must perceive *ma1* and *ma3* as different sounds but not necessarily understand that the former means 'mother' and the latter means 'horse.' As a result, the participants in the –instruction condition can perform the tasks without explicit knowledge of Mandarin. The stimuli will be modified to include speech from 20 speakers (thus matching the nonspeech hum stimuli). In addition to task accuracy, we will track time-on-task.

Protocol Details Specific to +instruction Learners. *Classroom Metrics.* We will collect metrics that include class attendance, self-reported study time, and self-reported classroom social network and interaction with Mandarin speakers. These metrics will allow us to compare and control for potential differences that may arise between +speech and +nonspeech groups. Importantly, the classroom instructor will be blind to each participant's assignment so as not to affect his/her performance. *Mandarin Tone Production.* Whereas -instruction, nonspeech learners train on perceptual representations relevant to Mandarin tone, they never explicitly encounter Mandarin

in the course of training. Thus, it is not possible to elicit Mandarin tone speech productions among –instruction learners. (However, speech production measurements will be possible in the Artificial Language Learning task administered post-training. This provides an across-group assessment of tone production). But, since all +instruction learners are instructed explicitly in Mandarin tone within the classroom environment it is possible to examine their tone productions across the course of the study. We will elicit speech productions from +instruction participants following the approach taken in Preliminary Study 2 using native-Mandarin raters, with associated acoustic analyses of the F0 contour to characterize these speech productions.