University of Tartu

Faculty of Social Sciences

Institute of Psychology

Holger Peelmann

THE TRAINING EFFECT ON THE PERCEPTION OF ESTONIAN QUANTITY BY CHINESE NATIVE SPEAKERS

Research Paper

Supervisors: Siqi Lyu, Nele Põldver

Running head: Training effect on quantity perception.

Tartu

2023

Abstract

This study examines the effect of short-term training on Chinese (tonal language) participants perception of Estonian quantity (quantity language). A total of 60 Chinese participants whose first language (L1) is Mandarin were tested on their ability to discriminate between different Estonian (L2) stimuli, using a behavioural discrimination task (AX task). Participants were randomly assigned to three groups, with each group containing 20 participants. One group served as the control group (Group 1), while the other two groups received L2 word stimuli training (Group 2) and L2 pure tone stimuli training (Group 3). Firstly, the results showed that the Chinese participants were able to effectively discriminate between the Estonian words in the pre-test (mean hit rate= 91.39%). This could be attributed to the fact that two thirds of the stimuli pairs used in the discrimination task included a pitch feature, which would be familiar to speakers of tonal languages. Secondly, we found that there was no significant difference in improvement between two groups (Group 2 & Group 3) that received different types of stimuli training and the control group (Group 1), which did not receive any training. The lack of training effect could be attributed to the use of passive training task, or other factors such as the length of the training period.

Keywords: Chinese lexical tone, Estonian quantity, AX discrimination task, short-term training, language perception.

Lühikokkuvõte

Käesolevas töös uuritakse, kuidas tajuvad hiina keelt (mis on toonikeel) emakeelena kõnelejad eestikeelseid (mitte emakeelseid ehk L2) vältesõnu. Kokku testiti AX-diskrimineerimise ülesande abil 60 osalejat, kelle emakeel (L1) on Mandariini hiina keel. Osalejad jaotati juhuslikult kolme rühma, igas rühmas oli 20 osalejat. Üks rühm oli kontrollrühm (rühm 1), ülejäänud kaks rühma said kas L2 lihtsalt sõna-stiimulite treeningut (rühm 2) või L2 puhast tooni sõna-stiimuli treeningut (rühm 3). Tulemused näitasid, et osalejad suutsid treenigu eelses eeltestis eestikeelseid sõnu tõhusalt eristada. Keskmine tabamusmäär oli 91.39%. Selle põhjuseks võib olla asjaolu, et kaks kolmandikku diskrimineerimisülesandes kasutatud sõna-stiimulipaaridest sisaldasid helikõrguse tunnust, mis on tuttav tonaalsete keelte kõnelejatele. Teiseks avastasime, et rühmade (rühm 2 ja rühm 3), kes said erinevat tüüpi sõna-stiimulitreeningut, ja kontrollrühma (rühm 1), kes ei saanud stiimulitreeningut, vahel ei olnud olulist erinevust tulemuste paranemise osas. Treeningefekti puudumine võib olla tingitud passiivse treeningülesande kasutamisest või muudest teguritest, näiteks treeningperioodi pikkusest.

Märksõnad: Hiina keel, eesti keele välted, AX diskrimineerimisülesanne, lühiajaline treening, keeletaju.

Introduction

Speech is constructed by a series of components that occur while conversing. Processing acoustic information of speech, such as speech-sound duration and pitch are the key features of speech perception. Early in childhood, categorical perception of speech sounds is thought to arise, and it is further influenced by learning one's first language (L1) (Kuhl et al, 1997). Children can learn their first native language simply by being exposed to it and experiencing social interaction with others in their imminent vicinity. It is thought that language learning has a sensitive period during which a person might master a new language structure and vocabulary (Hurford, 1991). Neurological maturation has been cited as the primary motivation for speech learning during a vital period. Many scientists feel that once a critical period has passed, a new kind of language cannot be mastered perfectly (Lantolf, 1990; Flege, 1995).

It is generally known that native language influences the perception and production of second language (L2) speech. The use of duration cues and pitch varies by language, which has implications for learning the phonological system of a second language (L2) (Han, 1992; Kato & Tajima, 2002; McAllister et al., 2002). The overall purpose of this study was to see if short-term training could improve adult listeners' perception of a non-native quantity stimuli. To this end, stimuli from a quantity language, Estonian, were applied to Chinese subjects who speak a tonal language and do not incorporate duration/quantity cues in their L1.

Chinese Lexical Tone

Languages that exploit phonologically contrastive variations in pitch at the word or syllable level are called tonal languages. The tone of such languages is produced by high or low pitch of the voice, altering the vibration of vocal cords, and pitch patterns (e.g., change from higher to lower pitch is known as falling tone) (Wang, 2013). To learn Mandarin Chinese (Mandarin or Chinese hereafter), one needs to learn how to produce four basic tones with their voice: high level, high rising, low falling-raising and high falling. This is known as using pitch contour with their voice. Changing of the pitch contour will result in the change of the meaning of the word.

Estonian Quantity

The importance of duration signals is particularly apparent in so-called quantity languages (languages that use duration as the principal cue for phonological distinction evaluation and pitch cue as a secondary one). Short (Q1), long (Q2) and overlong (Q3) quantity degrees are common in Estonian (e.g., [*sa:da*] ‘hundred’ (Q1), [*sa:.ada*] ‘to send’ (Q2) and [*sa;.ada*] ‘to get’ (Q3)) as opposed to tonal languages that rely on the use of tone to change meanings of the word when words have the same sequence of vowels and consonants.

Estonian as opposed to Mandarin is one of the so-called quantity languages where sound duration physically indicates quantity distinction, hence duration is an important component of its phonology. The three-way quantity distinction, which presents itself as a complex prosodic structure that incorporates durational and tone components, is a major aspect of Estonian word prosody.

Pärtel Lippus (2009) with colleagues gives examples of such quantity occurring in two-syllable words (disyllabic): short (Q1), long (Q2) and overlong (Q3) quantity degrees. In Estonian, verbal stress is placed on the first syllable. This is where phonologically temporal structure can be heard in the stressed vowel. For example, [*vilu*] means "chilly" and is short, [*vi:lu*] means "slice" and is long, and [*vi::lu*] also means "slice" and is overlong. Verbal stress is also placed on syllable-medial consonants, such as [*kal’i*] which means "kvass", [kal’li] which means "hug", and [*kal’:li*] which means "precious". Verbal stress can also be placed on a combination of a stressed vowel with a following consonant, such as [sate] which means "fall-out", [*sa:tte*] which means "get", and [*sa:t:te*] which means "broadcast" (Lippus et al., 2009).

Native Language and Perception of a Foreign Language

There are a few different ways that qualities given in one's native language might be easier to distinguish than qualities given in a foreign language. One reason is that there are likely to be stronger memory traces and greater accessibility for words and phrases that are learned and used frequently in the native language. This is because the brain is more efficient at processing and recalling information that is familiar and has been encountered repeatedly. Additionally, the brain tends to undergo a process called "long-term tuning" when learning a language, in which the neural networks that support language processing become more specialized and efficient over time. This process is thought to be more pronounced for the native language, which is learned and used on a daily basis, compared to a foreign language, which may be studied for a limited time and used less frequently (Gibson & Gibson, 1955).

It has been shown that native language background affects the auditory perception of a foreign language. In a 2001 study, it was discovered that there is a possible difference in lateralization between tonal (Mandarin) and non-tonal (English) languages, that tone does not carry the meaning of the word in non-tonal languages, and tone may be understood as prosodic information (Klein et al., 2001).

In Estonian L1 listeners, pitch is an essential indication in recognizing the overlong quantity (Q3), however it has little effect on people to whom Estonian is not a first language (i.e., L2 listerners). Without considering the pitch cue, Finnish and Russian L1 listeners correctly acquire the Estonian three-way quantity distinction. Latvian L1 listeners have some difficulty distinguishing between the long and overlong amount degrees, which could be explained by their native language's tonal and temporal contrasts (Lippus et al., 2009). Quantity language is difficult to master for L2 learners if speech-sound duration is less essential in their L1 (Lippus et al., 2009).

It is possible that having two-way length-system contrasts in one's L1 language will assist L2 listeners to distinguish equivalent two-way length contrasts in a non-native language (Lee et al., 2021). Lee and others in their study found that Japanese listeners were significantly better than Cantonese listeners at discriminating Estonian word length differences. Segment type and length contrast had substantial effects as well. This study showed that vowels had higher discriminating accuracy than consonants and indicates that the overlong vs. long contrast was more difficult to discriminate across participants native languages than the long vs. short contrast. The fact that the length contrast had a significant interaction with the participants L1 indicates that the effect of the length contrast was greater for Japanese listeners than for Cantonese listeners. The Japanese listeners may have been able to detect identical two-way length disparities in a non-native language since they have systematic two-way length contrasts in their L1.

Training Effect

There are different examples of auditory training effects on the perception of linguistic stimuli. When studying neurophysiological changes and behavioural perceptive capacities, nine people who spoke only English were taught to recognize a prevoiced labial stop sound (one that is not used phonemically in the English language) (Tremblay et al., 1997). The subjects were instructed to differentiate and identify a pre voiced alveolar consonant after training. Before and after training, mismatch negativity (MMN) cortical evoked responses were measured to both labial and alveolar stimuli. An age-matched control group that did not receive training was likewise tested for behavioural performance and MMNs. The experimental group's ability to discriminate and identify an unknown voice onset time improved after training. The findings show that auditory training can be applied to listening contexts other than those utilized in training sessions, and that auditory training can change the pre attentive central neurophysiology that supports perceptual learning (Tremblay et al., 1997).

In the cross-language perception of tones and training studies it was concluded that native tonal language (Chinese) listeners outperformed native non-tonal language (English) listeners under different inter-stimulus interval (ISI) conditions before and after training. This shows that persons whose native language is also tonal language find it easier to acquire a tonal language as a second language (Wayland et al., 2008; Wayland & Guion., 2004). Under two ISI circumstances (500 ms and 1500 ms), researchers tested two perceptual training approaches that might be utilized to improve native English and native Chinese listeners ability to discern the mid- vs. low-tone contrast in Thai. Participants were taught either a categorical same/different discrimination process or a two-alternative forced-choice identification procedure. Native Chinese listeners discrimination was better under the longer ISI before training, whereas Native English listeners were better after training (Wayland et al., 2008). These findings suggest that forced-choice identification and categorical same/different discrimination training procedures were equally effective in improving native English and native Chinese listeners' discrimination.

Kaan et al. (2007) conducted a similar experiment where they measured an event-related potential (ERP) to see if native speakers of tonal and non-tonal languages differ in their pre-attentive discriminating of Thai lexical tones, and if training had a distinct effect. In an unusual paradigm, EEGs were collected from 10 native Mandarin Chinese speakers, 10 English speakers, and 10 Thai speakers. Following that, the Chinese and English participants went through a two-day perceptual identification training on mid- and low-level falling tones, and then returned for a post-training EEG. Findings of this experiment imply that non-Thai speakers of another tonal language can identify Thai tones but that they are sensitive to distinct physical features of the tones depending on their native language. Early pitch disparities are more responsive among English speakers, but the later pitch contour is more sensitive in Mandarin Chinese native speakers (Kaan et al., 2007).

To better understand and determine the training effects, another similar study was conducted on native Finnish speakers, where the researchers Tamminen, Peltola, Kujala and Näätänen used behavioural identification, goodness rating, discrimination, and reaction time tests, as well as mismatch negativity brain response registrations (Tamminen et.al., 2015). The subjects were taught how to distinguish between voicing contrasts in fricative sounds. In Finnish, voicing does not distinguish fricatives, and voiced fricatives are not part of the phonological system. The native Finnish turned out to learn to recognize the difference after just three days of training, as results showed that the MMN reaction had changed dramatically (Tamminen et.al., 2015). The results suggest that participants can attain language-learning effects at the perceptual and pre-attentive brain levels as a result of brief listen and repeat draining.

Hypotheses

H1: Chinese participants will show a low sensitivity to Estonian quantity words in the pre-test.

H2: Chinese participants who received training on Estonian quantitative pure tone and word stimuli (group 2 and group 3) will outperform those who did not receive training (group 1).

Method

Study Design

The participants in the experiment visited the lab twice. Between the pre- and post-measurement visits to the lab, the participants underwent a 5-day training period. Depending on which group they were assigned to, the participants received either L2 pure tone training, L2 word stimuli training, or no training at all for the control group.

Participants

The participants of this study were all native Mandarin speakers, total of sixty (*M*=22.07 years, *SD*=1.74) (33 females and 27 males). The participants were randomly assigned to 3 different groups (see table 1): Group 1 (*N*=20) (being control group), Group 2 (*N*=20) and Group 3 (*N*=20). All the participants were right-handed. The highest level for completed education was high school (*N*=32) or bachelor's degree (*N*= 28).

Table 1

Description of the participants and groups.

	Training	Females	Males	Mean age	SD	R
Group 1	Control group	10	10	22.15	1.69	18-25
Group 2	Pure tone	12	8	21.50	1.50	20-25
Group 3	L2 stimuli	11	9	22.45	1.50	20-25

** SD=Standard deviation, R=Age range, L2 stimuli= language stimuli in Estonian*

Stimuli

AX experiment stimuli.

The stimuli used in the experiment were resynthesized from the Estonian word SADA, either in Q2 (long) or in Q3 (overlong) (see table 2), using Praat (Boersma, 2001). The length of the initial vowel was manipulated as either 170 milliseconds (short) or 290 milliseconds (long), resulting in four stimuli in total, i.e., Q2_170 (Q2, short), Q2_290 (Q2, long), Q3_170 (Q3, short), and Q3_290 (Q3, long). We also included non-speech stimuli (i.e., pure tones), which had the same parameters as the word stimuli.

Table 2

Parameters of the Estonian word SADA.

	Consonant 1	Vowel 1	Consonant 2	Vowel 2
	S	A	D	A
Q2_170	90 ms	170 ms	74 ms	115 ms
Q2_290	90 ms	290 ms	74 ms	115 ms
Q3_170	100 ms	170 ms	90 ms	68 ms
Q3_290	100 ms	290 ms	90 ms	68 ms

** Q3= overlong quantity, Q2=long quantity.*

The stimuli were combined into all possible pairs. Four were the same pairs repeated 15 times and the remaining 12 were different pairs repeated 5 times each. In total, there were 120-word pairs and an additional 120 pure tone pairs, resulting in a total of 240 pairs.

Table 3

Word pairs produced.

A	X	Repetitions
Q2_170	Q2_290	5
Q2_170	Q3_170	5
Q2_170	Q3_290	5
Q2_290	Q3_170	5
Q2_290	Q3_290	5
Q3_170	Q3_290	5
Q2_290	Q2_170	5
Q3_170	Q2_170	5
Q3_290	Q2_170	5
Q3_170	Q2_290	5
Q3_290	Q2_290	5
Q3_290	Q3_170	5
Q2_170	Q2_170	15
Q2_290	Q2_290	15
Q3_170	Q3_170	15
Q3_290	Q3_290	15

** A=first sound, X=Second sound, Repetition=number of times a pair of sounds was repeated,*

Training stimuli.

For Group 2 (*N*=20), five Estonian word triplets (short, long, overlong) in natural speech were used to reflect Estonian quantity changes as training stimuli. Words used were SAGI (short [*sagi*] - ´rush´, long [*sa:gi*] - ´catch, harvest´, overlong [*sa::gi*] - ´saw´), SAGE (short [*sage*] - ´frequent´, long [*sa:ge*] - ´receive´, overlong [*sa::ge*] - ´saws´), VIDE (short [*vide*]´ , long [*vi:ide*] ´to refer´, overlong [*vi::ide*] – ´number fifth, five´), LIGA (short [*liga*] - ´dirt´ , long [*li:ga*] - ´union´, overlong [*li::ga*] - ´too much´) , VIGI (short [*vigi*] , long [*vi:gi*] - ´to iron clothes´, overlong [*vi::gi*] -to ´bring´).[1]

For Group 3 (*N*=20), pure tone stimuli were created based on two of the above Estonian word triplets SAGI and VIDE.

[1] Note that the short variant of the words VIDE and VIGI do not have any meaning in Estonian, but as for the Chinese, none of the stimuli carry meaning and are interpreted as pseudowords, it should not affect the results.

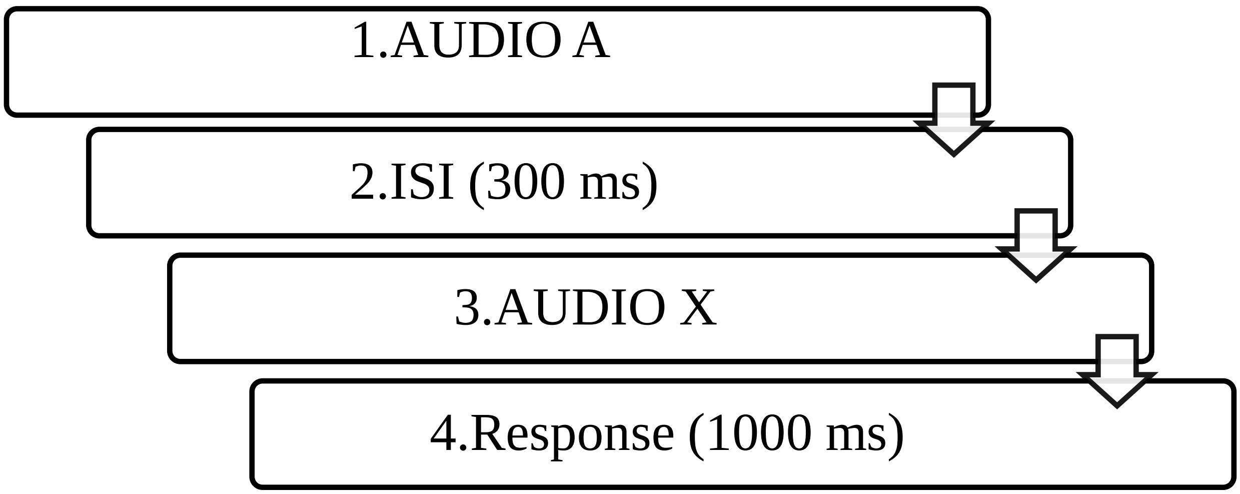
Procedure

AX Discrimination.

In speech research, the same – different design (AX) is extremely common. In this design two stimuli are presented together as a pair. Each stimulus presentation is separated by an interstimulus interval (ISI). The stimuli in the trial are paired so that they are either the same or different and the participants objective is to figure out which of them was introduced to them (McGuire, 2010; Gerrits & Schouten, 2004).

Figure 1

Procedure for each trial.



** Audio A= first word, ISI= interstimulus interval, Audio X=second word, Response=time to respond, ms=millisecond.*

In this AX experiment, all stimuli pairs were pseudo randomized. Word pairs and pure tone pairs presented are listed in Table 3. Subjects were sitting in a quiet and dimly lit room, used headphones to listen to the stimuli and a computer keyboard to give their answers. The instructions for the experiment were given orally before the experiment as well as on the computer screen in Chinese. E-Prime 3 (Psychology Software Tools, Pittsburgh, PA, United States) was used for presenting stimuli. Each experiment (figure 1) started with 10 practice trials where the participant was played a non-word called *tada*. A total number of 240-word pairs were played for each participant following the procedure in Figure 1. A participant in the experiment was given a stimulus (A stimulus), followed by interstimulus interval (300 ms) after that another stimulus (X stimulus) was given. Upon hearing both stimuli the participant had 1000 milliseconds to answer whether the two stimuli they heard were the same or different. Participants responded by pressing "F" or "J" on the keyboard (F represented “same” and J represented “different”).

Stimuli were presented in blocks, but in different orders for pre- and post-tests (see Table 4). Within each block, the same and different pairs were presented pseudo randomly.

Table 4

Word block and lists.

Pre-test	Words pairs	Post-test	Word pairs
Block 1	L2 pure tone (N=60)	Block 1	L2 (N=60)
Block 2	L2 (N=60)	Block 2	L2 pure tone (N=60)
Block 3	L2 pure tone (N=60)	Block 3	L2 (N=60)
Block 4	L2 (N=60)	Block 4	L2 pure tone (N=60)

* L2 pure tone= p*ure tone Estonian words, L2= Estonian stimuli words.*

Training

The training occurred on a web-based platform called LimeSurvey (https://survey.ut.ee). Participants were provided with a daily link to a 30-minute audio recording, which they were asked to listen to on their personal electronic devices in a quiet environment with earphones. Additionally, they were instructed to read a book in Chinese and ignore the sounds while listening to the audio.

Group 2 received one word triplet (SAGI, VIDE, SAGE, LIGA, VIGI) each day. The order of the five triplets on each day was randomized among subjects.

Group 3 received the pure tone stimuli created from one of the two word-triplets (SAGI, VIDE) each day. The order of pairs each day was randomized among subjects.

Every day, there were three training segments. Each training segment featured a pair of words with two possible lengths from the three quantities (i.e., Q1-Q2, Q1-Q3, and Q2-Q3). In each segment, the word pair was repeated 200 times with a 300 ms pause in between (e.g., Q1-300 ms-Q2-300 ms-Q1-300 ms-Q2…). There was a 1500 ms interval after the 200 repetitions, after which the order of the stimuli was switched (e.g., Q2-300 ms-Q1-300 ms-Q2-300 ms-Q1…). Each segment lasted for 10 min in total. After each segment, there was a 30-second pause, followed by two other segments with the other two word- pairs, i.e., Q1-Q3 and Q2-Q3.

Results

Due to their failure to respond to more than 50% of the trials, three participants' data were excluded from the results.

To calculate D-prime values, each subject's Hit and False alarm values were calculated in Microsoft Excel (Microsoft Corporation, 2018) by adding up hit and false alarm occurrences. Hit stands for when two words presented were correctly identified as being different, while False alarm stands for when two words presented were the same, yet the test subject heard them as different. This was followed by calculating D-Prime values in Excel following the instructions by Keating (2005). The descriptive statistics are shown in Table 5.

Table 5

D-Prime values, Hit & False alarm calculated values. Means are presented with SDs in paratheses.

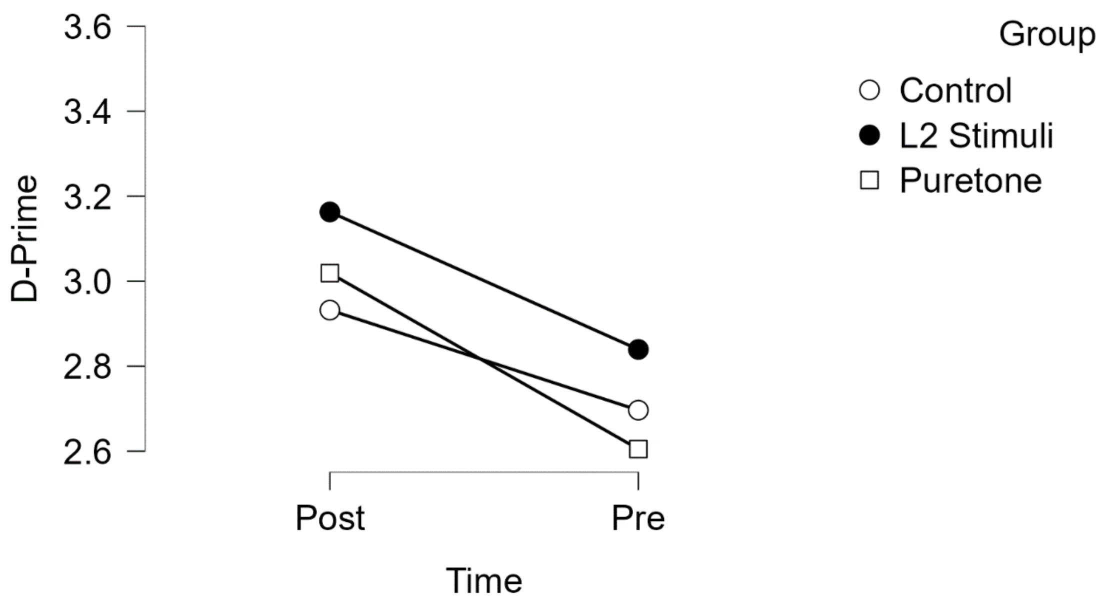
Time	Group	D-prime	Hit	False alarm	N
Pre	Control	2.69 (0.68)	91.2 (17.31)	4.65 (5.48)	20
	L2 stimuli	2.83 (0.35)	92.52 (12.60)	3.41 (3.31)	17
	Puretone	2.60 (0.78)	90.45 (15.84)	7.2 (15.29)	20
Post	Control	2.93 (0.69)	96.10 (17.70)	4.7 (5.84)	20
	L2 stimuli	3.16 (0.58)	100.76 (12.50)	3.58 (4.10)	17
	Puretone	3.01 (0.95)	96.7 (16.59)	5.56 (11.91)	20

** Post=post-test, Pre=pre-test, Control=control group, L2 Stimuli=pure tone Estonian words, Puretone= pure tone stimuli words, D-prime= values for D-prime, SD=standard deviation, Hit mean= average of correct answer, False alarm mean=average of wrong answer, N= sample size*

Data was then forwarded to JASP to be used for statistical analysis. An ANOVA (analysis of variance) is a statistical test that is used to compare the means of two or more groups. It is used to determine whether there are significant differences between the groups. In my case, the ANOVA was conducted using the JASP software. The D-Prime values, which represent the sensitivity of the participants in detecting the stimuli, were used as the dependent variable. Two fixed factors were also included in the ANOVA: Time (pre-test and post-test) and Group. The results of the ANOVA are presented in Table 6 and Figure 2. These results allow me to determine if there are significant differences in the D-Prime values between the different groups and conditions, and whether the training had an effect on the participants sensitivity to the stimuli.

Figure 2

D-Prime scores difference between pre-test and post-test.



** Pre=pre-test, Post=post-test, Control=Control group, Puretone=Estonian puretone stimuli training, L2 Stimuli= Estonian L2 Stimuli training.*

Table 6

Summary table for a two-way ANOVA of D-Prime results

Cases	Sum of Squares	Df	Mean Square	F	P
Time	2.981	1	2.981	5.891	0.017
Group	0.844	2	0.422	0.834	0.437
Time ✻ Stimuli	0.159	2	0.080	0.157	0.855
Residuals	54.642	108	0.506		

** F=Variation between the sample mean, p= Probability value, Mean Square= Average of the squares of all numbers, df=Degrees of freedom.*

Results showed that time between pre-test and post-test has an effect, with post-test scores being higher than scores in pre-test (<0.017). The probability of the outcome occurring by random is less than 0.017, according to a p-value of 0.017. This is normally seen as having a very low chance and would be significant statistically. In other words, it would suggest that the outcome is more likely to represent a real consequence and less likely to have happened by coincidence.

The analysis of the rest of data showed that there was no statistically significant effect of group on the participants performance on test. Specifically, it is found that the type of group participants was assigned to had no effect on their values (<0.437).

Additionally, results showed that there was no statistically significant interaction between time and group on the participants performance. It is found that the effect of time on the participants performance did not vary depending on the type of group used in the experiment (<0.855). This suggests that the training had the same effect on the participants performance, regardless of the type of the training stimuli. Overall, these results indicate that the short-term training had no effect.

Discussion

The purpose of this study was to test the following hypotheses: (H1) Chinese participants will show a low sensitivity in distinguishing different Estonian quantity words in the pre-test; (H2) Chinese participants who received training in Estonian quantitative pure tone and word stimuli will outperform those who did not receive training.

Chinese Native speakers' Perception of Estonian Quantity

The three participating groups did make distinctions between Estonian words. As a result, H1 was not confirmed. Based on the data presented in Table 5 and Figure 2, it appears that the mean D-prime scores for all groups are above 2.6. The D-prime score is a measure of sensitivity in a signal detection task, with higher scores indicating better performance, calculated based on the hit rate and false alarm rate. If we look at the hit rate and false alarm rate in Table 5, the mean hit rate in pre-test was above 90%, and the mean false alarm rate was below 10%. This is not as bad as we had initially expected.

This suggests that Chinese participants, despite having no background knowledge of the Estonian language, were able to effectively discriminate the Estonian quantity. One explanation for this is that two thirds of the stimuli used in the AX discrimination task involved a pitch cue. In the Estonian three-way quantity system, Q3 involves a pitch cue (Lippus 2009). According to Table 3 in this thesis, two thirds of the pairs included Q3, which means that tonal speakers would be familiar with the way these words are produced.

Mandarin speakers and native Estonian listeners both rely on pitch as a key cue in their native languages. This means that both groups use the changes in pitch of a speaker's voice to distinguish between different words or sounds. In Mandarin, pitch changes can indicate different tones, such as rising or falling tones, which are important for distinguishing between words with different meanings. In Estonian, pitch changes can also be used to distinguish between long and overlong quantities. Therefore, it is likely that Mandarin speakers and native Estonian listeners have similar pitch perception abilities. Because the Chinese language employs a tonal system, where a word's pitch can alter its meaning (Lin, 2007), Chinese speakers are often regarded as having high pitch discrimination skills. This skill is developed and refined through regular exposure to and use of the language, and it becomes an important part of their ability to understand and produce words and phrases in Chinese.

In previous studies, without considering the pitch cue, Finnish and Russian L1 listeners were able to grasp the Estonian three-way quantity distinction without relying on pitch cues. However, due to the tonal and temporal contrasts in their native tongue, Latvian L1 listeners exhibit some confusion between the long and overlong quantity degree and seem to be relying on more on a tone (Lippus, 2009). Using a tone queue in a manner similar to Latvian, native Chinese speakers who heavily rely on pitch cues for their L1 used their L1 expertise to complete the listening task for Estonian quantity (L2) in the current study.

Training Effect on Chinese Native Speakers' Perception of Estonian Quantity

It is shown that regardless of the type of training, individuals improved to the same degree even after being divided into three groups and given training tasks over a five-day period. In other words, even though they did not receive any training, the control group also displayed improvement, rendering H2 essentially unconfirmed.

In previous studies, the results of laboratory training suggest that using either a two-alternative forced-choice identification procedure or a categorical same/different discrimination training procedure were equally effective in improving the ability of native English and native Chinese listeners to discriminate between the mid-tone and low-tone contrast in Thai (Kaan et al. 2007). Additionally, these studies suggest that individuals with prior experience with a tone language may have an advantage in learning another tone language. This is because tonal languages, in which pitch changes convey meaning, require the listener to be able to discriminate between different pitch contours in order to understand the language (Wayland, 2008).

However, we did not discover any training effects on the native Chinese speaker's perception of Estonian quantity in the present research. This could be attributed to numerous reasons. The first reason could be the passive task used in the training. When speaking about a passive training task, it is unlikely to have an effect on an attentive AX discrimination task because passive training tasks do not require the participant to actively engage or perform any actions, and therefore may not be engaging or challenging enough to produce a meaningful learning experience. In such cases, the participants are not paying attention during the passive training task, it is unlikely that they will be able to learn from it and transfer the learning to the attentive AX discrimination task. On the other hand, if the passive training task is engaging and challenging enough, and the participants are paying attention.

Secondly, in our AX paradigm test, participants were presented with a series of repeated stimuli and asked to respond to each one. Over the course of the test, the participants may become familiar with the stimuli, which can have a positive effect on their performance when on task. A third plausible explanation is that since pitch cues are employed in two out of every three stimuli pairs used in both the AX discrimination task and the training task, Chinese speakers are simply adept at recognizing it because their own language employs similar cues.

Limitations and Implications

The five-day training session was a passive training task in which the participants did not consciously memorize the Estonian quantity. By switching to attentive training tasks that require the participants to practice during the 5-day period rather than letting them passively listen to the Estonian words, future research might achieve more accurate results.

The two groups that received training were allowed to train on their own, without supervision, in the comfort of their own homes. This means that they were not required to come to a specific location for training and were free to schedule their training sessions according to their own availability and preferences. The researchers did not monitor whether the participants in these two groups followed the training schedule, or if they deviated from it in any way. This lack of supervision and monitoring allows for greater flexibility and autonomy for the participants, but it also means that the researchers cannot be certain that the participants received the same amount and quality of training as intended.

My findings call for more data analysis. Additional research should be done using different language groupings. Looking at the short(Q1), long(Q2), and overlong(Q3) data separately would be helpful to determine if there are any noteworthy differences in recognition pattern.

To see if Estonians respond similarly to Chinese phrases that are similar to Estonian long and overlong tones, future research can include both Estonians and Chinese individuals and compare their responses to the phrases with similar long and overlong tones. By doing this, we can determine whether Estonians and Chinese individuals have similar responses to these phrases, or if there are differences in their responses due to cultural or linguistic differences. This will allow us to gain a better understanding of how language and culture can influence the perception of tones in speech.

Acknowledgements

I would like to express my sincere gratitude to Siqi Lyu and Nele Põldver, who served as my thesis supervisors during the writing of this document. I am truly grateful for their guidance, support, and encouragement throughout the entire writing process. Their expertise, insights, and constructive feedback were invaluable in helping me to develop and refine my ideas, and I greatly appreciate their patience and persistence in helping me to overcome the various challenges and obstacles that I encountered along the way. I would not have been able to complete this thesis without their support and guidance, and I am deeply grateful for their contribution to my academic development.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9), 341-345.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 233–277). York Press

Gerrits, E., & Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363-376. https://doi.org/10.3758/BF03194885

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: differentiation or enrichment? *Psychological Review*, 62, 32-41. https://doi.org/10.1037/h0048826

Han, M. S. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica*, 49(2), 102-127. https://doi.org/10.1159/000261906

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition*, 40(3), 159–201. https://doi.org/10.1016/0010-0277(91)90024-X

Kaan, E., Wayland, R., Bao, M., Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain research*, 1148, 113-122. https://doi.org/10.1016/j.brainres.2007.02.019

Kask, L., Põldver, N., Kreegipuu, K., Lippus, P. (2020). Processing of language specific stimuli among Estonian and Russian native speakers: an EEG study. https://www.frontiersin.org/articles/10.3389/fnhum.2021.612617/full

Kato, H., & Tajima, K. (2002). Native and non-native perception of phonemic length contrasts in Japanese: A categorization study. *The Journal of the Acoustical Society of America*, 112(5), 2387-2387. https://doi.org/10.1121/1.4779718

Keating, P. (2005) D-prime (signal detection) analysis. UCLA Phonetics Lab statistics. http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm

Klein, D., Zatorre, R.J., Milner, B., Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, 13(4), 646-https://653.10.1006/nimg.2000.0738

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684-686. https://10.1126/science.277.5326.684

Lantolf, J. P. (1990). A time to speak: a psycholinguistic inquiry into the critical period for human speech. *Studies in Second Language Acquisition*, Volume 12, Issue 1. 84 – 85. https://doi.org/10.1017/S0272263100008779

Lee, A., Shinohara, Y., Mut, T. C. (2021). Non-native length contrast perception by Japanese and Cantonese speakers. In Proceedings of Meetings on Acoustics 181 *Acoustical Society of America*. Volume 45, Issue 1. https://doi.org/10.1121/2.0001529

Lin, Y-H. (2007). The Sounds of Chinese. Cambridge University Press. p14.

Lippus, P. Pajusalu K., Allik, J. (2009). The tonal component in perception of the Estonian quantity, *Journal of Phonetics* 37. https://doi.org/10.1016/j.wocn.2009.07.002

McAllister, R., Flege, J. E., Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of phonetics*, 30(2), 229-258. https://doi.org/10.1006/jpho.2002.0174

McGuire, G. (2010). A Brief Primer on Experimental Designs for Speech Perception Research. UCSC. https://people.ucsc.edu/~gmcguir1/

Tamminen, H., Peltola, M. S., Kujala, T., Näätänen, R. (2015). Phonetic training and non-native speech perception—New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, 97(1), 23-29. https://doi.org/10.1016/j.ijpsycho.2015.04.020

Tremblay, K., Kraus, N., Carrell, T. D., McGee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, 102(6), 3762-3773. https://doi.org/10.1121/1.420139

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, 97(1), 144-160.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54(4), 681-712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wayland, R.P., & Lin, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250-267. https://doi.org/10.1016/j.wocn.2007.06.004

Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.

Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis DSpace.

Holger Peelmann