Tartu University
Faculty of Social Sciences
The Institute of Psychology

Hanna Pria Hulkko

LEARNING ESTONIAN QUANTITY AMONG SPEAKERS OF A TONAL LANGUAGE

Master's thesis

Supervisor: Siqi Lyu, PhD

Running head: Learning Estonian quantity by Chinese

Tartu 2026

Eesti keele kontrastide õppimine tonaalsete keelte kõnelejate poolt

Kokkuvõte

Leidub vähe uurimusi, mis käsitleksid eesti keele väldete õppimist tonaalsete keelte kõnelejate poolt. Käesoleva uuringu eesmärk oli seda uurimislünka täita. 60 hiina keele kõnelejat osalesid 8-päevases treeningeksperimendis. Osalejad jaotati juhuslikult treening- ja kontrollrühma. Uuringus kasutati AX-eristusülesannet pseudosõnadega, mis esindasid eesti keele vältekontraste. Nii 1. kui ka 8. päeval läbisid osalejad eeltest–treening–järeltest protseduuri ning 2.–7. päeval osalesid nad kaugtreeningus. Kontrollrühmas ei ilmnenud katse jooksul väldete taju paranemist. Treeningrühma puhul paranes nii kestus- kui ka helikõrguspõhise kontrasti taju pärast lühiajalist treeningut 1. päeval märkimisväärselt. Helikõrguse taju jõudis eesti emakeelega kõnelejate tasemele, samas kui kestuspõhine taju jäi emakeelsete tasemele alla. Pärast kuuepäevast treeningut paranesid nii kõrgus- kui ka kestuspõhine taju veelgi. Kõrguspõhine taju paranes väiksemas ulatuses ja ületas emakeelsete taseme. Kestuspõhine taju paranes kõrguspõhisega võrreldes rohkem, kuid ei jõudnud emakeelsete rääkijate tasemeni. Tulemused viitasid sellele, et kestuse ja kõrguse õppimistrajektoorid olid erinevad. Kestuspõhine taju algas madalalt tasemelt ja paranes ulatuslikumalt kui kõrguspõhine taju, kuid ei saavutanud emakeelsete taset. Kõrguspõhine taju algas kõrgel tasemel, paranes ning ületas treeningu tulemusel emakeelsete taseme. Tulemused viitavad seostele emakeelekogemuse ja teise keele fonoloogiliste tunnuste taju vahel ning illustreerivad seda, et tunnuste taju jääb treeningu käigus muudetavaks.

Märksõnad: eesti keele vältesüsteem, mandariini hiina keel, lühiajaline treening, üldistumine, helikõrgus, kestus

Learning Estonian quantity among speakers of a tonal language

Abstract

There is little research about tonal language speakers learning Estonian quantity degree system. This study aimed to address that gap. 60 Mandarin Chinese speakers participated in an 8-day training experiment. They were randomly assigned to training or control group. An AX discrimination task was used with pseudoword stimuli representing Estonian quantity contrasts. Participants underwent a pretest-training-posttest procedure on both day 1 and day 8, and on days 2-7, they underwent remote training. The control group showed no improvement throughout the experiment. For the training group, both duration and pitch perception improved significantly after the short-term training on day 1. Pitch perception reached an Estonian native listener level, whereas duration perception remained below native-like level. After the 6-day training, both pitch and duration perception showed further improvement. Pitch perception improved to a smaller extent and exceeded native-like performance. Duration perception showed a larger improvement but did not reach native-like level. Our results showed a different learning trajectory of duration and pitch. Duration perception started at a low level and improved to a greater extent than pitch, but still did not reach native level. Pitch perception started high, improved and exceeded native level as a result of extensive training. Results suggest that the L2 phonological feature perception is shaped by L1 experience and remains plastic.

Keywords: Estonian quantity, Mandarin Chinese, short-term training, generalization, pitch, duration

Learning Estonian quantity among speakers of tonal languages

The production and acquisition of a second language (L2) is known to be shaped by learners' native language (L1) (Best & Tyler, 2007; McAllister et al., 2002). Native language affects the sensitivity to phonetic variation which can make learning a new language challenging for adult learners (Chandrasekaran et al., 2010; Grenon et al., 2019; Saloranta et al., 2020). It can be especially difficult if the contrasts present in L1 are used differently or are absent in L2. These difficulties, however, can be improved through auditory training (Grenon et al., 2019; Kaan et al., 2007).

Research on second language (L2) acquisition overall is extensive. However, the Estonian language presents unique opportunities for understanding language learning that have not been fully examined yet. This thesis aims to investigate how well L1 tonal language speakers (e.g., Mandarin Chinese) perceive and integrate multiple phonetic cues while learning Estonian quantity contrasts.

Linguistic features of Estonian quantity and Mandarin tone

One of the distinctive characteristics of the Estonian language is the three-way quantity system. Estonian is linguistically rare with its three-way contrastive quantity system. In disyllabic words, there is a distinction between Q1 short, Q2 long and Q3 overlong contrasts whereas the majority of languages contrast short vs. long. The quantity contrast in disyllabic words can be realized through the vowel in the stressed syllable (e.g. Q1 vilu 'chilly', Q2 viːlu 'slice, gen. sg.', Q3 viːːlu 'slice, part. sg.'), consonants at syllable boundaries (e.g. Q1 kalji 'kvass', Q2 kaljli 'hug, nom. sg.', Q3kaljːli 'precious, gen. sg.') or a combination of vowel and consonant length (e.g. Q1 sate 'fallout', Q2 saːtte 'get, 2nd pl.', Q3 saːtːte 'broadcast, gen. sg.') (Lippus et al., 2013).

There are multiple temporal cues that determine the meaning of the word. One key factor is the duration ratio between the stressed and unstressed syllables which is 2:3 for Q1, 3:2 for Q2 and 2:1 for Q3 according to Lehiste (2003). Another determining aspect of the language is the pitch contour that differs between Q1/Q2 and Q3. The high level pitch falls at the end of the stressed syllable in Q1 and Q2 but peaks earlier in the stressed syllable in Q3. This pitch alignment is therefore an important cue for perceiving Q3 (Lippus et al., 2010) and Estonian L1 speakers seem to rely on that especially when the temporal cues are unclear (Lippus et al., 2009). Lippus and colleagues (2010) found that altering the pitch contour of Q2 can lead the listener to perceive it as Q3. This highlights the importance of pitch cues associated with different quantity levels. However, it is important to note that there was variability in the results and the perception change wasn't noted among all listeners.

Mandarin Chinese is a tonal language where the meaning of a word is determined by the pitch contour. There are four lexical tones: high-level (T1), mid-rising (T2), low-dipping (T3), and high-falling (T4). Therefore, a monosyllable /ma/ can mean /ma1/ (mother 妈), /ma2/(hemp 麻), /ma3/(horse 马), and /ma4/ (to scold 骂) depending on which tone is used. (Li et al., 2021) The length of vowels or constants is the secondary cue in Mandarin Chinese. This linguistic feature gives its speakers a heightened sensitivity to pitch but that may come at the expense of reduced sensitivity to other temporal cues in L2 perception such as duration (Li et al., 2021; Meister et al., 2015).

Native language influence on L2 perception

According to Feature Hypothesis, L2 speech perception is influenced by the learner's native language (Best & Tyler, 2007; McAllister et al., 2002). Previously acquired sensitivity to native sounds influences the perception of L2 phonetic details. The features that are present in both L1 and L2 are acquired more effectively in L2 as compared to features that are absent or language-specific in L1 (Murakami & Alexopoulou, 2015; Zobl, 1980). Yet, if the features present in both L1 and L2 are used differently, then the likelihood of the learner experiencing difficulties in acquiring L2 grows (Best & Tyler, 2007; Ellis, 2006). Additionally, the level of L1 influence depends on the learner's proficiency and exposure to L2. L1 transfer is initially more prominent, but L2 processing becomes more independent as the proficiency increases (Yamashita & Jiang, 2010).

A study by Kask and colleagues found that L1 Russian speakers relied heavily on duration and were less sensitive to pitch cues in their perception of Estonian stimuli. Duration is involved in lexical contrasts in Russian language. Therefore, that aligns with the idea of L1 background affecting the perception of L2. It was suggested that features that are present in L1 can have created long-term memory traces which can influence the decoding and comparison of different sound elements (Kask et al., 2021).

Leppik and colleagues (2023) examined the perception of Estonian quantity degrees by native Spanish listeners. The results showed that the listeners succeeded in distinguishing Q1 from Q2 and Q3, but they were unable to distinguish Q2 and Q3 from each other. That aligns with previous research on L2 which states that L2 speakers whose L1 lacks phonological quantity tend to struggle with distinguishing between the long Q2 and overlong Q3 contrast. L2 learners can therefore show a binary perceptual pattern of distinguishing between short vs long/overlong quantities. Additionally, Meister et al. (2015) have found that Estonian L2 listeners are influenced by their L1 prosodic system, with different L1 groups using different cue-weighting strategies (tonal vs. temporal).

One of the few studies on native tonal language speakers' perception of Estonian is by Lee et al. (2022). It compared the ability to distinguish non-native length contrasts between Japanese and Cantonese speakers. Japanese speakers outperformed the Cantonese speakers which was expected since Japanese is a language with two-way length distinction. This confirmed the Feature Hypothesis theory which suggests that if a particular feature is used in L1 then the benefit of the transfer can be helpful in acquiring L2. An important aspect about this study is that the stimuli were created synthetically and the pitch aspect was not considered, rather, the focus was on the correct duration cues in stimulus types. The pitch cue must be present to represent the Estonian language correctly. It could be suggested that the results would even apply to Cantonese speakers if the pitch cue was present.

Lyu et al. (2024) sought to understand how Mandarin Chinese as L1 affects the perception of duration only, pitch only and duration plus pitch information. Both behavioural discrimination tasks and MMN (mismatch negativity) were applied in the study to understand how L1 background affects both attentive and pre-attentive processing. It was found that the Chinese speakers were better at pitch perception but had difficulties with discriminating short and long contrasts. That aligns with precious research on L1 background influence on learning a language since the cues present in their L1 were easier to differentiate. These findings are relevant in the current study since all three conditions are present in Estonian contrasts. Therefore, it could be expected that contrasts where the duration is present will be more difficult to discriminate.

Perceptual plasticity

One of the key mechanisms in L2 learning seems to be training-induced changes in relevant cue perception. For example, when investigating individual differences in learning lexical tones, Chandrasekaran et al. (2010) found that while both good and poor learners improved during auditory training, only good learners showed sustained improvement in learning lexical tones. Good learners were consistently more sensitive to pitch direction both before and after training which is the most crucial cue in Mandarin Chinese. This indicates that language learning is affected by individual differences, such as pre-existing cue weighting ability, but can still be improved even for poor learners. In other words, learning is not just "better performance" but involves reweighting acoustic cues, selectively reshaping the perceptual space so that the focus is on the relevant contrasts (Chandrasekaran et al., 2010).

It has been found that learning is reflected in neural changes, not just performance improvement. An fMRI study by Yang and colleagues (2015) showed that successful learners had different brain activation patterns compared to poor learners. The former showed an even more integrated brain network after the training (Yang et al., 2015). In a study by Kaan et al. (2007), neural evidence supported the idea that different L1 speakers use different mechanisms in cue integration. Whereas early auditory sensitivity was recorded among English speakers, Mandarin speakers showed changes in later neural processing. This suggests that L1 background influences both which cues are processed and when.

Auditory training tends to improve the perception of L2 contrasts even if the learners differ in their L1 background (Wayland and Li, 2008). In a study focused on learning Thai contrasts, the native Mandarin speakers outperformed native English speakers both before and after training (Wayland and Li, 2008). These improvements were noted after a training period of two 60-minute sessions. Both Thai and Mandarin are tonal languages which indicates that experience with lexical pitch contrasts can provide an advantage in learning another tone system. However, both groups benefited from the training, which means that speech perception in adults can be improved with short-term training. Furthermore, the comparison of different training types found that both identification-based training and discrimination training were equally effective in improving perception.

It has been found that even though training improves the perception of a non-native language, the improvement might be asymmetric across categories (Cao et al., 2024; Lu et al., 2015; Saloranta et al., 2020). For instance, a three-day listen-and-repeat training showed an improvement on the perception of vowel duration contrast (Saloranta et al., 2020) whereas learning of consonant duration contrasts was not affected (Saloranta et al., 2022). This illustrates that the effectiveness of short-term auditory training may depend on the segmental properties targeted. It is important to note that independent of the contrast type, the subjects' stimuli discrimination sensitivity improved and the reaction time was reduced. Thus, short-term training has a general effect on perceptual processing in addition to potential contrast-specific improvements.

Generalization of the training effect

Beyond the improvement on a certain stimuli set, generalization of the learned information is important for acquiring a new language. Generalization means that the improvement is noted beyond a specific stimulus, context, speaker etc. (Lively et al., 1993) which is the goal of a successful learning process.

Two of the clearest demonstrations of generalization in non-native speech perception are from studies by Lively et al. (1993) and Wang et al. (1999). In the 1993 study, a robust generalization to unfamiliar talkers and novel stimuli was demonstrated. This applied to participants whose stimuli were produced by 5 different talkers. Additionally, during the 15 days from pre- to posttest, improvement in both accuracy and response speed was noted. It is important to note though, that the materials didn't generalize for participants who trained on stimuli produced by only one speaker. Similarly, in the Wang et al. (1999) study, after a high-variability training, a robust generalization of tone identification was shown among learners of Mandarin Chinese. The improvement gained in training generalized to unfamiliar talkers and new lexical items. Furthermore, the improvement was maintained 6 months after the training. These two studies illustrate that a robust generalization among adult L2 learners is possible and can be generalized to new lexical stimuli and new speakers.

Cao et al. (2024) examined the learning and generalization of the four Mandarin tones separately. Improvements from training to testing were noted. Additionally, participants' performance improved later on new Mandarin stimuli which confirmed successful generalization of the learned information. It was found that the level of improvement differed across tone categories. For example, even though tone 1 accuracy improved from training to testing, it wasn't generalized as successfully. Tone 2 and 4 were learned and generalized robustly, tone 3 didn't improve much from training and was not generalized. That illustrates that the learning and generalization of different stimuli can be influenced by the acoustic properties of the stimuli.

Lu et al. (2015) examined the effectiveness of perception versus perception + production training. It was found that both the perception-only and the perception + production groups showed improved tone discrimination abilities after the training session. Additionally, subjects in both groups generalized their improvements to the untrained stimuli. As the different training methods were equally effective, production was not required for perceptual learning. The study also confirmed that certain contrasts might be harder to acquire than others.

Current study

Differences across language groups and phonological skills that respective speakers require can make navigating in language learning difficult. According to Feature Hypothesis (McAllister et al., 2002), L2 learners are more successful in the perception of contrasts that include phonetic features already used in their L1. For instance, speakers of quantity languages tend to succeed in sensing the length of a vowel (McAllister et al., 2002). It is important to understand these advantages and disadvantages better to help support language learning.

The goal of this study is to investigate how tonal language speakers learn Estonian quantity. Since quantity depends on relative duration ratios and pitch alignment, it poses unique challenges for learners whose native languages (L1) only have tonal contrasts. To be more specific, the aim is to examine which phonetic features (pitch vs duration) are more effectively learned during the course of training.

In the present study, native Mandarin Chinese speakers participated in an 8-day experiment designed to assess the learning and generalization of Estonian quantity contrasts. The participants were divided into a training group and a control group. All participants underwent a pretest-training-posttest procedure on both day 1 and day 8. On days 2-7, the training group underwent language training and the control group did visual perception tasks. To be able to assess generalization, different sets of stimuli were used on days 1-7 (set A) and day 8 (set B). The Estonian contrasts the present study focused on were Q1 vs Q2, which represent duration contrast, and Q2 vs Q3, which represent pitch contrast. A group of Estonian native speakers was recruited as a baseline who underwent a one-hour experiment procedure on the entire stimulus set (both set A and set B).

The following hypotheses were formulated.

H1: Before training, the Chinese participants will be more successful in perceiving the difference between Q2 vs Q3 as compared to Q1 vs Q2 due to pitch contour change.

H2: As a result of the training, the Chinese participants in the training group will show greater improvement in differentiating Q1 vs Q2 as compared to Q2 vs Q3. Since the subjects had the advantage in pitch perception from the beginning, improvements in pitch-based contrasts may be limited. We might therefore see perception improvements in Q1 vs Q2 where only the duration cue is present.

H3: The Mandarin control group, which did not receive any training, will not show any improvement in perceiving the Estonian quantity.

H4: The Estonian native group will show higher performance across all stimuli as compared to Mandarin Chinese speakers.

Method

Participants

The sample consisted of 60 native Mandarin speakers (30 males, 30 females) recruited in China. Participants were aged between 19 and 27 years (*M* = 23.22, *SD* = 2.30). The subjects were native speakers of Chinese with no more than 2 years of formal musical training or prior linguistic background in Estonian. Additionally, 30 native Estonian speakers were included as a baseline group in Estonia. The group consisted of 15 males and 15 females. Their ages ranged from 18 to 27 years (*M* = 21.47, *SD* = 2.18).

The study in China was approved by the Institutional Review Board of the College of Foreign Languages of Zhejiang University of Technology in China, and the experiment in Estonia was approved by the Research Ethics Committee of the University of Tartu. All participants in both labs provided written consent. The procedure was introduced to the participants before the beginning of the experiment in their native language, and they had the option to stop the experiment at any point if they chose to. All collected data was pseudonymized so it was impossible to connect to a specific participant to ensure confidentiality. All participants were rewarded with a gift card after the experiment.

Stimuli

The stimuli were recorded by a native female speaker of Estonian in a soundproof recording booth using the software SpeechRecorder (Draxler & Jänsch, 2025). To avoid the recognition of any real words, Estonian pseudowords were chosen as the stimuli. All pseudowords respect Estonian phonotactic rules, avoiding unnatural combinations.

The stimuli were constructed by combining consonants and vowels in a controlled manner to ensure phonotactic plausibility in Estonian. First syllables were constructed by combining 11 consonants (h, j, k, l, m, n, p, r, s, t, v) with 9 vowels (a, e, i, o, u, ä, õ, ö, ü) which resulted in 99 combinations, and second syllables were constructed by combining 2 consonants (d, g) with 4 vowels (a, e, i, u) which resulted in 8 combinations.

Through the systematic combining of the syllables, a total of 792 (99×8) pseudowords were created. Additionally, all the pseudowords were modified to represent the three quantity degrees (Q1, Q2, Q3). The created stimuli variants were then tested to make sure none of them are actually existing words. Firstly, the stimuli were matched to the Eesti Keele Instituut Ühendsõnastik database and any found matches were excluded. Secondly, manual proofreading by three native Estonians was conducted to ensure that no pseudowords resembled real words. A total of 298 sets of segmental combinations, each being put into one of the three quantity degrees, were kept and sent for recording. The recordings were analyzed acoustically using the PRAAT software (Boersma & Weenink 2007) to ensure that the three quantities in each set represent the Estonian quantity system. In addition, a flattened Q3 was created in each set, which took the duration of Q3 but the pitch of Q2. In the end, a total of 120 sets of stimuli were selected for the experiment.

The stimuli were divided into two equal sets, i.e., set A and set B. The two sets of stimuli were matched on the segmental combinations to ensure equal perception difficulty. Each pseudoword in set A has a corresponding pseudoword in set B with similar consonant and vowel combination. For example, if set A has *haldu*, set B has *halge*. This ensures that both sets share similar phonological structures. Additionally, the differences between the sets are minimal and systematic, typically changing one segment while keeping the rest of the structure constant.

Each set contains four stimuli, i.e., Q1, Q2, Q3, and flattened Q3, which were paired into four different contrasts: Q1 vs. Q2, Q2 vs. Q3, Q3 vs. flattened Q3, and Q1 vs. Q3. Only the Q1 vs. Q2 (involving duration change) and Q2 vs. Q3 (involving pitch change) contrasts will be analyzed in this study.

Procedure

Experiment in China

The experiment in China consisted of eight consecutive days per subject and its goal was to assess performance changes over the period. The experiments on day 1 and 8 took place in the lab and included pretest, training and posttest (see Figure 1). On days 2–7, subjects underwent training remotely at home. To assess the generalization effect, different stimulus sets were used such that set A was used on days 1-7 and set B was used on day 8 (see Figure 1).

Figure 1

Experimental Outline in China

Day 1	Days 2-7	Day 8
In the lab	**At home**	**In the lab**
Pretest (set A)	Training (1/6 set A per day)	Pretest (set B)
Training (set A)		Training (set B)
Posttest (set A)		Posttest (set B)

In the lab, the pretest, training, and posttest adopted an AX discrimination task. Subjects performed an AX experiment where, after hearing two stimuli played after another, they chose whether they were the same or different. They had 2000 ms to respond by pressing either J (for same/different) or F (for different/same) on the keyboard. A no-answer response was recorded if the participant didn't respond in time. The pretest and posttest did not provide feedback. Participants received feedback on each trial (correct or incorrect) in all training sessions. Metrics such as accuracy and reaction time (RT) were recorded. The pretest, training, and posttest in the lab each took approximately 30 minutes to complete, altogether 1.5h in the lab.

At home, the set A stimuli were equally divided into six equal subsets where one subset was administered per day on days 2-7. The remote sessions used the same AX task with feedback but implemented an adaptive training paradigm. If the subject answered correctly, then the pair was removed from the test pool for that occasion. If the subject answered incorrectly, then the same pair would be played again until the subject answered correctly. The process would be repeated until the entire pool of stimuli was cleared or the experiment reached five blocks. The remote training each day took approximately 10–12 minutes to complete.

Experiment in Estonia

The Estonian participants only went to the lab once and didn't undergo training. The participants performed an AX experiment on both set A and set B, mixed together in a pseudorandomized order. This procedure ensured that all the stimuli the Chinese participants were exposed to were covered. Moreover, it helped to test that the two sets used were of the same perceptual difficulty. The experiment took approximately 50-60 minutes.

Statistical Analysis

The statistical analyses were conducted in three parts, each serving a distinct purpose. All analyses were conducted on trial-level accuracy data (correct = 1, incorrect = 0) using generalized linear mixed-effects models (GLMMs). Participant was included as a random intercept in all models to account for individual variability.

First, an analysis was conducted on Estonian native speakers' dataset. The aim was to verify that the two different stimulus sets were equal in perceptual difficulty. This model included set (A, B) and contrast (Q1 vs Q2 - duration, Q2 vs Q3 - pitch) as fixed effects (Model 1). The analysis tested whether there were any main effects of stimulus set or interactions between set and contrast.

Secondly, two separate but conceptually linked analyses were conducted on Chinese participants' data. The first analysis aimed to examine whether a single training session would lead to perceptual improvement and whether the improvement differed between the training and control group. A model was constructed with group (training, control), session (pretest, posttest) and contrast (duration, pitch) as the fixed effects on day 1's data (Model 2). The second analysis examined whether training effects would generalize to a new stimulus set after the training period. The pretest data on day 1 and day 8 was included to ensure that performance reflected generalization and not any training effects. Fixed effects included were group (training, control), day (1, 8) and contrast (duration, pitch) (Model 3).

Finally, additional analyses compared results of the Chinese participants with that of Estonian native speakers. The purpose of the analyses was to illustrate how close Chinese learners came to native-like perception after the training. The first analysis aimed to capture the immediate training effects relative to native performance was conducted. This model compared the day 1 posttest results of the Chinese training group, Chinese control group and Estonian native speakers. Group (control, training, Estonian) and contrast (duration, pitch) were included as the fixed effects (Model 4). The second analysis was conducted on day 8 pretest data to reflect the performance after extended training and on a novel stimulus set. Fixed effects were group (control, training, Estonian) and contrast (duration, pitch) (Model 5).

Results

Estonian dataset

Model 1 results showed that the main effect of set ($\beta = -0.10$, $SE = 0.14$, $z = -0.77$, $p = 0.444$) and the main effect of contrast ($\beta = -0.18$, $SE = 0.13$, $z = -1.37$, $p = 0.172$) were non-significant. Same as the set × contrast interaction ($\beta = 0.08$, $SE = 0.19$, $z = 0.42$, $p = 0.674$). Findings confirm that the two sets had no significant differences in perceptual difficulty level and can be used interchangeably.

Short term training effect (Day 1 Pretest vs. Day 1 Posttest)

Fixed-effects estimates for Model 2 are presented in Table 1. The visual comparison of the discrimination results for the training and control group per contrast is found in Figure 2.

A significant main effect of contrast was found indicating that overall, the pitch contrast was discriminated significantly more accurately than duration contrast. These findings are consistent with H1 stating that Mandarin Chinese speakers would be more successful in perceiving pitch contrasts as compared to duration contrasts before any training. Pitch discrimination remained high across both groups and sessions (pre: control $\hat{p} = 0.927$, $SE = 0.014$; training $\hat{p} = 0.928$, $SE = 0.014$; post: control $\hat{p} = 0.880$, $SE = 0.021$; training $\hat{p} = 0.950$, $SE = 0.010$) as compared to duration discrimination, which showed lower overall accuracy (pre: control $\hat{p} = 0.590$, $SE = 0.047$; training $\hat{p} = 0.628$, $SE = 0.045$; post: control $\hat{p} = 0.503$, $SE = 0.048$; training $\hat{p} = 0.694$, $SE = 0.041$).

A significant group × session interaction indicated that groups accuracy differed across sessions. Bonferroni-corrected pairwise comparisons showed improvement in the training group from pre- to posttest for both duration ($OR = 0.748$, $z = -3.93$, $p < 0.001$) and pitch ($OR = 0.669$, $z = -3.07$, $p = 0.013$). Contrarily, the control groups' performance declined from pre- to posttest for both duration ($OR = 1.424$, $z = 4.62$, $p < 0.001$) and pitch ($OR = 1.736$, $z = 5.16$, $p < 0.001$). The control group demonstrated non-improvement in discrimination accuracy as compared to training groups improvement, which supports H3.

The three-way group × session × contrast interaction was not significant, which indicates that the training benefit did not differ significantly between duration and pitch discrimination on day 1 among groups. Even though the training group improved in duration discrimination, the comparable improvement observed for pitch discrimination means that H2 is not fully supported.

Table 1

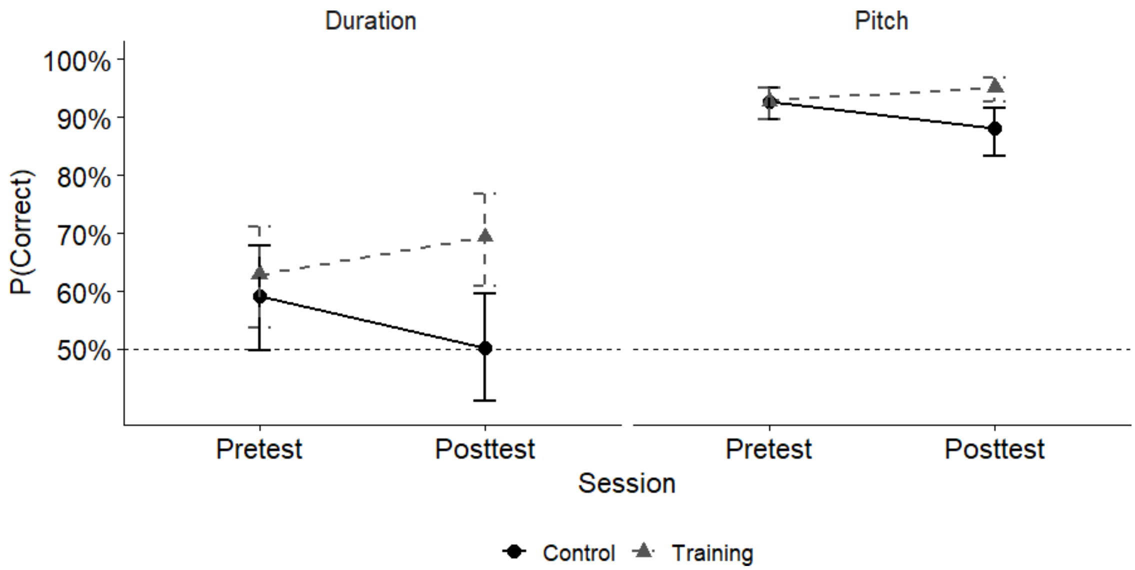
Fixed-Effects Estimates for Model 2 (Short-Term Training Effect, Day 1)

Predictor	β	*SE*	z	p
Intercept	0.366	0.193	1.899	.057
Group (training)	0.160	0.272	0.588	.557
Session (post)	−0.354	0.077	−4.623	< .001
Contrast (pitch)	2.181	0.101	21.697	< .001
Group × Session	0.645	0.106	6.055	< .001
Group × Contrast	−0.154	0.142	−1.079	.280
Session × Contrast	−0.198	0.131	−1.508	.132
Group × Session × Contrast	0.309	0.200	1.547	.122

*Note***.** N = 14,400 observations, 60 participants. Random effect variance = 1.022 (*SD* = 1.011). Reference levels: group = control, session = pre, contrast = duration. Coefficients are log-odds. AIC = 12,684.4.

Figure 2

Comparison of the Day 1 Pretest and Day 1 Posttest results by Groups and Contrasts.



Note. Correct response (p̂) is derived from generalized linear mixed-effects model. Error bars indicate 95% confidence intervals. The dashed horizontal line indicates chance level (50%).

Generalization of training (Day 1 Pretest vs. Day 8 Pretest)

The fixed-effects estimates for Model 3 are shown in Table 2. The comparison of the results for the training and control group per contrast is presented in Figure 3.

There was a significant main effect of contrast which indicates that the discrimination of the two contrasts was different overall. Discrimination accuracy for pitch was high for both the control and training group whereas accuracy for duration was considerably lower. This pattern remained similar on day 8. Pitch discrimination remained high (control: $\hat{p} = 0.915$, $SE = 0.015$; training: $\hat{p} = 0.981$, $SE = 0.004$) as compared to duration discrimination (control: $\hat{p} = 0.552$, $SE = 0.045$; training: $\hat{p} = 0.800$, $SE = 0.029$).

A significant group × day interaction was also observed. This suggests that the changes in discrimination accuracy from day 1 to day 8 are different based on the group the participant belongs to. The training group showed improvement across days whereas the control group showed no improvement. This interaction shows that training led to better perception performance on a new stimulus set, whereas only repeated testing did not.

A significant three-way group × day × contrast interaction was also found which indicates that training effects generalized differently across contrast types between the training and control group. For the duration contrast, the training group improved substantially from day 1 ($\hat{p}$ = 0.63) to day 8 ($\hat{p}$ = 0.80) (*OR* = 0.425, *SE* = 0.033, $z = -10.871$, $p = < 0.001$), while the control group showed no significant change (day 1 $\hat{p}$ = 0.59, day 8 $\hat{p}$ = 0.55; *OR* = 1.161, *SE* = 0.087, $z = 1.994$, $p = 1.000$). For the pitch contrast, the training group improved from day 1 ($\hat{p}$ = 0.93) to day 8 ($\hat{p}$ = 0.98) (*OR* = 0.240, *SE* = 0.042, $z = -8.095$, $p = < 0.001$), while the control group remained stable (day 1 $\hat{p}$ = 0.92, day 8 $\hat{p}$ = 0.92; *OR* = 1.104, *SE* = 0.123, $z = 0.888$, $p = 1.000$). In conclusion, the control group showed no improvement either in pitch or duration contrast discrimination which supports H3. The training group showed greater improvement in differentiating duration contrast as compared to pitch contrast, which supports H2.

Table 2

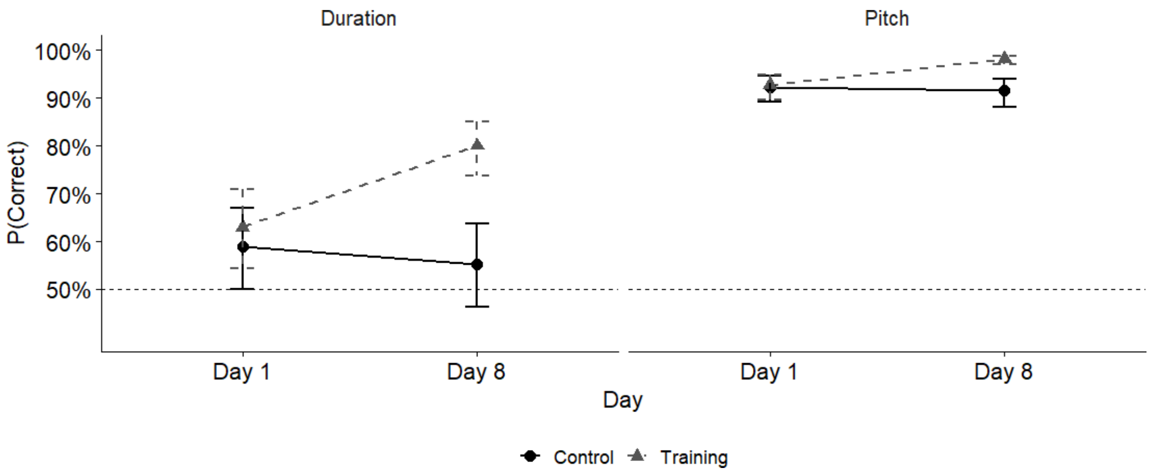
Fixed-Effects Estimates for Model 3 (Generalization, Day 1 vs. Day 8 Pretests)

Predictor	*β*	*SE*	*z*	*p*
Intercept	0.357	0.181	1.973	.049
Group (training)	0.175	0.255	0.686	.493
Day (Day 8)	−0.149	0.075	−1.994	.046
Contrast (pitch)	2.115	0.099	21.432	< .001
Group × Day	1.004	0.109	9.245	< .001
Group × Contrast	−0.111	0.141	−0.788	.431
Day × Contrast	0.051	0.134	0.378	.705
Group × Day × Contrast	0.521	0.235	2.219	.027

*Note***.** N = 14,400 observations, 60 participants. Random effect variance = 0.892 (*SD* = 0.945). Reference levels: group = control, day = Day 1, contrast = duration. Coefficients are log-odds. AIC = 11,930.8.

Figure 3

Comparison of the Day 1 Pretest and Day 8 Pretest results by Groups and Contrasts.



Note. Correct response (p̂) is derived from generalized linear mixed-effects model. Error bars indicate 95% confidence intervals. The dashed horizontal line indicates chance level (50%).

Comparison with Estonian speakers

To be able to illustrate, if and how close the Chinese participants had come to the discrimination level of native Estonian speakers, additional analysis was conducted. Three groups (Chinese control, Chinese training and Estonian native speakers) were compared. Only interactions where group was involved are reported as the main interest was in group-level differences.

Short-term training results compared to Estonian native speakers

The posttest results on day 1 from both the Chinese training group and the Chinese control group were compared with the Estonian dataset. The pairwise contrasts of Model 4 are shown in Table 3. The comparison of the group differences in contrast discrimination is presented in Figure 4. A significant Group (Estonian) × Contrast interaction ($\beta = -2.21$, $SE = 0.13$, $z = -16.37$, $p < 0.001$) indicated that the native speakers' advantage was different in the duration contrast than in the pitch contrast.

For the duration contrast, all three groups differed significantly from one another in duration discrimination. The training group outperformed the control group, but both Chinese groups remained far below the Estonian baseline, with estimated probabilities of $\hat{p}$ = 0.50, 0.70, and 0.95 for control, training, and Estonian groups respectively.

For the pitch contrast, the training group ($\hat{p}$ = 0.95) performed at an equivalent level to the Estonian native speakers ($\hat{p}$ = 0.95). This shows that a single day of language training was sufficient to bring pitch discrimination to a native-like level. The control group ($\hat{p}$ = 0.88) underperformed when compared to both the training group and the Estonian group.

Findings show that Estonian participants exceeded the results of control group participants in duration and pitch perception, which supports H4. Estonian participants' results exceeded the results of the training group in duration contrast. However, the pitch perception results were equivalent regarding the Estonian native speakers and training group. This means that H4 is not fully supported.

Table 3

Pairwise Contrasts for Model 4: Group Differences by Contrast Type (Day 1 Posttest)

Comparison	*OR*	*SE*	*z*	*p adj*
Duration (Q1 vs. Q2)				
Control / Training	0.438	0.123	−2.936	.010
Control / Estonian	0.050	0.014	−10.446	< .001
Training / Estonian	0.114	0.033	−7.575	< .001
Pitch (Q2 vs. Q3)				
Control / Training	0.405	0.121	−3.020	.008
Control / Estonian	0.454	0.132	−2.719	.020
Training / Estonian	1.121	0.335	0.382	1.000

*Note***.** OR = odds ratio; SE = standard error; p adj = Bonferroni-adjusted p-value.

Figure 4

Group Differences in Duration and Pitch Discrimination on Day 1 Posttest

Note. Correct response (p̂) is derived from generalized linear mixed-effects model. Error bars indicate 95% confidence intervals. The dashed horizontal line indicates chance level (50%).

Training generalization results compared to Estonian native speakers

The pretest results on day 8 from both the Chinese training group and the Chinese control group were then compared with the Estonian dataset. The results of Model 5 pairwise contrasts are shown in Table 4. The comparison of the group differences in contrast discrimination is presented in Figure 5. There was a significant Group (Estonian) × Contrast interaction ($\beta = -2.33$, $SE = 0.14$, $z = -16.88$, $p < 0.001$) which shows that the success of contrast discrimination varies based on the group the subject belongs to.

For the duration contrast, the training group ($\hat{p} = 0.80$) significantly outperformed the control group ($\hat{p} = 0.55$; $OR = 0.32$, $p < 0.001$). Despite that, both Chinese groups remained substantially below the Estonian baseline ($\hat{p} = 0.95$). Pairwise contrasts confirmed that both Chinese groups differed significantly from Estonian speakers (control vs. Estonian: $OR = 0.063$, $p < 0.001$; training vs. Estonian: $OR = 0.20$, $p < 0.001$).

For the pitch contrast, the training group ($\hat{p} = 0.98$) significantly exceeded the Estonian native speakers ($\hat{p} = 0.94$; $OR = 2.86$, $p = 0.001$). The control group ($\hat{p} = 0.92$) was significantly below the training group ($OR = 0.22$, $p < 0.001$) but did not differ significantly from the Estonian group ($OR = 0.64$, $p = 0.278$).

H4 which states that Estonian participants will exceed the results of Chinese participants was supported for the duration contrast. However, Chinese training group participants' results in the pitch contrast exceeded the results of the Estonian native speakers. This means that the results don't fully support H4.

Table 4

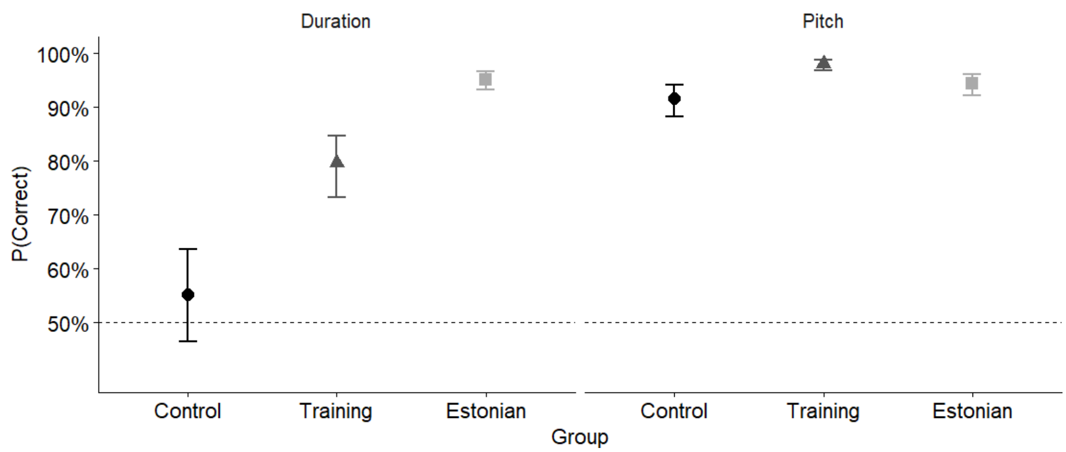
Pairwise Contrasts for Model 5: Group Differences by Contrast Type (Day 8 Pretest)

Comparison	*OR*	*SE*	*z*	*p adj*
Duration (Q1 vs. Q2)				
Control / Training	0.315	0.080	−4.577	< .001
Control / Estonian	0.063	0.016	−10.791	< .001
Training / Estonian	0.199	0.051	−6.257	< .001
Pitch (Q2 vs. Q3)				
Control / Training	0.224	0.067	−5.015	< .001
Control / Estonian	0.642	0.169	−1.681	.278
Training / Estonian	2.862	0.844	3.563	.001

*Note***.** OR = odds ratio. Bonferroni correction applied for 3 tests per contrast type. OR < 1 indicates the first group has lower odds of a correct response.

Figure 5

Group Differences in Duration and Pitch Discrimination on Day 8 Pretest



Note. Correct response (p̂) is derived from generalized linear mixed-effects model. Error bars indicate 95% confidence intervals. The dashed horizontal line indicates chance level (50%).

Discussion

The current thesis aimed to understand how targeted linguistic training changes tonal language speakers' perception of Estonian quantity contrasts. Chinese participants were randomly divided into a training group or a control group. The whole experiment consisted of 8 days. On days 1 and 8, subjects underwent a pretest-training-posttest procedure in the lab and training at home on days 2-7. Short-term training results were assessed on the basis of day 1 pretest and posttest results after just one training experience. Generalization of the training was evaluated by comparing day 1 pretest and day 8 pretest results, with a whole new dataset used on day 8. As the interest was perceptual training of Estonian quantities, two types of stimuli were assessed - pitch contrast and duration contrast.

Perception of Estonian quantity prior to training

We found that without any training, all Chinese native speakers demonstrated superior pitch discrimination ability. This was confirmed through the day 1 pretest where the discrimination was assessed prior to any training. A main effect of contrast was found which showed that pitch contrast was overall perceived more successfully than the duration contrast. This finding aligns with Feature Hypothesis theory (Best & Tyler, 2007; McAllister et al., 2002) which states that sensitivity to phonetic features is affected by which sounds are present in native language. These findings align additionally with the results of the study by Lyu et al. (2024), where it was found that Chinese speakers were most successful at pitch perception when it came to Estonian quantity contrasts. Several studies have suggested that learners rely initially more on the cues that are present in their L1 (Kask et al., 2021; Meister et al., 2015). Lee et al. (2022) showed a similar result to this papers' where participants whose L1 had similar features to the L2 had success in perceiving those features in the context of another language.

Effect of short-term training on learning Estonian quantity

The results of this study showed that short-term training can improve the discrimination of L2 features, both if they are present in L1 or not. The training group improved significantly after one training session and the improvement was noted in both duration and pitch contrast. Our results are consistent with previous findings. Wayland & Li's (2008) study found similarly to this one that perceptual performance can be improved by short-term training. Both native English and native Chinese speakers showed improvement in discriminating tonal contrasts. Improvement was therefore noted in speakers with similar tonal language (Mandarin Chinese) and dissimilar (non-tonal language) background. It has previously been found that short-term training can induce perceptual training in adult L2 learners, even after just a one-hour training (Lu et al., 2015).

We had the expectation that since pitch perception would be superior from the start, it wouldn't improve as much as duration contrast perception. This expectation was not met as the improvement was not selective to one contrast. This means that even though short-term training improved perceptual discrimination generally, it didn't improve duration contrast more than the pitch contrast as H2 suggested it would.

Contrary to the training group, the control group's performance on day 1 did not show improvement after the one-time training. These findings support H3. In fact, the control group's performance declined from pre- to posttest on day 1. It could be because the control group participants completed a demanding visual task between the pre- and posttest. Therefore, a realistic explanation to this would be task-related fatigue.

Previous research on Estonian quantity has mainly focused on perceptual differences in L1 backgrounds (Kask et al., 2021, Lee et al., 2022, Leppik et al., 2023). The present study is one of the first to examine the learning of Estonian quantity contrasts (see also Lyu et al., 2026) by applying a perceptual training paradigm. Current study therefore contributes to the growing literature on the less-studied Estonian language.

Perception of Estonian quantity on novel stimuli

One of the key questions of this study was whether training would lead to generalization of trained items to novel stimuli. The design allowed us to assess the learning of phonetic cues rather than memorization of certain stimuli, as there were two separate stimuli sets used. Evidence for successful generalization in the training group was found. The results showed that the training group performed better on day 8 (pretest) than on day 1. That is important since novel stimuli were used on day 8. This illustrates that the improvement generalized to novel stimuli and was not limited to trained items.

The findings are consistent with earlier research which has shown that auditory training can help generalize to new stimuli (Cao et al., 2024; Lively et al., 1993; Lu et al., 2015; Wang et al., 1999). For instance, Lively et al. (1993) found that training could lead to the generalization of the stimuli to untrained items and unfamiliar speakers. Wang et al. (1999) and Lu et al. (2015) showed that training effects can transfer to untrained stimuli in lexical tone learning. Cao et al. (2024) reported successful generalization but showed that its extent can vary across contrasts.

The results showed that the generalization differed across contrast types. Even though training led to significant improvement for both pitch and duration, generalization was stronger for duration than for pitch. The training group showed a substantial increase in duration perception and smaller increase for pitch perception. On day 1, participants' pitch perception started very high, whereas their duration perception started quite low. Therefore, there was more room for improvement in duration contrast as compared to pitch contrast.

Previous studies found that the success of generalization can vary depending on the properties of the stimuli (Cao et al., 2024). Not all stimuli can be generalized equally (Lu et al., 2015). Current study's results provided proof for that since there were differences in how the duration and pitch contrasts were learned and generalized.

Note that the improvement of duration and pitch perception was similar after the day 1 short-term training. On day 8, however, improvement in duration perception was more substantial than in pitch perception. A possible explanation for that can come from previous studies which state that previously acquired sensitivity influences which phonetic details are perceived more naturally (McAllister et al., 2002; Zobl, 1980). Pitch contrast is more familiar for Mandarin Chinese L1 speakers. Familiar features can therefore have been perceived more successfully initially, but training may have changed the relevancy in the perceived cues (Chandrasekaran et al., 2010).

Learning outcome in comparison to Estonian native speakers' baseline results

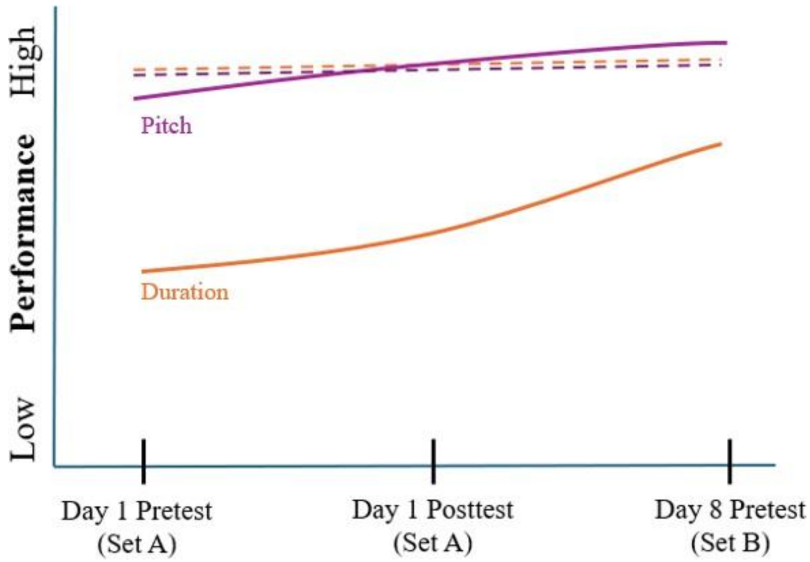
The results of Mandarin Chinese speakers were compared to those of Estonian native speakers. The findings illustrate how auditory training can lead to results that are comparable to native speakers' results. For the pitch contrast, after the short-term training on day 1, the training group performed at an equal level to the Estonian native speakers. Furthermore, on day 8, the training group outperformed the Estonian native speakers. For the duration contrast, there was an improvement equal to that of the pitch contrast, which occurred after the short-term training on day 1. However, the results didn't reach the level of Estonian native speakers. On day 8, the training group improved more in duration perception than in pitch perception. Despite that, the results in duration contrast perception still didn't reach the results of those of the Estonian baseline.

The results demonstrated different learning trajectories for duration and pitch. The trajectories are presented in an illustrative figure (see Figure 6). Duration perception started at a low level but showed clear improvement after short-term training on day 1. The performance, however, was not close to native-like. After the at-home training, on day 8, duration had improved a lot although it remained below native-like performance. Pitch perception started at a high level. Similarly to duration, a clear improvement was noted and the perception reached a level comparable to native listeners. On day 8, pitch perception improved further, but the amount of improvement was smaller as compared to improvement in duration perception. Pitch perception on day 8 exceeded the results of the native listeners.

The differences in learning trajectories can be explained by the influence of L1. As found by Lyu et al. (2024), pitch-based contrasts of the Estonian quantity are more easily perceived than duration-based contrasts for Mandarin Chinese speakers. This provides an initial advantage in perceiving and learning pitch-based contrasts. Duration, however, can be more difficult to learn as it is a less-familiar cue to Mandarin Chinese language.

Figure 6

Illustration of Learning Trajectories for Duration and Pitch perception



Note. This is an illustrative figure that shows the trajectory of improvement in duration and pitch perception across different timepoints. The dotted lines represent the Estonian native baseline. Pitch is represented by the color purple and duration is represented by color orange.

Limitations and future research

One limitation of this study is the relatively short duration of the training period. Even though the training significantly improved the discrimination of Estonian contrasts, a longer training period could provide further insight into the long-term effects of language training.

Another limitation is that short-term learning and generalization were tested across novel stimuli but not with new speakers or speech rates. Study by Lively et al. (1993) is one of the best examples of generalization through auditory training where the stimuli was recorded by 5 different speakers. This allowed the stimuli to generalize to different speakers so that the training is not only beneficial when listening to one specific speaker. Future studies could therefore include several speakers in the training and testing phase so that more robust generalization could be assessed.

The age group of participants was restricted. Current study focused on adult subjects and specific age-range was decided upon so that Mandarin Chinese and Estonian groups could be compared. In the future, a wider range of ages could be included to provide further insight into auditory training effects across different age groups.

Something that could provide further understanding about the process of learning Estonian quantity would be including reaction time in the data analysis. Since the aim of this study was to observe the differences in accuracy throughout the training process and to maintain analytical focus, reaction time was not included. However, it could provide valuable information about the changes in perception speed throughout the training process in the future.

Future research could introduce additional kinds of language training such as production to the experimental setup. Previous research indicates that production might not significantly alter the results (Lu et al., 2015) but it has not yet been confirmed in the context of a quantity language. Future designs could even include different tonal language groups in the training group to see whether all tonal languages perceive Estonian contrasts similarly or if there are some notable differences.

Conclusion

The current study examined how Estonian quantity contrasts are learned and how the learning can be generalized to new stimuli among Mandarin Chinese native speakers. Prior to training, the participants showed a strong advantage for pitch-based contrasts over duration-based contrasts. Pitch perception improved significantly after the short-term training and reached a native-like level. After the 6-day training period, pitch perception improved further and exceeded the performance of Estonian native speakers. Duration perception also improved significantly after the short-term training but remained below native-like level. After the 6-day training, duration perception showed a large improvement. The improvement in duration perception was more substantial than the improvement in pitch perception. Despite this, the duration perception did not reach native-like level.

Both the training and control group completed the same amount of pre- and posttests and differed in the fact that the training group completed auditory training tasks. The results therefore suggest that the improvement can be attributed to training and not to memorization of stimuli or repeated testing.

In conclusion, findings of this study demonstrate that L2 perception of Estonian quantity is impacted by a complex interaction between L1 experience, cue integration processes and training-induced plasticity.

The use of Artificial Intelligence

Artificial intelligence such as Claude was used to provide help in programming scripts for R Studio. Microsoft Copilot was used as aid for brainstorming and to provide feedback.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Siqi Lyu for support and valuable guidance throughout the completion of this thesis.

References

Anthropic. (2026). Claude (March 30 version) [Large language model]. https://claude.ai

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In Language Experience in Second Language Speech Learning (pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Boersma, Paul and David Weenink. 2007. Praat: Doing Phonetics by Computer (Version 4.6.31) [Computer program]. http://www.praat.org/.

Cao, M., Pavlik, P. I., & Bidelman, G. M. (2024). Enhancing lexical tone learning for second language speakers: effects of acoustic properties in Mandarin tone perception. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1403816

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, 128(1), 456–465. https://doi.org/10.1121/1.3445785

Draxler, C., Jänsch, K. (2025). Speech Recorder (Version 6.9.2) [Software]. Available from https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics*, 72, 17–34. https://doi.org/10.1016/j.wocn.2018.10.005

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience*, *9*(1). https://doi.org/10.1186/1471-2202-9-53.

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., Lippus, P., & Kreegipuu, K. (2021). Perceptual Asymmetries and Auditory Processing of Estonian Quantities. *Frontiers in Human Neuroscience*, 15. https://doi.org/10.3389/fnhum.2021.612617

Lehiste, I. 2003. Prosodic change in progress: From quantity language to accent language. In: Fikkert, P., & Jakobs, H. (eds.), Development in prosodic systems, 47–65. Berlin/New York: Mouton de Gruyter

Leppik, K., Lippus, P., & Asu, E. L. (2023). The perception and production of Estonian vowels and vocalic quantity contrasts by Spanish L1 learners. *Ampersand*, 100147. https://doi.org/10.1016/j.amper.2023.100147

Li, Y., Tang, C., Lu, J., Wu, J., & Chang, E. F. (2021). Human cortical encoding of pitch in tonal and non-tonal languages. *Nature Communications*, 12(1). https://doi.org/10.1038/s41467-021-21430-x

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, 41(1), 17–28. https://doi.org/10.1016/j.wocn.2012.09.005

Lippus, P., Pajusalu, K., & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, 37(4), 388–396. https://doi.org/10.1016/j.wocn.2009.07.002

Lippus, P., Pajusalu, K., & Allik, J. (2010). The Role of Pitch Cue in the Perception of the Estonian Long Quantity. In Prosodic Categories: Production, Perception and Comprehension (pp. 231–242). *Springer Netherlands*. https://doi.org/10.1007/978-94-007-0137-3_10

Lippus, P., Pajusalu, K., & Allik, J. (2007). The tonal component in perception of the Estonian quantity. In J. Trouvain, & Barry, W. J. (Eds.), The proceedings of the 16th international congress of phonetic sciences: 16th international congress of phonetic sciences, Saarbrücken, Germany, 6–10 August 2007 (pp. 1049–1052)

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, 1624, 28–44. https://doi.org/10.1016/j.brainres.2015.07.014

Lyu, S., Põldver, N., Kask, L., Wang, L., & Kreegipuu, K. (2024). Native language background affects the perception of duration and pitch. *Brain and Language*, *256*, 105460. https://doi.org/10.1016/j.bandl.2024.105460

Lyu, S., Põldver, N., & Kreegipuu, K. (2026). Perceptual learning via passive exposure: Neural and behavioral evidence from learning Chinese tone and Estonian quantity [Manuscript in preparation]. Faculty of Social Sciences, Tartu University

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

Meister, E., Nemoto, R., & Meister, L. (2015). Production of Estonian quantity contrasts by Japanese speakers. *Journal of Estonian and Finno-Ugric Linguistics*, 6(3), 79–96. https://doi.org/10.12697/jeful.2015.6.3.03

Microsoft. (2026). Copilot (April 29 version) [Large language model]. https://copilot.microsoft.com/

Murakami, A., & Alexopoulou, T. (2015). L1 influence on the acquisition order of English grammatical morphemes. *Studies in Second Language Acquisition*, 38(3), 365–401. https://doi.org/10.1017/s0272263115000352

Nick C. Ellis, Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning, Applied Linguistics, Volume 27, Issue 2, June 2006, Pages 164–194, https://doi.org/10.1093/applin/aml015

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, 147, 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

Saloranta, A., Heikkola, L. M., & Peltola, M. S. (2022). Listen-and-repeat training in the learning of non-native consonant duration contrasts: influence of consonant type as reflected by MMN and behavioral methods. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-022-09868-6

Sõnaveeb. (n.d.). Sõnaveeb. Retrieved September 22, 2025, sõnaveeb.ee

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, 44(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Language, Brain, and Gene: The Chinese Context*, 33, 29–49. https://doi.org/10.1016/j.jneuroling.2014.09.004

Zobl, H. (1980). The formal and developmental selectivity of L1 influence on L2 acquisition. *Language Learning*, 30(1), 43–57. https://doi.org/10.1111/j.1467-1770.1980.tb00150.x

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36, 250–267.

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