university of groningen

faculty of science and engineering

artificial intelligence

# The asymmetry in Cross-linguistic phonological perception

Bachelor's Project Thesis

Jiayun Zhang, s3541312, J.zhang.44@student.rug.nl,
Supervisors: Mi Tang & Dr Jennifer Spenader

**Abstract:** This study looks at the research question of whether or not the phonological perception is asymmetric in a cross-linguistic context. Chinese native speakers and Dutch native speakers participated in two experiments that evaluated the patterns of perceptual identification of tone categories and trilled alveolar categories, respectively. To investigate whether or not the Chinese and the Dutch groups perceive those speech stimuli differently, we looked at the slope and the category boundaries of the logistic regression curve fitted to the experimental data. The results showed that Dutch native speakers have sharper category boundaries on trilled alveolar sounds than those of Chinese native speakers. However, there was no significant between-group difference in identifying tone categories. In conclusion, we observed the asymmetric pattern that is shaped by native language experience only in trilled alveolar sounds categorization but not in tone categories identification. The possible reason is that the Dutch participants may have been exposed to the intonation changes through their native language experience, while the Chinese native speakers have no language experience with trilled sounds at all.

## 1 Introduction

It is well known that people's native language or language experience can affect many aspects of vocal perception. A study by van der Bij et al. (2017) indicated that participants from different language backgrounds had a significant difference in their perception of word stress. In their study, they selected a meaningless vocal "tatata", and by changing the three features (pitch, duration, intensity) of the syllables, the participants were asked to decide which syllable "ta" tends to be the stress in their perception. The results show that participants whose native language is Singapore English are more tend to locate the stress on the syllable with a longer duration and higher intensity. Mandarin Chinese subjects are more likely to notice the changes in pitch and duration of syllables. One possible reason why Mandarin native speakers are more sensitive to pitch changes is that Mandarin Chinese is a tonal language. The change of pitch plays a significant role in the tonal language (McCawley, 1978). In tonal language, in addition to syllables, pitch changes can also be used to express some information (also called tonal information). Mandarin Chinese is a typical tone language, but English and Dutch are not.

How pitch changes over time in a speech sound is shown as the F0 contours (Chao & Chao, 1968). There are four tones in Mandarin Chinese, with each of them has a distinct F0 pattern, namely Tone 1 (horizontal), Tone 2 (rises moderately), Tone 3 (falling-then-rising), and Tone 4 (falling sharply)(Tseng, 1981). Different tones can give different meanings to the same syllable. For example, the vowel /i/ with Tone 1 means the number one ( 一 ). While with Tone 2 means movement ( 移 ), means ant ( 蚁 ) with Tone 3, and means pandemic ( 疫 ) with Tone 4. If people want to perceive a tonal language correctly, they must possess a categorical perception of the tones. Therefore, the categorical perception of tones in Mandarin is necessary for every person who is able to perceive Mandarin Chinese. As a result, the topic related to the categorical perception of tone has been widely investigated.

There are various factors that affect people's classification and recognition of tones. According to Wang et al. (2017)'s research, the elderly have significantly different perception thresholds for the changing of tones than young people. Age reduces the capability of native Chinese speakers to perceive the changes in pitch contours. According to another research (Tillmann et al., 2011), congenital amusia also affects patients' ability to perceive and learn tonal information. But among so many factors, the most crucial one is a person's language experience. There is a lot of evidence to suggest that native Chinese speakers perform better than native speakers of other non-tonal languages in the categorical perception of tones. Wu & Lin (2008) 's research showed that English speakers' perception of tones was not categorical between Tone 1 and Tone 4. Lee et al. (1996) 's research showed that Cantonese (also a tonal language with five tones) speakers were better at classifying Cantonese tones than speakers of Mandarin and English. When it came to classifying Mandarin tones, native Mandarin speakers performed the best, followed by Cantonese speakers, and English speakers were the worst.

All these results showed that a person is better at classifying tones in their native languages. However, most of these previous studies collected data through an experimental paradigm method called the AX (same/different) discrimination test (Rogers, 2017). In this paradigm, two sounds are played and subjects are asked to judge whether the tones of the played sounds are the same, and then the accuracy or error rate of subjects' responses was calculated to evaluate their ability to perceive the tones categorically. However, this kind of accuracy or error rate is just a simple representation of the categorical perception. According to Hallé et al. (2004) 's research, Chinese native subjects have a quasi-categorical perception of tone, and they can clearly distinguish the tone contours of two sounds. But tone perception of French speakers is psychophysically based. It was shown that most of the French speakers in the study could not actually distinguish the tones but had a certain sensitivity to the tones and could realize that the two sounds are a little bit different. Thus, it can be inferred that there may be sophisticated differences between native and non-native tone perception and even between any native and non-native phonetic perception. In order to reveal the differences in more detail, we have to employ a fine-grained experimental paradigm and apply it to participants from different language backgrounds. The AXB (forced-choice) identification test has been widely used to measure the threshold of categorical perception(Rogers, 2017). Compared with accuracy and error rate, the threshold is more likely to represent the features of categorical perception and the sensitivity of the category boundary. The threshold, namely the category boundary, can be estimated by fitting responses to a continuum of speech stimuli to a sigmoid function. This is a more finely-grained method compared with AX because it demonstrates changes in the perceptual identification of two categories.

Although it looks clear that a language user has a more clear-cut categorical perception of native speech sounds (Wu & Lin, 2008; Lee et al., 1996; Hallé et al., 2004), whether people from different native language backgrounds have significantly different category boundaries for changes in native and non-native speech sounds is still a meaningful question which is worth discussing. Investigating native speakers' advantage in perceiving native sounds against a cross-linguistic context can draw a fuller picture for understanding the pattern of native versus non-native phonological perception. So, the research question of this paper is: Is there any difference between native and non-native language speakers when perceiving native and non-native speech sounds?

To answer this question in a cross-linguistic context, we conducted two experiments with the same procedures using Dutch and Mandarin materials, respectively. We chose to use the alveolar trill as the experimental material for the Dutch experiment as the cross-linguistic comparison of the tone in Mandarin. Dutch is a non-tonal language. But the alveolar trill is ubiquitous in Dutch and does not exist in Mandarin or even English. Alveolar trill is produced by holding the tongue near the alveolar ridge. When the tip of the tongue gets the contact with the alveolar ridge, the flowing of air will cause it to vibrate repeatedly(Kummer, 2001). Typically, for Dutch people, this kind of trill is straightforward to recognize. The results of the two experiments will demonstrate how native Mandarin speakers and Dutch speakers differ in their sensitivity to perceiving alveolar trills and pitch contours of lexical tones.

# 2 Method

## 2.1 Subject

There are a total of 42 subjects participating in this experiment. 20 of them are native speakers of Mandarin Chinese and the rest of them are native speakers of Dutch. All Chinese subjects and 10 Dutch subjects are recruited from the University of Groningen (RUG). The rest 12 Dutch subjects are recruited from Prolistic, which is an online psychology experiment platform. We applied a T-test on the data of online subjects and offline Dutch subjects. The result shows that there is no significant difference between these 2 groups ($p = 0.204$). None of them reported a history of hearing damage, neurological illness or injury. All subjects were asked via a questionnaire about how long they living in the Netherlands, their proficiency in Dutch and Mandarin, the frequency of exposure to Mandarin and Dutch in their daily lives, whether they had any experience with learning any musical instruments or vocal music, and how often they practised in the last half year. All of the native Chinese subjects use English as their second language. In their daily life, the frequency they get in touch with Chinese through various channels is 65% and the rest of the time they use English. Only 30% of native Chinese speakers have a fundamental knowledge of Dutch. In addition, almost all native Dutch speakers speak Dutch and English and have no knowledge of Mandarin.

## 2.2 Stimuli

### 2.2.1 A continuum of Mandarin tones

From the four tones in Mandarin, we choose to use the rising Tone 2 and the falling Tone 4 as the experimental stimuli. Since Mandarin tones contrast from each other in pitch contour (Tseng, 1981), we generated the stimuli by manipulating the pitch of the speech sounds. The stimulus to be manipulated is a vowel */i/* recorded by the experimenter with a duration of 0.5s. The frequency of the speech was first adjusted to 150 Hz (the pitch of a normal male voice) by PRAAT to ensure that there was no pitch change in the speech at this point. Then, we maintained the pitch at 150hz for the first 0.1 seconds and changed the pitch from 150hz to 200hz (rising tone) and 150hz to 100hz (falling tone) at the endpoint. Maintaining 0.1 seconds of 150hz is to make the sound more like real Mandarin pronunciation. Now, two standard raising tone and falling tone speeches are generated. Afterwards, we constructed a continuum by generating 21 tokens ranging in equal steps from the rising tone to the falling tone. Within the continuum, each token was slightly different from the other ones in frequency. The resulting stimuli all had the same onset pitch frequency (i.e., 150 Hz) and differed in offset frequency only.

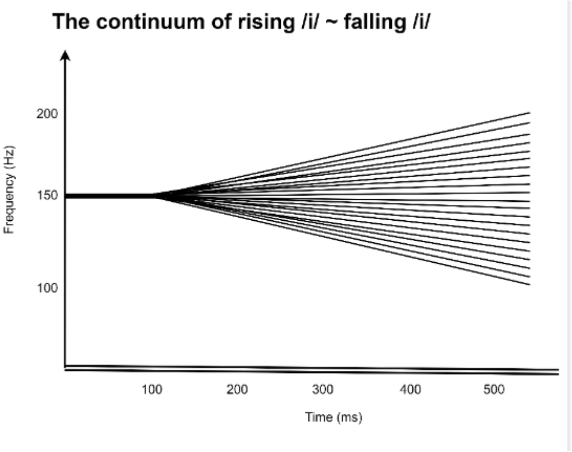

**Figure 2.1: The continuum of Tone stimuli consists of 21 tokens.**

### 2.2.2 A continuum of Dutch alveolar consonants

The second experiment is about precepting the trills in Dutch, where the independent variable was the extent of the trill present in the speech. The stimuli used in this experiment were Dutch sounds */ri/* and */li/* in the male voice. They were generated via Voicebooking (*Stemmenbureau voicebooking in Amsterdam*, 2022). The consonant */r/* in */ri/* is an alveolar consonant with a trill, while the */l/* in */li/* is an alveolar consonant without a trill. A morphing function in Tandem STRAIGHT (Kawahara, 2014) was used to generate a continuum of 21 steps between */ri/* and */li/*. By inspecting formant contours for the two sounds in Praat, we observed a difference in the third format (F3), starting from the onset of the speech sounds (0 ms) till approximately the end of the consonants (90 ms). This difference may account for the distinction between the alveolar with and without trill sound. We put 4 anchors in the duration span from 0 ms to 90 ms to align the acoustic properties of constants in the temporal and frequency domains. Another 3 anchors were also used for aligning the vowel part to make each stimulus in the continuum sound natural.

## 2.3 Procedure

The experiment was conducted in The Alice lab at the University of Groningen, a professional lab of cognitive research. The experimental procedure was made by OpenSesame. First, there is an interface showing the experiment's instructions. The text on this page told they were going to hear some speech sounds, with each sound played only once. After a sound was played, subjects were told to make a decision between two choices to indicate the speech sound belonged to which category (i.e., In Dutch, either */li/* or */ri/*. In Chinese, either a rising tone or a falling tone). The speech-playing interface presented two arrows, which told subjects where the responding keys were, together with the label text of the two categories. For example, in a trial where a Dutch speech sound was played, subjects saw a left arrow labelled with the text "Q, */li/*" and a right arrow with "P, */ri/*". The key press here was counterbalanced across the stimuli type. At the beginning of each trial, there was a one-second buffer phase, and a cross shape appeared on the screen to remind subjects to stay focused.

As the "Stimuli" part describes, we had Dutch and Chinese materials. The two languages were presented in separate blocks. The procedure of both blocks was identical to what we described above. The block order was counterbalanced as well.

Each block has a total of 15 rounds. Each round with 21 trails corresponding to the 21 different continua, and each continuum will play once and in random order in a round. Before the formal experiment, there are 3 practice trials to help the subject be familiar with this experiment. The experimental programs were run on a Mac mini and were presented on a Benq XL2420-B monitor. Responses were recorded by keyboard input, and the audio files were played through a Sennheiser HD201 headset. Participants who are native Chinese speakers were presented with the interface and instruction in Chinese. Meanwhile, native Dutch participants used the Dutch version. This is for ensuring that the subjects were primed with their native languages.

## 2.4 Statistical analysis

We first fit a Logistic regression for each subject's data in Python and took the slope of the regression curve and the categorical boundary as the dependent variables. The category boundary is the threshold of the subject's perception between two speech sounds. In another word, the point at which the probability of classifying a speech stimulus as one category is equal to the probability of the other category. The slope of the logistic regression curve at the categorical boundary represents the clarity or sharpness of the threshold. The larger the slope, the clearer the threshold, and vice versa. The formula for this logistic regression is as follows:

$$y = 1/(1 + exp(b * (x - c)))$$

In this formula, where $y$ refers to the fraction of responses that identified a tone as a falling tone or an alveolar sound as trilled, and $x$ is the proportion of the falling tone or the alveolar trill in a stimulus. $c$ gives the stimulus parameter corresponding to 50% of the psychometric function (i.e., the categorical boundary) and $b$ determines the slope of the logistic regression curve.

When analyzing the $b$ and $c$ values of all participants, we used the LME (Linear Mixed Effects) model(J. C. Pinheiro & Bates, 2000) from the R package NLME(J. Pinheiro et al., 2023). LME is a statistical method to analyze data. Compare with other methods, it is a type of mixed-effects model that allows for both fixed and random effects. It is useful for analyzing data that has a nested structure, such as repeated measures or hierarchical data. In the present study, we constructed an LME model that has the language of stimulus (Alveolar trill, Tone) and group (Dutch, Chinese) as the fixed effects and a by-subject random intercept. For analysing the interaction effect between all the variables in this experiment, we chose a multiple comparisons method called Tukey.

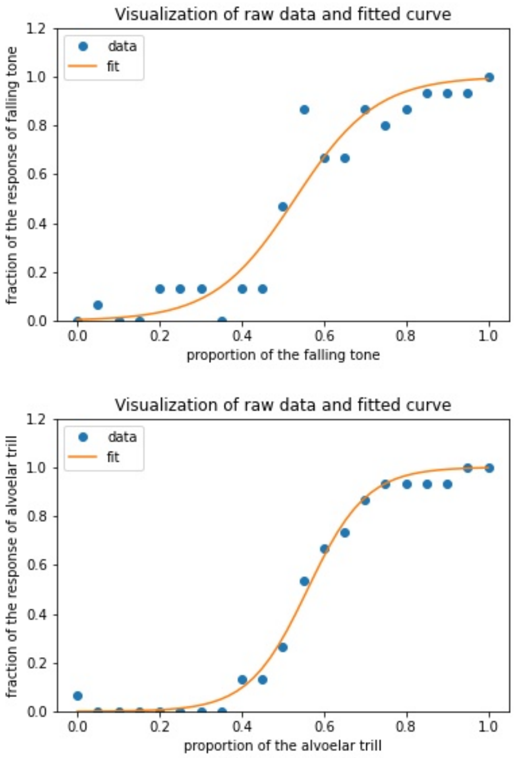

**Figure 3.1: Visualization of raw data and fitted curve from No.15 native Dutch speaker on alveolar and tone classification task**

# 3 Result & Discussion

Figure 3.1 respectively shows the data from a native Dutch speaker on alveolar and tone classification. The blue data points indicate the proportion of responses that classify a stimulus as an alveolar trill or with a falling tone, and the orange curves are the logistic regression curve generated by fitting these data points. The b-value (curve slope) and c-value (category boundary) of each subject can be obtained by taking the parameters of these logistic regression curves. According to the above two figures, it can be seen that the category boundary for both alveolar trills and tones is about 0.5, but the clarity of the category boundary (slope of the curve) on alveolar trills is obviously higher than that on tone.

Figure 3.2 shows the data averaged by all subjects. There are a total of four experimental conditions (1) native Dutch speakers identifying tone stimuli (light purple line), (2) native Chinese speakers identifying tone stimuli (dark purple line), (3) native Dutch speakers identifying alveolar stimuli (orange line), (4) native Chinese speakers identifying alveolar stimuli (yellow line). The average value of each data point of each condition was calculated.

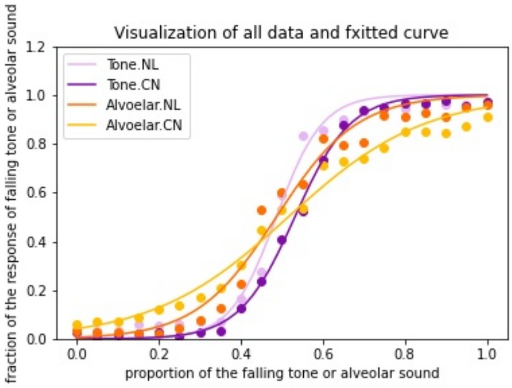

**Figure 3.2: Visualization of raw data and fitted curve for the four experimental conditions.**

We modeled the data using the LME model to find the statistical differences on b-values and c-values. The formula we used is:

$$Perceptual\ Slope(b)\ or\ Category\ Boundary(c) \sim Language * Group + (Language|Participants)$$

*Language* in this formula is the within-subject variable, which is the material that the subject listened to in the experiment, tones, and alveolar trills. *Group* is the between-subject factor, referring to the native language of the subjects, Dutch native speakers and Chinese native speakers. As for the random effect structure, we have the *Participants*, which refers to the sequence of subjects, and the random slope of *Language* on *Participants*. The model was fit on either Perceptual $Slope(b)$ or $CategoryBoundary(c)$.

The results show that there is no significant difference in the b-value between Dutch and Chinese native speakers ($\beta$ = -0.2092898, $SE$ = 0.12313158, $t$ = -1.69972, $p$ = 0.0978). But there is a significant difference in the c-value ($\beta$ = 0.0317941, $SE$ = 0.014054334, $t$ = 2.26223, $p$ = 0.0298). This means there is a group difference in thresholds of perception of tones and alveolar trills between Dutch native speakers and Chinese native speakers, but not for the clarity of the threshold.

Afterwards, the Tukey method was used to perform multiple comparisons on both the c-value and the b-value data. The results showed that the b-values of Chinese native speakers and Dutch native speakers were significantly different in classifying alveolar trills ($\beta$ = 0.5510, $SE$ = 0.1861, $z$ = 2.961, $p$= 0.01621 ), while there is no significant between-group difference in classifying tones ($\beta$ = -0.4101, $SE$ = 0.1861, $z$ = -2.204, $p$ = 0.12204). On the other hand, in the results derived from the c-value, neither tone classification($\beta$ = -0.037507, $SE$ = 0.021493, $z$ = -1.745 $p$ = 0.300) nor alveolar trills classification ($\beta$ = -0.024085, $SE$ = 0.021493, $z$ = -1.121, $p$ = 0.677) showed a significant between-group difference. This means that the subjects from different native language backgrounds only have significant differences in the clarity of the category boundary when classifying alveolar trills. But the values of category boundary of perceiving alveolar trills and tones do not differ between native and non-native language speakers. Figure 3.3 is the Box-plot of the four different experimental conditions, it shows the direction of the differences indicated by the positive or negative $\beta$ values in the results of the c-value, and can be inferred that Chinese native speakers have a lower threshold for rising tones and are more likely identify alveolar trills as not trilled, while Dutch speakers tend to classify the stimuli as falling tones and trilled.

Regarding the potential reason why our Dutch participants did not differ from Chinese speakers in tone classification, it could be that Dutch native speakers have been exposed to many changes in tones in their daily life. For example, when there are changes in emotion, people's voices will become higher when they are excited or angry, and lower when they are depressed. In addition, the change of tone is actually the change of pitch contours, and the change of pitch generally exists in music or songs. Moreover, a study has found that the Dutch nuclear pitch accent is very similar to Tone 2 and Tone 4 in Mandarin (Chen et al., 2015). Even if Dutch native speakers have never learned the rising and falling tones in Mandarin, they have actually been exposed to quite a lot of changes of tone in their daily life and have some language experience with such changes. As a result, the performance of Dutch native speakers in the classification of tones is similar to Chinese native speakers. However, although there is no statistically significant difference, it can be seen from $\beta$ = -0.4101 that the Dutch native speakers have slightly shallower slopes than the Chinese native speakers when classifying tones.

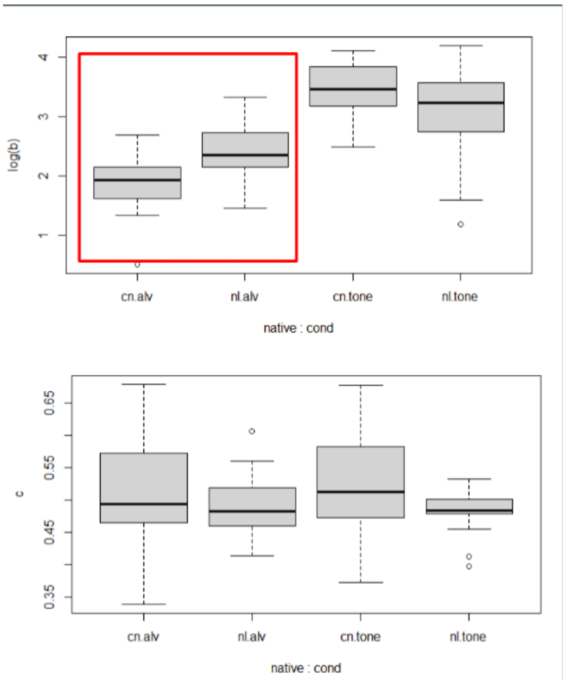

**Figure 3.3: Box-plot of the four different experimental conditions. The red box is the group with a significant difference.**

In addition, there have been studies on tone perception showing that native Mandarin Chinese speakers have significant differences in many aspects when perceiving tones compared to non-native speakers (Wu & Lin, 2008; Lee et al., 1996; Hallé et al., 2004). For the stimuli in these experiments, tones were given to some meaningful Chinese words. In the experiments using words as stimuli, native Chinese speakers were more likely to recognize the stimuli as a part of meaningful language, while non-native Chinese speakers recognized them as meaningless voice sounds. However, in the present study, the chosen tone stimulus /i/ was described as a nonsense syllable according to our experimental instruction. As a result, both native Chinese and Dutch speakers only recognized the stimulus in our experiment as a nonsense monosyllable, not a meaningful part of the language. The fact that no meaning was attached to our stimuli may reduce the mother tongue effect between native speakers and non-native speakers. This can explain why in our experiments, the participants showed no significant difference in identifying the tone continuum.

The possible reason why Chinese native speakers have a significant difference in the clarity of category boundary from Dutch native speakers is that there are no trilled speech sounds in Mandarin Chinese. To produce the alveolar trill speakers should have a good amount of muscle control of the tongue so that there could be multiple contacts between the tongue and the alveolar ridge. Most Chinese native speakers basically feel difficult to articulate that. This difficulty in production may lead to the insensitivity of Chinese native speakers to trill changes. In terms of the category boundary (the c-value) of perceiving alveolar trills, $\beta =$ -0.031773 it can also be seen that Chinese native speakers are less likely to categorize the stimuli as trilled than Dutch native speakers.

# 4 Further Research & Conclusions

The major finding of the present study was that the significant difference between Dutch and Chinese native speakers was only in the sharpness of the category boundary of classifying the alveolar trilled stimuli. Chinese native speakers showed lower sensitivity to alveolar trills than Dutch native speakers. According to some Chinese participants who shared their thoughts with us after the experiment, there was actually a development in how the alveolar trills sound to them. At the beginning of the experiment, it was difficult for them to distinguish between the trilled sounds and the untrilled ones, and after several rounds, the differences between the two categories became easier for them to catch. Therefore, in order to illustrate the potential development in the sharpness of the category boundary described by our Chinese participants, we conducted a further but preliminary analysis. First, the data of all Chinese native subjects were grouped by the order of experiment rounds. Recall that we had 15 rounds in each block. That is saying, we grouped the data from Chinese participants in the alveolar trill block into 15 subsets. We then averaged the data across participants in each subset and fit them to the logistic regression function. This way, we are able to see the changes in data along rounds. It can be seen from Figure 4.1 that the data in round 1 did not fit the function well. The fitted curve shows a very shallow slope. However, in round 15, the curve stretched out with a sharper slope.

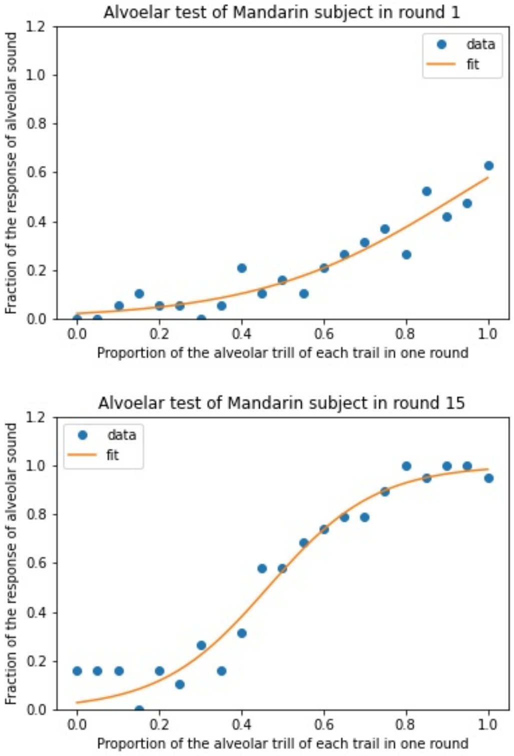

**Figure 4.1: The Average logistic regression curve of all Chinese subjects when they perceive alveolar trill in round 1 and round 15.**

Most Chinese native speakers have never produced or noticeably perceive the sound of alveolar trill in their daily life. It was difficult for them to classify such sounds at the beginning. Interestingly, the changes in the round-by-round analysis indicate that our participants were learning to identify these sounds within the experiment. In the present thesis, we did not go further to reveal more evidence of this fast learning process. But it is worthwhile to look at how statistically significant the round-by-round progress is. This can be a very good direction for further research.

To summarise this research on the asymmetry in cross-linguistic phonological perception, we designed an experiment to answer our research questions that were there any differences between native and non-native language speakers when perceiving native and non-native Speech sounds? We chose Tone in Mandarin as the experimental condition and Alveolar trill in Dutch as the control condition. We were interested in seeing whether or not there were significant differences between Chinese and Dutch speakers in the category boundary and its sharpness in identifying Mandarin tones and Dutch trills. The results showed that participants from different native language backgrounds only have significant differences in the sharpness of the category boundary when classifying alveolar trills. Such difference was not shown when it comes to tone categorizing. There are many reasons that can explain our findings. For example, Dutch native speakers could be exposed to prosodic sounds that are similar to tones in their daily life, while it is rare for native Chinese speakers to be exposed to alveolar trill sounds. In addition, the tone stimuli in the continuum were derived from a monosyllable, and they were described as non-meaningful syllables in the present experiments. So, such a design may also lead to the loss of the native language management effect in tone categorization.

## References

Chao, Y. R., & Chao, Y.-R. (1968). *Language and symbolic systems* (Vol. 260). Cambridge University Press Cambridge.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of mandarin tone sandhi. *Language Sciences*, *48*, 62–69.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of mandarin chinese tones by mandarin chinese vs. french listeners. *Journal of phonetics*, *32*(3), 395–421.

Kawahara, H. (2014, 12). Tutorial: Straight analysis..

Kummer, A. W. (2001). *Cleft palate and craniofacial anomalies: the effects on speech and resonance.* Taylor & Francis US.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in cantonese and mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*, 527–542.

McCawley, J. D. (1978). What is a tone language? In *Tone* (pp. 113–131). Elsevier.

Pinheiro, J., Bates, D., & R Core Team. (2023). nlme: Linear and nonlinear mixed effects models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=nlme` (R package version 3.1-162)

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus.* New York: Springer. doi: 10.1007/b98882

Rogers, L. (2017). *Discrimination testing in sensory science: a practical handbook.* Woodhead Publishing.

*Stemmenbureau voicebooking in amsterdam.* (2022, Nov). Retrieved from `https://www.voicebooking.com/nl`

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in psychology*, *2*, 120.

Tseng, C.-y. (1981). *An acoustic phonetic study on tones in mandarin chinese.* Brown University.

van der Bij, L., Gilbers, D., & Kehrein, W. (2017). Perception of word stress. *Mining for Parsing Failures*.

Wang, Y., Yang, X., Zhang, H., Xu, L., Xu, C., & Liu, C. (2017). Aging effect on categorical perception of mandarin tones 2 and 3 and thresholds of pitch contour discrimination. *American Journal of Audiology*, *26*(1), 18–26.

Wu, X., & Lin, H. (2008). Perception of mandarin tones by mandarin and english listeners. *J. Chin. Lang. Comput.*, *18*(4), 175–187.