university of groningen / faculty of science and engineering

Investigation whether using a game improves the learning of Mandarin Chinese tones in comparison to traditional training

Bachelor's Project Thesis

Tessa Zwart, s3683850, t.zwart.1@student.rug.nl,
Supervisors: Stephen Jones & Mi Tang

Abstract: This research examined what a more effective way of learning Mandarin Chinese tones is: training using a game or a more traditional method. Playing games is an effective way of learning, as the level of involvement and the motivation of the player might increase. The participants in this experiment did a training using a game, a test before this training and after the training. In addition, a generalization test was performed. Evidence was found that the traditional training method outperformed the training method with the game in distinguishing the Mandarin tones. This could be explained by participants first focusing on learning the game rather than learning the tones directly. Also, the game was based on implicit learning and the traditional learning method was not, so the participants did not know what to expect. An example of future research is to extend the period of learning so that the participants have more time to first learn the game.

1 Introduction

Mandarin Chinese is the language with the highest number of native speakers (Eberhard et al., 2021). The language is a tonal language, which means that using different pitches of the same syllables can result in different meanings of the syllable. Mandarin distinguishes four different tones: 1) tone with high-falling pitch, 2) tone with high-rising pitch, 3) tone with low-dipping pitch, and 4) tone with high-falling pitch. The syllable 'ma', for example, means either 'mother', 'hemp', 'horse', or 'scold' in the four different tones (Jongman et al., 2006).

Huang (2000) stated in their paper that one of the biggest difficulties of learning the language Mandarin Chines is making the correct pronunciation. Participants of their experiment had, for example, difficulties pronouncing the difference between the words 'shisi' (fourteen) and 'sishi' (forty). The participants use three strategies to try to overcome this difficulty: 1) making notes when they were corrected, 2) using the dictionary if they were reading, and 3) using Pinyin to write down the Mandarin tones. Pinyin is putting the Chinese characters into the letters of the Latin alphabet. Even when the participants used these strategies, learning the pronunciation of Mandarin Chinese remained difficult. The participants needed multiple corrections for the same mistake they made, in order to understand the mistake and to say it correctly.

However, there are multiple ways to learn the different tones in Mandarin like a native speaker of a non-tonal language. Wang et al. (1999) examined if auditory training is useful to become able to distinguish the different tones in Mandarin. They asked eight native speakers of American English, a non-tonal language, to participate in their experiment and to learn the Mandarin tones in eight training sessions over two weeks. In these eight training sessions, the participants did a discrimination task in which they needed to choose the correct tone according to the sound that they heard. Direct feedback was given directly after each stimulus. The American speakers showed a 21% increase in the overall tone perception accuracy.

Even if the auditory training shows an increase of 21% in the perception accuracy, a question arises: Is there a way to improve the training of distinguish-

ing the different tones in Mandarin? In 2005, De-Haan (2005) already stated that 'language learners may benefit from video games'. The author wrote that by using a game the motivation of the learner might increase, as well as the level of involvement, which seems to be beneficial in learning a language. Next to that the repetition in games creates more chances for the language learner to learn the language, according to the paper. All these points show that there seems to be some advantage to use a game to learn Mandarin Chinese instead of using a traditional method, such as a training using flashcards.

Wade and Holt (2005) made a video game to teach four different sounds to participants. They created four different characters, each with its own sound. The sounds were not speech sounds, so the sounds were no existing words. The player of the game had to capture two of the characters and shoot the other two, based on the sounds. Further into the game, the player moved faster through the game, and the characters were shown further away from the centre of the game. This way, the player had to recognize the sounds of every character to be able to shoot or capture them on time, and thus to stay alive in the game. The participants, who played the game for 30 minutes, did score higher on a discrimination task after the game training than the chance level, which shows that they learned some of the sounds. However, there should be mentioned here that this game was based on teaching four different sounds, not teaching four different tones. Learning the different four different tones might be more complicated than using four different sounds.

This brings us to the research question: Does the use of a game outperform the learning of the tonal language Mandarin Chinese by non-tonal language speakers compared to learning by a traditional method, like learning using flashcards?

It is expected that the training with the game outperforms the traditional training in distinguishing the different Mandarin tones. It is shown that native speakers of a non-tonal language are able to learn the different tones (Wang et al., 1999), and using a training with a game the level of involvement and motivation of these speakers might increase, which seems to improve the learning (De-Haan, 2005). Also, the training will be similar to the training of Wade and Holt (2005), and their training also showed an improvement in learning different sounds.

2 Method

The purpose of the experiment was to learn the distinction between the four different Mandarin tones. The experiment consisted of four phases: a test before the training (pre-test), the training, a test after the training (post-test), and a generalisation test. The experiment was made in Psytoolkit (Stoet, 2010, 2017), which was also used to make the different tests (pre-test, post-test, and generalization test). The game was made in Unity Technology (2021) and posted online using the website of Simmer Industries (2021). The link to this game was implemented in the experiment in Psytoolkit.

The procedure of the experiment is shown in Figure 2.1.

2.1 Participants

Eighteen English speakers participated in the experiment. They only knew the English language and thus do not speak a tonal language. The participants were recruited via the website Prolific Academic Ltd (2021). The age of the participants ranges from 18 to 40 with a mean of 25.2. 15 of them said that they have some game experience in the past and 9 of them said that they still play games regularly.

2.2 Stimuli

Three different syllables (pei, kou, and bi) were used in the learning phase of the experiment. These three syllables were also used for the pre-test and the post-test. In the generalization test, three additional syllables (tai, du, and ge) were used. All four tones were generated for each syllable in both a female and male voice. So in total 48 (6 syllables * 2 voices * 4 tones) different stimuli were created. Each stimulus had a duration of 1000 milliseconds.

The stimuli were recorded via Audio-Technica AT2020 and processed via Adobe Audition CS6. Two voice actors were invited to do the recordings of the sounds, one female native Mandarin speaker and one male native Mandarin speaker. The stimuli were associated with a character depending on the

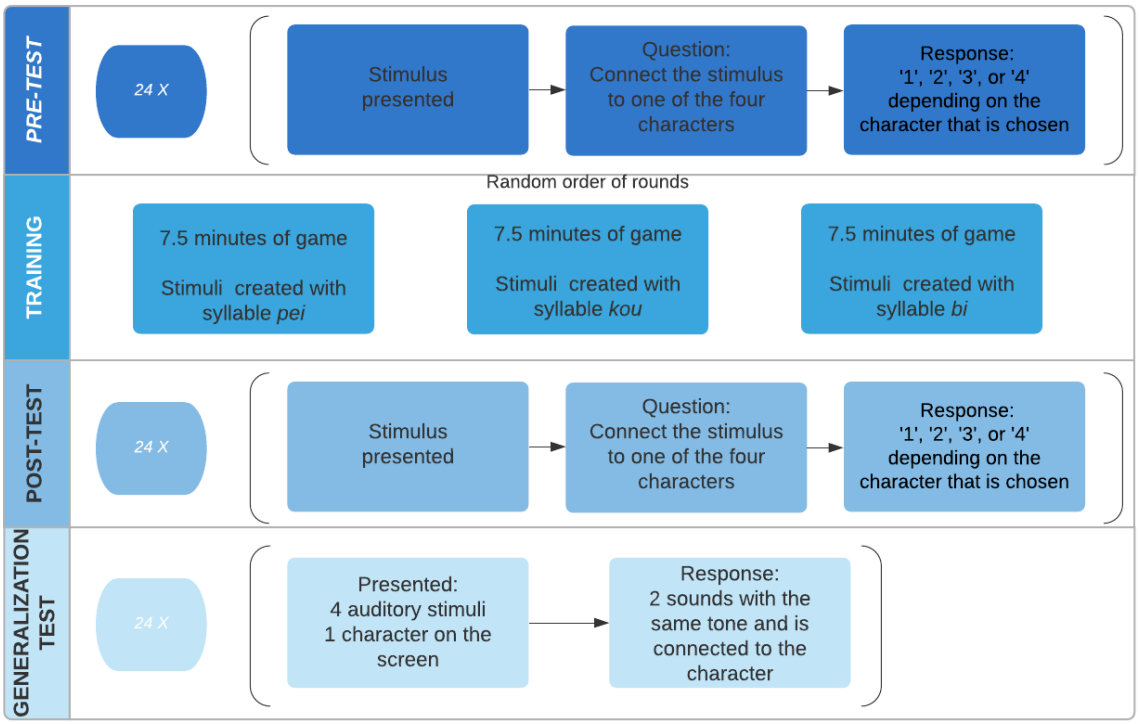

Figure 2.1: Procedure of the experiment

tone. The shape of the characters is similar to the pitch of tones. Chun et al. (2012) showed in their experiment that their participants said that seeing the pitch curves of the different tones helps them to improve their own use of the tones. So, the visualization of the pitches might help participants to learn the different tones. The characters are shown in Figure 2.2.

2.3 Pre-test

In order to find out the knowledge about the tones of the participants before the experiment, a pre-test was done. This way, knowledge can be compared with the knowledge of the test after the training and the improvement in tone distinction can be examined.

In this test, the stimuli were used that were created using the syllables pei, bi, and kou in the male and female voices. As the stimuli were made in all the four different tones, this resulted into 24 stimuli (3 syllables * 2 voices * 4 tones). The options of the characters as is shown in Figure 2.2 were displayed on the screen at a time for 5000 milliseconds or till

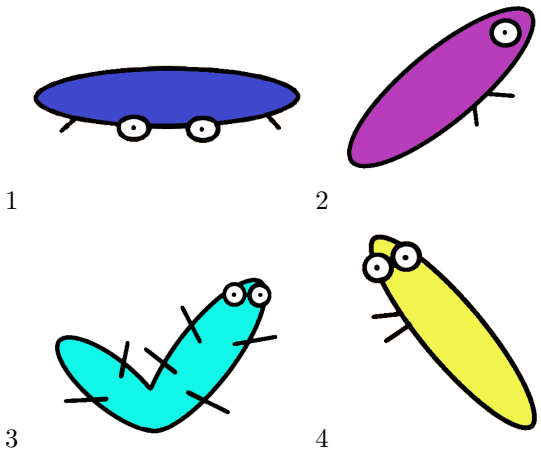

Figure 2.2: The different characters that are connected with the different tones: 1) the flat tone, 2) the rising tone, 3) the lowdipping tone, and 4) the falling tone. The shape of the characters shows the pitch of the associated tone.

the participant chose an answer. Since each stimulus lasted 1000 milliseconds, participants had 4000 milliseconds to respond. The participants had to connect the stimuli they heard to one of the characters on the screen, which they thought belonged to the sound.

2.4 Training

2.4.1 Game method

To train the distinction between the Mandarin tones an online game was created. The game was based on the game that is described in the paper of Wade and Holt (2005). This idea of their experiment was used and implemented in the training for this experiment. However, some adjustments were made to make the game more suitable for this experiment. The game was modified to a 2D environment.

The goal of the experiment was to learn the distinction between Mandarin tones. However, the explicit goal in the game was to score as many points as possible by killing enemies. The enemies that were used are the characters shown in 2.2. Each character represented its own tone. The game was played three rounds, each with the stimuli of one of the syllables (pei, bi, or kou). Each round was 7.5 minutes.

The players started with five lives and they lost a life when they failed to kill an enemy. If the players lost all their lives, the game started again after 4 seconds, and the round only stopped once the 7.5 minutes were reached. After each round, a code was given. The participants needed to fill in this code in Psytoolkit to continue the experiment. This way, we could check if the participants completed all different rounds. The score started at zero and the player earned 10 points if an enemy was shot.

Once an enemy appeared, the player could kill it, by first selecting it and then shooting it. The enemy was selected if the player moved the mouse over it, then the enemy turned into the colour red. After that, once the enemy was selected, the players had to click on the left side of their mouse to kill the enemy.

Each of the four sides of the game was consistently associated with one tone, and therefore one of the four characters. The enemy that represented tone 1 always occurred on the top of the screen, the enemy with tone 2 always on the left side of the screen, the enemy with tone 3 on the bottom of the screen and the enemy with tone 4 always on the right side of the screen. When an enemy appeared a stimulus was played. This stimulus was in the tone that was represented by this specific enemy in either the female or the male voice. The enemies were shown for 4 seconds in the beginning, but this time decreased every time after an enemy showed up by 0.05 seconds to a minimum of 2 seconds. So the enemies appeared for a shorter amount of time when the player was further in the game. When an enemy showed up, the stimuli were played repeatedly till the enemy was shot or disappeared.

Every time a certain enemy was shot by the player, this enemy appeared the next time 0.5 pixels further away from the player till a maximum of 5 pixels. Later in the game, this means that the player could not see the enemy as it has moved out of the screen and the player had to focus on the sound they heard. The player had to recognize the tone of the sound in order to know in what direction to move and to kill the enemy in time. So, the player had to connect the correct tone to the correct enemy to be able to kill the enemy on time. This is how the player was able to learn the different Mandarin tones. The player could move through the space using the WASD-keys, W for moving up, A for going to the left, S for moving down, and D for going to the right. When the player failed to shoot an enemy, the next time this enemy occurred closer to the player again, although one of the player's lives was lost. The procedure of the game is visible in the pseudocode 2.1. How the game looked is shown in Figure 2.3.

2.4.2 Traditional method

An experiment to evaluate the effectiveness of the traditional training method was performed by Osinga (2021). In this experiment, 18 participants did the same tests (the pre-test, post-test and generalization test), but the training was different. The training consisted of three rounds. In each round, the participants were shown three times 8 stimuli of the same syllable (kou, bi, or pei) together with the image of the corresponding character. After each set of 8 stimuli, they heard the stimuli again and they had to choose the corresponding character of each stimulus. When this stimulus was played, the

Algorithm 2.1 Procedure game

```
highscore ⇐ 0
while 7.5 minutes are not finished do
  time ⇐ 4
  for all enemy do
    distance[enemies] ⇐ 2
  end for
  score ⇐ 0
  lives ⇐ 5
  while lives > 0 do
    enemy ⇐ one of the four enemies
    enemy appears at distance[enemy] from the
    player
    if player hits enemy faster than time then
      enemy disappears
      score ⇐ score + 10
      if distance[enemy] < 5 then
        distance[enemy] ⇐ distance[enemy] +
        0.5
      end if
    else
      enemy disappears
      lives ⇐ lives − 1
      distance[enemy] ⇐ 2
    end if
    if time > 2 then
      time ⇐ time − 0.05
    end if
  end while
  if score > highscore then
    highscore ⇐ score
  end if
end while
```

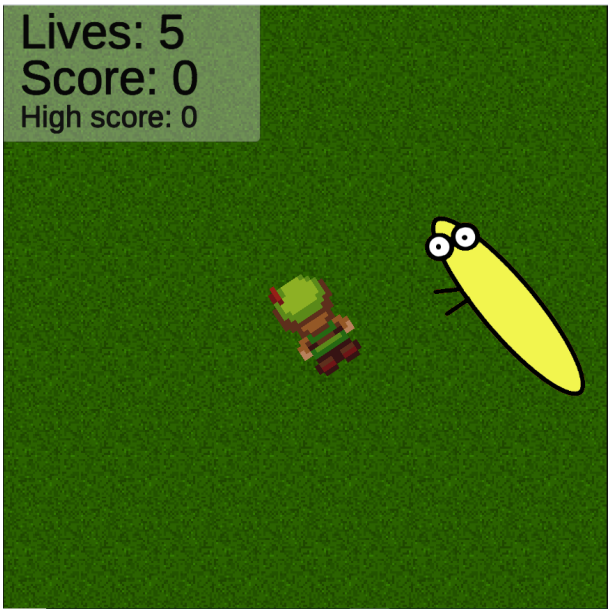

Figure 2.3: The game that the participants need to play during the training with the game. On the left top corner, the score, high score, and the number of lives are visible. The enemy with tone four is shown. This enemy always appears on the right side of the player.

screen showed the options of the characters as is shown in Figure 2.2. The participants had to press '1', '2', '3' or '4' depending on the tone they heard. The participants received feedback on whether they chose the correct character according to the tone or not.

2.5 Post-test

A post-test was done after the training. This way, the improvement of each participant could be examined. The post-test was the same as the pre-test.

2.6 Generalization test

A generalization test was performed to see whether the participants were also able to distinguish the different Mandarin tones for different syllables. Also, in this test, we wanted to look if the participants made a distinction between the syllables rather than the tones.

In the generalization test stimuli were used made out of six different syllables (pei, bi, kou, tai, du,

and ge). Four stimuli were played in each trial to the participant, and the participant had to pick the two stimuli that had the same tone. The participants had to fill in the two numbers of the stimuli with the same tone. So, for example, if the participant fills in '12', it means the first and the second stimuli have the same tone according to the participant. Two of the four stimuli had the same tone but different syllables, and two of them had the same syllable but different tones. The last was done to see if the participants made a distinction between the syllables rather than the tones. The test comprised 24 presentations of four stimuli.

3 Results

After the experiment, the data of both the traditional and the game methods were analyzed and compared. First, the improvement in the distinguishing of the tones was computed for both the traditional method and the game method separately. Then, the two methods are compared using all three tests (pre-test (test before the training), post-test (test after the training), and generalization test).

3.1 Learning of tones

Figure 3.1 is a boxplot showing the number of correct identified tones. As the pre-test and the post-test are the same for both the game method as the traditional method, the results can directly be compared. When looking at the boxplot, the number of correct assigned characters in the pre-test of the game method (Mean = 6.056; SD = 2.071) is similar to the post-test of the game method (Mean = 8.056; SD = 3.903). There seems to be a difference between the number of correct assigned characters in the pre-test (Mean = 6.556; SD = 2.874) and post-test (Mean = 15.778; SD = 5.976) of the traditional method.

A two-way ANOVA was conducted that examined the effect of the different tests (pre-test and post-test) and the different methods (traditional and game) on the number of correct identified tones. This test shows that the interaction between the methods and the tests is significant ($F_{1,71} = 14.79, p = 0.000268$). The interaction effect was examined using two simple effect analyses. One will be discussed in this section and the other in section 3.2.

To compute the difference between the pre- and post-tests of the two methods separately, the first simple effect analysis is done. This analysis shows that there is no difference in correct assigned characters between the pre- and the post-test of the game ($F_{1,73} = 2.268, p = .137$). However, for the traditional method, there is a significant difference between the pre- and the post-test ($F_{1,73} = 48.219, p = 1.79e - 9$). This means that the participants of the traditional method did show an improvement in categorizing the different tones, while the participants of the game method did not show an improvement.

3.2 Comparison pre- and post-test

Figure 3.1 it appears that the pre-test in both the traditional method and the game method is equivalent. To confirm if this is the case, another simple effect analysis was done. This test shows that there is no significant difference between the pre-test in the game method and the pre-test in the traditional method ($F_{1,73} = .142, p = .708$).

The same simple effect analysis was done to compute the difference between the post-test in the game method and the post-test in the traditional method. In Figure 3.1, there seems to be a difference between the two post-test and the simple effect analysis supports this with a significant difference ($F_{1,73} = 33.809, p = 1.79e - 7$).

3.3 Comparison generalization test

In the generalization test, the participants were presented with four different sounds of which two had the same tone and two had the same syllable. The participants had to identify the two sounds that had the same tones. The participants' responses are categorised into four groups: 1) successful generalization: the participant chose the correct sounds with the same tone, 2) confusion caused by syllables: the participant chose the two sounds with the same syllable, 3) no generalization: the participant only filled in the sound they recognized from the training, and 4) no learning: the participant chose neither the two sounds with the same tone nor the two tones with the same syllable nor the sound that they recognized from the training.

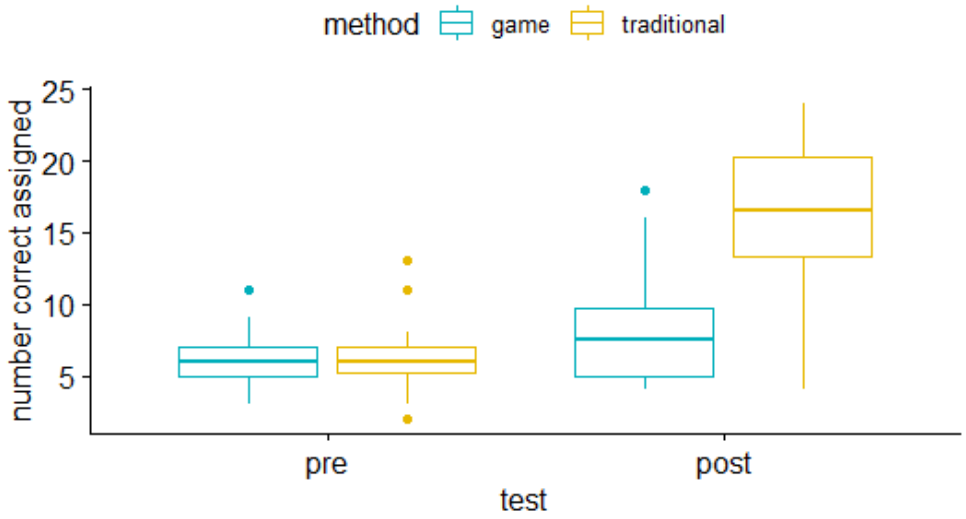

Figure 3.1: The number of correct assigned characters divided into the pre- and post-test for both the traditional and the game method.

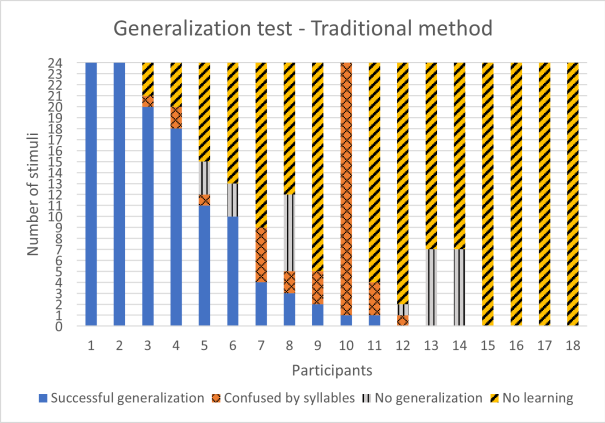

Figure 3.2: The results of the generalization test after the traditional training.

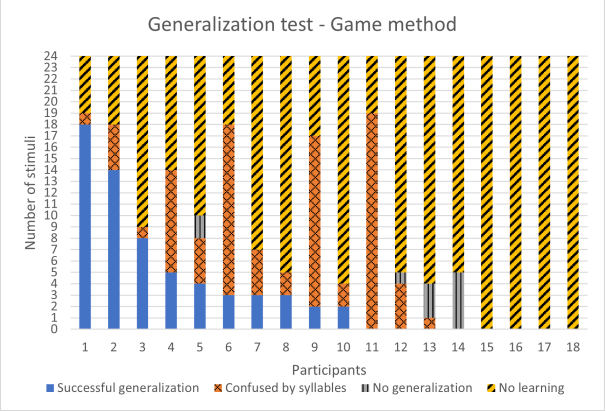

Figure 3.3: The results of the generalization test after the training with the game.

There is a considerable variability between the participants. It appears that some participants did not understand the task fully, which resulted in answers that did not consist of only two numbers, but sometimes also letters or four numbers for example. Because the responses of all participants differ a lot from each other, the decision is made to not build a statistical model. The data from the generalization test of the traditional method is shown in Figure 3.2, and the data from the game method is shown in Figure 3.3. The graphs are ordered in descending order first in category 1 (successful generalization), then 2 (confused by syllables), then 3 (no generalization), and finally 4 (no learning).

In both figures, the left-hand side shows the participants that seem to understand the test, as they answer at least some questions correctly. As the participants on the right-hand side of the figures seemed to not understand the test, we will focus on the participants that correctly identified at least one of the generalization questions. By looking at these participants in Figure 3.2, which is done after the traditional training, it can be seen that some participants got (almost) all answers correct. In Figure 3.3, the generalization test result after the training with the game is shown. The participants that did this method seemed to show more 'confused by syllables'-responses than the participants that did the traditional training.

4 Discussion

In this project, we looked into the best way to learn the distinctions between the Mandarin tones: using a training with a game or a more traditional training using flashcards. We expected that the training with the game would outperform the traditional training, but the results were against our prediction.

The results of this experiment show a significant difference between the training using a game and the training using traditional training. The traditional training outperforms the training with the game. For the game, the comparison between the test before the training and the test after the training shows that there was no improvement in knowing the distinction between the different tones, while there was an improvement after the traditional training. Some participants were able to generalize the distinction between the tones. However, more participants that did the traditional training seemed to be able to generalize than the participants that did the training with the game. So, the traditional training showed more improvement in learning the Mandarin tones than the training using the game in this experiment.

We could think about some reasons why the traditional training showed more improvement in distinguishing the Mandarin tones than the training with the game. The first reason is that the training using the game might be more difficult than the traditional training. The training with the game is based on implicit learning, so the participants did not know what to expect, while this was not the case for the traditional training. Wiener et al. (2020) found evidence that giving explicit instructions about the rise of the pitch improves the production of the second and fourth tone. So, if explicit instructions were given to the participants that played the game, this might have increased the ability to distinguish the Mandarin tones for these participants.

Another reason could be that the cognitive load of the training using the game is higher than the traditional training. The cognitive load means that people can only process a limited amount of new information, as Sweller (2011) described. People are likely to be familiar with a traditional training method, while the game was completely new to them. This way, the participants that played the game focused more on learning the game than on learning the distinction between the Mandarin tones. The cognitive load of the training using the game is even higher, as there are more dimensions in the game. The traditional training only focused on the different tones and the shape of the enemies. The participants that did the training with the game also focused on the tones and the shapes, but also the direction that the participants had to go to play an important role. These participants had to remember the distinction in tones, the different shaped enemies and the places where each enemy appeared. So, the cognitive load of the training using the game was higher than the traditional method.

Thirdly, in the traditional training method, the participants had to focus on the sound immediately once they heard them. The participants that did the training with the game could play the game

in the beginning without listening to the sound. Because in the beginning, the enemy was visible in the game without moving, so the participants knew that they had to shoot that enemy without focusing on the sound that it was making. So overall, the participants that played the game had less repetition in listening to the tones than the participants that did the traditional training. Huang (2000) stated that the participants thought that learning the pronunciation of the Mandarin tones was one of the hardest things to learn Mandarin Chinese. The participants only resolved their mistake when they receive a correction multiple times. This makes it possible that the improvement of distinguishing the Mandarin tones was less in the training with the game than the traditional training.

Finally, the game training might have been too short to show an improvement in the distinction between different Mandarin tones. The training length of the game experiment of Wade and Holt (2005) was 30 minutes, which was 7.5 minutes more than the training time in this experiment. Lim and Holt (2011) created a game to teach native Japanese speakers the /r/-/l/ sounds in English in which the participant had to play 2.5 hours over five days. Also, Wiener et al. (2020) performed their experiment that was 30-minute testing and training combination over four days. So if the training phase using the game was extended, there might be an improvement in the distinction between the tones.

One thing that could be improved in this experiment is the explanation of the game. The game was explained once at the beginning of all three rounds, but as the game was fully new for the participants, they might need to practice it more before doing the experiment. In the generalization test, a lot of participants did not understand the test, as they did not fill in two numbers, although this was in the instruction. This could also have been practised more before, as this test might have been new for people. Another way to improve the explanation is to carry out the experiment in a supervised environment rather than doing it online so that the participants could be corrected immediately if they miss some information that was needed for the game or the tests.

Future research could extend the period of learning, maybe do it over multiple days as Lim and Holt (2011) and Wiener et al. (2020) did. This way, the participants have more time to understand the game and there might be a learning effect visible. Additionally, this could determine if the participants can still make the distinction between the Mandarin tones over a longer period and if there is a difference in improvement in distinguishing the Mandarin tones between the traditional training and the training with the game.

Next to extending the period, the generalization test can also be adjusted, as this test was not entirely clear for the participants of this experiment. This test can have a similar set-up as the pre-test and the post-test, only with stimuli that were not used during the training. The disadvantage of this is that it is not possible to check this way if participants tend to connect two sounds with the same syllable rather than two sounds with the same tone. However, if the participants are able to generalize and distinguish the tones in a generalization test that is comparable with the pre-test and the post-test, then it might not be needed to check this connection. In the end, the experiment is only about learning the different tones, and in real life, people know that they have to focus on the syllables in combination with the tones.

References

Chun, D. M., Jiang, Y., and Ávila, N. (2012). Visualization of tone for learning Mandarin Chinese. In *Proceedings of the 4th pronunciation in second language learning and teaching conference*, pages 77–89.

DeHaan, J. (2005). Learning language through video games: A Theoretical Framework, an Evaluation of game Genres and questions for Future Research. *SP Schaffer & ML Price (Eds)*, pages 229–239.

Eberhard, D. M., Simons, G. F., and Fenning, C. D. (2021). *Ethnologue: Languages of the World.* Twenty-fourth edition. Dallas, Texas: SIL International. Online version: https://www.ethnologue.com/guides/most-spoken-languages. (Accessed: 06.20.2021).

Huang, J. (2000). Students' major difficulties in

learning mandarin chinese as an additional language and their coping strategies. *ERIC*.

Jongman, A., Wang, Y., Moore, C. B., and Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. Cambridge University Press.

Lim, S.-j. and Holt, L. L. (2011). Learning foreign sounds in an Alien World: Videogame training improves non-native speech categorization. *Cognitive science*, 35(7):1390–1405.

Osinga, T. (2021). Unpublished project report. University of Groningen.

Prolific Academic Ltd (2021). *Prolific | Online participant recruitment for surveys and market research*. (Accessed: June 2021).

Simmer Industries (2021). *Simmer | Discover, Share, Play Anywhere*. (Accessed: June 2021).

Stoet, G. (2010). Psytoolkit: A software package for programming psychological experiments using linux. *Behavior research methods*, 42(4):1096–1104.

Stoet, G. (2017). Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1):24–31.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation*, volume 55, pages 37–76. Elsevier.

Unity Technology (2021). Unity. (Version 2020.3.7f1) [Computer software]. https://unity.com.

Wade, T. and Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, 118(4):2618–2633.

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

Wiener, S., Chan, M. K., and Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' mandarin tone productions? *The Modern Language Journal*, 104(1):152–168.