Tonal production by Dutch teenagers: How Dutch secondary school students improve in the production of Mandarin tones over a course of three years

Student name: Lianne Veen

Student number: S1367471

Program: MA East Asian Studies

Supervisor: Prof.dr. R.P.E. Sybesma

Date: 15-07-2020

Abstract

Secondary schools in the Netherlands have been offering Chinese Language and Culture as an official exam subject for three years now. Studies on Dutch students SLA of Mandarin are very limited in number, and established literature often studies participants in their twenties with English as their L1. This study aims to look at improvement shown in tonal production by Dutch teenagers aged 13-17 who follow the relatively new Mandarin course in secondary school, to provide deeper insight into SLA of Mandarin by Dutch teenagers, and into the effectiveness of the course.

There were two groups of participants: students with four months of experience in Mandarin, and students from the same school with three and a half years of experience. They were asked to produce a selection of both monosyllabic and disyllabic words. Their tones were visualized using Praat, and compared to both the tonal production of native speakers to identify errors. The production of the two different grades was then compared to one another to identify points of improvement.

The results showed that female students showed improvement mainly in pitch contour, and to some extent even acquired tonal coarticulation. The male students showed improvement in pitch height, and very little improvement in pitch contour. For both genders the production of tone 3 appeared most difficult, followed by tone 4, then tone 2. Tone 1 was the easiest to produce.

Table of Contents

1.0 Introduction

The secondary school subject Chinese Language and Culture has officially been an optional exam subject for three years in The Netherlands now. Students have class for one hour each week, in which they learn about Chinese culture, and are taught Mandarin. Classes are often taught by Dutch teachers of Mandarin, who use different teaching methods to let their students achieve the same level of Mandarin in either four or five years' time. Mandarin as a language is substantially different from any other languages normally taught at Dutch secondary schools (i.e. French, German, and English), and likely harder to perfect as a language within those 4-5 years of studying. Especially at the start, Mandarin would need more instruction than European languages so that the students get used to a different system of writing, and tones in pronunciation. It is likely that in secondary schools this process of learning the basics will take rather long, as they cannot offer enough hours of class to jumpstart good pronunciation and understanding of the characters.

Before Mandarin became an official exam subject in 2018, pilots were carried out on a few secondary schools starting in 2010. While research has been done on contents of the program, no research has been done on the development of Mandarin proficiency for the students individually. Besides a dissertation written by Zou (2017), there are very few studies that look at students with a Dutch mother tongue (L1) learning Mandarin (L2). Many studies look at English learners of Mandarin who study Mandarin at a university, and are thus older than secondary school students. This study looks at Dutch L2 learners of Mandarin in the age group of 13-17, and therefore it covers a gap in the literature, and adds to the existing body of literature on Chinese second language acquisition (SLA).

Since the scope of what can be researched, analyzed and discussed in this thesis is limited, the focus will be on tonal acquisition. The decision to focus on tones was based on several reasons: firstly, because it is measurable with a program called Praat (Boersma & Weenink, 2020) which enables for objective research; secondly because this processing of data would allow for clear visualization and analyzation; and thirdly because ABN (standard Dutch) is not a tonal language, and the languages Dutch students are normally taught do not include any tones, which therefore makes tonal acquisition one of the most interesting aspects to study. The research question central to this study is:

In what ways do Dutch second language learners of Mandarin improve
in Mandarin tonal production over a course of 3 years?

To provide an answer to the question this study will make use of Praat to analyze the students' tonal production acoustically. This entails that the tones will be visualized according to pitch contour and pitch height, to identify mistakes the students make. Participants will include students who have received around four months of Mandarin classes, and students with around three and a half years of experience in Mandarin within the same school. The students will be asked to pronounce a total of twenty-one words, thirteen of which are relevant to this research. The word list includes monosyllabic words and disyllabic words, written in pinyin with tonal diacritics. The students will be given a short survey to ensure there are no variables that might influence their results (i.e. a

student who has a Chinese background may have an advantage in tonal production over the students without a Chinese background).

In addition to these students, a control group of native Mandarin speakers will be asked to pronounce the same list of words. To determine whether the students' tonal production is accurate or not, the tonal production of native speakers will be compared to that of the students.

This thesis is divided into five sections. The first section establishes a framework for this study by discussing the established relevant literature in this field. This will be divided into literature on tonal production, L2 tonal acquisition, and challenges which L2 learners face. After discussing the literature, it is followed by the methodology and any variables involved. The results will be presented in the section thereafter, followed by a discussion of these results, and lastly the conclusion.

2.0 Literature review

In Mandarin, tone is lexical (Li & Thompson, 1981). Just like other phonemes, tone is used to distinguish lexical meanings. There are four tones. These four tones are a level high tone, referred to as tone 1; a high rising tone, referred to as tone 2; a low falling and rising tone, referred to as tone 3; and a high falling tone, referred to as tone 4. Mandarin also makes use of a neutral tone, sometimes referred to as the fifth tone, but since the production of the neutral tone is mostly dependent on the preceding tone, it will not be used in my research, nor will it be discussed in this section.

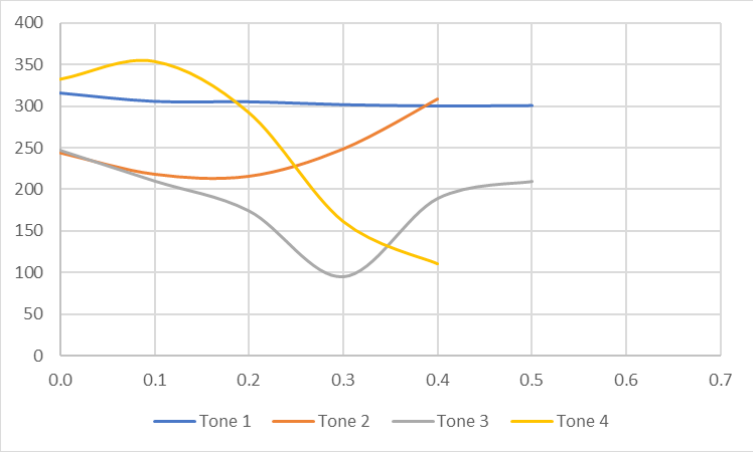


Figure 1: Visual representation of Mandarin tone 1, tone 2, tone 3, and tone 4, shown in frequency (Hz) over time (s).

Figure 1 shows the standard visual representation of these tones. The tones shown are the average frequency of two females from the Beijing area, both of whom have been used as a control group for the current study. One of the most common examples to show that tone is lexical is the syllable *ma.* First tone *mā* means "mother", second tone *má* means "hemp", third tone *mǎ* is "horse", and fourth tone *mà* means "to scold". Because of the semantic differences these tones cause, proper L2 acquisition of the tones is crucial to become fluent in Mandarin. To provide a framework for this study on L2 acquisition, this section will first discuss tonal production, then look at specific challenges L2 learners face, and lastly discuss previous research on L2 tonal acquisition.

2.1 Tonal production

There are many different factors that contribute to the correct production of any of the four tones. As can be deduced from Figure 1, pitch contour, the shape of the tone, is one of those factors. Other important factors are pitch height and pitch duration (Gandour, 1984; Jongman, Yue, Moore, & Sereno, 2006; Yang, 2015).

Jongman et al. (2006) discuss the perception and production of Mandarin tones. They note that both pitch height and contour are the most important cues in tonal perception. For tone 2 and 3, however, the turning point specifically is also of importance. The turning point is where the direction of the pitch changes from falling to rising. In the standard diagram of Mandarin tone there is no visible turning point for tone 2, but as can be seen in Figure 1, which shows the tones

produced by a native speaker, it does have a turning point. The turning point of tone 3 is later than that of tone 2. For clear differentiation between tone 3 and tone 4 the duration is also of importance. Both have a falling pitch, but tone 4 has a steeper and faster fall, while tone 3 takes more time to reach its lowest point. In addition, the starting pitch of tone 4 is higher.

Liao (1994) looked at tonal production in both phrases and syllables of seven different native Beijing Mandarin speakers. As others before her, she observed that tone 3 was often not produced according to the standard contour: the last rising part was often cut off, which made it a low level tone in contour. It was not completely predictable when the full contour would be used and when it would be cut off, but most often the full tone was pronounced when followed by a pause. This means that in single syllables produced in isolation the tone is likely to be realized fully, but in disyllabic words starting with a third tone, the rise of that third tone is often not realized. Liao showed that this was not the only tonal coarticulation applicable to the third tone. Besides studying monosyllabic and disyllabic words in isolation, she also focused on phrases to study intonation and pitch undulation. Liao found that tones are affected by preceding and following tones, resulting in tonal coarticulation. Tone 3 is affected by tonal sandhi when followed by any other tone. Tone sandhi occurs when the production of a tone contour or pitch changes due to surrounding tones. In Liao's study, this was shown when a third tone was followed by another third tone, and the one on the first syllable was produced as a second tone instead of a third tone; when followed by the other tones it will have a low level contour. Besides tone 3, tone 4 is also affected by neighboring tones; it ends less low when followed by other tones than when pronounced in isolation.

Pitch undulation, the tendency of tones in sentences to undulate instead of jump up and down, is partly caused by coarticulation, another cause is emphasis. Stressed syllables in sentences often have fully pronounced tones instead of them subjecting to coarticulation, while syllables with weaker stress cause the tones to adjust to pitch undulation (Liao, 1994: p. 161). This means that e.g. the pitch height of tone 1 does not always have the same frequency within a sentence, yet native speakers can still perceive the right tone because of the contours.

There are several other studies that suggest that pitch contour is slightly more important than pitch height (Gandour, 1984; Jongman, Qin, Zhang, & Sereno, 2017). Moore and Jongman (1997) specifically, looked at the turning point of tones 2 and 3 when researching speaker normalization. Speaker normalization means that a listeners adjust their expectations for relative pitch height, and thus the height of the four tones, depending on the speaker's natural pitch. Native speakers listened to two different native speakers, one of which had a high F0 range, while the other had a lower F0 range. Their range differed to the point where the low dip of tone three of the higher-ranged native would have the same pitch as the lower-ranged native's tone 1. Their results suggested that listeners do indeed adjust their expectations of tone height according to the speaker's range, but context speech is needed for it.

Considering these studies there seems to be a consensus that pitch contour is most important in tonal production, followed by pitch height. Pitch duration also plays a role, but a less significant one than the other two. This study will therefore mainly focus on pitch contour and pitch height in evaluating the tones students produce, but duration mistakes will be pointed out when clearly applicable.

2.2 L2 tonal acquisition

Research on Chinese foreign language (CFL) tonal acquisition is abundant, but only a very select few look at Chinese L2 learners with Dutch as mother tongue. Most research focuses on English learners of Mandarin, with an occasional focus on Japanese or Korean learners. This subsection will focus on research done on L2 tonal acquisition and discuss their methods and findings. Most of these based their research on Western learners of Chinese. These have been selected specifically to keep the L1 language as closely related to Dutch as possible.

Pitch height and contour were tested both acoustically and auditorily in the study done by Chun, Jiang, Meyr, and Yang (2015). They studied the effect that tone visualization might have on English learners of Mandarin. Tones can be visualized by plotting their change in frequency in a chart, as shown in Figure 1. The participants of this study, all young adults, studied a comparison of the visualization of their own and a native speaker's tonal production and tried to improve on it over a course of nine weeks. The words tested were all disyllabic words to include every possible tone combination. The auditory analysis was done by native speakers, trained to indicate whether they thought mistakes made were due to wrong pitch contour or pitch height, or whether both had no mistakes. The acoustic analysis made use of Praat (Boersma & Weenink, 2020), which visualized both pitch contour and height. Chun et al. (2015) discovered that the use of visualized training does positively affect tonal production, and that pitch height was more often incorrect than pitch contour. In addition, they showed that the native speakers that acted as judges were not always correct in pointing out the exact error (height or contour), suggesting that teachers might also not always be able to correct their students perfectly.

Lu (1992) studied both production and perception of L2 learners in her dissertation, and tested the participants over a period of ten weeks with monosyllabic words, disyllabic words, and sentences. The participants were young adults with English as their L1, and were at the beginning trajectory of learning Mandarin. All words tested were words the students had studied before, to make sure the results were not affected by them stumbling over new words. The production was evaluated auditorily by three Taiwanese judges. They only had to state whether the tone they heard was understandable to them, no specifics about the error were supposed to be noted, as opposed to the study by Chun et al. (2015) mentioned above. Lu's (1992) results showed that tone production for monosyllabic words was significantly easier for the participants than for sequences of words, for which he reasoned that students can focus on one segment and tone at the time, not having to worry about the next, which makes it easier than with disyllabic words or sentences. An interesting outcome is the fact that participants did better in tonal production in sentences than in disyllabic words, a cause of which is probably the familiarity with the sentences the participants already had. Lu also noted that the imitation test, where the participants had to repeat words after a native, was significantly easier, likely because it concerns instant processing and mimicking of information. Over the ten-week term, perception of the tones by the students increased significantly, yet production of the same tones did not. For the imitation task this can be attributed to the ceiling effect, but for the reading task the cause is not clear.

The relationship between a learner's awareness of tonal categories and his/her tonal production was researched by Li (2013). 'Tonal categories' refers to the label each tone has, instead of the specific acoustic features of each tone. In this case the tonal categories are thus the labels tone 1, tone 2, tone 3, and tone 4. Once again the participants had English as their L1, and were in their twenties. The tests were done with monosyllabic words with the pinyin, without tonal diacritic, provided. The participants were familiar with the words. For the first test they had to read the words out loud, for the second they had to write down the tone number that belonged to each word. Native speakers from the Beijing area evaluated their tones. In the end, they were interviewed to find out whether they thought awareness of tonal categories was important or not. Students performed better on the production test than on the tonal category test, which shows that the majority of the participants did not effectively make use of these categories when producing tones. They did, however, think that knowledge of the tonal categories was important, stating it made them feel more confident in their choices.

A similar study to Li (2013) is a study conducted by Wang, Jongman, and Sereno (2003). They researched the effect tonal perceptual training has on tonal production. Their participants were young adults with English as mother tongue who had just started studying Mandarin. Half of them were given perceptual training, the other half was the control group and was not given any extra training. Monosyllables were used for the pre- and posttest, half of which did not occur during the perceptual training to test generalization of the progress the participants might show. Native speakers from the Beijing area acted as judges on the production test. In addition, visual acoustic representations were also used to compare native speakers' tones with the participants' tones. The results show that perceptual training does indeed positively affect tonal production. Because acoustic analysis was done as well, it was shown that pitch contour and pitch height improved.

Hao's (2012) research included both English and Cantonese participants with an average age of 23. She used both mono- and disyllabic words to test production and perception in a tone identification task, an imitation task, and a reading out loud task. For the production tasks she relied on auditory analysis by native speakers from the Beijing area. She discovered most mistakes were made in distinguishing between tone 2 and tone 3 in both production and perception for both the Cantonese and English participants. In addition, distinguishing between tone 1 and 4 was difficult for those with Cantonese as mother tongue. The Cantonese learners of Mandarin did not seem to benefit from having a tone language as mother tongue when studying Mandarin.

The dissertation by Zou (2017) will be discussed specifically since it researches Dutch learners of Mandarin. Her dissertation does, however, have a phonological approach, which is not the intention for the current study. Still, the results can enable a deeper understanding on Dutch learners, specifically.

Zou (2017) makes use of four different experiments to study tonal coarticulation, segment-tone integration, the phonological processing of tonal contrasts, and the relative parts segmental and tonal information play in speech recognition. All four experiments included beginning and advanced Dutch learners of Mandarin, and a native control group. For the segment-tone integration experiment she also used a control group of Dutch native speakers with no experience in Mandarin.

Because of the different nature of her dissertation, I will not go into more detail about the experiments themselves, as they are irrelevant to this study. The results, however, provide information on Dutch learners of Mandarin. She shows that the advanced learners of Mandarin have better results in all tests, partly because they have started to process tones on a phonological level similar to native speakers. Her results also show that tones 2 and 3 are the most difficult tones for Dutch learners of Mandarin to perceive correctly, while tone 4 seemed most difficult to produce.

When looking at the studies mentioned above several differences and similarities are important to note. Almost all studies made use of monosyllabic words to look at tonal production in isolation, although several also made use of disyllabic words to take tone sandhi into consideration; whereas Hao (2012) and Lu (1992) also looked at full sentences. Since half of the participants of the current study are only at a beginner level, who are not expected to know tone sandhi, this study will look at both monosyllabic and disyllabic words, but the disyllabic words exclude words where the first syllable starts with tone 3. As mentioned in Liao's (1994) study earlier, tone 3's pitch contour changes easily when followed by another tone. Tonal coarticulation does occur with some of the other tones, but the change is significantly less drastic, which still allows for the comparison of the results between the native speakers and students properly.

The aforementioned studies also had two different ways of analyzing tonal production: auditory analysis and acoustic analysis. Auditory analysis involved native speakers listening to tones and identifying whether they were produced correctly or not, acoustic analysis involved the use of software to visualize tones. The study by Chun, Jiang, Meyr, and Yang (2015) used both methods, and showed that acoustic analysis is more reliable than auditory analysis, which is why this study will make use of acoustic analysis only.

The participants of the studies mentioned above are all in and around their twenties. This is an age group that is looked at most frequently, perhaps because finding participants at the researcher's university is easier than finding participants at high schools. In addition, the speed of learning is also considered to be higher at universities. The age group in this study consists of teenagers ranged 13-17, which would make an interesting contrast to the other age groups normally studied. It should be taken into consideration that, although children are often regarded to learn languages faster than adults (Yang, 2015: 30), the number of hours of Chinese high school students in the Netherlands are taught per week is considerably less than at university level.

2.3 L2 challenges

Challenges that L2 learners of Mandarin face differ depending on their L1. Most studies revealing difficulties have been conducted with English as L1, but overall there is no reason to believe the difficulties discussed below will not apply to students who have Dutch as their L1. This is mainly because the cause of these difficulties for English speakers is likely to be the same cause of difficulties Dutch speakers might encounter.

First of all, Wang, Jongman, and Sereno (2006), have shown that one of the main difficulties of learning Mandarin is the difference in pitch range between the L1 and L2. The pitch range needed for Mandarin is wider than that of standard non-tonal languages, which means L2 learners

will have to increase the pitch range they are comfortable with. This is also likely to be the reason that pitch height is more difficult to master than pitch contour.

Secondly, L2 learners of Mandarin tend to apply the intonation of their own language to their sentences when speaking Mandarin (Chen, 1993). This use of intonation influences the tones, thus creating errors in their production. However, since the beginner group of participants of this study are not familiar with full sentences yet, this is unlikely to be an issue for them. It would be interesting in further research to look at these effects for those at a more advanced level of Mandarin.

Additionally, tone 2 and tone 3 have been shown to be the most difficult tones to both perceive and produce for L2 language learners of different language backgrounds (Chen, 1993; Hao, 2012; Yang, 2015; Chun, Jiang, Meyr, & Yang, 2015; Lu, 1992). The cause is likely because they are acoustically similar (Lu, 1992), which causes much confusion for L2 learners.

Furthermore, both Li (2013) and Yang (2015) show that the development of tone categories is also important for tonal acquisition. Without the context and contrast of all the tones it is hard to learn to identify and produce them correctly.

Lastly, it is interesting to note that there is no consensus yet on whether or not L2 learners of Mandarin with a tonal background acquire correct tonal production earlier or not. Both Hao (2012) and Yang (2015) argue that it makes no difference, while Chan and Leung (2019) argue that those with a tonal background do acquire tones sooner. Chan and Leung also mention that having a musical background does not affect the speed at which one can acquire the tones.

Based on all studies mentioned in this section, the research question “In what ways do Dutch second language learners of Mandarin improve in Mandarin tonal production over a course of 3 years?” can be divided into four similar sub-questions:

1. What mistakes do Dutch female secondary school students with four months of Mandarin experience make in terms of pitch contour and pitch height when producing the four Mandarin tones in both monosyllabic and disyllabic words?
2. What mistakes do Dutch female secondary school students with three and a half years of Mandarin experience make in terms of pitch contour and pitch height when producing the four Mandarin tones in both monosyllabic and disyllabic words?
3. What mistakes do Dutch male secondary school students with four months of Mandarin experience make in terms of pitch contour and pitch height when producing the four Mandarin tones in both monosyllabic and disyllabic words?
4. What mistakes do Dutch male secondary school students with three and a half years of Mandarin experience make in terms of pitch contour and pitch height when producing the four Mandarin tones in both monosyllabic and disyllabic words?

With the results of these questions the last question can be answered:

5. In what aspects do the students with three and a half years of experience show improvement in tonal production as compared to the first year students?

As mentioned before, pitch contour and pitch height are important factors in tonal production. Testing both monosyllabic and disyllabic words provides for more data, and could show whether those in their fourth year already have the ability of using correct tonal coarticulation.

As mentioned in the background section, acoustic analysis provides for more reliable data analysis than auditory analysis. Therefore, in this study audio files of native speakers and the

students will be processed in Praat (Boersma & Weenink, 2020) to compare pitch contour and pitch height and identify deviations from the target.

3.0 Methodology

This section discusses the methodology of this study. The important aspects to provide an answer to the research question are, firstly, the participants; secondly, the words tested; thirdly, the control group of native speakers; and lastly, the method of data analysis itself.

The participants were volunteers from a secondary school in Noord-Holland that has been teaching Chinese for ten years. Due to time restraints it was not possible to use the same students for a 'before and after experiment', as several studies mentioned in the literature review did, so instead this study utilizes different age groups to study general improvement. One group of participants was around the age of 13, of which three males and four females. They all had around four months of experience in Mandarin, taught by a Dutch female teacher. That same teacher taught the Mandarin classes in all five grades, and has used the same teaching method throughout the past five years. The other group of participants was in the fourth year of Chinese and around the age of 17. Two females and two males participated in the tests. Both groups were not informed that tones specifically would be tested, to prevent them from practicing tones beforehand.

Initially there were ten volunteers (five male and five female) from both groups, but when the schools were closed due to the corona virus some complications occurred. Instead of going to the school in person to record the audio of the students and keep control of potential variables, the students were asked to record audio by themselves and send their files over the mail. In the email they were instructed to:

1. First fill in the survey.
2. Not practice the list of words beforehand. To open it and start recording immediately.
3. If they made a mistake, simply repeat that word again from the beginning.
4. Send their recording to the researcher by email.

The survey they were given was to establish background information and to ensure there were no differences between the students that would influence their L2 level (i.e. someone of Chinese background). The questions asked can be found in Appendix A.

Since the students now had to record at home instead of at school with the researcher, the responsibility of recording now lay with them. It took approximately six weeks total before the current participants had responded, the others never responded. This meant that some participants, had a few hours extra of Chinese class. This difference in number of hours of class is not likely to make a difference, however, since they have only one hour of Chinese each week, and the focus of these classes is not only on language but also on culture, meaning the difference in tonal training specifically is minimal. In addition to this discrepancy in hours, this situation also meant that there was no certainty whether they indeed followed the instructions, and whether they were the ones in the recording. This latter issue was resolved by asking the teacher whether the voice on the recording also matched the name, the former issue could not be resolved. After listening to the recordings, however, it did not appear as if any of them had practiced the words before recording. This assumption is based on the fact that all of the students had words they clearly struggled with, whether it be the pinyin itself, or the tone.

The list of words tested can be found in Appendix B. As stated, it included monosyllabic and disyllabic words. There were four monosyllabic words included to test their isolated tone

production. These words were four variants of *ma*: *mā, má, mǎ, mà*. There were nine disyllabic words included to test their production when not produced in isolation. These words were chosen based on tonal combinations: tone 1 followed by tone 2, 3, and 4; tone 2 followed by tone 1, 3, and 4; and tone 4 followed by tone 1, 2, and 3. As explained earlier, tone 3 was avoided on the first syllable because of the heavier tone sandhi it undergoes. After these words were selected, they were mixed into a list with a total of twenty-one words, to prevent students from realizing during recording that the focus was on tonal production.

The native speakers of the control group are from the Beijing area, and all speak Standard Mandarin. Two males and two females within an age range of 23-27 years old sent me their recordings of the same list of words as the students received. They also had to record at home, with their own equipment, because of the corona virus.

Lastly the method chosen to analyze the data was acoustic analysis. As mentioned in the literature review, acoustic analysis is more precise and more reliable than auditory analysis by native speakers (Chun et al., 2015). The drawback of the latter method is that, although a tone might not look entirely correct in acoustic analysis, a native speaker might still identify that tone as correct. To use both methods of data analysis, like Chun et al. (2015) did, was not possible due to time restrictions.

The acoustic analysis was done using Praat (Boersma & Weenink, 2020). Per word the visual contour was extracted manually, from which the frequency values were obtained, which were then plotted in an Excel chart. The values were measured for each tenth of a second.

4. Results

The results section is divided into five sections. First, the results of the monosyllabic words will be shown. These are divided into the control group's results, the production of females in their first year, the males' production in their first year, the fourth year results for females and, lastly, the males in their fourth year. In the section after that the results of the disyllabic words as produced by the control group will be shown, and then the relevant results produced by the students.

4.1 Native speakers

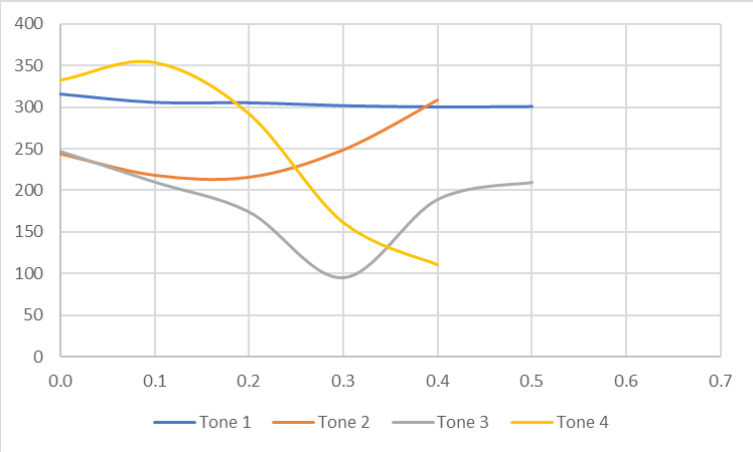


Figure 2: Average of four Mandarin tones measured by frequency (Hz) over time (s) as produced by two female native speakers.

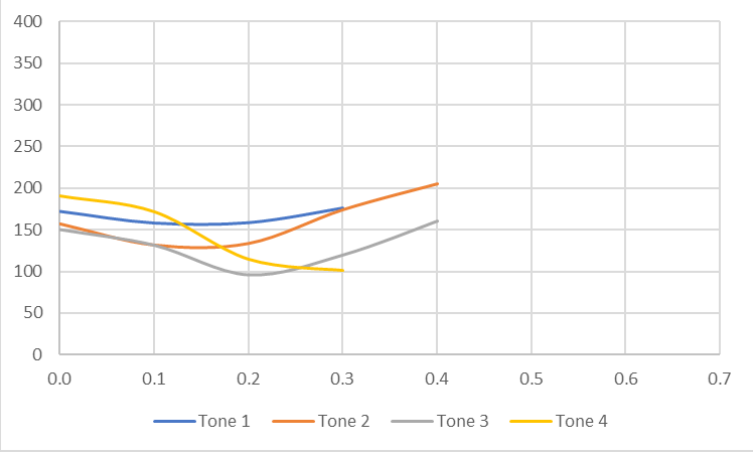


Figure 3: Average of four Mandarin tones measured by frequency (Hz) over time (s) as produced by two male native speakers.

Figures 2 and 3 show the average tonal production of the control groups used for this experiment. Pitch height, pitch contour, and pitch duration can be identified clearly in both visualizations. The contour of the female native speakers is more clearly defined than the contour of the males. The duration of tones 1, 3, and 4 is slightly longer in the females' production. The females have a higher pitch than the males.

4.2 Class 1 females

The production of the female beginner level speakers of Mandarin tone 1 is shown in Figure 4. In comparison to the control group it shows that the overall contour and duration is similar, yet the pitch height of two of the participants is lower.

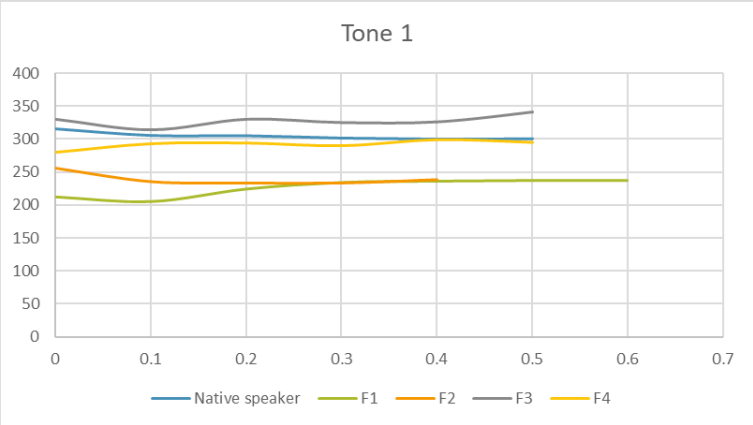


Figure 4: Tone 1 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

Production of tone 2 is shown in Figure 5. Females 1 and 2 have a shorter duration than the native speakers' but a similar contour, the register is slightly lower. Female 3 has a longer duration and the contour is slightly off. Female 4's register, contour and duration do not resemble the standard for tone 2.

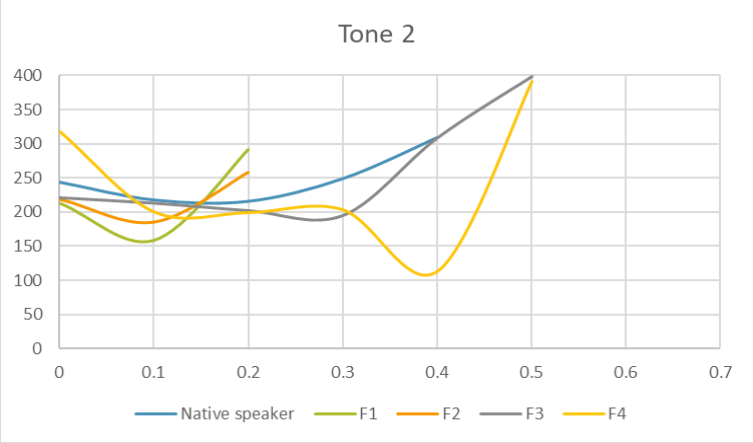


Figure 5: Tone 2 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

Figure 6 shows the production of tone 3. The contours of females 1, 2, and 4 do show both the falling and rising of pitch, but females 1 and 2 do not lower their pitch enough to get the right contour. Female 4's pitch, to the contrary, falls too fast and rises too high to get the contour right. Female 3 barely shows a decline in pitch, then rises too high.

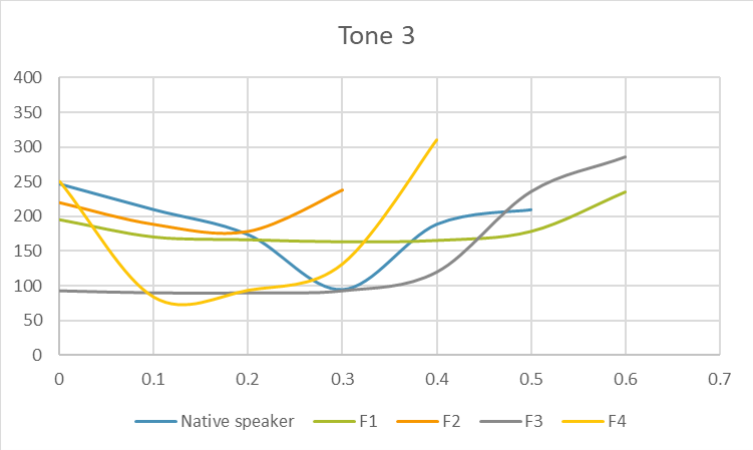


Figure 6: Tone 3 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

Tone 4 is shown in Figure 7. All four females start at a lower register than the native speaker, their falling pitch is not fast nor deep enough, resulting in a wrong contour. Their durations differ, but are all shorter than the native speakers' duration.

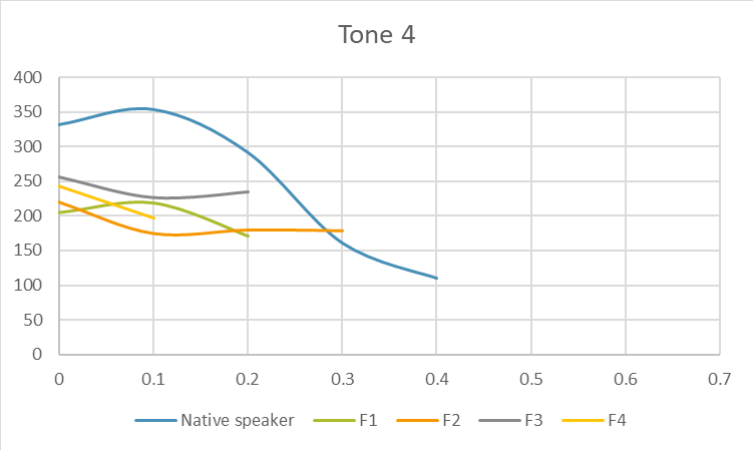


Figure 7: Tone 4 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

4.3 Class 1 males

Figure 8 shows the production of tone 1 by male beginner level speakers of Mandarin. The duration is longer than that of native speakers. The overall contours are similar, but the contour of male 3 falls too far. The pitch height of all participants is higher than the native speakers', the pitch height of male 2 is especially high.

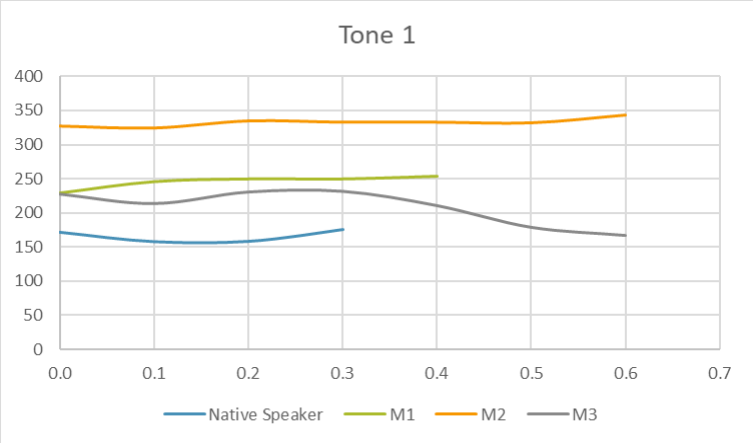


Figure 8: Tone 1 measured by frequency (Hz) over time (s) as produced by three Dutch male students at a beginner level of Mandarin, compared to the native speakers' production.

The production of tone 2 is displayed in Figure 9. The contour of male 2 is exemplary, yet the pitch height itself is too high. The pitch of the other participants is too high as well, and their contours rise first instead of falling. For male 3 the contour even shows a downward curve after the initial rise.

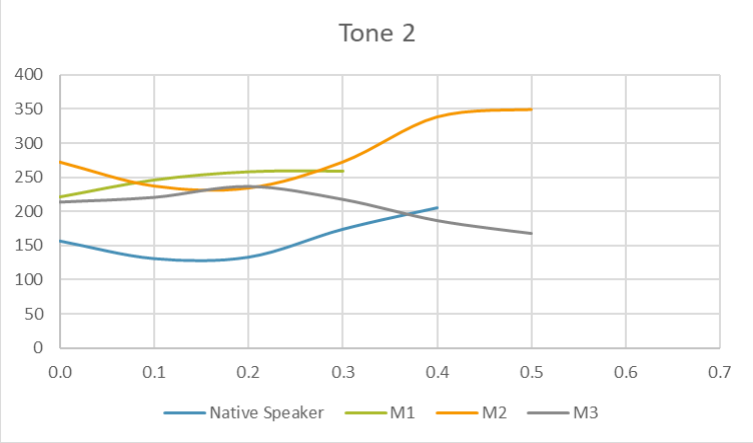


Figure 9: Tone 2 measured by frequency (Hz) over time (s) as produced by three Dutch male students at a beginner level of Mandarin, compared to the native speakers' production.

In the production of tone 3 the pitch height of all three participants is too high, as shown in Figure 10. The contour of both male 1 and male 2 show the falling and rising of a third tone, but the duration is too long to form the correct contour. The contour of male 3 first rises, and then falls.

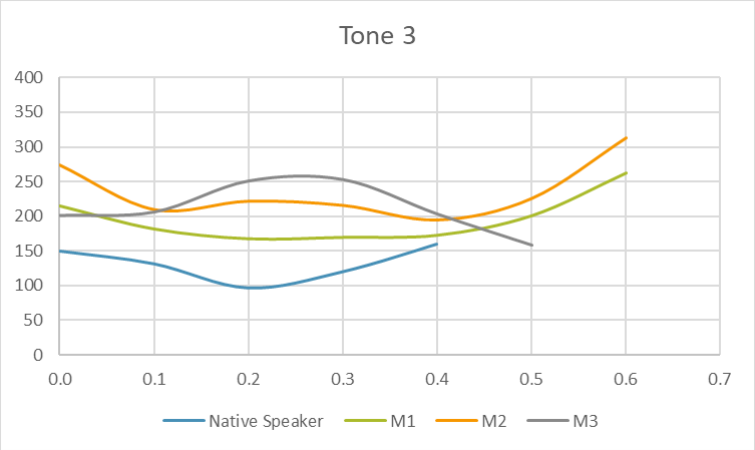


Figure 10: Tone 3 measured by frequency (Hz) over time (s) as produced by three Dutch male students at a beginner level of Mandarin, compared to the native speakers' production.

Figure 11 shows the production of tone 4. The pitch height of the participants is slightly higher than that of the native speakers. None of the contours match the native speakers' contour. Male 1 does have an initially falling pitch, but it is not steep enough and flattens out. Male 2's pitch falls slightly but then rises again. Male 3 first shows a rise in pitch, then a fall.

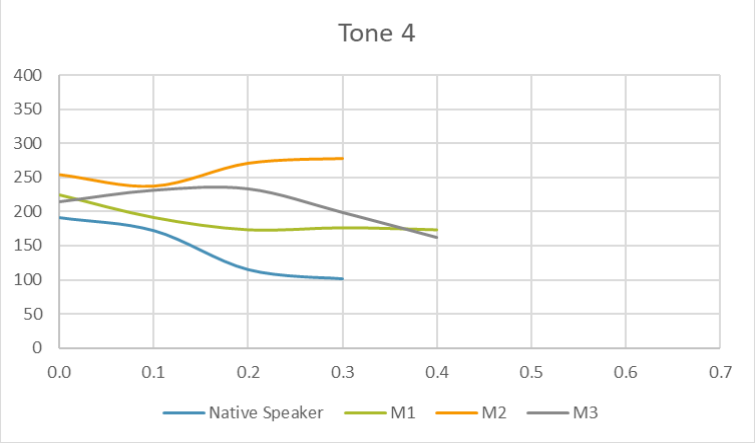


Figure 11: Tone 4 measured by frequency (Hz) over time (s) as produced by three Dutch male students at a beginner level of Mandarin, compared to the native speakers' production.

4.4 Class 3 females

The production of tone 1 by two female intermediate level speakers of Mandarin is shown in Figure 12. Both the pitch contour and pitch height of female 1 are similar to the native speakers'. Female two is at a lower pitch height, and the contour first falls, then flattens out, and lastly rises slightly.

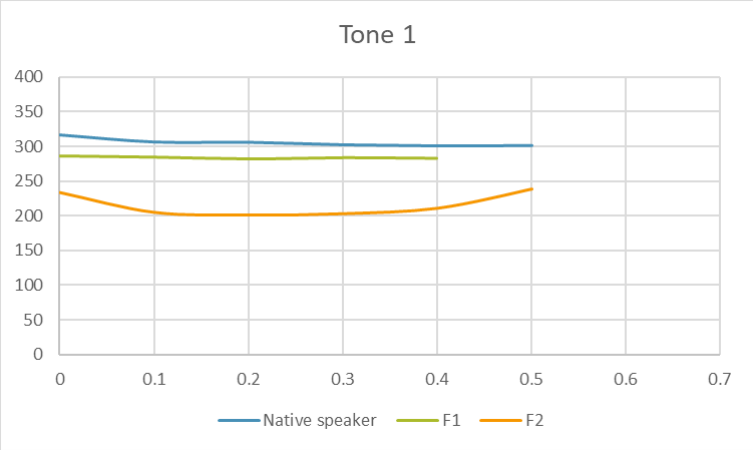


Figure 12: Tone 1 measured by frequency (Hz) over time (s) as produced by two Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

Figure 13 shows the production of tone 2. Both female 1 and female 2 show a pitch contour and pitch height similar to the native speakers. Their rise, however, happens later than the native speakers' production of the same tone, giving the contour a more flat appearance in the beginning.

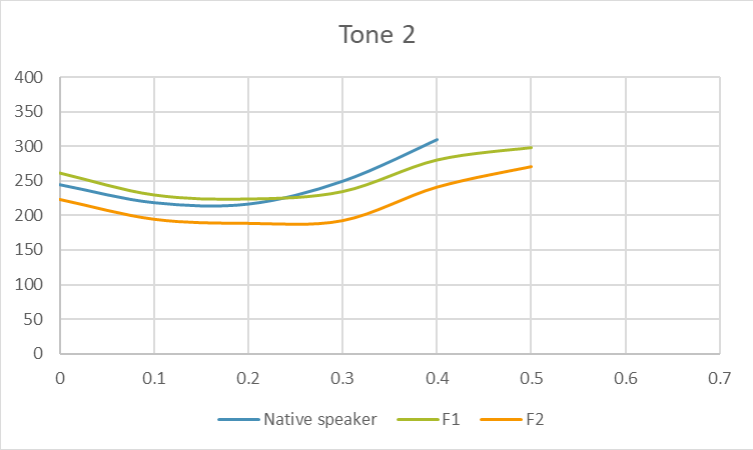


Figure 13: Tone 2 measured by frequency (Hz) over time (s) as produced by two Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

The production of tone 2 is portrayed in Figure 14. Female 1 does have a falling and rising contour at a similar pitch height as the native speakers, but the pitch falls too fast, and rises back too high. Female 2, on the contrary, does not fall low enough nor rise high enough in pitch to form the right contour. Her starting pitch is at a similar level as that of the native speakers' pitch.

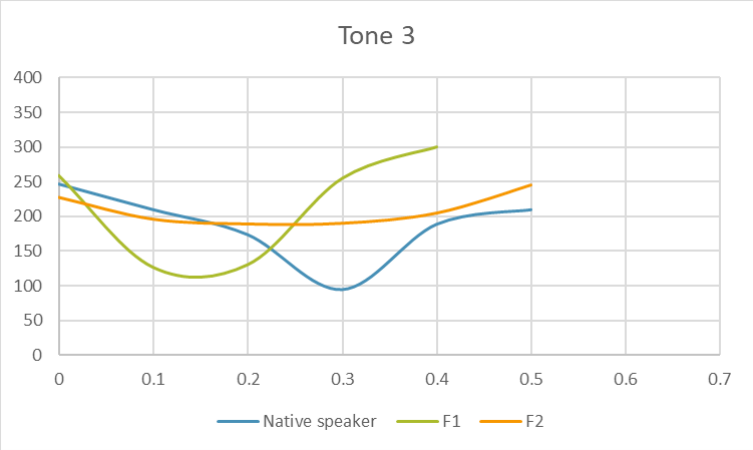


Figure 14: Tone 3 measured by frequency (Hz) over time (s) as produced by two Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

Figure 15 shows the production of tone 4. Female 1 starts with the right amount of raise in pitch, but does not fall in pitch fast enough to get the exact right contour. Female 2 first drops her pitch slightly, then rises again.

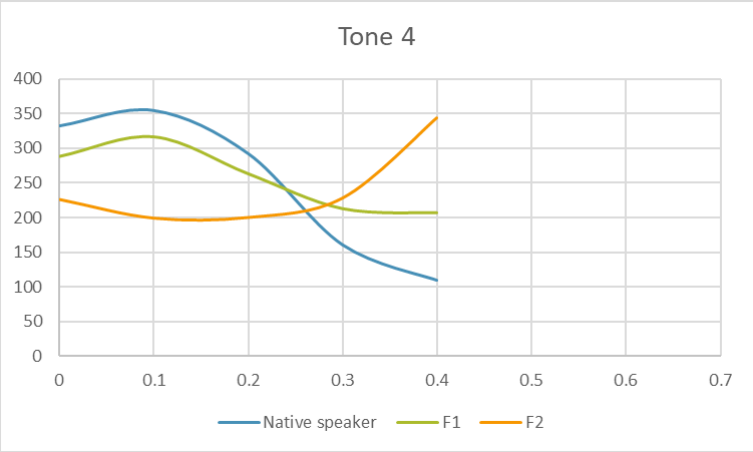


Figure 15: Tone 4 measured by frequency (Hz) over time (s) as produced by two Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

4.5 Class 3 males

Figure 16 shows the production of tone 1 by male intermediate level speakers of Mandarin. The overall contour of the two participants is similar to the native speakers' contour. Male 1 is at the same pitch height as the native speakers, male 2 has a slightly lower pitch height.

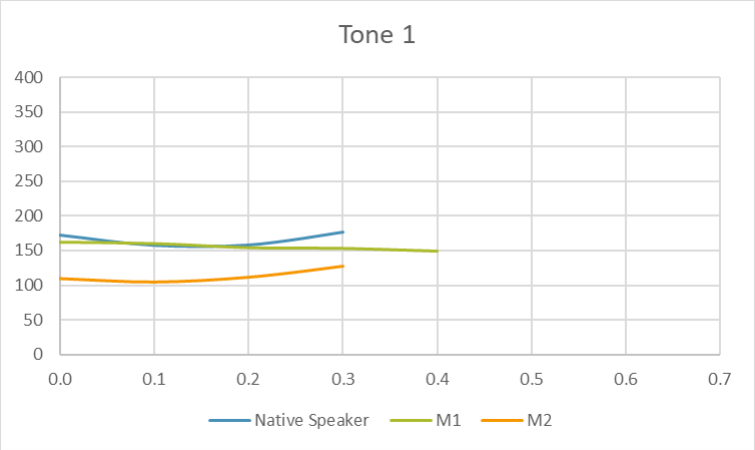


Figure 16: Tone 1 measured by frequency (Hz) over time (s) as produced by two Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

The production of tone 2 is shown in Figure 17. The pitch height of both participants is at a similar height as the native speakers. The contour of male 1 is very similar to the native speakers' average, but he starts the rise in pitch slightly earlier. Male 2 shows a similar slight drop in pitch, but does not start the rise in pitch after.

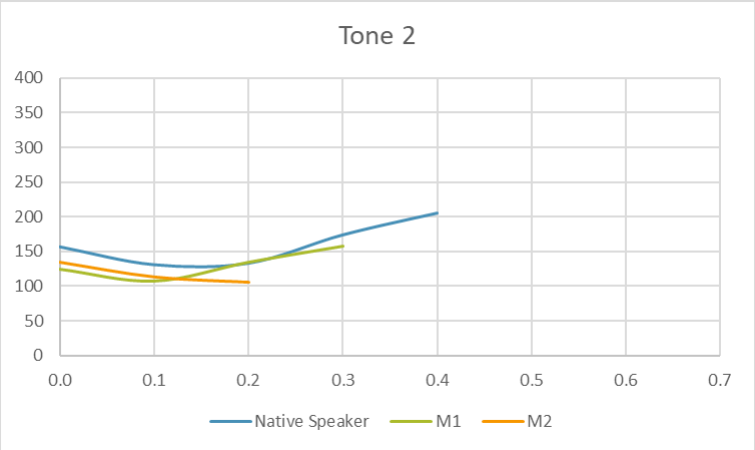


Figure 17: Tone 2 measured by frequency (Hz) over time (s) as produced by two Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

Figure 18 shows the production of tone 3. Male 1's pitch contour does show the rise and fall typical of the third tone, but the fall in pitch is too fast and too deep as compared to the native speakers. In addition, he starts off at a higher pitch. Male 2 does not fall low enough in pitch to create the proper contour, and neither does his pitch rise again.

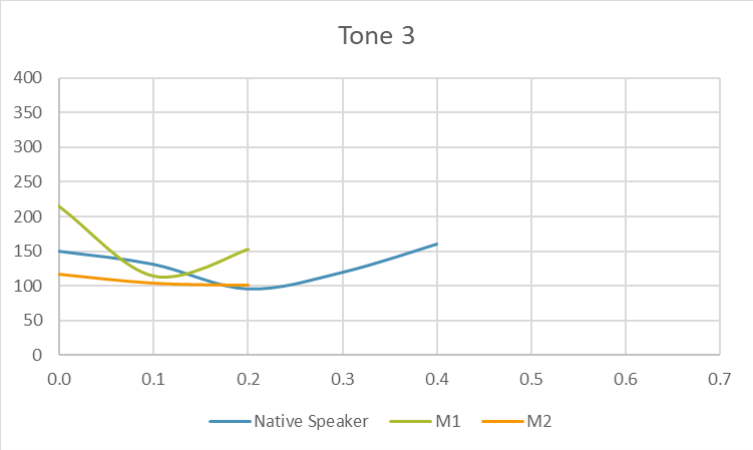


Figure 18: Tone 3 measured by frequency (Hz) over time (s) as produced by two Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

Tone 4 is portrayed in Figure 19. Neither of the participants have the right contour. They do both show a slight drop in pitch, but it is not fast nor low enough to get the right contour. Their pitch height is lower than the native speakers' pitch height.

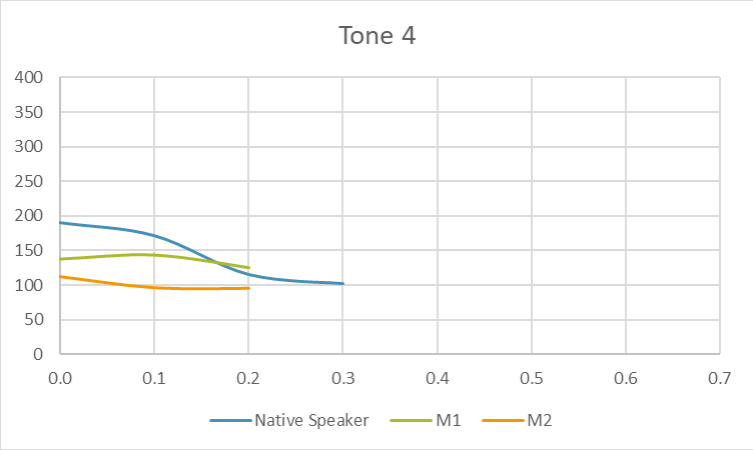


Figure 19: Tone 4 measured by frequency (Hz) over time (s) as produced by two Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

4.6 Disyllabic words control group results

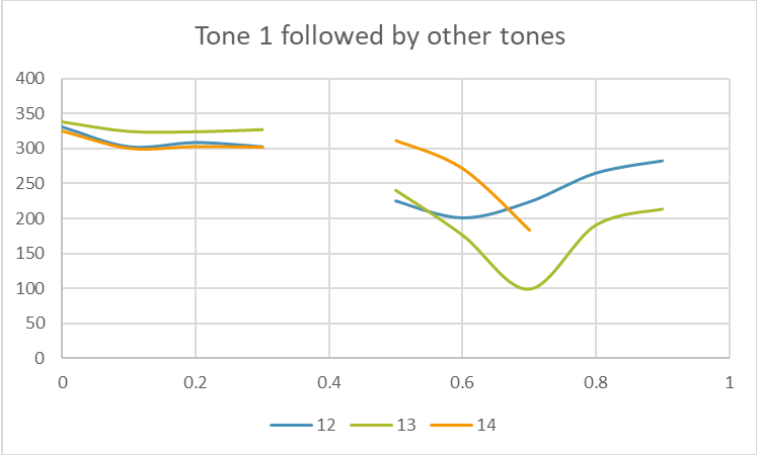


Figure 20: Average of Mandarin tone 1 followed by tone 2, 3 or 4 measured by frequency (Hz) over time (s) as produced by two female native speakers.

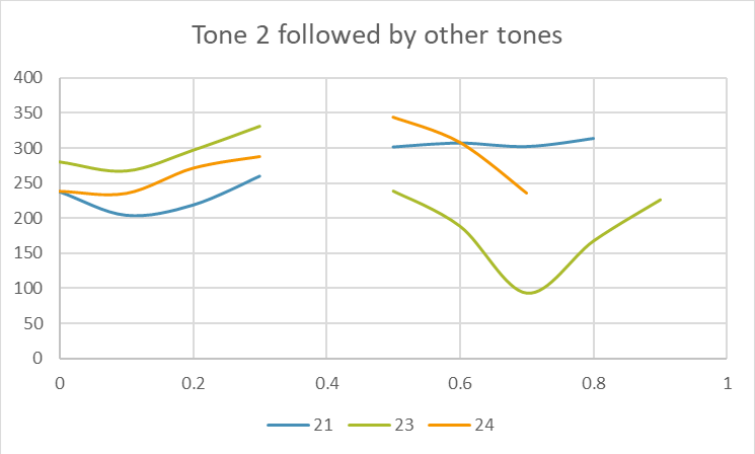


Figure 21: Average of Mandarin tone 2 followed by tone 1, 3 or 4 measured by frequency (Hz) over time (s) as produced by two female native speakers.

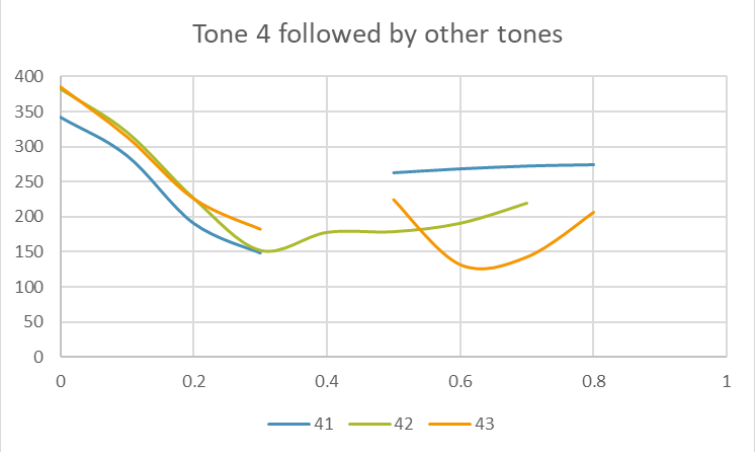


Figure 22: Average of Mandarin tone 4 followed by tone 1, 2 or 3 measured by frequency (Hz) over time (s) as produced by two female native speakers.

First of all it should be noted that, in plotting these tones for both the males and females, the gap portrayed between two tones was often 0.1 seconds long, in order to show the gap in the Excel

chart, however, one point had to be left empty, resulting in it showing on the chart as a 0.2 second gap.

From all three figures it shows that the first tone produced in the disyllabic words often has the same pitch height, contour, and duration for all three tones examined. The contours of the tones following the first tone produced are also similar in all cases shown. The pitch height, however, does show some changes. Tone 4 was shown to have a starting pitch of around 350 to 400 Hz, yet the tone 4 following tone 1 starts at a frequency of 325, at around the same height tone 1 ended, showing some coarticulation of the tones. Tone 1 following tone 4 is also lower than the 300+ Hz it is produced at when on the first syllable or in a monosyllabic word. Tone 2 is also interesting to look at, since there was no gap between tone 4 followed by tone 2. In all other instances there was a gap. This is, first of all, presumably due to the word they were asked to produce: *fàláng*, where consonant 'l' provides for a more continuous pronunciation of the two syllables, thus resulting in no gap. In addition, continuing into the rising tone 2 after the ending point of falling tone 4 seems more effortless than raising pitch in between the two. This does result in tone 2 having a lower pitch than when produced on the first syllable or after tone 1.

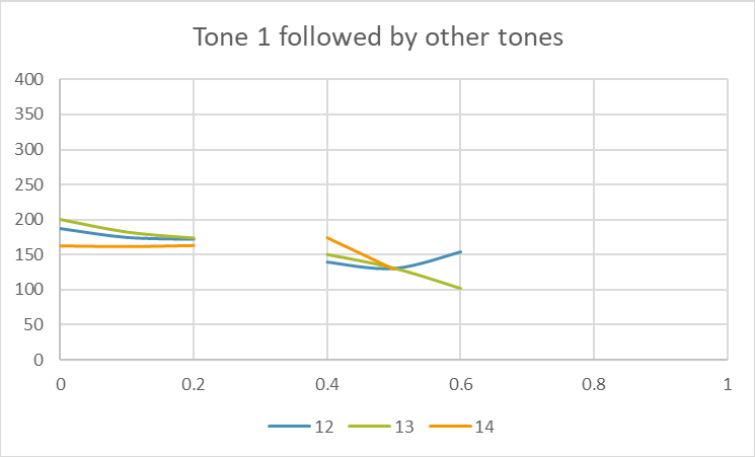


Figure 23: Average of Mandarin tone 1 followed by tone 2, 3 or 4 measured by frequency (Hz) over time (s) as produced by two male native speakers.

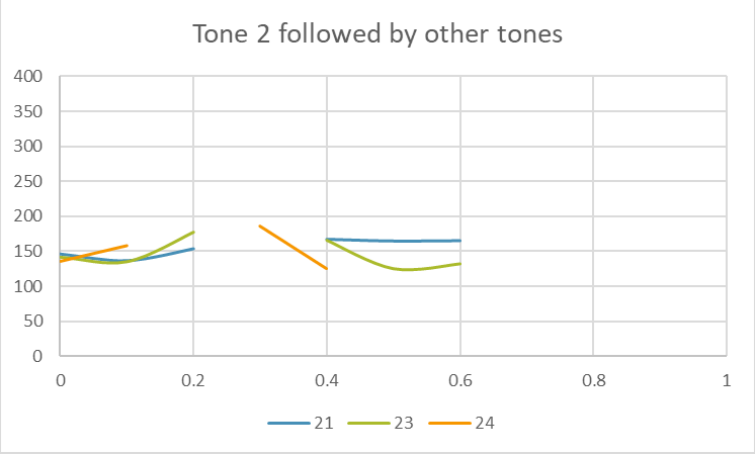


Figure 24: Average of Mandarin tone 1 followed by tone 1, 3 or 4 measured by frequency (Hz) over time (s) as produced by two male native speakers.

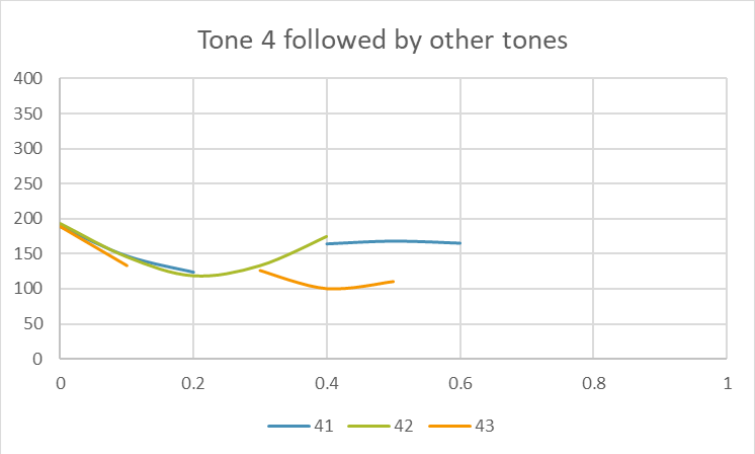


Figure 25: Average of Mandarin tone 1 followed by tone 1, 2 or 3 measured by frequency (Hz) over time (s) as produced by two male native speakers.

When looking at the tones the male native speakers have produced, less conclusions can be drawn than with the female productions. As with the monosyllabic words, the males' production is less clearly defined than the females'. Tone 3 following tone 4 seems to be at a slightly lower frequency compared to the other two tones, but just from these results it cannot be concluded whether this 25-50 Hz difference is significant enough to state that coarticulation took place. Tone 4 followed by tone 2, however, does show the same overall contour the females also had: tone 2 continues out of the tone 4 contour. The pitch height of the male tone 2, however, is not much lower than the tone 1 – tone 2 combination, nor is it much lower than tone 2 produced on a monosyllabic word.

4.7 Disyllabic words student results

This section will look at some of the visualizations of the disyllabic words produced by the students. Appendix C includes all other results of the disyllabic words by students. When looking at all of these visualizations, it is worth noting that, except for the native speakers, the gaps between the two tones portrayed were almost always longer than the 0.2 seconds shown. Most students struggled with the disyllabic words, and did not follow up with the second syllable as fast as the natives. The gap between these syllables in the charts is reduced to allow for easier comparison in both pitch contour and pitch height between both the native speakers' production as well as their peers'.

In addition, it should be noted that this task of disyllabic word production was clearly considerably more difficult for the students than the monosyllabic word production task. In the first year results of both males and females almost all tones have not been produced correctly. In the fourth year more tones were produced correctly, especially by the females. Selections have been made from all results to be represented below, based on the students' production in general. In example, if tone 2 was produced well in all instances by the third year females, then a selection of one or two charts was made to represent this conclusion. All results that are not shown below, can be found in Appendix C. The results shown there will still be used to draw generalized conclusions.

4.7.1 First year students' production of disyllabic words

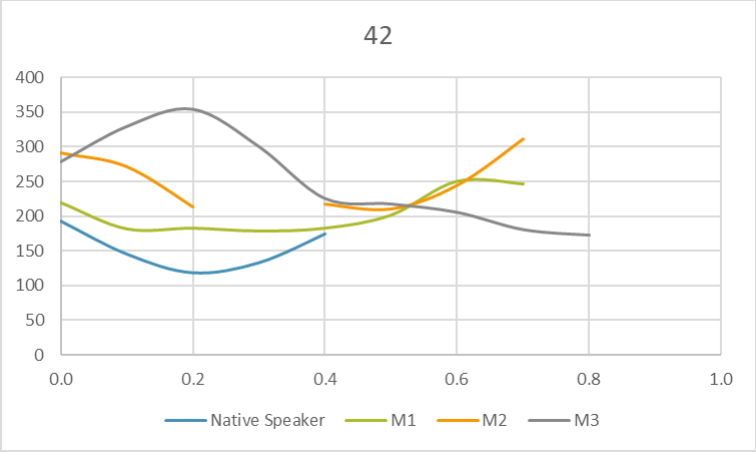


Figure 26: Tone 4 followed by tone 2 measured by frequency (Hz) over time (s) as produced by three Dutch male students at a beginning level of Mandarin, compared to the native speakers' production.

First of all, Figure 26 shows the production of tone 4 followed by tone 2 of males in their first year. This is one of the few results where the production was comparable to that of the native speakers. The pitch contours of both male 1 and male 2 are almost correct; both could produce a steeper slope for tone 4 to improve, tone 2 of male 2 is produced well, tone 2 of male 1 is slightly off because it flattens out at the end. For both, the pitch height is higher than the native speakers', and even though male 1 starts off with a similar pitch height, he ends much higher.

The rest of the results in Appendix C show that only one of the males produced tone one right fairly consistently, but otherwise all other tones are consistently produced wrongly by all males. Their pitch height was too high most of the time, and their pitch contours were off.

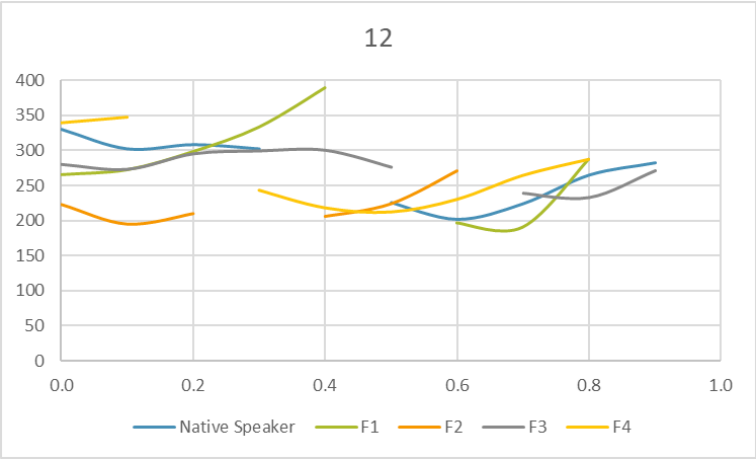


Figure 27: Tone 1 followed by tone 2 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

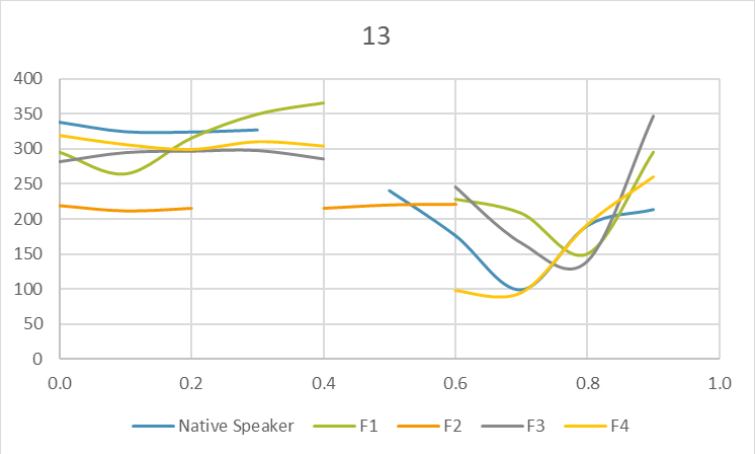


Figure 28: Tone 1 followed by tone 3 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

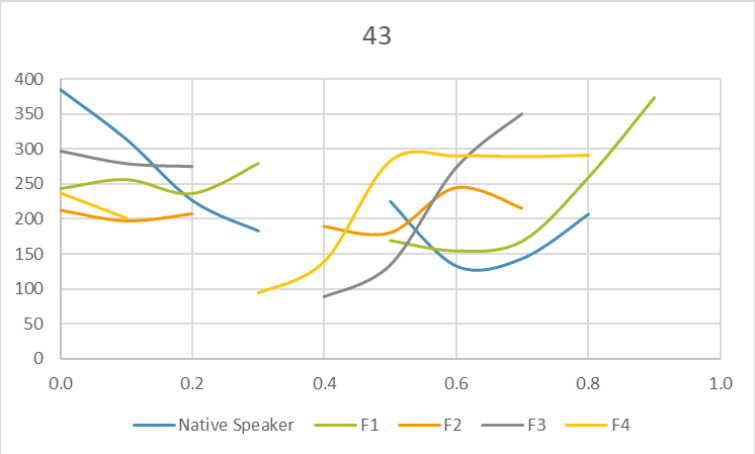


Figure 29: Tone 4 followed by tone 3 measured by frequency (Hz) over time (s) as produced by four Dutch female students at a beginner level of Mandarin, compared to the native speakers' production.

The females in the first year showed more accurate tonal production than the males. They produced more tones correctly, although they were not always consistent. Especially in the series of disyllabic words starting with tone 1 they produced the tone 2 and tone 3 well. Figure 27 shows they all got a correct rising contour in their tone 2, and most of their pitch heights are correct as well. The production of tone 1 went less well for female 1, who did not get the right contour, and the other females are not quite at the same pitch height as the native speaker. Figures 28 and 29 show the inconsistency mentioned earlier. Female 1 and female 3 produced a correct tone 3 when preceded by tone 2, but none of the females produces a correct tone 3 when preceded by tone 4. From Figure 29 it is also clear they had difficulties with tone 4, where they got neither the pitch height nor the pitch contour correct. The other results also portray this difficulty they have in producing tone 4, as only female 3 manages to produce it almost right, and only in one of the instances (tone 4 followed by tone 1).

Overall tone 1 is often produced correctly when it comes first, more pitch contour and height mistakes were made in its production when it occurred on the second syllable.

4.7.2 Fourth year students' production of disyllabic words

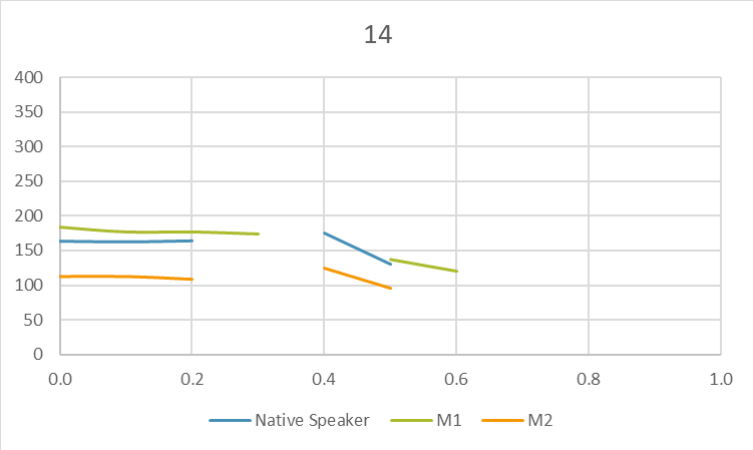


Figure 30: Tone 1 followed by tone 4 measured by frequency (Hz) over time (s) as produced by 2 Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

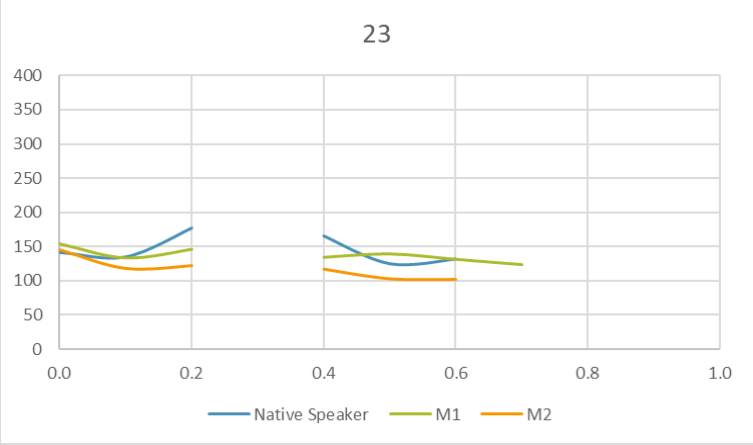


Figure 31: Tone 2 followed by tone 3 measured by frequency (Hz) over time (s) as produced by 2 Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

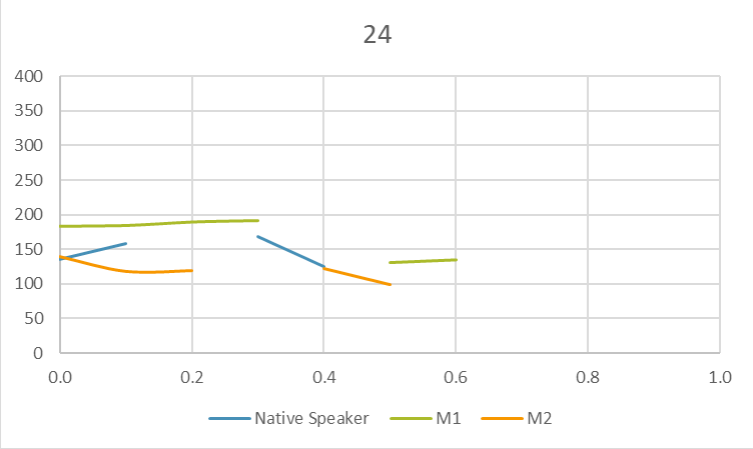


Figure 32: Tone 2 followed by tone 4 measured by frequency (Hz) over time (s) as produced by 2 Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

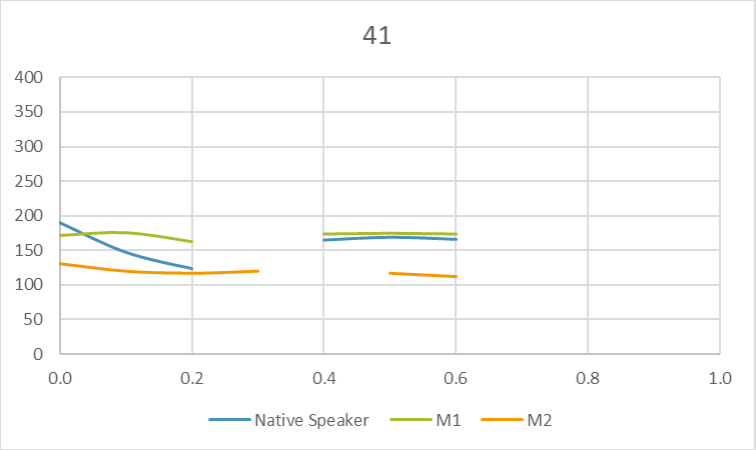


Figure 33: Tone 4 followed by tone 1 measured by frequency (Hz) over time (s) as produced by 2 Dutch male students at an intermediate level of Mandarin, compared to the native speakers' production.

The Figures above show a selection of the results of the males in their fourth year. Figure 30 is the only instance where both tones were produced correctly by both males, although male 1's tone 4 should be slightly steeper in contour. Male 2's pitch height is lower than the native speakers', but he is very consistent in this pitch height, so it might be due to his personal vocal range.

Figures 31 and 32 show the difficulties they have with tone 2. Tone 2 was not produced right in any of the results produced by them, and Figure 31 shows the one instance where they got closest to producing the right pitch contour. Here, still, they would need a higher and steeper rise to produce it correctly. Their tones are generally around the same pitch height as the native speakers.

Both males were not consistent in their production of tone 4, as shown in Figures 30, 32 and 33. The mistakes are mainly made in pitch contour; their pitch height is similar to the native speakers'. In Figure 30 both males do have a falling contour, but it is not quite steep enough. In most other instances their tone 4 pitch contour looks more similar to a tone 1.

Tone 3 was produced with a wrong pitch contour most of the time, one of which instances is shown in Figure 31. Similar to what often happened for tone 4, the pitch contour is often relatively flat, instead of low falling. Once again, they produce a pitch height similar to the native speakers.

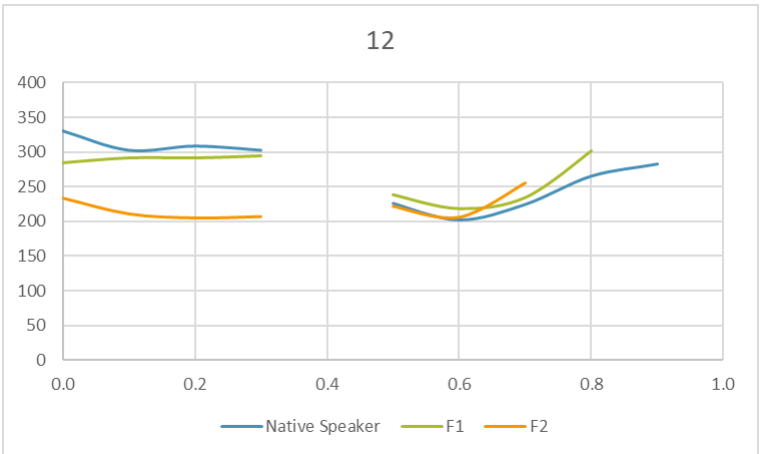


Figure 34: Tone 1 followed by tone 2 measured by frequency (Hz) over time (s) as produced by 2 Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

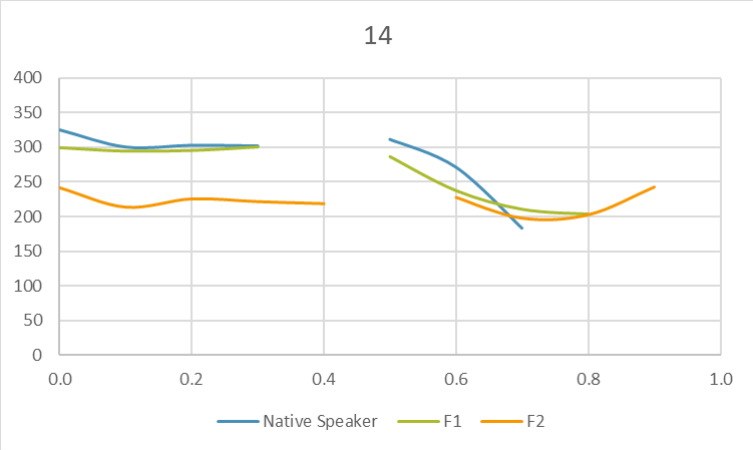


Figure 35: Tone 1 followed by tone 4 measured by frequency (Hz) over time (s) as produced by 2 Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

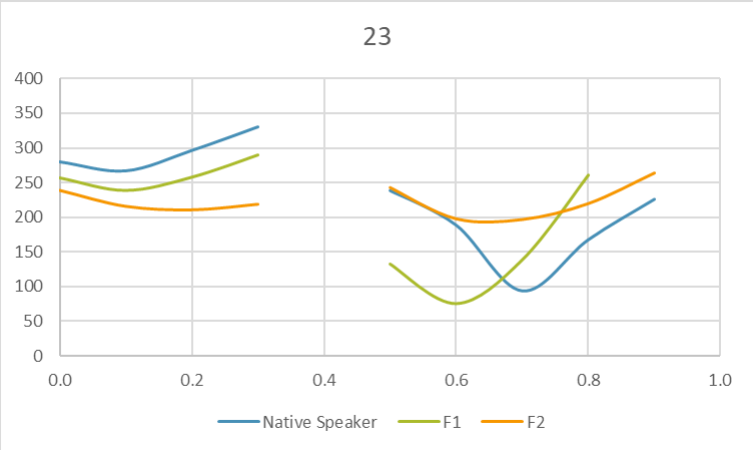


Figure 36: Tone 2 followed by tone 3 measured by frequency (Hz) over time (s) as produced by 2 Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

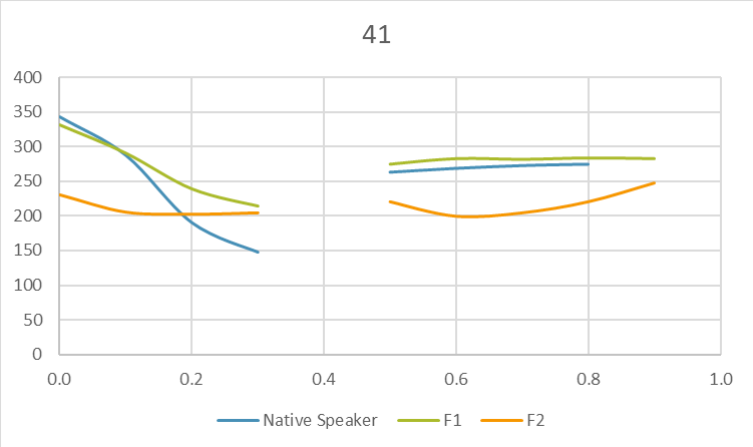


Figure 37: Tone 4 followed by tone 1 measured by frequency (Hz) over time (s) as produced by 2 Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

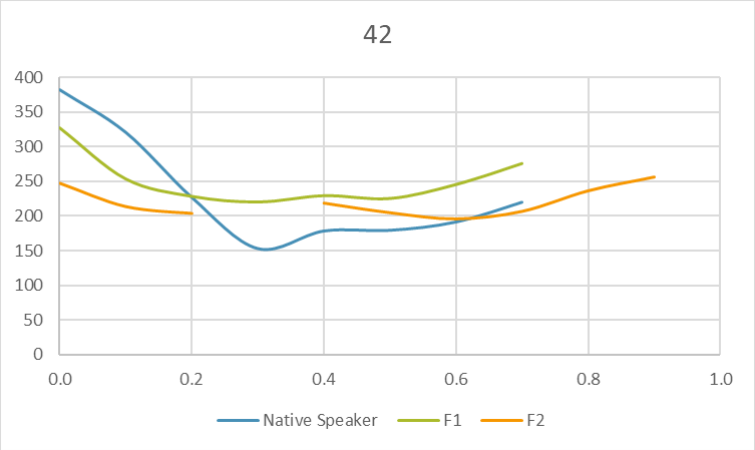


Figure 38: Tone 4 followed by tone 2 measured by frequency (Hz) over time (s) as produced by 2 Dutch female students at an intermediate level of Mandarin, compared to the native speakers' production.

Similar to the first year students, the females in their fourth year produced tones more accurately than the males. This is especially visible in their productions of both tone 1 and tone 2. Female 2's pitch height is consistently too low for tone 1, but her pitch contour is correct, as demonstrated in Figures 34, 35, and 37. These same Figures show that female 1 produces both the pitch height and pitch contour correctly. In Figure 37 it is especially noticeable that she adjusted the pitch height of her tone 1 the same way native speakers did. Her standard pitch height for tone 1 seems to be between 300 and 350 Hz, yet here she lowered it slightly to around 275 Hz, thus showing coarticulation. In Figure 37 her starting pitch height of tone 4 is also lower than in the other instances, which also shows coarticulation similar to the native speakers. In Figure 38, however, the other instance where the native speakers had coarticulation of the tones, she has not adjusted her tone 2 pitch height the way the native speakers have. The pitch contour and height of her tone 2 in general are produced well. If she were to let tone 4 drop lower she might achieve the same results as the native speakers, since it would be more natural to start tone 2 off from tone 4's endpoint. Female 2 shows no evidence of coarticulation yet, and as compared to female 1 her tone 2 pitch contour could still show improvement by rising a bit higher

Figure 36 shows the contours of tone 3 the two females produce, which they seem to do quite consistently at the same pitch height with the same pitch contour. The pitch contour of both is not quite perfect yet, but they do show the falling and rising contour. Female 1 should start off with a slightly higher pitch contour and keep the falling contour up longer, and then rise less after the dip. Female 2 should lower her pitch farther into the dipping point; at this point her contour is still too flat.

Instances of tone 4 are shown in Figures 35, 37, and 38. It would seem that female 2 specifically has difficulties with this tone, although female 1's pitch contour is also not yet perfect. Female 2 does show a falling contour, but the contour has to be significantly steeper to match the native speakers' contour. In addition, her pitch height should start off higher. Female 1 has the same starting pitch height as the native speakers', but her pitch contour should be slightly steeper still.

5.0 Discussion

Relating the results back to general improvement of tonal production in Mandarin, the results will now be discussed per gender, while comparing the differences between the two grades. I will first discuss the results of the female participants, and then the results of the male participants. In both sections the first conclusions will be based on the monosyllabic word test results per tone category, to then look at whether the disyllabic word test results support these conclusions, and whether it revealed any new information as well. After these results I will discuss some differences between the male and female results.

5.1 Improvement by females

The participants of the first year show that tone 1 was not too difficult even for beginner learners. All participants have an overall correct pitch contour. The pitch height, however, is too low for two out of the four participants.

In year 3 the results for tone 1 are similar. The contour of one of the two participants is slightly off, and her pitch is also too low. The other participant has a correct pitch contour and pitch height.

There seems to be no specific improvement for tone 1 considering these results. This is mainly because the pitch contour of the first year students was already correct and thus could not necessarily be improved. Pitch height had different results in both years, thus for tone 1 specifically no conclusions can be drawn.

Tone 2 appeared to be harder for the first year students to produce. All their pitch contours do show the rise typical of tone 2, but most lower their pitch too far before the rise, and rise too abruptly. The initial pitch height of all but one participant, however, is similar to the native speakers' and can be considered correct.

The two fourth year students have a pitch height similar to the native speakers', and their pitch contours also show a shallow dip before a higher rise. Their rise is closer to the native speakers' than the first year students', but they could still improve by initiating the rise slightly faster.

Overall pitch height was mostly correct in both years, but the fourth year showed a clear improvement in pitch contour as opposed to the first year students.

None of the pitch contours of tone 3 produced by the first year students is correct. Two of them do not dip low enough, one lets her pitch fall too fast, and the other shows no dip at all. For most the initial pitch height is similar but slightly lower than the native speakers', which can still be considered an appropriate height, but one of the participants starts too low.

The two females in year 3 both start with a correct pitch height. Their contours are still off, but do seem to show improvement. Both pitch contours show a fall and rise; one, however, does not fall low enough to form a proper third tone contour, the other falls too quickly and rises too high back.

Tone 3 is clearly a difficult tone to master for the female participants. Both the initial pitch height and the pitch contour do show a slight improvement, but even in the fourth year the tone is not produced correctly quite yet.

Both pitch height and pitch contour of tone 4 are produced incorrectly by all four participants in the first year. Their initial pitch height is considerably lower than that of native speakers, and their fall in pitch is neither low nor fast enough.

In year 3 one of the participants has a pitch height that is much closer to the native speakers as compared to the first year students, the other participant has a pitch height that is just as low as the ones in the first year. The contour of the participant with better pitch height also shows improvement: it has a deeper and steeper fall in pitch than the first year students' contours showed. She could still improve, however, by falling even lower in pitch faster. The other participant's contour rises instead of falls.

Both pitch height and contour were better in the fourth year for one participant, showing that it can be improved upon. The other participant, however, shows no improvement at all.

The results of the disyllabic words seem to support the conclusions drawn above. As Lu (1992) also discovered during her study, the students did clearly struggle more in producing the disyllabic words. Especially the first year students needed more time before producing the second syllable.

Tone 1 was produced quite well by the first year students when on the first syllable, but significantly less well when on the second. In the fourth year both females seemed to have no trouble producing a correct tone 1 in both positions, showing improvement.

Tone 2 was produced right by the first year students when preceded by tone 1, yet in other instances the pitch contour was often off. The females in the fourth year show improvement in overall contour and consistency, as their tone 2 contour mostly stays similar throughout all tonal combinations.

For tone 3 there were also varying results in the first year, where most got it right at least once, but made mistakes at the other instances. The fourth year students are once again more consistent in their production, and their contours are better overall. Yet, both can still improve on their contour.

Besides tone 3, tone 4 also seems difficult to master. In the first year tone 4 was produced wrong most of the time, with mistakes in both pitch contour and pitch height. The fourth year students do show improvement, but they have not mastered the tone yet. Both have a falling pitch contour, but both should produce a steeper contour to improve more.

As overall improvement there is the consistency mentioned for tone 2 and 3 specifically, while this could also be mentioned for the other two tones. In the first year the females' pitch contours differed in different instances, whereas the pitch contours of the females in the fourth year were stable throughout. In addition, one of the females seems to show tonal coarticulation similar to that of the native speakers', which also shows better mastery of the tones and language as a whole.

Studies discussed in the literature review showed that tone 2 and tone 3 were considered most difficult to both perceive and produce for L2 learners of Mandarin (Chen, 1993; Hao, 2012; Yang, 2015; Chun, Jiang, Meyr, & Yang, 2015; Lu, 1992). Zou (2017), on the other hand, showed that for Dutch students specifically tone 4 was most difficult to produce. Considering the results above,

although based on a limited number of participants, tone 4 is difficult to produce for female Dutch students, which is similar to Zou's results. The students show, however, that tone 3 is the most difficult tone to produce for them, which is in accordance with the literature based on students with English as their L1. Although the production of tone 2 by the female students was not consistently accurate, the improvement shown is better than that of tone 4, suggesting that tone 4 is more difficult to produce than tone 2 for Dutch female students. Except for tone 1 being the easiest to master, the results of this study neither contradict nor verify results of the established literature. More research on Dutch L2 learners of Mandarin could provide a deeper insight into their SLA of Mandarin tones.

5.2 Improvement by males

The pitch contour of tone 1 is produced well overall by the males in their first year. One male starts off well, but falls in pitch height too much at the end. The pitch height of all males is too high.

In the fourth year both participants have a proper pitch contour. Their pitch is also lower than the first year students', bringing them at about the same level as the native speaker. One male is rather far below the native speakers in pitch height.

Pitch height seems to have improved considerably, possibly due to the vocal changes boys undergo in their teenage years, making their pitch height lower. Pitch contour was produced well overall in both the first and fourth year.

The pitch contour of tone 2 has been produced well by only one of the first year males. The pitch height of all three males is too high.

The pitch height of the fourth year students is at the same height as the native speakers. One of the participants also produces an adequate pitch contour.

As with tone 1, the pitch height has improved considerably. To state pitch contour has improved, more data is needed. In both years one participant had an adequate pitch contour, but since the fourth year only had two participants, it is hard to conclude whether contour has improved as well.

Neither pitch contour nor pitch height was produced well for tone 3 by the first year males. Their contour does show a fall and rise, but it does not quite match the native speakers.

Production by the fourth year students is done slightly better considering pitch height, but the pitch contour is still not correct. One of the participants did produce a falling and rising pitch contour, but his fall in pitch was too fast and too low. The other participant produced a level tone.

As with the females, it would seem that tone 3 is a difficult tone to produce. There were slight improvements in pitch height, but pitch contour does not seem have been improved.

Tone 4 was produced incorrectly by all three first year participants. Their pitch height was too high, and none of their contours match the native speakers.

The pitch height of the fourth year students is too low instead of too high, and neither of the two have a correct pitch contour for tone 4.

The fourth year students do not show any considerable improvement. There is a difference in pitch height between the two years, but neither have the pitch height correct. The pitch contour is also incorrectly produced in both years.

The disyllabic word test shows similar results as the monosyllabic words test. Where the pitch height has improved for all four tones, the pitch contour does not show much improvement. The fourth year males seem to show more consistency in the correct production of tone 1 regarding both pitch height and pitch contour, but the other three tones are still produced inconsistently with differing contours.

Wang et. al (2006) stated that pitch height was more difficult to master than pitch contour. The results of these Dutch male students, however, suggest that pitch contour is more difficult. Little improvement was shown in pitch contour, whereas pitch height was improved upon considerably. As mentioned, this might be due to the vocal changes teenage males undergo, while the test subjects of Wang et. al were in their twenties. The females of this study, however, also seemed to have more struggles with pitch contour as opposed to pitch height.

Based on the results of this limited number of participants it seems that the most difficult tone for young male Dutch students is tone 3, followed by tone 4 and then tone 2, similarly to the female students. As discussed above, this is partly in agreement with the established literature, and partly a new outcome.

5.3 Males compared to females

The results between both genders seem to differ to such an extent that it is worth discussing them here. Where the male students only seem to have improved in pitch height, the female students showed improvement mainly in pitch contour; pitch height was often correctly produced by the first year female students already. The fact that pitch height was mostly produced correctly by first year females but not first year males is likely to be due to two causes. The first, as mentioned, is a males vocal range, which changes during puberty. The younger participants have likely not had their voice transition yet, whereas the older ones already naturally have a lower voice, thus letting them produce the correct pitch height as compared to male native speakers. Secondly, their Mandarin teacher is (a Dutch) female. As was shown from the native speaker results, most males produce the tones at a lower frequency than females, yet the initial example the participants have is the pitch height of a female speaker. This might also be an explanation for the higher pitch height.

When looking at the fourth year results the differences are particularly striking. Where the female participants produce almost everything (close to) correct and show consistency, the male participants rather show inconsistency and only get tone 1 correct. This might show that females master the tones faster than the males, but it could also have to do with the set-up of this experiment. As mentioned in the methodology, the audio had to be recorded by the students themselves, on their own. Since they are still in their teenage years, their motivation to record earnestly for this experiment might not have been high. The fourth year males specifically needed several reminders before they responded with their audio. Their motivation to study Chinese in the first place, however, is not much different from the females. Male 2 and female 2 mention they

chose Chinese to not have to take German or French; female 1 says she picked it for fun initially, but changed her attitude to take it more seriously soon after; male 1 says he chose Chinese because it would look good on his résumé. These reasons do not seem to provide an explanation for mentioned differences. Further research on male and female differences in the acquisition of Mandarin tones could provide us with more results worth examining to be able to adapt teaching methods to their needs.

5.4 Summary

Males and females seem to show different levels of improvement over the years. Pitch height in general seems to show improvement for the females for tone 3 and tone 4, whereas the males show most improvement in pitch height for tones 1 and 2. Pitch contour of the females is clearly improved upon for tones 2, 3, and 4. The males showed no specific improvement in pitch contour overall.

For both genders, it would seem that tone 3 is most difficult to learn, with limited improvement shown in both pitch height and pitch contour. Tone 4 was produced better by the female fourth year students, but they still showed the same errors in pitch contour throughout the results, suggesting tone 4 is also a tone that is difficult to master for Dutch secondary school students.

6.0 Conclusion

The main goal of the current study was to determine in what ways Dutch L2 learners of Mandarin improve in tonal production over a course of 3 years. It was discovered that pitch contour was mainly improved upon by the females, whereas their pitch height was quite accurate in the first year already, thus not being able to show improvement on that aspect. The males did not show any specific improvement regarding pitch contour. Their pitch height was too high in the first year, but was improved upon by the fourth year students by lowering it. This could be due to vocal changes throughout the years, or due to the fact that they had a female teacher with a higher pitch.

In addition, it would seem that both tone 3 and tone 4 are most difficult to learn for teenage Dutch learners of Mandarin. They seem to struggle the most with the pitch contour specifically, although the starting pitch height of tone 4 also seems to be a difficulty.

I am aware of the limitations of his study, caused, in particular by the low number of participants. This being an MA Thesis, which must be completed within a limited period of time, I had already planned the experiments with fewer participants than would be acceptable for a professional researcher. As mentioned in the methodology section, the corona outbreak made it impossible to visit the school as planned (as schools were closed) as a result of which I had to contact the students online and rely on their willingness to take part on their own accord. Not all students who had originally agreed to participate responded to my mails, thus lowering the number of participants even further. For those who did respond and who made their recordings at home, I could not check whether they followed all instructions strictly, which could have influenced some of the results. More importantly, however, I felt that the students' motivation to participate in this experiment was lessened by these circumstances. This is based on the fact that most took several weeks and reminders before replying with their recordings. Had the tests been in person, on the other hand, it would have been easier to motivate them, and the fact that someone was monitoring them might have increased their motivation as well. Less motivation to participate could have caused them to take it less seriously, which I suspect might have happened with the third year male students specifically.

In addition to their motivation, in retrospect it would seem that the words tested could have been more simple. In the recordings it is noticeable that even some of the third year students still struggle with the pronunciation of pinyin, thus giving them easier words to pronounce might make it easier for them to produce the tone correctly.

Lastly the age of the control group can still be questioned. The participants of the control group were in their twenties, whereas the students were aged 13-17. Apart from the vocal changes that males undergo in their teenage years, however, the results seem to have been fairly consistent.

Despite these limitations and reservations we must have with respect to how representative the data in this thesis are, I think this thesis has provided a deeper insight into Dutch male and female high school students learning Mandarin, and any differences between them. As discussed, there is very little literature on Dutch L2 learners of Mandarin, and this study helps provide a more thorough understanding of the difficulties these learners encounter. In addition, it shows the level of effectiveness of Mandarin classes in Dutch secondary schools. Some studies have been done in the first few years that Mandarin was taught at secondary schools, but these were all regarding the level

they wanted the students to achieve, and whether the envisioned level was achievable (i.e. Folmer, Tigelaar, & Sluijsmans, 2013). None of these look at the students individually, however, and they do not test any specifics such as grammar, tonal production, etc. This study therefore contributes to our overall view and understanding of Dutch Mandarin classes in secondary schools.

A question raised by this study is the reason for the discrepancy shown in improvement between males and females. Some potential explanations were offered for these differences, but it would be interesting do further research on these differences. In addition, due to the limited number of participants in this study, further research on general improvement on tonal production by Dutch secondary school students would still be valuable in this field. More results on tonal production can provide us with deeper insight on specific difficulties Dutch students of Mandarin face, and help improve Mandarin education in the Netherlands.

References

Boersma & Weenink (2020)	Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 14 Jan 2020 from http://www.praat.org/
Chan & Leung (2019)	Chan, R. K. W., & Leung, J. H. C. (2019). Why are lexical tones difficult to learn. *Studies in Second Language Acquisition*, *42*, 33–59. doi: 10.1017/S0272263119000482
Chen (1993)	Chen, Q. (1993). Toward a sequential approach for tonal error analysis. *Deseret Language and Linguistic Society Symposium*, *19*(1), 68–77. Retrieved May 6, 2020 from https://www.semanticscholar.org/paper/Toward-a-Sequential-Approach-for-Tonal-Error-Chen/97c94783375eaa210a22255b246553225de1f2d1
Chun, Jiang, Meyr, & Yang (2015)	Chun, D. M., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of second language pronunciation*, *1*(1), 86–114.
Folmer, Tigelaar, & Sluijsmans (2013)	Folmer, E., Tigelaar, D., & Sluijsmans, L. (2013). *Curriculumevaluatie. Chinese taal en cultuur. Chinese taal en cultuur eindrapportage*. Enschede: SLO.
Gandour (1984)	Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese linguistics*, *12*(2), 235-261.
Hao (2012)	Hao, Y. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of phonetics*, *40*(2), 269-279.
Jongman, Yue, Moore, & Sereno (2006)	Jongman, A., Yue, W., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In Li, P., Tan, L., Bates, E., & Tzeng, O. (Eds.), *The handbook of East Asian psycholinguistics* (Vol. 1, pp. 209–217). doi: 10.1017/CBO9780511550751.020
Jongman, Qin, Zhang, & Sereno (2017)	Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The journal of the acoustical society of America*, *142*(2), 163–169. doi: 10.1121/1.4995526

Li (2013)	Li, J. (2013). Chinese as a foreign language (CFL) students' lexical tonal development: An investigation of tonal production and awareness of tonal categories. *Journal of the national council of less commonly taught languages, 13*, 129-158.
Li & Thompson (1981)	Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference for grammar.* University of California Press
Liao (1994)	Liao, R. (1994). *Pitch contour formation in Mandarin Chinese: A study of tone and intonation.* (Doctoral dissertation). Ohio State University.
Lu (1992)	Lu, J.-M. (1992). *On the perception and production of Mandarin tones by adult English-speaking learners.* (Doctoral dissertation). Michigan State University
Moore & Jongman (1997)	Moore, C., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The journal of the acoustical society of America, 102*(3), 1864-1877.
Wang, Jongman, & Sereno (2003)	Wang, Y., Jongman, & Sereno. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The journal of the acoustical society of America, 113*(2), 1033-1043.
Yang (2015)	Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer-Verlag Berlin Heidelberg
Zou (2017)	Zou, T. (2017). *Production and perception of tones by Dutch learners of Mandarin*. (Doctoral dissertation). Leiden University

Appendix A

Survey questions

Je bent

1. Man
2. Vrouw
3. Anders: _____

2. Hoe oud ben je?
    1. _____
3. Wat is je moedertaal?
    1. _____
4. Spreek je meerdere talen vloeiend? Zo ja, vul in:
    1. _____
    2. _____
    3. _____
5. Hoe lang heb je Chinees gehad (in jaren of halve jaren)?
6. Heb je al eens Chinees geleerd voordat je aan dit vak begon? Zo ja, voor hoe lang?
    1. _____
7. Hoeveel tijd besteed je ongeveer aan Chinees buiten de les en huiswerk om? (Denk aan series kijken, muziek luisteren, etc.)
    1. _____
    2. Waaraan:
        1. _____
        2. _____

3. _____
4. _____

8. Waarom heb je dit vak gekozen? (Voor de lol, als serieuze overweging voor de toekomst, etc.)

Appendix B

Word list

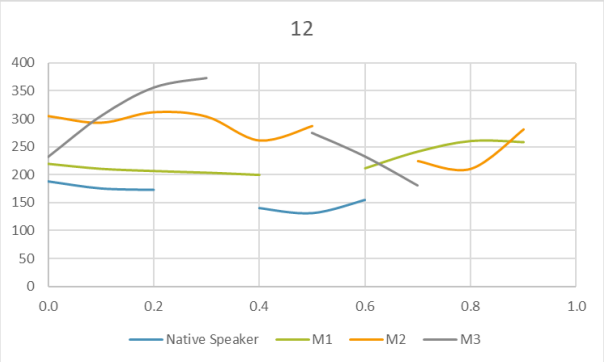
The words in bold are the words used for the results.

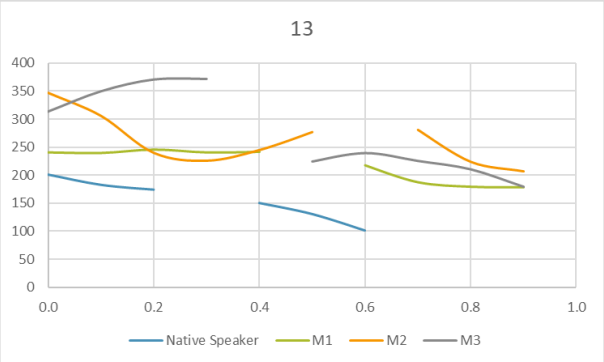
1. **Búdàn**
2. Kāifàng
3. Bào
4. **Dājiù**
5. **Mǎ**
6. **Huāshí**
7. Shuōmíng
8. Díquè
9. **Má**
10. **Càidān**
11. **Shúdiǎn**
12. Yǔyán
13. **Fàláng**
14. **Mà**
15. **Mā**
16. Shīqù
17. **Xiāngshuǐ**
18. Gōnglǐ
19. **Zàofǎn**
20. **Fángjiān**
21. Chā

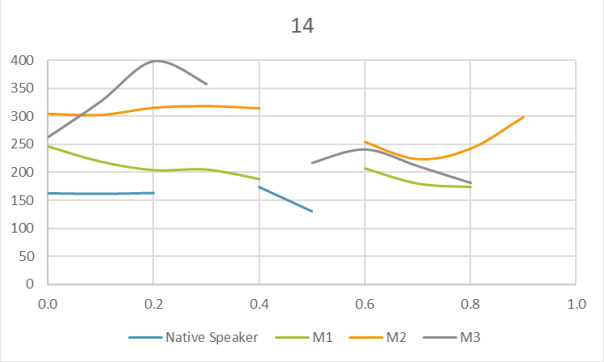
Appendix C

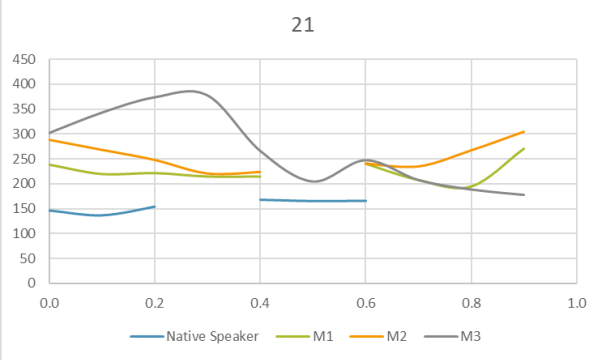
Results of disyllabic words by students

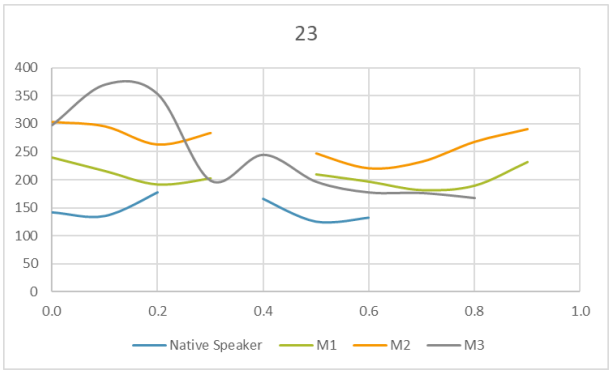
First year males results:

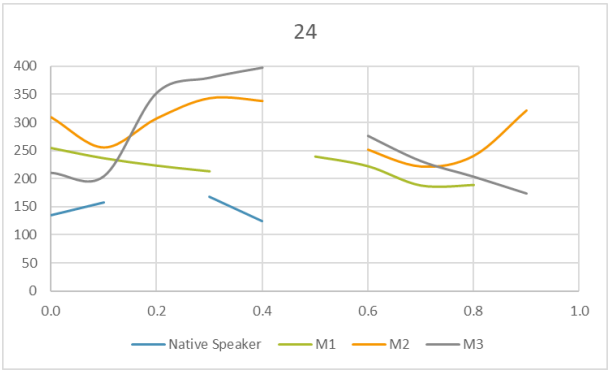


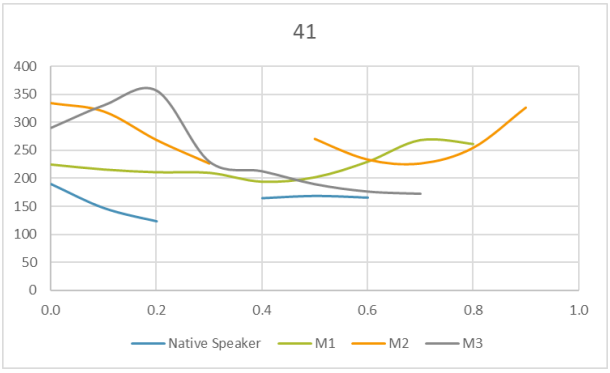




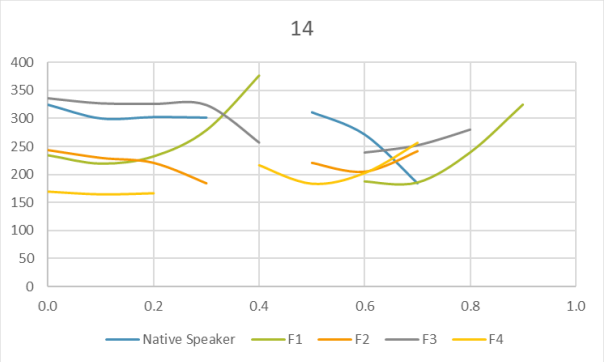


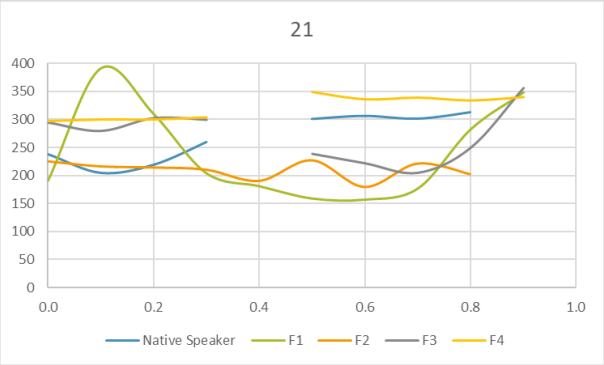

23
400
350
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Native Speaker
M1
M2
M3

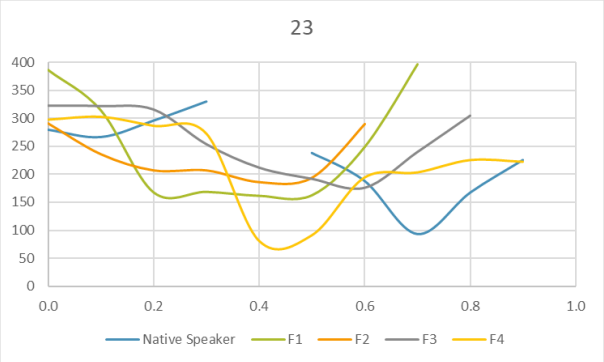

24
400
350
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Native Speaker
M1
M2
M3

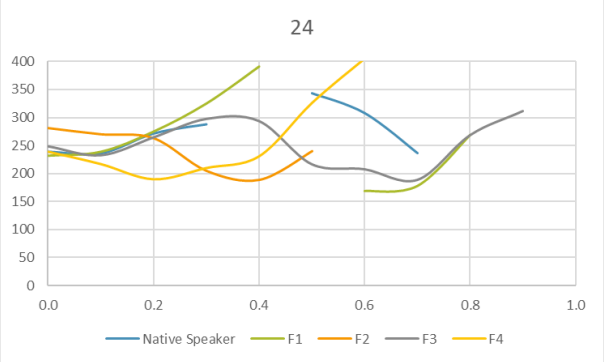

41
400
350
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
Native Speaker
M1
M2
M3

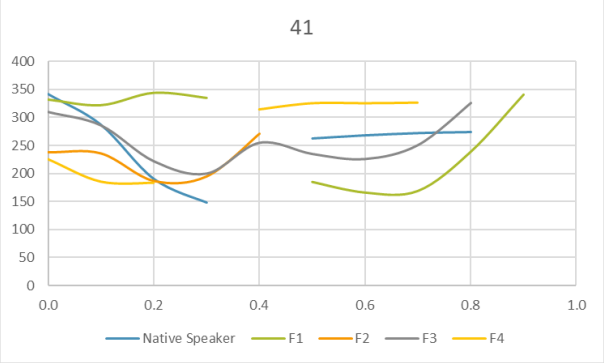
First year females results:

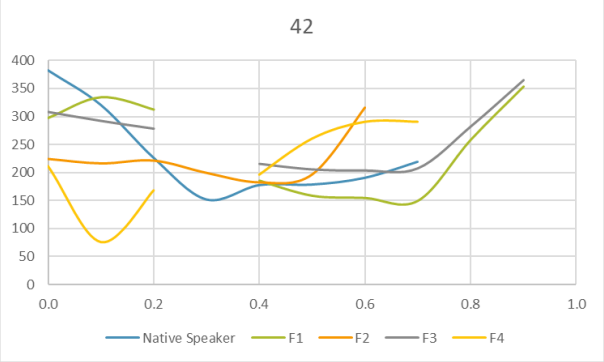


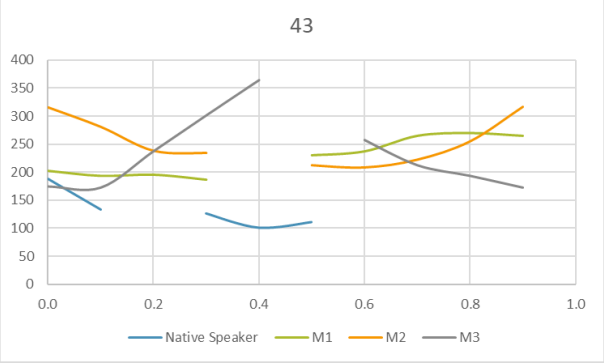




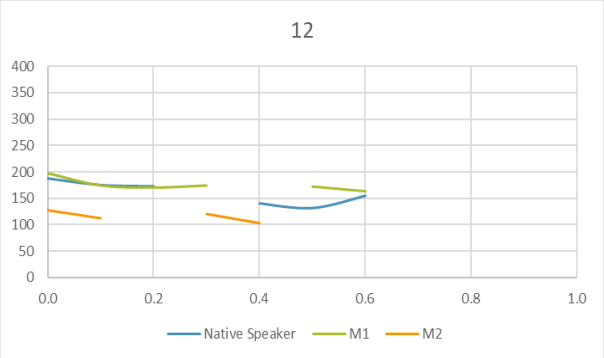


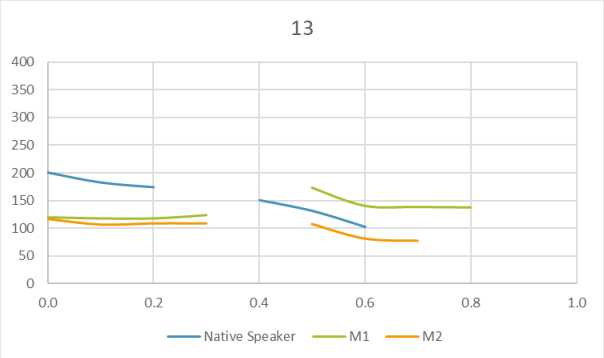


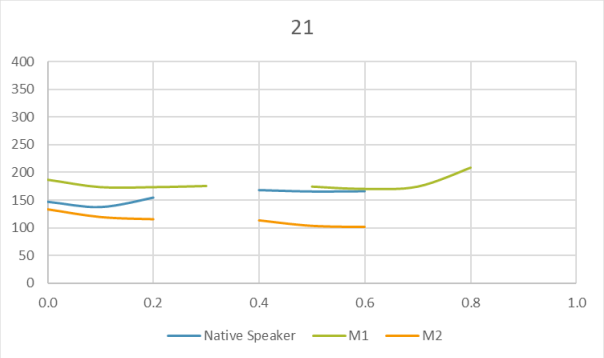


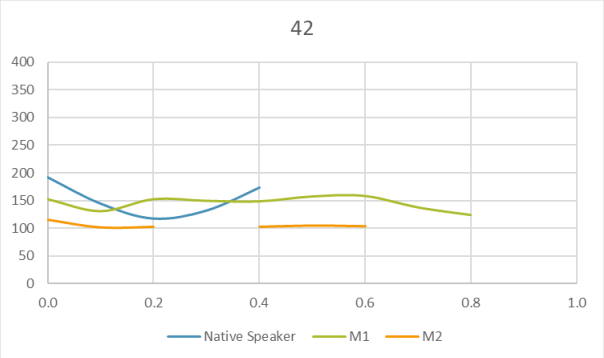


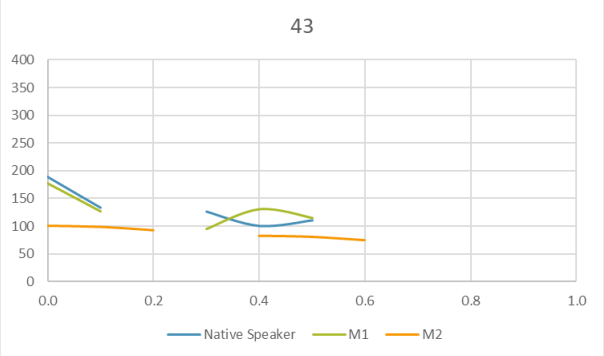
Third year males results:











Third year females results: