The prosodic realization of Y/N-questions in Mandarin Chinese

Mira Grubic
Matrikelnummer 719113
grubic@ling.uni-potsdam.de
Magisterarbeit im Bereich Computerlinguistik
Universität Potsdam, Institut für Linguistik
28.07.2008
1 Introduction

Most strategies employed in natural languages to syntactically mark Y/N questions do so either sentence-initially (e.g. sentence-initial particles, subject-verb-inversion) or medially (e.g. with special verb morphology) (Sadock and Zwicky, 1985, p. 181). In Mandarin Chinese, however, a Y/N question is formed by adding a sentence-final particle to the corresponding statement. This kind of question formation is interesting, because, to be able to respond correctly and with minimal delay, which in turn is important for adequate turn-taking management (Sacks et al., 1974, p. 719), the hearer must perceive whether it is a question or a statement before the particle is actually uttered. There must be some kind of non-syntactical cue which marks these kinds of utterances as questions. This is further supported by the fact that in colloquial Mandarin, one can also hear a second kind of Y/N question which is syntactically identical to the corresponding statement, without any syntactical question cue.

The literature about the prosodic realization of statements and questions in Mandarin Chinese suggest a) that a prosodic difference does exist, both between questions and statements and between the two kinds of questions, b) that fundamental frequency is the most important cue to recognize this difference, and c) that the difference becomes more noticeable towards the end of the sentence. There is disagreement on the nature of prosodic differences which are necessary to mark questions, as well as on where this difference becomes salient. The first part of this thesis paper is devoted to introducing these theories. The second part describes my own attempt to answer this question. In an analysis based on Yuan (2004), 441 sentences by three speakers are studied. The sentences were selected to investigate the effect of the three sentence types, sentence length, focus, and tones of the final two syllables. The aim of this section is to isolate the prosodic characteristics of these two kinds of questions that mark them as questions. It is further investigated whether these characteristics are the same for both the particle question and the syntactically unmarked question, and whether the prosodic marking of the syntactically unmarked question is stronger due to the fact that this is the only differentiating feature. The resulting characteristics are tested in a perception experiment, the main aim of which is to investigate at what point in the sentence the difference becomes audible. The long sentences of one speaker were prepared using a ‘gating’ procedure (Grosjean, 1983): final parts of different length of the last word were truncated, so that only the prosodic information of the sentence preceding it was available to the hearers. Twelve subjects were presented with 252 stimuli each, and were asked to choose whether what they heard belonged to a statement, an unmarked question, or a particle question. This perception experiment and its results are described in the third part of this thesis paper.
The most basic way of asking for the truth of a proposition cited by Mandarin Chinese grammar textbooks is to use the *ma*-particle question. As in all kinds of questions in Mandarin, the surface word order is the same as in the corresponding statements (Yip and Rimmington, 1997, p. 347). It is constructed by adding a *ma* particle to the proposition in question (for example Yip and Rimmington, 1997, p. 347).

\begin{align*}
\text{ta1 mai3 dian3 xiang1jiao1. (Statement)} \\
\text{She buy a-bit banana.} \\
\text{‘She’s buying/bought/will buy some bananas’}
\end{align*}

\begin{align*}
\text{ta1 mai3 dian3 xiang1jiao1 ma0? (marked Y/N question)} \\
\text{she buy a-bit banana q-particle?} \\
\text{‘Is she buying/Has she bought/Will she buy some bananas?’}
\end{align*}

It is to be expected that due to the fact that the *ma* particle is final, some additional marking may take place to ensure that the hearer realizes it is a question. This is further supported by the fact that in informal speech, a second kind of Y/N question can be found which is syntactically identical to the corresponding statement sentence. This kind of question usually goes unmentioned in Mandarin grammar books, with the exception of Li and Thompson (1981, p. 520). Li and Thompson do not treat it as a separate category, nor do they elaborate on its frequency or function.

\begin{align*}
\text{ta1 mai3 dian3 xiang1jiao1? (unmarked Y/N Question)} \\
\text{she buy a-bit banana?} \\
\text{‘Is she buying/Has she bought/Will she buy some bananas?’}
\end{align*}

Such declarative questions, utterances that have the syntactic form of a declarative statement, but are used as questions, seem to exist in most, if not all, languages (Ultan, 1978, p. 219). According to Ultan, they are distinguished from declarative statements by intonation. Li and Thompson note that the Mandarin Chinese syntactically unmarked Y/N questions are produced with an intonation different from that used for statements, and that it is easy to recognize these utterances as questions ‘in context’ (Li and Thompson, 1981, p. 520). For different languages, it has been noted that questions do not necessarily make use of ‘question intonation’. Not only is it not consistent across languages (Sadock and Zwicky, 1985), (Ultan, 1978), but it is also not consistent within a language (Cruttenden, 1986, p. 59, 96ff), (Geluykens, 1988, p. 468). There were claims that any contour found for questions can also be found for statements and vice-versa (Bolinger, 1978, p. 503), (Pike, 1945, p. 163). Li and Thompson’s statement that these questions are easy to recognize in context is therefore

\footnote{The examples in this study are mine unless stated otherwise.}
\footnote{In the romanized transcription of the examples, every syllable is followed by its accompanying tone, for example ta1 would be the syllable ta pronounced with first tone.}
not without relevance for the current discussion. Does this mean that they are difficult to recognize without context? There were claims that context and lexical items may be more important cues for question recognition than previously realized, while the significance of question intonation as a cue is overrated (Beun, 1990). Therefore the first research question to be asked is:

- Do *ma*-particle questions and syntactically unmarked questions differ prosodically from statements?

As noted above, there are various kinds of ‘question intonations’ in one language. Since the *ma*-particle question and the syntactically unmarked question are also syntactically indistinguishable before the start of the particle, it is to be expected that they also differ from each other. I assume, with (Weber, 2003, p. 16), that different syntactical forms are used for a reason, i.e. that the functions of the two kinds of Y/N questions are not the same. I also assume that the speaker would want to communicate this difference in meaning to the hearer early enough so that the hearer can incorporate it in her answer. Therefore the following questions would also be of importance:

- Do they differ prosodically from each other?
- If there are differences between the three sentence types, at what point in the sentence are they recognizable?

For Dutch declarative questions, Haan proposed a *Functional Hypothesis* which states that prosodic marking of questions is most salient in questions which are syntactically unmarked, and becomes weaker the more marked the question is (Haan, 2001, p. 56). In Mandarin Chinese, there is a third kind of Y/N question which is typically called A-not-A question.

```
ta1 mai3-bu-mai3 xiang1jiao1? (A-not-A question)
she buy-not-buy banana? ‘Is she buying bananas?’
```

Considering that this kind of Y/N question, which is syntactically marked as question early in the sentence, is not marked by the final rise which is typically seen as ‘question intonation’ (Shen, 1990, p. 20), a further relevant research question is:

- Is Haan’s Functional Hypothesis correct for Mandarin Chinese Y/N questions?

In the following, previous research on this topic will be presented. Consider that in what follows, the terms ‘syntactically unmarked question’ and ‘syntactically marked question’ may be abbreviated to ‘unmarked question’ and ‘marked question’, respectively. These names refer only to the *syntactical* marking of the questions.
2.2 Question Intonation in Mandarin Chinese

2.2.1 Is there a distinct question intonation?

Mandarin Chinese is what Yip (2002, p. 2) calls a lexical tone language: Tone distinguishes the meaning of lexical items. Tone is associated with the syllable (Xu, 2001, p. 6), (Xu, 1999, p. 97).\(^3\) It is realized mainly through pitch (Hartmann, 2006, p. 163), (Yip, 2002, p. 5), (Howie, 1976, p. 145), but there is evidence from research on whispered tone that duration and intensity might play a role as well (Kratochvil, 1968, p. 37), (Kong and Zeng, 2006, p. 7), (Whalen and Xu, 1992, p. 31), (Liu and Samuel, 2004, p. 132).\(^4\) A very generalized characterization of the citation forms of the four tones, i.e. the realization of the tones when they occur in isolation, can be seen in Table 1. Additionally, there is a so-called neutral tone, whose pitch height depends on the preceding tone. This is not a tone category per se, but a feature of weak syllables.

<table>
<thead>
<tr>
<th>Tone</th>
<th>relative height</th>
<th>relative duration</th>
<th>relative intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>level</td>
<td>middle</td>
<td>remains the same</td>
</tr>
<tr>
<td>Tone 2</td>
<td>rising</td>
<td>slightly longer</td>
<td>rises towards the end</td>
</tr>
<tr>
<td>Tone 3</td>
<td>low or dipping</td>
<td>very long</td>
<td>lightly falling, then rising</td>
</tr>
<tr>
<td>Tone 4</td>
<td>falling</td>
<td>very short</td>
<td>falls sharply towards the end</td>
</tr>
</tbody>
</table>

Table 1: Generalized characterization of the four citation tones of Mandarin Chinese, according to Chao (1968, 25), Hunold (2005), and Kratochvil (1968).

Because intonation is realized through the same means, the early studies on Chinese phonology came to the conclusion that Mandarin Chinese cannot have intonation (according to Chao, 1933, p. 130). In fact, many tone languages have obligatory question intonation, some even lack other means to distinguish questions (Yip, 2002, p. 114). Recent studies on intonation in Mandarin Chinese found question intonation for both types of questions discussed here. Fundamental frequency was the most important cue found for marking questions. The different findings on prosody in questions can be summarized as follows:

**F0 height**  Several studies found the f0 curve of questions to be shifted upward in the speaker’s voice range throughout the whole sentence (Yuan et al., 2002), (Yuan, 2004), (Chang, 1998, p. 53), (Shen, 1990), (DeFrancis, 1963, p. xxx), (Ultan, 1978, p. 227) (Ho, 1976), (Ho, 1977), (Chuang et al., 2007), (Tseng, 2003). A schematized example can be found in figure 1.

\(^3\) There is much disagreement on its exact scope, i.e. whether tone is a segmental or suprasegmental (i.e. prosodic) feature. For the current discussion, this is of little relevance. See for example Duanmu (2000, p. 233), Hartmann (2006, p. 164) or Yip (2002, p. 28) for a discussion of this.

\(^4\) There is also evidence to the contrary, for example Chang and Yao (2007, p. 4) found that in contrast to non-whispered speech, intensity differences were non-significant in whispered speech, except for the difference between tone 1 and tone 3. Since intensity and F0 seem to be highly correlated (Duanmu, 2000, p. 232), changes in intensity might be due to changes in pitch height. See the discussion on this in Whalen and Xu (1992, p. 45). Durational differences between tones were found by Duanmu (2000, p. 219) to be insignificant, except for final tone 3, which is considerably longer. Howie (1976, p. 203-214), in contrast to the pattern in table 1, found tone 1 to be even shorter than tone 4. If one considers the general tendency that rising tones take longer to be produced (see Ohala, 1978, p. 31), Howie’s findings would be in agreement with the theory that more complicated tones take longer to produce.
Another f0 behavior commonly mentioned for question intonation involves a level or gradually rising pitch, in contrast to statement intonation, which shows a falling tendency (Duanmu, 2000), (Lee, 2005), (Shen, 1990). In statement sentences, the f0 contour gradually falls due to declination. Declination refers to the gradual lowering of F0 during an utterance (Shih, 2001, p. 17), (Yip, 2002, p. 9), (Ohala, 1978). It is generally assumed to be a feature of all languages, at least in statements; in questions it can be suspended in some languages (Connell, 2001). This might be the case in Mandarin Chinese. Figure 2 illustrates the phenomenon.
The third mechanism often found in Mandarin Chinese involves a local pitch rise starting at or near the end of the sentence, as shown in figure 3 (Chao, 1933), (Gårding, 1987), (Lee, 2005), (Yuan et al., 2002) (Yuan, 2004), (Chang, 1998, p. 53), (Chuang et al., 2007), (Tseng, 2003), (Liu and Xu, 2005). In cross-linguistic studies, this final rise is one of the most frequently found methods of marking questions (Ultan, 1978, p. 220), (Sadock and Zwicky, 1985, p. 181), (Bolinger, 1978, p. 501).

**F0 range**  Another possible mechanism for indicating questions is an expansion of pitch range, as found for example by (Lee, 2005, p. 155), (Ho, 1977, p. 451), or (Yuan, 2004, p. 102).
Intensity/Duration  Differences in intensity and duration were also found. The duration of all syllables except the final one seems to be shorter in questions (Yuan, 2004, p. 80), while the final syllable is longer in unmarked questions (Yuan, 2004, p. 80), (Tseng, 2003, p. 483), (Ho, 1977, p. 449). One of the four speakers recorded by Gårding (1987, p. 27) showed an entirely different way of forming questions than the others: his question intonation is level rather than rising, while the questions are shortened compared to the statements. In the two studies which took intensity into account, intensity seems to be generally higher in questions, and to become higher towards the end (Yuan, 2004, p. 80), (Ho, 1977, p. 451). This parallels the behavior of the f0 curve found in these two studies.

Tone-specific changes The interaction of final tones with question intonation is a frequently studied topic. Question intonation seems to cause the high level tone 1 to rise slightly or to be raised slightly (Schack, 2000, p. 32f), and to steepen the rise of the rising tone 2 and the dipping tone 3 (Shen, 1989), (Zeng et al., 2004), (Ho, 1977), (Schack, 2000, p. 32f). The fall of the falling tone 4 was found to either be flattened (Zeng et al., 2004), (Ho, 1977), (Yuan, 2004, p. 102), (Schack, 2000, p. 32f) or to be left unchanged (Shen, 1989, p. 65). This form of combination is called simultaneous tonal addition, its resulting tones are ‘algebraic sums’ of the tones and the sentence intonation (Chao, 1933, p. 131). The realization of the third tone in connected speech is more complicated. In non-final position it is always low (Duann, 2000, p. 220), (Yip, 2002, p. 30), a fact that is traditionally captured by a tone sandhi rule (Li and Thompson, 1981, p. 8):

- When a third-tone syllable is followed by a syllable with any tone other than another third tone, the third-tone syllable changes to a low-tone syllable.

In final position, some speakers pronounce it as a low tone in statements, but it is always pronounced as a dipping tone in questions (Shih, 1997, p. 102) (Lee, 2005, p. 115), (Yuan, 2004, p. 82) (Gårding, 1987, p. 21). Yuan (2004, p. 76) found that it is about the same height in questions and statements, in contrast with the other tones which are higher in questions. Ho (1977, p. 453) comes to the same conclusion, while Ho (1976, p. 8) notes that utterance-final third tones end higher in statements than in questions, but the turning point is lower. The duration of all tones increases in unmarked questions (Ho, 1977, p. 452), (Yuan, 2004, p. 81), while their relative order of duration stays the same. The greatest increase in length is for tones 3 and 4, then tone 1, and finally tone 2 (Ho, 1977, p. 452). Lee is one of the few who analyzed sentences with several final neutral tone syllables. They do not necessarily rise, a fact which Lee sees as support for the claim that a larger unit, namely the last constituent, carries the final rise (Lee, 2005, p. 111).

For Yuan, tone specific phenomena at the end of the utterance are very important for question recognition in the problematic cases of falling and rising tones (Yuan, 2004). Similarly, Schack claims that her speaker did not show any significant differences between questions and statements except for the tone-specific changes on the final syllable, which she attributes to a sentence-final boundary tone (Schack, 2000, p. 36).

---

5But note that according to Zeng et al., the duration of final tones is shorter in syntactically unmarked questions than statements.
Summary  Based on the findings on question intonation in ma-particle questions as well as in syntactically unmarked questions, our hypothesis regarding question intonation is the following:

- Hypothesis 1: Ma-particle questions and syntactically unmarked questions differ prosodically from statements.
- Research Question 1: Which of the following possibly differentiating features are relevant?
  - Global raising of the pitch contour
  - Gradually rising or level pitch contour
  - Increase of pitch range
  - Local rise in the last syllable
  - Shorter syllables, except for the final one in unmarked questions
  - Higher and rising intensity values
  - Tonal changes in the last syllable

While most of the researchers cited above found that several of these distinguishing features apply together, the results of Chuang et al. (2007) indicate that it is either a global raise or a final rise, depending on the degree of astonishment of the speaker, and Tseng (2003) claims that unmarked questions are globally raised, and ma questions involve a final rise.

2.2.2 Are there different kinds of question intonation?

Generally, ma-particle questions and syntactically unmarked questions seem to be marked in a similar manner, compared to other kinds of questions (Chang, 1998, p. 57), (Shen, 1989, p. 19f), (Schack, 2000, p. 34), but the prosodic features assumed for questions are usually said to be more prominent in unmarked questions. For example, Shen (1990, p. 19) observes a higher final rise, and Lee (2005, p. 90) a wider pitch range. This would suggest that Haan’s Functional Hypothesis (see p. 4) is correct for Y/N questions in Mandarin Chinese. Chang’s findings, however, suggest that this is not always the case. The f0 difference between ma questions and unmarked questions was not significant for one speaker, and for the other two, it was significant only initially (for one speaker), or medially (for the other) (Chang, 1998, p. 57). Both of them uttered their ma questions with a higher pitch than their unmarked questions. This is also in contrast to Tseng (2003, p. 483), who claimed that the difference between the two kinds of questions is a difference in method: unmarked questions show global raising of the pitch contour, ma questions exhibit only a final rise.

This means that evidence for a hypothesis such as the following was found:

- Ma-particle questions and syntactically unmarked questions differ prosodically from each other;

but not for a hypothesis along the lines of:

- The Functional Hypothesis is correct for Mandarin Chinese Y/N question.

Note that Yuan (2004) and Ho (1977) did not investigate ma questions, and Ho (1976) did not investigate unmarked questions.
The Y/N question particle *ma* has neutral tone (Shen, 1990, p. 60), it is therefore expected that its realization will depend strongly on the preceding tone. Yip (2002, p. 181) cites table 2 from Shih (1987).

<table>
<thead>
<tr>
<th>Preceding tone</th>
<th>Neutral tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>tone 1</td>
<td>starts high, then falls</td>
</tr>
<tr>
<td>tone 2</td>
<td>starts high, then falls, but not as low as after tone 1</td>
</tr>
<tr>
<td>tone 3</td>
<td>starts fairly low, then rises</td>
</tr>
<tr>
<td>tone 4</td>
<td>starts fairly low, then falls even lower</td>
</tr>
</tbody>
</table>

Table 2: Realization of neutral tones after tones 1, 2, 3, and 4.

Essentially, the weak syllable receives a falling tone in all cases except for after tone 3. Yip (2002, p. 182) suggests that the third and fourth tone spread to the subsequent weak syllable, i.e. the last part of the tone is realized there. The rise after third tone syllables occurs because of an added feature H, i.e. the tone occurs high on the lower register. This spread of tone has been used as an argument for the suprasegmental nature of tone. For the realizations of the neutral tone after syllables with first and second tone, she proposes that weak syllables generally have lower register tones, they are on the lower register in her feature system. This explains Schack’s results. According to Schack (2000, p. 34ff), in *ma* questions, the high tone is realized similarly to that in unmarked questions, but the following *ma* is low; in sentences with final rising tone, the *ma* carries a rise like the final rising tone in statements, then falls about 30Hz. In the other tones, the *ma* carries part of the intonation which the last syllables carry in unmarked questions. Lee argues that in *ma* particle questions (MQs), ‘the *ma* particle does not stand as an independent prosodic unit that could bear the terminal rise in MQs. Rather, it is cliticized onto the preceding full-toned syllable and forms part of the preceding prosodic unit. Therefore, the terminal rise in MQs needs to be viewed as a raised pitch contour over the prosodic unit followed by the *ma* particle’ (Lee, 2005, p. 122). The previous findings thus suggest that the difference in intonation is not carried by the *ma* particle alone.

When looking at the difference between marked and unmarked questions before the start of the *ma* particle, an important distinction is that of duration. In unmarked questions, the duration of the final syllable is longer than in statements. In *ma* particle questions, the durations of test word syllables, independent of tones, is shorter for *ma* questions than for unmarked questions or statements (Zeng et al., 2004), (Ultan, 1978, p. 219), probably because there is no final lengthening.

Summary A summary of the few findings on the difference between unmarked questions and *ma* questions would be the following:

- Hypothesis 2: *Ma*-particle questions and syntactically unmarked questions differ prosodically from each other. The difference is not carried mainly by the *ma* particle.

If Hypothesis 2 is correct, the findings described above lead to two new research questions, the answer of the first of which will also concern the Functional Hypothesis.

---

6Note that Zeng et al. (2004) and Ho (1976) attribute the first tone to the *ma* particle.
• Research Question 2: Are the two kinds of questions marked the same?
  – If yes,
    * is the marking stronger for syntactically unmarked questions?
    * is the marking stronger for ma particle questions?
  – If no,
    * are unmarked questions marked by global raising and ma questions with a final rise?

• Research Question 3: Is the shortening of the last word in ma questions, and its lengthening in unmarked questions a relevant differentiating feature?

2.2.3 At what point does the distinction start?

Initial  Early theories of global raising, under the influence of the impression that tone language does not have intonation, usually proposed that the f0 contour of questions is parallel to that of statements, albeit raised (for example DeFrancis, 1963, p. xxix). Current advocates of this viewpoint generally believe global raising to be combined with level pitch or a rise or a pitch range expansion at some point of the sentence. Global raising of register contrasts with the other features in that it suggest that there are initial differences between questions and statements. Yuan found an initial difference of about 10 Hz (Yuan, 2004, p. 73), while Shen speaks of 57-81Hz. Note that Schack also speaks of initial pitch height differences of about 20Hz, but concludes that the difference is too small as to be perceivable and therefore decides to ignore it in her study (Schack, 2000, p. 29). Chang and Ho do not give any precise values, but the mean initial pitch values depicted in their graphs do not lie more than 50Hz apart (Chang, 1998, p. 56f),(Ho, 1977, p. 450). Concerning the differentiation between unmarked questions and ma-particle questions, there was not much to be found in the literature. If Chang and Tseng are right in their differentiation of these two kinds of questions, the difference must be noticeable sentence-initially: either ma-particle questions are raised to a further extent (Chang, 1998, p. 57), or unmarked questions are raised, while ma questions stay similar to statements until the final rise (Tseng, 2003, p. 483).

Medial/Final  Most studies mentioned here seem to indicate that the differences between statements and questions is clearest at or near the end of the utterance. The studies can be classified according to which point the divergence starts at. The results of (Duanmu, 2000), (Lee, 2005), and (Shen, 1990) indicate that the rise begins sentence-initially, while Liu and Xu’s findings indicate that it begins after the focused word (Liu and Xu, 2005, p. 77), and gets steeper exponentially or even double-exponentially towards the end (= the ‘typical’ question intonation) (Liu and Xu, 2005, p. 79, 81). When it comes to final rises, Gårding found that for three of her four subjects, questions showed a rise beginning with the predicate, i.e. in the last three syllables of her pentasyllabic test sentences, with the exception of final low tone (Gårding, 1987, p. 21). The other studies which concluded that questions are indicated by a final rise found it to take place on the last two syllables, becoming most salient in the last syllable (average final difference: 47.6 Hz, according to Yuan) (Chang, 1998, p. 53), (Yuan, 2004, p. 73),(Ho, 1977, p. 450). The last two syllables do not always constitute a single syntactic unit, in Ho’s study, they were two words, while in Chang’s study, they were the last part of the final trisyllabic word. Yuan used final words of one, two, or
three syllables. Lee found a gradual expansion in pitch range and f0 rise which reached its greatest extent towards the end, namely the utterance-final NP that contains most important information. According to Lee, this expansion and rise is the greatest on the last high tone target, i.e. the high part of the last tone 1, tone 2, or tone 4 of the sentence (Lee, 2005, p. 155). Ho found a pitch range difference in the final three syllables. Interestingly, pitch range is not uniformly broader in questions, but it was smaller than in the corresponding statements in the third-to-last and second-to-last syllable, and greater in the last syllable (Ho, 1977, p. 451). This may mean that pitch range expands in questions only in the last word, but it may also be due to some disturbances due to focusing of the third-to-last word in Ho’s statement sentences. According to Yuan, the strength increases towards the end. This includes expanded pitch range (Yuan, 2004, p. 102). When it comes to differentiating between the two kinds of questions, if lengthening or shortening of the last syllable is the most relevant perceptual cue, the differences would only become clear during the course of the last syllable.

**Depending on Focus**  Mandarin Chinese is one of the tone languages in which prosodical means are used to express narrow focus (Xu, 2001, p. 15). This kind of focus is used when a constituent is meant to contrast with a set of alternatives (Schaffar, 2000, p. 13). Most researchers agree that pitch range expansion on the focused element and post-focal pitch range reduction are the main cues for focus in Mandarin Chinese, for example (Shih, 2001, p. 18), (Surendran et al., 2005), (Jin, 1996), (Xu, 1999, p. 94), (Yuan, 2004, p. 92), (Liu and Xu, 2005, p. 71,76), (Gårding, 1987, p. 23). The f0 range of the pre-focal constituents does not appear to be significantly different from that in sentences without focus (Surendran et al., 2005). According to Shih, the pitch range reduction leads to post-focal pitch values that are lower than they would be in sentences without focus (Shih, 2001, p. 42). The duration of a syllable increases when it is focused (Jin, 1996), (Xu, 2001), (Liu and Xu, 2005, p. 72), (Yuan, 2004, p. 96). Intensity increases on the focused syllable, and decreases strongly afterwards (Yuan, 2004, p. 99). The tones of focused syllables are more clearly recognizable, while the f0 curve of the following syllable falls steeply, so that its actual tone is barely distinguishable (Xu, 1999, p. 95).

Focus realization interacts with and sometimes competes with question realization. Liu and Xu are most explicit about this. They looked at initial, medial and final focus and concluded that in sentences with initial or medial focus, the pitch range of questions begins to rise after the focused part (Liu and Xu, 2005, p. 77). In sentences with final focus or without focus, questions are distinguished by a final rise in the last word. This has several consequences for question recognition: (a) Sentence-final focus in statements makes it difficult to distinguish them from questions (Liu and Xu, 2005, p. 82), (Yuan, 2004, p. 99); it is usually less strong compared to medial or initial focus, so as not to be taken for question intonation (Liu and Xu, 2005, p. 86); (b) Sentence-initial or -medial focus makes it difficult to distinguish questions from statements with neutral focus, because of the following reduced pitch range (Liu and Xu, 2005, p. 82); (c) That other questions (for example A-not-A-questions) do not show the same rise at the end of the utterance might be due to focus. Liu and Xu explain that these questions have an inherent narrow focus, thus leading to a post-focal decrease in pitch range (Liu and Xu, 2005, p. 85). There are also contrasting views. Lee claims that the pitch raising in pre-focal position may be enough to indicate question intonation (Lee, 2005, p. 145). She also claims that pitch range expansion on the focused component is larger in questions...
Previous work Grubic, 13

than in statements (Lee, 2005, p. 150). Additionally, the post-focal pitch is higher at least for unmarked questions (Lee, 2005, p. 150), and pitch range expansion to signal question still takes place, although to a lesser extent than in sentences with neutral focus (Lee, 2005, p. 159). Post-focal compression of pitch range is more salient in ma questions than in unmarked questions, and the terminal rise is more reduced (Lee, 2005, p. 151).

Summary

• Research question 4: At which of the following locations do questions and statements start to differ:
  – sentence-initially?
  – after the predicate?
  – during the last two syllables, independently of whether they form a lexical unit or not?
  – after the focused constituent?

• Research question 5: Do ma-particle questions and syntactically unmarked questions begin to differ from each other sentence-initially or finally?

2.3 Summary

It is now possible to summarize the working hypotheses on Mandarin Chinese question intonation. The first hypothesis concerns the existence and nature of the prosodic difference between questions and statements. The working hypotheses of this study concern the existence of differences between the three kinds of sentence types:

• Hypothesis 1: Ma-particle questions and syntactically unmarked questions differ prosodically from statements.

• Hypothesis 2: Ma-particle questions and syntactically unmarked questions differ prosodically from each other. The difference is not carried mainly by the ma particle.

If the difference between questions and statements can be assumed, the first important new questions concern the nature of question marking. There were seven possibly relevant sentence type distinguishing features, which can occur together or separately.

• Research Question 1: Which of the following possibly differentiating features are relevant?
  – Global raising of the pitch contour
  – Gradually rising or level pitch contour
  – Increase of pitch range
  – Local rise in the last syllable
  – Shorter syllables, except for the final one in unmarked questions
  – Higher and rising intensity values
  – Tonal changes in the last syllable
As soon as the question-distinguishing features are established, the next relevant research question concerns the differences between unmarked and marked questions. Here, if the questions are marked the same, but the differentiating features are more salient for unmarked questions, Haan’s Functional Hypothesis would be correct for Mandarin Chinese. The lengthening/shortening of final syllables in questions may be relevant to distinguish them from each other, even if it is not a relevant feature to distinguish questions from statements.

- Research Question 2: Are the two kinds of questions marked the same?
  - If yes,
    - is the marking stronger for syntactically unmarked questions?
    - is the marking stronger for ma particle questions?
  - If no,
    - are unmarked questions marked by global raising and ma questions with a final rise?

- Research Question 3: Is the shortening of the last word in ma questions, and its lengthening in unmarked questions a relevant differentiating feature?

The question of the location of the difference is strongly related to that of the nature of the difference. The reason for investigating it separately is that this is a question that cannot be answered solely by analyzing the sentences produced. It might be that there are clear and systematic prosodic differences at an early point in the sentence, which are nevertheless not perceived by the hearer.

- Research question 4: At which of the following locations do questions and statements start to differ:
  - sentence-initially?
  - after the predicate?
  - during the last two syllables, independently of whether they form a lexical unit or not?
  - after the focused constituent?

- Research question 5: Do ma-particle questions and syntactically unmarked questions begin to differ from each other sentence-initially or finally?

In the following chapter, a production experiment is conducted to further investigate the research questions with respect to how questions are produced, while an attempt to answer how questions are perceived will follow in chapter 4.
3 Production

In the previous chapter, literature on the intonation of two kinds of Y/N questions in Mandarin Chinese was presented, and research questions were formulated which can be summarized as follows:

- Is there a distinct question intonation for Y/N questions?
- Are there different kinds of question intonation for the two kinds of questions?
- If the three sentence types can be distinguished, at what point does the distinction start?

In this chapter, a production experiment is presented in an attempt to answer these questions. Sections 3.1 to 3.3 explain how the experiment was performed, how the materials were constructed, and how the data was prepared for analysis. Section 3.4 presents the results of the analysis. First some general features of the test sentences are described, then the relevance of each of the possibly sentence type distinguishing features found in the literature presented in chapter 2 is tested on the production data. In section 3.5, the hypotheses of chapter 2 are reevaluated under the light of the findings of this chapter, and preliminary answers to the research questions are presented.

3.1 Procedure

To test the hypotheses of chapter 2, we recorded 441 sentences by three speakers. All three speakers are female, well educated, in their twenties or early thirties, and speak Mandarin Chinese as their first language. They have all lived in Germany for less than five years, and speak at least two other languages fluently. The speaker referred to as ‘speaker 1’ grew up in Xi’an, with parents from Beijing, ‘speaker 2’ went to university in Beijing, and ‘speaker 3’ visited high school and university there. All live in Germany at the moment. The recordings were done in a soundproof recording booth at the ‘Zentrum für allgemeine Sprachwissenschaft’ (ZAS, http://www.zas.gwu-berlin.de/index.html?phonlab) in Berlin.

3.2 Materials

Three kinds of sentences were recorded: statement, ma-y/n-questions and unmarked y/n-questions. They are preceded by a context sentence, and are all the same except for the last word:

ta1 xi3huan1 mei3tian1 mai3 dian3 X.
she like daily buy a-bit X.
‘she likes to buy a bit of / some X every day.’

In these sentences, X is one of the words in table 3.

The last word is varied, in order to test the influence of its tones on the intonation pattern. The tones of the sentence preceding the test word are not controlled. Sentence length is also varied, because of the possibility that differences may become clearer in longer sentences (Lee, 2005, p. 35). In a pre-study with sentences of different length, it was found that the longest sentence that still is spoken as a single intonational phrase was a 14-syllabic sentence:

7The combination of two third tones was not used because of tone three sandhi (see chapter 3.4.1 below).
Table 3: Test words constructed by combining a low and a high tone with all four tones

<table>
<thead>
<tr>
<th>Tone</th>
<th>Word</th>
<th>Tone</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>xiang1jiao1</td>
<td>2</td>
<td>xiang1you2</td>
</tr>
<tr>
<td>3</td>
<td>xiang1cao3</td>
<td>4</td>
<td>xiang1liao4</td>
</tr>
</tbody>
</table>

*Banana*  
*Sesame oil*  
*Leek*  
*Spices*

<table>
<thead>
<tr>
<th>Tone</th>
<th>Word</th>
<th>Tone</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nai3gao1</td>
<td>2</td>
<td>nai3you2</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
<td>4</td>
<td>nai3liao4</td>
</tr>
</tbody>
</table>

*Baby food*  
*Cream*  
*Cheese*

Both short and long sentences were varied for focus. In short sentences, the focus was initial or final, while in long sentences there was an additional medial focus. The focus was controlled by the context sentence so it would be on *ta1*, on *hang2*, or on the test word. At the end, this yields 147 sentences (7 different sentences x 7 different last words x 3 kinds of sentences). The sentences and their contexts can be found in appendix A. Each of the three speakers read each sentence twice, in randomized order. Only the second versions of each sentence were used, because the speakers tended to speak more naturally after getting used to the sentences and the situation.

### 3.3 Data Preparation

The 441 sentences were prepared using Praat (Boersma and Weenink, 2007). First, the recordings were segmented into sentences, and the context sentences removed. Then, using a Praat script by Yi Xu, the syllables were labeled manually on one tier, and the syllable rhymes on another. The script can be found in appendix B on the CD accompanying this thesis paper. The labeling conventions used were those of (Turk et al., 2006). For nasals, laterals and fricatives, the abrupt changes in the spectral characteristics of the sound were used as defining feature, stops were defined from the offset of F2 and decrease of amplitude to the (first) release burst or, if no release burst existed, to the F2 onset. Affricate onsets were treated like those of stops, and offsets like those of fricatives. The vowels were not labeled separately, rhymes were defined to begin at the offset of the syllable-initial consonant and to end at the onset of the next syllable-initial consonant. In the case of the syllable *you*, the syllable and rhyme were identical. Figure 5 shows an example of the labeling procedure.

There were several cases where the segmentation was difficult due to the similarity or sameness of neighboring sounds. Such cases were the transitions between the words *xi3huan1* and *mei3tian1*, *mei3tian1* and *mai3*, *dian3* and *nai3*, and *nai3* and *you2*. In these cases, if there was no visible change in the formant structure, the point of lowest sound pressure in the waveform was used. When there were several possible boundaries, the first one was used (see Turk et al., 2006, p. 6, p. 16). Figure 6 shows an example of this. The labeling of these problematic cases was checked several times to ensure consistency. The assumption behind this was that as long as these cases were labeled consistently, it would not be problematic for the data analysis.

Another difficulty due to the choice of test sentences was that of the initial voiceless stop. Because of the preceding pause, it was impossible to find the onset of the stop (see Turk et al., 2006, p. 15). Due to this, the durational features of the first syllable are not reliable in this study.

---

8retrieved from [http://www.phon.ucl.ac.uk/home/yi/downloads.html](http://www.phon.ucl.ac.uk/home/yi/downloads.html)
Figure 5: Example: labeling based on spectral features

A Praat script was written to extract the duration of each syllable, and, departing slightly from Yuan’s analysis, to extract the intensity and the f0 values of the rhyme. Yuan had measured the maximal intensity and mean f0 of the whole syllable. In this study, the rhyme was chosen instead to minimize the influence of the tone of the previous syllable, which often carries over to the beginning of the next syllable (Chao, 1933, p. 128), (Xu, 1999, p. 6). This script can be found in appendix B on the data CD. The script opens each sound file and its textgrid. For each syllable, the syllable duration (in milliseconds) is extracted, and the f0 and intensity values are extracted via Pitch and Intensity objects. In creating the Pitch object, the pitch ceiling was set to 500 Hz, which was recommended in the Praat manual for female speakers, and the pitch floor was set lower than usual to 50 Hz, which is recommended for creaky voice, which is typical of tone 3 syllables. The timestep, i.e. the distance between two points of measurement, was set to 0.01 seconds. This is the standard value for Pitch object creation when the pitch floor is 75 Hz, but becomes undesirably larger if the pitch floor is lower. Note that this method of pitch extraction yields values that are slightly different to those obtained by the measures that are taken from the SoundEditor, where the values are computed based on all visible time points. For each syllable rhyme, the mean, minimal and maximal F0 values were computed, and the F0 values at ten points in time. Due to the labeling conventions, the onset and offset of the rhyme are not necessarily voiced (see Turk et al., 2006, p. 3). For this reason, the rhyme-initial and -final F0 values were not always measurable. The intensity object was created with a minimum pitch of 100 Hz, the standard value in Praat. The smoothing of the curve, so that it is independent of f0, can depend on this value. It has to be set as highly as possible so that the values are not smoothed too much, but low enough so that the values become independent of pitch. As can be seen in the section on intensity in chapter 3.4, the intensity values extracted are sufficiently pitch-independent. The mean intensity values were calculated for each syllable rhyme. The results can be seen in appendix C on the data CD.

3.4 Analysis

3.4.1 General Characteristics of the Sentences

Using spontaneous speech to study intonation is of course preferable to using read speech, because the naturalness of the data may be affected by the difference in cognitive processing of read and spontaneous speech, and the reading aloud competence of
Nevertheless, for several reasons we chose read speech in this thesis paper. Corpora of spoken Mandarin Chinese of the mainland variety are very difficult to obtain (McEnery and Xiao, 2004), and for studies on intonational features, reliable results would be very difficult to extract from corpus studies due to the fact that there are numerous factors that influence the intonation contour other than question intonation. Features that are controlled in the present study but are uncontrollable in spontaneous speech include segmental effects on the pitch of a vowel due to its height or to the nature of the consonants preceding it (Xu, 2001), (Hombert, 1978), (Ohala, 1978), tone and influences of the tones on each other and on toneless syllables (Chao, 1933) (Xu, 2001), (Xu, 1999), (Yip, 2002); stress, rhythm and timing which produce some of these toneless syllables and influence the time and effort spent on each tone (Cao, 2001), (Duanmu, 2000), (Lin, 2001); topic (Shih, 2001, p. 18); effects of discourse/sentence intonation and downtrends (Shih, 2001); and the current emotional state of the speaker (Pao et al., 2005), (Hartmann, 2006, p. 165), (Murray and Arnott, 1993). Due to the fact that the sentences were all the same except for the test word, and the sentences were read without emotion, these factors can be neglected in the current study, and it can be expected that all remaining differences are due to the difference between the three sentence types. This does not mean, however, that the exact contour found for these statements and questions is generalizable to all questions and statements. The individual curves are still influenced by the factors mentioned above. Two factors worth mentioning here in greater detail which highly influence the contour of the test sentences are downdrift and tone sandhi. They are both due to the influence of neighboring tones on each other. The following is an example of the pitch...
What can be seen here is that in statements, the third syllable, even though it carries a high tone like the first, is considerably lower than the first. Similarly, the fifth syllable is lower than the third. This effect, that a low tone triggers lowering of a following high tone (Shih, 2001, p. 17), is known as automatic downstep or downdrift (Yip, 2002, p. 262). According to Xu, downdrift is due to carryover influences of the low tones on the following high tones, while in a sequence of high tones, there is almost no lowering in Mandarin Chinese (Xu, 1999, p. 99) (Liu and Xu, 2005, p. 73). In some languages, there is downdrift in declaratives, but not in question intonation (Yip, 2002, p. 262), (Ohala, 1978, p. 31), (Schuh, 1978, p. 245). Duanmu even tentatively proposes a high final boundary tone for Mandarin questions that prevents downdrift (Duanmu, 2000, p. 236). This is not expected to be the case, as it has been found that tone 3 syllables pull down the f0 curve to statement level (Yuan, 2004, p. 76), (Ho, 1977, p. 453); and as we see in the example sentence in figure 8, this is not the case for the kinds of questions studied here (see also chapter 3.4.2)

Duanmu actually calls it ‘downstep’, but this is probably due to terminological confusion. (Non-automatic) downstep, a lowering of the high tones after preceding high tones, does not occur in Mandarin Chinese (Xu, 1999).
Another characteristic of the test sentences which can be seen in figure 7 is third-tone sandhi in the sequence *mai3 dian3*. The term sandhi originally refers to a systematic change in the phonological form of a word due to its surrounding words. For Chinese, however, sandhi is also used to describe changes that take place within words, in the domain of the syllable (Yip, 2002, p. 180). In Standard Chinese, when two third tones come together, a dissimilatory change takes place, changing the first syllable’s tone into a second tone (for example Li and Thompson, 1981, p. 8).

- When a third-tone syllable is followed by another third-tone syllable, the first one changes into a second tone.

This occurs before all low tones. Gandour cites a study by Cheng (1968), testing sentences by bilingual speakers, who tend to mix English words into their Chinese sentences. Tone 3 was changed into tone 2 before English words starting with syllables with reduced stress, which are pronounced with low pitch. He concluded that low pitch triggers tone 3 sandhi (Gandour, 1978, p. 56), (also described in Shih, 1997, p. 83). An application of third-tone sandhi on the sequence *mai3 dian3* would yield *mai[2] dian3*.

If several third tones succeed each other, as in the sequence *mai3 dian3 nai3 you2*, which ones are changed depends on phonological grouping (Shih, 1997). According to Shih, tone sandhi application is cyclic, but in the initial cycle, it can also apply simultaneously, although this rarely occurs since disyllabic feet are preferred as initial cycle. Shih notes that measure words like *dian3* are cliticized onto the preceding verb. Therefore, the sequence *mai3 dian3 nai3 you2* is grouped as *(mai3 dian3) (nai3 you2)*. Note that syntactically, *dian3* would be closer to *nai3 you2* than to *mai3*. The grouping often ignores syntactic structures (see Shih, 1997, p. 97). If the foot is taken as the initial cycle, third-tone sandhi would initially change the third tone of *mai3* into a second tone, then in a later cycle, optionally, the third tone of *dian3*. Therefore two different realizations are possible, *mai[2] dian3 nai3 you2* or *mai[2] dian[2] nai3 you2*. The latter could also be the result of a larger initial cycle causing simultaneous application of tone sandhi, but this is highly unlikely in this case, because it is an infrequent sentence, and a four-syllable structure as an initial cycle is very rare.

![Figure 9: Pitch contour of a statement with three successive third tones without focus](image)

Emphasis can change the prosodic grouping which is the basis for third-tone sandhi, because the emphasized element is usually initial in a foot or superfoot (Shih, 1997, p. 112). In the sentences presented here, focus on the final word (for example *nai3 you2*)
would not change anything, since it forms a foot already. Initial and medial focus does not occur close to a tone sandhi domain.

The aim of this section was to include a word of caution before the results are presented in the next section. Not only are the individual curves of the sentences which will be included below for visual comparison highly sentence-specific, but the outcome of the production experiment may also be influenced by the choice of the test sentence. In reading this description of the results, it must for instance be kept in mind that the sentences chosen for this experiment contain many tone 3 syllables, which are said to pull down the f0 curve to statement level (Yuan, 2004, p. 76), (Ho, 1977, p. 453). While this probably does not affect the general relevance of the distinguishing features found in this section, it may be that the distinction would be stronger if other test sentences were chosen. The following sections will test the theories on the f0 curve, pitch range, duration, intensity, focus and tones.

3.4.2 Is there a distinct question intonation?

F0 Speaker’s voice range and height influence f0 range and height, and are again influenced by individual factors, for example sex and age of the speaker (Xu, 2001, p. 6), (Shih, 2001, p. 35) (Hartmann, 2006, p. 165). For this reason, the data was speaker-normalized using the following formula (Truckenbrodt, 2004, p. 325):

\[
\text{normalized value} = \frac{\text{original value} - \text{mean}_s(\text{Point}_2)}{\text{mean}_s(\text{Point}_1) - \text{mean}_s(\text{Point}_2)}
\]

where \(\text{mean}_s(\text{Point}_x)\) is the mean value at point \(\text{Point}_x\) for each speaker, and \(\text{Point}_1\) and \(\text{Point}_2\) are reference points in the sentence which were chosen for the normalization. These reference points are chosen to reflect the speaker’s voice range, \(\text{Point}_1\) is typically an initial peak in the sentence, while \(\text{Point}_2\) is a lower point near the end of the sentence. In (Truckenbrodt and Féry, 2003, p. 11), it was the low boundary tone, while in (Truckenbrodt, 2004), it was one of the last prenuclear peaks. Here, \(\text{Point}_1\) will correspond to the initial peak \(\text{ta}_1\) in short unfocused statement sentences, while \(\text{Point}_2\) was chosen to be the lowest trough in the same sentences, namely the valley \(\text{mei}_3\), which is the lowest point for each individual speaker. Although this point is relatively early in the sentence, it is preferable to later low points, because these show large variations both between and within speakers, due either to the fact that they are test syllables, or that their tonal realization is influenced by the neighboring test syllables. All f0 values are normalized so that the resulting values are usually between 1 (the mean value at \(\text{Point}_1\)) and 0 (the mean value at \(\text{Point}_2\)).

F0 height Concerning f0 height of questions and statements, three features were mentioned in chapter 2. They are repeated here:

- Research Question 1: Which of the following possibly differentiating features are relevant?
  - Global raising of the pitch contour
  - Gradually rising or level pitch contour
  - Local rise in the last syllable
When comparing short statements without focus to their corresponding syntactically unmarked questions in figures 10(a) / 10(d), our results seem to confirm the first two hypotheses: Syntactically marked questions are higher than statements in all positions. While the mean f0 values for statements drop after the high tone syllable tian1, the values for questions stay level. The f0 difference between short ma questions and statements, shown in figures 10(b) / 10(e) parallels that of the difference between unmarked questions and statements. These ma particle questions are raised even higher compared to the statement sentences than unmarked questions. In both kinds of questions the difference gets larger after tian1, and slightly larger at xi3.\(^1\)

In long sentences, the results look different (see figures 11(a)/11(d)). Unmarked questions are lower than statements in the initial part of the sentence, and seem to be only marginally higher except for the last two syllables, the final syllable even exhibiting a slight rise. Long marked questions in figures 11(b)/11(e), on the contrary, show similar results to those in short sentences, with global raising, and a level pitch curve without final rise.

\(^1\)One final low tone value was undefinable due to creaky voice, the mean values for the final syllable were computed based on the remaining values.
Figure 10: Differences in mean f0 of short sentences without focus
Figure 11: Differences in mean f0 of long sentences without focus
The research question concerning the similarities and differences of the f0 curves of the two kinds of questions is repeated here.

- Research Question 2: Are the two kinds of questions marked the same?
  - If yes,
    - is the marking stronger for syntactically unmarked questions?
    - is the marking stronger for ma particle questions?
  - If no,
    - are unmarked questions marked by global raising and ma questions with a final rise?

- Research Question 3: Is the shortening of the last word in ma questions, and its lengthening in unmarked questions a relevant differentiating feature?

When the curves for the two kinds of questions are compared in figure 10(c) and 10(f), it seems that Tseng’s claim that unmarked questions are globally raised while marked questions only exhibit a local final rise can be refuted here, as well as the claim that the marking is stronger for syntactically unmarked questions. Even though the differences are only slight, the f0 values of ma particle questions is higher in seven out of nine syllables in figure 10(c). The two syllables where this is not the case are not the final ones, but the two syllables preceding the test word, mai3 and dian3. The differences are greatest for the second syllable xi3, then for the final syllable, but never exceeds 0.09. It seems highly likely that the two kinds of questions cannot be differentiated by f0 alone. In longer sentences, the syntactically marked questions were raised higher than the syntactically unmarked questions in all positions, with the difference being largest at li4, then in the first test word syllable. Here, the opposite of Tseng’s findings can be seen: unmarked questions exhibit a final rise, and marked questions are globally raised.

A repeated-measures analysis of variance was performed for the mean f0 values of each syllable, with sentence type as fixed factor, and speaker as random factor. To do several t tests was undesirable because of (a) an increased probability of making a Type I error, i.e. the probability that this study might wrongly reject a null hypothesis, and (b) the dependency of the results on each other due to computing the tests several times (Hays p. 403). Due to the fact that the results of an ANOVA do not show between which levels of the factors a difference exists, a post-hoc TukeyHSD test was administered for each comparison. The result is given below in tables 4 and 5. They are transformed using the single-step method. For an analysis of variance of this kind, there are certain assumptions about the treatment population:

- Normality assumption
- Homogeneity of variance
- Independency assumption
- Sphericity assumption

The analysis of variance is relatively robust concerning violations of the first two assumptions, but failure to comply to the independency and sphericity assumption can
lead to severe changes in the results. Because tests to establish whether the sphericity assumption is fulfilled rely heavily on normal distribution of data, usually other ways are found to adapt to this assumption. One of these ways is to adjust the degrees of freedom in the F test to account for possible nonsphericity in the data.

The results can be seen below. The number of stars indicates the level of significance (. = p < 0.1; * = p < 0.05; ** = p < 0.01; *** = p < 0.001). The results for short sentences in table 4 indicate that the global raising found in short unmarked questions was not statistically significant. The global raising found for short marked questions, however, was significant in eight out of nine syllables. The difference between statements and unmarked questions seems to be due to the level pitch at the end. It starts from the syllable \( \text{mai}3 \). In long sentences, however, the theory of global raising was refuted for both kinds of questions. No significant difference was found between statements and unmarked questions except for the last syllable (p < 0.01). Long marked questions are not globally raised, but the statistical significant difference to statements starts earlier in the sentence than with unmarked questions. The difference between unmarked and marked questions manifests itself in the first two high tone syllables \( \text{ta}1 \) and \( \text{huan}1 \), then in \( \text{li}4 \), then in \( \text{shi}4 \). This is different to what would be expected from the discussion above, but this difference may be due to the fact that the random error associated with the speaker of the sentence was taken into account in the computation of the ANOVA.

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>***</td>
<td>*</td>
<td>**</td>
<td>.</td>
<td>**</td>
</tr>
<tr>
<td>St. - M.Q.</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: Mean F0: Significance level of adjusted p-values, short sentences

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>St. - M.Q.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>***</td>
<td>****</td>
<td>.</td>
<td>*</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>*</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>*</td>
<td>**</td>
<td>.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Mean F0: Significance level of adjusted p-values, long sentences

**Summary**  Hypothesis 1 can be confirmed, \( ma \)-particle questions and syntactically unmarked questions differ prosodically from statements. Global raising was found for short \( ma \)-particle questions, while short syntactically unmarked questions were marked by level pitch after \( \text{mai}3 \). In long sentences, the differences between questions and statements appeared later in the sentence, with \( ma \)-particle questions being raised starting with \( \text{chao}1 \), and unmarked questions showing a final rise in the last syllable of the sentence. All kinds of questions showed a level pitch contour compared to falling pitch in statements. While all three kinds of findings concerning the differences between questions and statements in Mandarin Chinese (i.e. global raising, level pitch, local
rise) seem to occur at least in some environments, the findings concerning the f0 value differences between the two kinds of questions were not found to be true. Not only are *ma*-particle questions not consistently raised above their unmarked counterparts, but there was also no evidence for the Functional Hypothesis. The two questions seem to be marked differently, although there is no consistent pattern across different sentence lengths as to how they are marked. The marking of *ma* questions is more prominent earlier in the sentence in both cases. It remains to be seen whether these kinds of prosodical question marking are perceivable.
**Pitch range** Another possibly relevant feature concerning the prosodic marking of Y/N questions involves an expanded pitch range. According to Lee, apart from a rising F0 curve, questions are also characterized by a gradual expansion in pitch range. She claims that this expansion is greatest on the last high tone target of the sentence. She also claims that in disyllabic test words the final syllable remains unstressed, and the pitch range of the final stressed syllable is expanded (Lee, 2005, p. 136). Figures 12 and 13 show the pitch ranges of questions and statements. The pitch ranges are represented by one line each for the top and the bottom line. In these figures, we see that it is not the case that the pitch ranges of the questions are uniformly larger than those of the statement sentences: the average pitch ranges of questions and statements seem to be similar. In short syntactically unmarked questions (figures 12(d) and 13(d)) there seems to be no systematic difference, while in marked questions (figures 12(e) and 13(e)), and long unmarked questions (figure 13(d)) the pitch ranges of the last syllables are compressed compared to statement sentences. The two kinds of questions also do not exhibit any systematic pitch range differences (see figures 12(f) and 13(f)), apart from the lower pitch range of marked questions on the last syllable.

While Lee’s hypothesis on the expanded pitch range for questions was not confirmed by the results of this study, figures 12(a) to 12(c) and 13(a) to 13(c) reaffirm the findings made above about f0 height. Figures 12(a) and 12(b) show a globally raised pitch range for short questions, while figure 12(c) shows an at least initially higher pitch range for marked questions. In contrast, the pitch range of long unmarked questions (figure 13(a)) is very similar to that of statements. In this kind of question, there are no differences until the later part of the sentence, where the topline of the pitch range for questions rises in the last syllable and the baseline stays level. The pitch range of long marked questions (figure 13(b)), albeit slightly raised, stays parallel to that of statements in the first part of the sentence. In the last two syllables, while the pitch range of statements is falling, the pitch range of questions stays level.

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6: Pitch Range: Significance level of adjusted p-values, short sentences

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Pitch Range: Significance level of adjusted p-values, long sentences

The tables 6 and 7 show the results of a second repeated-measures ANOVA with the same fixed factor (sentence type) and the same random factor (speaker), performed on
the pitch range of each syllable. The results for short sentences confirm the judgement made by visual comparison. The only significant difference found was the difference between the final pitch range of ma particle questions and the two other sentence types, and it is only significant on the 10% level (p<0.1). Therefore, it will be ignored here. It would not confirm Lee’s hypothesis anyway, since the pitch range of marked questions is narrowed and not expanded at this point. Long unmarked questions show a pitch range expansion at li4 which is statistically significant (p<0.05), and marked questions show one at mai3 (p<0.05). These results cannot be explained at present. Again, the larger differences seen in figure 13 towards the end of the sentences seem to be due to speaker error.

Summary The data in this study do not support the hypothesis that questions are marked by an expanded pitch range which gets gradually larger up to the last high tone target. The mean pitch range of the test words was similar in unmarked questions and statements, and the last syllable was narrower in marked questions than in statements. The data does, however, support the findings above that pitch range is globally raised for at least some questions, and that it does not decline finally like the pitch range of statements does.
Figure 12: Differences in pitch range of short sentences without focus.
Figure 13: Differences in pitch range of long sentences without focus
**Duration** Since the absolute durations of the respective syllables may vary from test sentence to test sentence, and from labeler to labeler, the relative length of the syllables of a sentence to each other will not be discussed here. The durational values were normalized using a z-score transformation: For each speaker, each sentence length and each syllable type of statement sentences without focus,

\[
\text{normalized value} = \frac{\text{absolute value} - \text{mean}}{\text{sd}}
\]

Peters et al. had used the formula to investigate the differences between phrase-internal and phrase-final segments, and had computed the mean values accordingly (Peters et al., 2005, p. 149). Here, the interest lies in the difference between the length of syllables in questions, and the length of the same syllables in statements. Therefore, the means and standard deviations are calculated for syllables instead of for segments. The resulting values are the measure of number of standard deviations that this value departs from the norm, the norm in this case being defined by the values for unfocused statement sentences. All values between -1 and 1, i.e., within the standard deviation, are classified as unchanged, all values higher than 1 are elongated, all values lower than -1 are shortened. In figure 14, these thresholds are marked by a dotted line. Note that in order to be able to calculate an appropriate mean value and standard deviation, the values of each of the two test syllables are formed by considering all kinds of syllables in this position.

It is expected, according to the findings of the previous chapter, that the length of all syllables except for the final one is shorter in questions than in statements. The last syllable is expected to be shortest for *ma* questions, because there is no final lengthening, and longest for unmarked questions. As we see in figures 14(a) and 14(d), however, the syllables of the unmarked questions are not elongated or shortened compared to those of statements of the same length. In figures 14(b) and 14(e) it becomes clear that the last syllable is shorter in *ma* questions than in statements (2.83 sds and 1.77 sds shorter, respectively), as predicted, but that not all syllables are shorter than the statement syllables. Furthermore, the average length of the word *ta1*, is slightly shortened (1.26 sds) in short *ma* questions. In figures 14(c) and 14(f) the normalized values for the unmarked questions were subtracted from those of the *ma* question to get the difference between the duration of unmarked and marked questions. Again, the last syllable of the *ma* question is shortened in comparison with the corresponding syllable in unmarked questions (2.93 sds shorter).

A repeated-measures ANOVA with syllable duration as independent variable, sentence type as fixed factor and speaker as random factor yielded the following results. In short sentences, there was no statistically significant difference between statements and unmarked questions, but the difference between marked questions and statements was significant in the first syllable, in the syllable *dian3*, and in the final syllable, which were all shorter in marked questions than in statements. The duration difference between the two kinds of questions was significant at the syllable *tian1*, where the unmarked questions was shorter, and in the last syllable, where it was longer. In long sentences, only the duration differences of the final syllables were significant. The final syllables of unmarked questions were significantly longer than those of statements, while that of marked questions was significantly shorter.
Summary There were three predictions for the durations of question syllables, (a) that the pre-final syllables are shortened compared to those of statements, (b) that the final syllable is lengthened in syntactically unmarked sentences, and (c) that the final one is shortened in ma questions. The data only support the last claim.
Figure 14: Differences in duration of sentences without focus
Intensity  Similar to the formula used for duration, the intensity values were normalized using a z-score transformation:

\[
\text{normalized value} = \frac{\text{absolute value} - \text{mean}}{\text{sd}}
\]

The main difference to the formula for duration lies in the fact that the mean and standard deviation here were calculated over all syllables of unfocused statements of a speaker, because the mean intensity value of a speaker can only be reliably calculated with a greater amount of data.\(^\text{11}\)

Intensity is expected to be generally higher in questions than in statements, and to rise towards the end. Although it is the case that the intensity is slightly higher in questions than in statements on almost all syllables, and although there is a rise, it is usually less than one standard deviation higher. Only in the last syllable of *ma* questions is the difference more than one standard deviation. For short sentences, it is 1.12 sds higher (see figure 15(e)), and for long sentences, 1.14 sds (figure 16(e)). However, the results of a repeated-measures ANOVA on the intensity values do show some significant changes. In short sentences, the results seem to indicate that questions are produced with higher intensity than statements during most of the sentence. For unmarked questions, there was significant raising, at least at the 10% level, starting from the second syllable, excluding the fifth syllable *tian1*. For marked questions, the difference starts at the fourth syllable *mei3*, also subsides at *tian1*, and after that is more significant than in unmarked sentences. It is highest in the final word. In long sentences, unmarked questions are not pronounced with higher intensity until the last syllable. In marked questions, however, throughout the sentence there are syllables spoken with significantly higher intensity. Although intensity is not higher in all points in questions, there seem to be at least some points during the course of the sentence where the difference is significant. Furthermore, the intensity of *ma* particle questions seems to get higher in the last or last two syllables.

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>***</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>-</td>
<td>**</td>
<td>*</td>
<td>*** ***</td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 10: Intensity: Significance level of adjusted p-values, short sentences

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>-</td>
<td>**</td>
<td>*</td>
<td>-</td>
<td>**</td>
<td>-</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Intensity: Significance level of adjusted p-values, long sentences

\(^{11}\)Gina-Anne Levow, personal communication.
Summary  The claim that intensity values of questions are higher throughout the sentence could only be partly confirmed. In short sentences, significantly higher intensity values appear in some syllables, but the only consistently significant difference is that of the final syllables. Both unmarked and marked questions show significantly higher values for the final syllable, in both long and short sentences.
Figure 15: Differences in intensity of short sentences without focus
Figure 16: Differences in intensity of long sentences without focus

(a) Visual comparison: statements and unmarked questions

(b) Visual comparison: statements and marked questions

(c) Visual comparison: unmarked and marked questions

(d) Difference: statements and unmarked questions

(e) Difference: statements and marked questions

(f) Difference: unmarked and marked questions
Tones  The realizations of the tones in natural speech are highly variable (Chao, 1933, p. 125), (Kratochvil, 1968, p. 36), (Yip, 2002, p. 21), especially that of tone 3 (Gandour, 1978, p. 55). These realizations depend on several factors, including for example segmental effects of vowels and consonants, stress and rhythm, focus, topic, downtrends and speaker-specific features. Tones are also influenced by their surrounding tones. There may be carryover variation from the preceding tone in the first part of the syllable, and, although less noticeable, (dissimilatory) anticipatory variation at the end of a syllable (Xu, 2001, p. 11f), (Chao, 1933, p. 128). Yip notes that the hearers compensate for these interactions between the tones: they do not hear them (Yip, 2002, p. 10). In the preceding paragraphs of this section, the segmental and tonal effects of the test words were ignored. In this section each final syllable will be investigated in turn, to see what effect question intonation has on the f0 values of the syllable. The syllables are grouped according to their own tonal values and segmental characteristics, and according to their preceding tones as well as the length of the sentence.  

- It is expected that in questions:
  - tone 1 will rise slightly and/or be raised
  - tone 2 will rise more steeply and be raised
  - tone 3 will have a rise, if it has one too in statements, that of questions will be steeper. It will not be raised.
  - tone 4 will fall flatter or be unchanged and/or be raised
  - the duration of all tones will be longer

In the first column of figure 18 (i.e. figures 18(a), 18(d), 18(g) and 18(j)) the behavior of the high tone is shown. One can see that it is raised and level after a high tone (figures 18(a), 18(d)), and raised and slightly rising after the low tone (figures 18(g), 18(j)). In contrast, the values of final tone 1 seem to fall slightly near the end of the syllable. In ma questions, a following ma particle shows a falling tendency before becoming level (see figures 19(a), 19(d), 19(g), 19(j)). In figures 18(b), 18(e), 18(h) and 18(k), the behavior of the rising tone is shown. After a high tone, the first part of the syllable is used to reach the lower point from which the rise should start. In ma questions, the syllable ends here. Statements and unmarked questions rise, but contrary to expectation, after high tones the rising tone of questions does not rise higher or faster than that of statements. After low tones, the rising tone of the ma question only rises slightly, then stops, and is continued on the ma syllable. Here the findings are as expected, with rising tones of unmarked questions rising faster and/or higher than those of the corresponding statements. In figures 19(b), 19(e), 19(h) and 19(k) it is shown that the rise is fulfilled on the initial part of the ma syllable, the syllable rhyme is high, then rising or slightly rising throughout. For low tones, due to tone three sandhi, only the tones following a high tone could be used for the analysis. Here, as in the rising tone syllables, the first part of the syllable was used to fall to the low target (see figures 18(m) and 18(n)). Only unmarked questions, which reached this target earlier, showed a subsequent rise. In ma questions, the rise carries over to the ma syllable (see figures 19(m), 19(n)). As predicted, falling tones, in figures 18(c), 18(f), 18(i) and 18(l) seem to fall steeper and faster in statements than in the question sentences. After high tones, the tones of question sentences are raised above those of the statement.

Note that for the tone combination 13 (HL), there were two undefined final syllables both in short and in long sentences.
sentences and show a slight fall, while after low tones, all three tone curves must first reach the high point from which the fall must start. In *ma* questions, the *ma* syllable continues the fall (see figures 19(c), 19(f), 19(i), 19(l)). The durations can also be seen in the figures, they are however calculated from the absolute values, not the normalized ones. The most striking feature is the shortening of the final syllable in *ma* questions compared to those of statements or unmarked questions.

In a repeated measures ANOVA with speaker as the replication factor, the interaction of type and tone was studied. It was found to be significant for the last three syllables in short sentences (with \(p<0.05\), \(p<0.001\), and \(p<0.01\), respectively), and for the last two in long sentences (\(p<0.001\) in both cases). The interaction plots for these syllables is shown in figure 17. Due to limited time and space, the individual interactions will not be discussed here, but the high variability of tone due to sentence type shown in this section indicates that tone is one feature that may help hearers distinguish between different sentence types. It will be further discussed in chapter 4.

**Summary** The predicted changes on the tone of the last syllable due to question intonation were found in the current study. The first tone is raised in questions, a following *ma* falls. The second tone rises faster in unmarked questions, at least after low tone. In *ma* questions the rise is partly or fully realized on the *ma* particle. The third tone is low in statements and in *ma* questions, and rising in unmarked questions. Following *ma* particles are rising. The fourth tone is raised and falls less steeply in both kinds of questions, a following *ma* continues the fall. Unmarked questions therefore show differences in all four tones, while the final syllables of *ma* questions only differ from statements in the first and fourth tone, and by not displaying a high rise in second tones. They are however shortened for tones one, two, and three.
Figure 17: Interaction plots: tone - sentence type for the last syllables

(a) Syllable dian3 in short sentences
(b) The first test word syllable in short sentences

(c) The second test word syllable in short sentences
(d) The first test word syllable in long sentences

(e) The second test word syllable in long sentences
Figure 18: Differences in sentence-final tones of sentences without focus
Figure 19: Last syllables of ma questions with the ma particle
Summary  In summarizing the results so far we can say that there is a prosodic difference between questions and statements; Hypothesis 1 is thus confirmed. An investigation of the mean height of the F0 curve showed that (a) the pitch contours of the questions stay level or rise slightly while those of statements decline; (b) there is global raising of pitch in short marked questions; (c) in long unmarked questions, where there is no global raising, there is a local rise in the last syllable. The pitch range, however, was not found to be changed in questions. Duration measurements showed no consistent changes except for the final syllable of the test word in ma questions, which was shortened compared to both its counterparts. In investigating intensity, some syllables were found to be louder in questions than in corresponding statements, but the changes were only systematic for the last syllable. There were significant interactions between sentence type and tone in the last syllables similar to what was found in previous studies, but it is as yet not certain whether they are as important to question recognition as postulated by Yuan (2004). There is therefore still no convincing answer to how questions are marked in Mandarin Chinese.

As for the differences between the two kinds of questions (research question 2), the results are even less precise. Contrary to what was predicted by the Functional Hypothesis, ma particle questions are raised slightly higher than unmarked questions, but this difference is not significant for all syllables. No significant pitch range difference was found. The duration of the last syllable may be more helpful to distinguish ma particle questions from unmarked questions, the difference in duration is larger than that between ma questions and statements. The intensity difference is insignificant. The tones of ma particle questions are realized differently, often part of the tone of the last syllable is fulfilled on the particle. Due to the difference in pitch, duration, and intonation, Hypothesis 2, that the two kinds of questions differ prosodically, can be verified. The Functional Hypothesis, however, does not seem to hold for Mandarin Chinese. Ma particle questions seem to be marked earlier in the sentence by raising of f0, and in the cases where both are significantly raised, the ma particle question is raised higher. It remains to be seen, however, whether these differences in raising are sufficient to distinguish the two kinds of questions, or whether the hearer needs the sentence-final cues of tone changes and final shortening or lengthening.

3.4.3 At what point does the distinction start?

Summary of previous findings  In the previous chapter, evidence for initial, medial or final differences was found: Short marked questions were found to be globally raised, while long marked questions became significantly higher after chao1. The contour of short unmarked questions showed significant differences after mai3 in short sentences, but in long unmarked sentences, only the rise on the final syllable was significant. The differences in duration and intensity manifested themselves mainly on the final syllable of ma questions. In long sentences, where there is a significant difference between ma particle questions and syntactically unmarked questions, it starts initially, but is not consistent throughout the sentence. In the final word, changes due to tonal combinations may be perceivable, while in the final syllable, duration may be a relevant feature. Before this is tested in a perception experiment, it is necessary to test another theory on the location where questions and statements start to differ, namely the theory of Liu and Xu, that the placement of this point is dependent on the placement of focus.
Focus
This section is devoted to testing Liu and Xu’s hypothesis that statements and questions start to differ from the point of the focus. The opposing hypothesis would be that both focus intonation and question/statement intonation modify the sentence in the way they usually do, leading to a possibly conflicting overlay. In this case, question and statement intonation may differ before the focus in the way they would if no focus were present, but be compressed after the focus.

For this study, the context was controlled to yield a contrastive focus on the first and on the last word for short sentences, and additionally on the eighth syllable hang2 in long sentences. The exact place of focus in the disyllabic test words was not controlled. This way, the speaker was given the choice at what point in the last word to articulate the focus, to adapt it to the constraints of question intonation and stress pattern.

![Table 12: Meanfo: Significance level of adjusted p-values, short sentences, initial focus](image)

![Table 13: Meanfo: Significance level of adjusted p-values, long sentences, initial focus](image)

Initial focus was placed on the first syllable ta.13 It was mainly realized by highly significant elongation of the first syllable (p<0.001), and by moderate lowering of the postfocal pitch height, which was significant on the syllable mai3 in short sentences, and on huan1, zal and shi4 in long sentences (all four: p<0.05). The claim made by Liu and Xu that sentences with initial focus start to differ from the point of focus is not supported by the data at hand. In contrast to their counterparts with neutral focus, short unmarked questions with initial focus only show a significant raising above statement level in the last syllable. Marked questions are slightly raised, but the difference is smaller than in neutrally focused sentences; as table 10 shows, this raising is not significant until the syllable mai3. In the last syllable, the difference between questions and statements gets larger, a fact that is not due to a rise in the questions, but due to a substantial lowering in the statement sentences. The ma question is again higher than the unmarked questions in all positions, and the difference gets significant earlier in the sentence. In long sentences, similar to long unmarked questions without focus,

---

13One final low tone value was undefinable in short sentences due to creaky voice, the mean values for the final syllable were computed based on the remaining values.
the difference between questions and statements starts to get significant at the syllable shi4, where the contour of the question sentences stays level, while the statement sentences decline. In contrast to unfocused questions, however, the unmarked questions are slightly, but insignificantly, higher than ma questions in most positions, and higher than statements in all positions. Ma particle questions are not globally raised above statement level. In the last syllable, however, marked questions are higher than unmarked questions, due to a slight fall in the unmarked questions. The later marking of questions with initial focus may be due to post-focal lowering of f0.
Figure 20: Differences in mean f0 of short sentences with initial focus
Figure 21: Differences in mean f0 of long sentences with initial focus.
The data for *medial focus* are different than expected. Although the context was constructed in a way to enforce a narrow focus placement on *hang*₂, and although the syllable was marked and the speakers were asked to emphasize it, it seems as though this kind of focus placement disagreed with the prosodic structure of the sentence, and prominence on the whole word with greater emphasis on *li*₄ was chosen instead. Both syllables were strongly marked in long sentences, using pitch range expansion, longer duration (p<0.001) and higher intensity (p<0.001), but the pitch range expansion was more significant for *li*₄ (p<0.001) than for *hang*₂ (p<0.01), and it was additionally marked with significantly higher f₀ (p<0.01). If Shih is right that emphasized syllables are initial in a foot (Shih, 1997, p. 112), this kind of shift would be predicted, because *li*₄*hang*₂ forms a foot, and even though it is a nonsense word, it is presented to be a word, one sense-unit, namely the name of a supermarket.

Liu and Xu’s analysis would predict for these sentences that the difference between questions and statements would begin after the focused element. As can be seen for unmarked questions in figure 22, however, the pitch values of prefocal syllables that do not have tone 3 are also raised. Table 12 shows that these differences are significant. The data therefore disconfirms Liu and Xu’s hypothesis. Instead, the raising of the f₀ curves of questions above statement level starts earlier than in unfocused sentences, and that of unmarked questions is raised more consistently than that of marked questions. However, unmarked questions with final focus show a rise on the last syllable, where marked questions show a fall. The pitch contour of both kinds of questions is again very similar, showing no significant differences.

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>**</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>*</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>***</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 14: Meanf₀: Significance level of adjusted p-values, long sentences, medial focus

---

14Two values of the same sentence, one of *xi*₃ and one of the final syllable *caos*₃, were undefinable in long sentences due to creaky voice, the mean values were computed based on the remaining values.
Figure 22: Differences in mean f0 of long sentences with medial focus
**Final focus** was placed on the test word.\textsuperscript{15} The test words in this experiment build a disyllabic foot. The initial syllable of the test word receives the focus, due to the fact that the initial element of a foot is usually stressed in Mandarin Chinese (Yip and Rimmington, 1997, p. 281). Depending on the dialect of the speakers, the last syllable may therefore be unstressed and lose its inherent tone. According to Duanmu, this is the case in Beijing Mandarin and Standard Chinese (Duanmu, 2000, p. 232). As seen in the section above about tone, though, this does not seem to be the case in the sentences tested.

A visual comparison with sentences with neutral focus shows that in both cases, questions are raised slightly above statement level, but while the difference becomes larger after tian1 in sentences without focus, the difference stays small until the last syllable, where questions are level or slightly rising, and statements drop. This is due to the fact that the pitch contour of statements with final focus does not decline as much as in neutrally focused sentences, but stays parallel to the curve of questions until the first syllable of the test word, then falls. The focus seems to lie on the first syllable of the test word. In long sentences, the situation is different. While both kinds of questions are raised above statement level in most syllables, the difference stays small until the last syllable. In unmarked questions, the difference gradually gets larger towards the end, but gets smaller again in the focused syllable. In both long and short sentences, the differences between marked and unmarked questions are very small.

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td><strong>-</strong></td>
<td>-</td>
<td><strong>-</strong></td>
<td><strong>-</strong></td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>St. - M.Q</td>
<td><strong>-</strong></td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>-</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Meanf0: Significance level of adjusted p-values, short sentences, final focus

<table>
<thead>
<tr>
<th>Syllables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. - Unm.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>***</td>
<td>*</td>
<td>-</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>St. - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Unm.Q - M.Q</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 16: Meanf0: Significance level of adjusted p-values, long sentences, final focus

Looking at the significance tables 13 and 14, we can see that the raising over statement level in short questions, even though it is slight, is significant in the initial high tone syllables ta1, huan1, and tian1, even though it is only marginally significant on huan1 for marked questions (p<0.1). The difference at the syllable mai3 and in the final syllable are also significant for both kinds of questions, and for marked questions, there is an additional significant difference at dian3. In long sentences, there is also

\textsuperscript{15}One final low tone value was undefinable in short sentences due to creaky voice, the mean values for the final syllable were computed based on the remaining values.
a pitch raising, but it occurs later in the sentence, starting with *chaol*. In all kinds of contrasts, the mean f0 value on the focused second-to-last syllable is not significantly different.
Figure 23: Differences in mean f0 of short sentences with final focus
Figure 24: Differences in mean f0 of long sentences with final focus
Summary

The aim of this section was to answer research question 4, namely to find the point where the prosodic realizations of questions and statements start to differ. The places found in the literature were:

- sentence-initially
- after the predicate
- during the last two syllables, independently of whether they form a lexical unit or not
- after the focused constituent

The possible places that were found to be statistically significant were the following. The level pitch was found to start with or after the last high tone, after tiān1 in the short sentences, and with chǎo1 in the long sentences.

- sentence-initially
- after the last syllable with high tone preceding the test word
- during the last syllable or the last syllables

Focus seems to have an influence on the method of question formation, but not in the way described by Liu and Xu. The opposing hypothesis, namely that the influence of focus and sentence type merely combine, was also not confirmed for all cases. In medial focus, global raising for questions was found even in long sentences, preceding the focused item.

The possible places where the realization of the two kinds of questions may start to differ are:

- during the last syllables, due to interaction with tone
- during the last syllable, due to differences in duration

In the next chapter, it will be seen which of these possibilities is relevant.

3.5 Summary

In the current section, we saw that the two hypothesis on Mandarin Chinese question intonation could be confirmed:

- Hypothesis 1: Ma-particle questions and syntactically unmarked questions differ prosodically from statements.
- Hypothesis 2: Ma-particle questions and syntactically unmarked questions differ prosodically from each other. The difference is not carried mainly by the ma particle.

Research question 1, on the relevant features differentiating questions and statements, was partly answered. For our data, there were the following features:

- Global raising of the pitch contour
- Gradually rising or level pitch contour
• Local rise in the last syllable
• Shortening or lengthening of the final syllables, depending on question type
• Higher intensity values
• Tonal changes in the last syllable

The most consistent of these differentiating features is the level pitch in questions, starting after tian1 in short questions, and after chaol or sometimes li4 in long sentences. In short questions, global raising of at least the high tone syllables seems very common, too. It looks as though there is a high tone target at the end of the sentence, which may be reached by different methods. In short sentences, it may lead to higher f0 values at an early point in the sentence, while in longer sentences, the raising can be postponed to a later point in the sentence. This may interact with focus intonation, for example in cases where the pitch is lowered at the point where the level pitch usually starts, due to medial focus, prefocal raising of some high f0 values may be a more appropriate strategy. In these sentences, level pitch begins from the point of focus. In sentences with initial focus, the focus prevented the global raising in short sentences, but the level pitch was distant enough to be realized as usual, at least in long sentences. In short sentences it was realized, but the difference only became significant in the last syllable in unmarked questions, due to the fact that both question and statement curves are lowered after the focus. In cases where the focus is final, there was a similar problem. Both question and statement curves were raised finally due to focus, and the differences were not significant in all cases. In short questions, global raising was used as a second relevant cue to indicate questionhood. In long sentences, there was only level pitch. However, the second-to-last syllable was in focus, so the last syllable of statements showed a considerable drop, while the last syllable of questions was level or slightly rising. It can be concluded that level pitch is the preferred method of marking questions, and rises or global raising are the solutions chosen when level pitch would not mark the question sufficiently. Additionally, there are significant sentence-final distinguishing features, namely tone, duration, and intensity.

The differentiation of the two kinds of questions was also not consistent. In sentences without focus, the marked questions were marked earlier in the sentence, while unmarked questions were more strongly marked in the last syllable. However, the difference in mean f0 was not significant in any cases except for long, neutrally focused questions. The only significant differences between questions that were found were highly significant differences in duration on the last syllable, and differences due to interaction with tone. Due to the fact that in some tone combinations, the final part of the final tone in marked questions is realized on the particle, which changes the tone contour substantially, and due to the fact that there is no final lengthening in these questions, the f0 contour and duration of the final syllables may be the most important cue to distinguish the two types of questions. In the next chapter, it will be investigated whether hearers can identify the different kinds of questions without these cues. The Functional Hypothesis could not be confirmed; the prosodical marking of syntactically unmarked questions does not seem to be stronger than that of syntactically marked questions.
In the current chapter, multiple different ways of marking questions were found. Therefore, it is not yet possible to say at what point the difference between the three sentence types becomes perceivable. However, Liu and Xu’s hypothesis that it is after the point of focus could not be confirmed for our data. The hypotheses raised by the discussion above are the following:

- In most questions, there is level f0 starting with a late peak
- In short sentences or sentences with medial focus, there is raising in the initial part of the sentence
- In longer sentences, a final rise may occur
- During the last syllables: tone-specific features
- During the last syllable: durational features

This means that in addition to the features mentioned in the literature, it was found here that longer sentences yield different results. In the following chapter, longer sentences are investigated in a perception experiment, to see whether the distinguishing features of the last two syllables are necessary to correctly identify the three sentence types.
4 Perception

When differences are produced, even when they are significant, it does not necessarily mean that they will also be perceived as differences. For this reason, a perception experiment was prepared to test whether the recognition of Mandarin Y/N question intonation is achieved locally or globally, and, if locally, at which point in the last word it becomes audible. In the previous section, the following results were found for long questions in Mandarin Chinese:

- Level pitch contour starting with chao1
- Additional local rise in the last syllable in some more complicated cases
- Global raising with medial focus
- Shortening or lengthening of the final syllables, depending on question type
- Higher intensity values
- Tonal changes in the last two syllables

The experiment thus tests (a) whether these differences are heard, and (b) whether the abundance of features on the final syllable or final word (rise, durational changes, intensity changes, tonal changes) are necessary or whether the level pitch contour or raising is significant enough so that hearers perceive the difference.

4.1 Procedure

Using the WebExp2 toolbox (WebExp2 version 1.0 alpha, www.webexp.info ), a computer experiment was constructed. It was conducted mainly at the philological library of the Freie Universität Berlin, and took about twenty minutes to complete. It was originally planned to be an online experiment, but a controlled experiment on the experimenter’s computer was found to be a better solution. This was due to the fact that some of the subjects reported technological difficulties either with the java applet or the display of Chinese characters, or, more importantly, difficulties in understanding the experiment instructions (some reported that ‘the sound file was broken’, because they didn’t hear the last word). The experiment was completed by twelve speakers of Mandarin Chinese between the age of 20 and 31, of which three were speakers from other regions of China that had learned Mandarin at a very young age. The subjects were a very homogenous group, most of them being female (nine), right-handed (eleven), university students (nine) and residents of Berlin or Potsdam (ten). Of those that were not university students, one was still in school, one worked as engineer, and one did not supply any information. The two speakers that did not live in or near Berlin at the time of the experiment lived in China and were in Berlin for a visit.

4.2 Materials

Due to limited space and time, only the long sentences of one speaker could be tested in the perception experiment. From a small random selection of sentences from the corpus of the production experiment, an informant chose those of speaker 2 to be the most natural. The ma questions were prepared so that the ma was not audible, to ensure that all three kinds of sentences were syntactically the same. Then varying parts of the
disyllabic test word were removed based on Grosjean’s *gating* procedure (Grosjean, 1983). The word was cut off either fully (0%), or after the first syllable (50%), or left intact (100%).

ta1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang1jiao1? (100%)  
ta1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang1...? (50%)  
ta1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 ...? (0%)

If this is applied to all long sentences of speaker 2, this yields 252 sentences for the perception experiment (7 test word tone combinations x 4 focus positions x 3 sentence types x 3 gating conditions).

### 4.3 Data Preparation

WebExp2 presented all 252 stimuli in randomized order to each subject. The subjects were asked to identify these sentences according to their sentence type. Each stimuli could be heard only once, and the three possibly corresponding full sentences were displayed as clickable buttons at the bottom of a blank page (see figure 25).

![Figure 25: Screenshot of the WebExp2 perception experiment](image)

The results can be seen in Appendix D on the data CD.
4.4 Analysis

4.4.1 Is there a distinct question intonation?

This section is devoted to testing whether the subjects identify the sentences correctly, and which factors may have an influence on this identification process. The null hypotheses are that the factor sentence type and any combination of sentence type with tone or focus do not affect sentence type recognition.

Correct identification of sentence types: 100% Gating As can be seen in table 17, sentences in which the last word remained intact were perceived correctly in 86% of the cases. Statements were identified most accurately, then unmarked questions, then marked questions. However, questions were recognized as questions in 90% of the cases, although marked questions were often heard as unmarked questions. The relatively low recognition rate of marked questions probably reflects the fact that compared with the other sentences in this category, marked questions are already truncated at this stage: The ma particle is removed. Nevertheless, these results show that hearers are capable of differentiating ma particle questions from the two other sentence types even though the particle is missing. In a three-way analysis of variance, the accuracy of the identification was tested. The factors were sentence type, tone combination of the last syllable, and focus, and the independent variable was the mean accuracy of identification of each stimulus. Because the mean accuracy is a percentage, the homogeneity of variance assumption is necessarily violated. The ANOVA procedure is relatively robust concerning these kinds of violations, especially if the numbers of observations are the same for each condition, as is the case here, but nevertheless it seemed advisable to transform the data using an arcsine transformation. The following formula is from Winer (1962, p.220).

\[ \text{newvalue} = 2 \times \text{arcsine}(\sqrt{\text{oldvalue}}) \]

The resulting values have less heterogeneous variances. All values in this section have undergone this transformation for statistical calculation, the values below are however given in percentage values. The ANOVA in this section tested the main effects of sentence type and the interaction effects of sentence type-focus position and sentence type-tone combination on the correct identification of the sentences with full last word. For each effect or interaction, the null hypothesis is assumed that they do not influence the correct identification of the sentence. The main effect of the sentence type was highly significant (p<0.001). The interactions with tone or focus were not significant on the 5% level, therefore null hypotheses concerning the influence of the interaction of focus or tone with sentence type could not be rejected. A Tukey Honestly Significant Difference procedure revealed that, unsurprisingly, the differences found between marked questions and the other two kinds of sentences are highly significant (p<0.001), while the difference found between statements and unmarked questions was not significant. See table 18.

These results indicate that the differences between questions and statements are perceivable. The difference between the two kinds of Y/N questions was more difficult to perceive. Ma-questions were identified significantly less accurately than both unmarked questions and statements. They were often wrongly classified as unmarked questions. In the next section, we investigated which factors contribute to this wrong classification.
Table 17: Accuracy of perception of sentences with full last word

<table>
<thead>
<tr>
<th>Statement type</th>
<th>Statement</th>
<th>Unmarked question</th>
<th>Marked question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heard as statement</td>
<td>96%</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td>Heard as unmarked question</td>
<td>4%</td>
<td>89%</td>
<td>19%</td>
</tr>
<tr>
<td>Heard as marked question</td>
<td>0%</td>
<td>0%</td>
<td>72%</td>
</tr>
</tbody>
</table>

Table 18: Significant differences between the effects of the levels of Type on sentence identification

<table>
<thead>
<tr>
<th>Comparison</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement - unmarked question</td>
<td>0.3114206</td>
</tr>
<tr>
<td>Statement - marked question</td>
<td></td>
</tr>
<tr>
<td>Unmarked - marked question</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

Incorrect identification of sentence types: **100% Gating**

In Table 17 we saw that incorrect identification of questions seems to follow an asymmetric pattern. While 4% of statements and 11% of unmarked questions were wrongly classified as belonging to the respective other group, neither statements nor unmarked questions were heard as marked questions. Marked questions, however, where identified incorrectly as unmarked questions in 19% of the cases, and as statements in 9% of the cases. To test the effect of the factors mentioned above on incorrect identification, three analyses of variance were performed, one for each level of the factor Type. For example, all items not belonging to the category Statement were grouped together. An ANOVA was conducted on this group, with the same factors as before, but with the independent variable Heard as statement instead of Correct answer. The null hypotheses are, accordingly, that the different factors and the interactions do not influence the wrong classification of a question as statement. On the 5% level, the null hypotheses could not be rejected. For the next category, the ANOVA was not necessary, since there were no items wrongly classified as marked questions. For sentences falsely heard as unmarked question, however, the main effect of sentence type was highly significant (p < 0.001), and the interaction between type and tone was significant, too (p < 0.05). As seen above, 19% of marked questions were heard as unmarked questions, in contrast to only 4% of statements, so it is not very surprising that type was significant here. Table 19 is a table of the type-tone interactions together with their mean Heard as unmarked question value. The stimuli that have a value of 0.00, i.e. those that were never heard as unmarked question are highly significantly different (p < 0.01) to those which have a value of 0.21 and above, and significantly different (p < 0.05) to those which have a value of 0.17 and above. The interactions with value 0.02 are significantly different (p < 0.05) to those which have a value of 0.21 and above. All other differences are not significant. What is interesting here is that marked questions with rising tone seem to be more unlikely to be classified as unmarked questions, while statements with a high-rising tone combination are more likely to be mistaken for an unmarked question.

Thus, the results indicate that the fact that marked questions are categorized as unmarked questions much more often than statements is statistically significant, and that there are certain tone combinations which were found to co-occur with the wrongly classified marked questions significantly more often than other tone combinations,
4.4.2 At what point does the distinction start?

In the previous section, it was found that questions and statements could be differentiated when the last word was intact, but that marked Y/N questions were identified significantly worse, especially when the last word did not end with a rising tone. In this section, the parts of the last word of the sentence were removed to test whether the difference between the different sentence types was still perceivable.

**Correct identification of sentence types: 50% Gating**  
Sentences in which only the first syllable of the test word remains were perceived correctly in 53% of the cases, which might be due to chance. There is however a difference between statements, which were identified correctly in 89% of the cases, and questions, which were heard as questions 62.5% of the time. The accuracy of perception of the correct question type was even worse. Unmarked questions were identified correctly in 51% of the cases, and marked questions in 20% (see table 20).

<table>
<thead>
<tr>
<th>sentence type</th>
<th>statement</th>
<th>unmarked question</th>
<th>marked question</th>
</tr>
</thead>
<tbody>
<tr>
<td>heard as statement</td>
<td>89%</td>
<td>35%</td>
<td>40%</td>
</tr>
<tr>
<td>heard as unmarked question</td>
<td>7%</td>
<td>51%</td>
<td>40%</td>
</tr>
<tr>
<td>heard as marked question</td>
<td>4%</td>
<td>14%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Table 20: Accuracy of perception of sentences with 50% of the last word

For correct identification of these sentences, the main effect of sentence type was highly significant (p<0.001), and the effect of the interaction between type and focus was also significant (p<0.01). The Tukey comparison showed that *p* was significant for all differences between sentence types, and therefore the null hypothesis that there is no sentence type effect can be rejected. This means that the difference between the correct identification of unmarked questions and statements becomes significant when the last syllable is missing. The identification of both kinds of questions becomes considerably worse. Table 21 shows the interaction between sentence type and focus. While statements are least easily recognized when occurring with final focus, probably because the declination is weaker in these sentences, the two types of questions are most easily recognized with medial focus, which was shown in the last section to have

<table>
<thead>
<tr>
<th>tone</th>
<th>sentence type</th>
<th>statements</th>
<th>marked questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>2%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>15%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>17%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0%</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>0%</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>2%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>6%</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

Table 19: **Heard as UMQ** - value of type-tone interactions

namely all tone combinations with non-rising tone.
global raising, or final focus, in which the final f0 values are raised, too. Unmarked questions seem to associate more with medial focus, and marked questions with final focus. The difference between the recognition of unmarked questions with medial focus and unmarked questions without or with initial focus is significant (p<0.05 and p<0.01, respectively). The other differences within sentence types are not significant, however. The trend that statements are identified with the most ease, then unmarked questions, then marked questions, is continued here.

<table>
<thead>
<tr>
<th>Focus</th>
<th>Statement</th>
<th>Unmarked Question</th>
<th>Marked Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>No focus</td>
<td>96%</td>
<td>38%</td>
<td>20%</td>
</tr>
<tr>
<td>Initial focus</td>
<td>91%</td>
<td>33%</td>
<td>14%</td>
</tr>
<tr>
<td>Medial focus</td>
<td>91%</td>
<td>75%</td>
<td>21%</td>
</tr>
<tr>
<td>Final focus</td>
<td>77%</td>
<td>56%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 21: Accuracy of perception: type and focus interaction, 50% Gating

The results of this section indicate that the differentiation between questions and statements is very difficult when the last syllable of the sentence is missing. Questions with medial or final focus are more easily recognized, while questions with initial focus are most difficult to identify. Unmarked questions with medial focus and marked questions with final focus are identified most easily. However, it seems as though the differentiation between the two questions is not possible when the last syllable is missing.

Incorrect identification of sentence types: 50% Gating Questions in which only half of the last word remains are often not identified correctly as either unmarked or marked question, but they are still identified as questions in more than 60% of the cases. This section describes the effect of sentence type and its interaction with focus or tone on the wrong characterization of the stimuli sentences. This might help identify features which distinguish the three sentence types from each other.

In sentences with 50% gating, the sentence type was not significant for wrong characterization of questions as statements, but the interaction between sentence type and tone (p<0.01) and sentence type and focus (p<0.05) were significant. Table 22 shows the interaction between type and tone for incorrect identification of questions as statements. In these sentences, only the first tone of the tone combination is heard, i.e. a high tone or a low tone. Nevertheless, the Tukey HSD procedure revealed a significant difference due to tone-type interaction, namely that between unmarked questions with low-high tone (tone ‘31’), which were heard as statements very often, and marked questions with high-falling tone (tone ‘14’), which were seldom heard as statements. This difference was significant on the 5% level. This might be due to the fact that in low-high tone sequences in long sentences, the low tone is almost as low in unmarked questions as in statements, while the f0 height of the first tone of a high-falling sequence is much higher in long marked questions than in corresponding statements (see figure 17(d) in the last chapter). All other differences are not significant.

Unmarked questions with no or initial focus were most often identified as statements, while for marked questions, it was initial focus. This is expected due to the fact that initial focus lowers the following sentence, and long unmarked questions were found in the previous chapter that were not significantly different from statements, until
the last syllable. The only differences which were found to be significant were that between unmarked questions with medial focus and marked or unmarked questions with initial focus (both p<0.05). Again, questions with initial focus seem to be identified as statements more often, due to the fact that their pitch values are lowered.

Table 22: *Heard as ST* - value of type-tone interactions, 50% Gating

<table>
<thead>
<tr>
<th>tone</th>
<th>sentence type</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unmarked questions</td>
<td>marked questions</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>42%</td>
<td>58%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>15%</td>
<td>58%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>48%</td>
<td>15%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>23%</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>58%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>21%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>40%</td>
<td>46%</td>
<td></td>
</tr>
</tbody>
</table>

Table 23: *Heard as ST* - value of type-focus interactions, 50% Gating

For incorrect identification as marked question, however, sentence type was the only significant factor (p<0.001), with unmarked questions being more than twice as likely to be wrongly identified as marked questions. For incorrect identification as unmarked question, both the effect and the two interactions were significant, the effect of type on the 0.1% level, and that of the interactions on the 1% level. The high significance of the main effect was expected, due to the fact that forty percent of marked questions were heard as unmarked questions, and only seven percent of statements. When it comes to interaction between tone and type, the marked questions with a high-falling tone combination (‘14’) were identified as unmarked questions most frequently, while all kinds of statements and marked questions with rising (‘12’ or ‘32’) or low-falling (‘34’) tone combination were seldom heard as unmarked questions. This reflects the fact that statements were again significantly less frequently identified as unmarked questions. The fact that the tone combinations with rising tone were seldom heard as unmarked questions cannot be explained with pitch height differences. Marked questions with a high-low tone combination (‘13’), the second most frequent group which was wrongly classified as unmarked questions, were identified as unmarked questions significantly more often than all kinds of statements except those with a high-rising tone combination (‘12’). Statements with a low-high tone combination (‘31’), which were never identified as unmarked questions, were significantly different to all marked questions except those with high-rising tone.

All types of statements except those with final focus are highly significantly different to all types of marked questions except those with initial focus (p<0.001), meaning
that these kinds of statements were significantly more unlikely to be heard as unmarked question than the kinds of marked questions. This is expected due to the fact that statements with final focus show less declination than the other kinds of statements, while marked questions with initial focus show more. Nevertheless, statements with final focus were significantly less frequently identified as unmarked questions compared to marked questions without focus (p<0.05) or with medial focus (p<0.01). Marked questions with medial focus were classified as unmarked questions significantly more often than marked questions with initial focus (p<0.05), probably again due to global raising instead of a final rise.

<table>
<thead>
<tr>
<th>focus</th>
<th>statements</th>
<th>marked questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>2%</td>
<td>46%</td>
</tr>
<tr>
<td>init</td>
<td>5%</td>
<td>24%</td>
</tr>
<tr>
<td>med</td>
<td>7%</td>
<td>52%</td>
</tr>
<tr>
<td>fin</td>
<td>16%</td>
<td>37%</td>
</tr>
</tbody>
</table>

Table 25: *Heard as UMQ* - value of type-focus interactions, 50% Gating

This section indicates that some differences can be perceived due to focus position or tone combination, even differences between the two kinds of question. Differences between questions and statements are recognized due to pitch height of the first syllable of the test word, and due to interaction with focus. Medial focus, due to its global raising, helps distinguish questions from statements, while initial focus on questions make them more similar to statements. Statements with final focus are heard as questions more often. When looking at the differentiation of questions, marked questions with final rising tones -even if they are not heard here- are significantly less likely to be classified as unmarked questions, while marked questions with medial or no focus are more likely to be classified as unmarked questions. These findings indicate that both the f0 height or realization of the first test word syllable and the focus placement have an influence on the differentiation of questions and statements, and of marked and unmarked questions, if a difference is perceived.

**Correct Identification of sentence types: 0% Gating**

In a next step, the whole last word was removed. It was tested whether questions and statements were still distinguishable even though the distinguishing features on the last two syllables were miss-
Sentences in which the test word was completely removed were only perceived correctly 44% of the time. Statements were identified correctly in 85% of the cases, and questions in 45.5%. The two types of questions were identified correctly 34% and 14% of the time, respectively (see table 26). The behavior of statements is similar in all three gating categories. It is either the easiest sentence type to recognize, or the default value in difficult cases. The identification of all three sentence types gets less accurate the more of the word is missing.

<table>
<thead>
<tr>
<th></th>
<th>statement</th>
<th>unmarked question</th>
<th>marked question</th>
</tr>
</thead>
<tbody>
<tr>
<td>heard as statement</td>
<td>85%</td>
<td>54%</td>
<td>55%</td>
</tr>
<tr>
<td>heard as unmarked question</td>
<td>10%</td>
<td>34%</td>
<td>31%</td>
</tr>
<tr>
<td>heard as marked question</td>
<td>5%</td>
<td>12%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Table 26: Accuracy of perception of sentences with no last word

For correct identification of these sentences without last word, all differences between different sentence types were highly significant \(p < 0.01\), and the interaction between type and focus was significant on a 5% level. The usual order of identification accuracy was maintained, with statements being identified more easily than unmarked questions, and unmarked questions more easily than marked questions. Statements with initial focus were most easy to identify, followed by neutral and medially-focused ones. For unmarked questions, sentences with medial and final focus were more easy to identify correctly, while for marked questions, sentences with initial focus seemed to be most difficult. Again, there were significant differences between unmarked questions with different focus, namely between those with medial focus and those without or with initial focus \(p < 0.05\), but no significant differences within the other sentence type groups. See table 27.

<table>
<thead>
<tr>
<th></th>
<th>statement</th>
<th>unmarked question</th>
<th>marked question</th>
</tr>
</thead>
<tbody>
<tr>
<td>no focus</td>
<td>85%</td>
<td>18%</td>
<td>16%</td>
</tr>
<tr>
<td>initial focus</td>
<td>95%</td>
<td>19%</td>
<td>11%</td>
</tr>
<tr>
<td>medial focus</td>
<td>86%</td>
<td>60%</td>
<td>13%</td>
</tr>
<tr>
<td>final focus</td>
<td>75%</td>
<td>42%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Table 27: Accuracy of perception: type and focus interaction, 0% Gating

The results of this section show that when the whole word is missing, questions are not recognized as questions. Hearers seem to rely on the question marking features on the last two syllables of the sentence for question recognition. Sentences with medial focus were more easily to identify as unmarked questions due to global raising. In the following section, the effect of sentence type and its interaction with tone and focus on incorrect identification of sentences is tested.

**Incorrect Identification of sentence types: 0% Gating**  The main effect of type was again irrelevant for statements in which the last word was missing, but the interaction of type and focus was significant \(p < 0.01\). Unmarked questions were significantly more frequently classified as marked question than statements. The erroneous classification of syntactically unmarked questions with medial focus as statements was significantly
less frequent than that of marked questions with initial focus (p < 0.05), or of unmarked questions with initial focus (p < 0.01), or without focus (p < 0.05).

<table>
<thead>
<tr>
<th>focus</th>
<th>sentence type</th>
<th>unmarked questions</th>
<th>marked questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td></td>
<td>71%</td>
<td>54%</td>
</tr>
<tr>
<td>init</td>
<td></td>
<td>74%</td>
<td>69%</td>
</tr>
<tr>
<td>med</td>
<td></td>
<td>25%</td>
<td>39%</td>
</tr>
<tr>
<td>fin</td>
<td></td>
<td>44%</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 28: *Heard as ST* - value of type-focus interactions, 0% Gating

For identification as marked question, all three effects were significant. The main effect of sentence type was highly significant (p < 0.001), the interaction of sentence type and focus was significant on a 5% level, and that of sentence type and tone on a 1% level. In table 29, the interaction between focus and sentence type is shown. The group which was most often wrongly identified as marked question, namely unmarked questions with medial focus, showed significant differences to statements with initial or medial focus (p < 0.001, p < 0.05, respectively). The wrong identification of statements with initial focus as marked questions was also significantly different to that of unmarked questions with final focus (p < 0.01).

<table>
<thead>
<tr>
<th>focus</th>
<th>sentence type</th>
<th>statements</th>
<th>unmarked questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td></td>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td>init</td>
<td></td>
<td>1%</td>
<td>7%</td>
</tr>
<tr>
<td>med</td>
<td></td>
<td>4%</td>
<td>16%</td>
</tr>
<tr>
<td>fin</td>
<td></td>
<td>7%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Table 29: *Heard as MQ* - value of type-focus interactions, 0% Gating

When looking at the interaction of sentence type and tone in figure 30, we see that although the whole last word is missing, there are some differences in identification between the different tone combinations. Additionally, it is not the case that the first tone of the tone combination makes a significant difference, as would be expected if the differences in identification were due to changes in the syllable *dian*, for example through tone sandhi. Statements which were never heard as marked questions, namely those with high-low and low-high tones, were significantly different to unmarked questions with high-rising (‘12’, p < 0.05), high-falling (‘14’) and low-rising (‘32’) tone (both p < 0.01). Unmarked questions with low-high tone (‘31’) were heard as marked questions significantly less often than unmarked questions with high-falling (‘14’) or low-rising (‘32’) tone combination (both: p < 0.05). There was however no satisfactory explanation for these effects. In the realization of long sentences in chapter 3.4, the interaction between sentence type and tones was found to only be significant in the last two syllables. These syllables are missing in this condition, therefore the significant interaction effect found here between sentence type and tone is probably due to chance.

For those sentences falsely classified as unmarked questions, sentence type (p < 0.001) and sentence type-focus (p < 0.05) were the relevant factors. The false identification of marked questions with medial focus as unmarked questions was significantly more fre-
Table 30: *Heard as MQ* - value of type-tone interactions, 0% Gating

<table>
<thead>
<tr>
<th>tone</th>
<th>statements</th>
<th>unmarked questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>12</td>
<td>10%</td>
<td>15%</td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>10%</td>
</tr>
<tr>
<td>14</td>
<td>6%</td>
<td>19%</td>
</tr>
<tr>
<td>31</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>32</td>
<td>8%</td>
<td>21%</td>
</tr>
<tr>
<td>34</td>
<td>4%</td>
<td>8%</td>
</tr>
</tbody>
</table>

The identification of marked questions without focus as unmarked questions was additionally significantly more frequent than that of statements with initial focus.

Table 31: *Heard as UMQ* - value of type-focus interactions, 0% Gating

<table>
<thead>
<tr>
<th>focus</th>
<th>sentence type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>statements</td>
<td>marked questions</td>
</tr>
<tr>
<td>no</td>
<td>6%</td>
<td>31%</td>
</tr>
<tr>
<td>init</td>
<td>4%</td>
<td>20%</td>
</tr>
<tr>
<td>med</td>
<td>11%</td>
<td>48%</td>
</tr>
<tr>
<td>fin</td>
<td>18%</td>
<td>24%</td>
</tr>
</tbody>
</table>

This section showed that the global raising of medially focused questions had a significant effect on the recognition of sentence types even when the last word was missing. Unmarked questions with medial focus were seldom heard as statements, but often heard as marked questions. Marked questions with medial focus were often heard as unmarked questions. Hearers seem to recognize global raising at least in some cases, but level pitch, which is significant before the last word starts, is not heard at this point. Statements with final focus were often mistaken for unmarked questions, probably due to the fact that the f0 curves of these statements do not decline immediately.

### 4.5 Summary

This chapter investigated whether the differences found between unmarked and marked questions and statements in the previous chapter were actually perceivable by native speakers of Mandarin. The differences found for long sentences were the following:

- Level pitch contour starting with *chao1*
- Additional local rise in the last syllable in unmarked questions without focus
- Global raising with medial focus
- Shortening or lengthening of the final syllables, depending on question type
- Higher intensity values
Tonal changes in the last syllable

Since many differences were realized on the last two syllables, additional tests were conducted to see whether the different sentence types were distinguishable even when these syllables were missing. There were three conditions: stimuli with the whole last word, stimuli with the first syllable of the last word, and stimuli without the last word. In the first condition, both the difference between statements and questions and between the two kinds of questions were recognized. When the last word is removed, the recognition of statements does not get considerably worse, but the recognition of questions worsens at every step. For the second condition, the hearers did not perceive the difference between the two types of questions anymore, even though they perceived the difference between questions and statements. For the third condition, the difference between questions and statements was not perceived. There is nevertheless some interaction with focus. In the second and third condition, statements are recognized most accurately when not realized with final focus, while unmarked questions are recognized best with medial focus, and marked questions with final focus.

Mistake types

In a second step, it was analyzed whether sentence type, or the interactions of sentence type with tone combination of the test word or focus position have an influence on the type of mistake made in identification. Statements are more often identified wrongly as unmarked questions than as marked questions in all three cases. Although the identification of questions is at chance level and below in the cases where half of the last word was heard, for both question types around sixty percent of the question stimuli were actually identified as questions, even though the hearers often erred in the kind of question they chose. In contrast, in the cases where last word was not heard, more than fifty percent of questions were selected to be statements in both cases. Statement seemed to be the default value in unsure cases. The fact that both kinds of questions were heard as unmarked question considerably more often then as marked question is interesting, because, as noted above in chapter 2, marked questions are often though of as the most basic type of questions in Mandarin Chinese. In the sentences with full last word, the reason may be that the marked question is the only sentence which is already truncated, due to the fact that the ma particle is missing. In the sentences in which half or the whole of the last word is missing, however, there must be other reasons for the preference of the other sentence types. There may be a biasing property of the sentences chosen for the perception experiment, lying in the word dian, meaning ‘a bit of’, ‘some’. According to Beun, textual indicators are very important cues for distinguishing questions and statements in difficult cases in Dutch (Beun, 1990, p. 55). Although the verbs (for example say, think), pronouns (for example I or you), and particles (for example Oh, so, and) he mentioned as possibly biasing were avoided in the present study, it may be that the classifier dian is an assertive form, which is regularly used in statements, while neutral questions use nonassertive forms (compare for English: Quirk et al., 1985, p.808):

- Someone called last night.

- Did anyone call last night?

However, even if dian were a nonassertive form, these kinds of forms can be used in questions which are biased towards a possible answer. Compare (Quirk et al., 1985, p. 808):

- Did someone call last night?
Unmarked questions, like all declarative questions, are usually biased towards a possible answer (Geluykens, 1988, p. 478). Therefore it may be that they are the preferred question type due to the use of dian3 in the test sentences. Ma-particle questions can also show a speaker bias at least in some cases (Li and Thompson, 1981, p. 552), (Chao, 1968, p. 800), (Beutel, 1996, p. 168). Nevertheless, the possibility that the use of dian influenced the hearers to choose statements or unmarked questions should not go unmentioned.

Wrong classification as statement. It was found that for questions which were wrongly heard as statements, the interaction between sentence type and focus was of great importance for the conditions in which the last word was not complete. In these two conditions, unmarked questions with initial or neutral focus and marked questions with initial focus were identified as statements significantly more often than unmarked questions with medial focus. In the condition were half of the last word was audible, there was an additional significant interaction of sentence type and tone. Unmarked questions with a low-high (‘31’) tone combination were heard as statements significantly more often than marked questions with a high-falling (‘14’) tone combination. In all three conditions, the main effect of sentence type was not significant. All significant differences found here have to do with f0 height: Lowered pitch is correlated with statements.

Wrong classification as marked question. In sentences with complete last word, unmarked questions and statements were never falsely classified as marked questions. In the other two conditions, unmarked questions were identified as marked questions significantly more often than statements. In the condition in which the last word was missing, the interactions of sentence type with focus and tone were also significant. In this condition, unmarked questions with medial or final focus were heard as marked questions most often, while statements with initial or medial focus were heard as marked questions least often. Due to the fact that the last word is missing, and the question marking on the syllables preceding it were said to be unaffected by its tone combination, the sentence type - tone interactions found here were judged to be irrelevant. The two focus positions which are most easily associated with questions, namely final and medial focus in questions are also those in which either the sentence final f0 height is highest, or in which there is global raising, respectively.

Wrong classification as unmarked question. In all three conditions, there was a significant main effect of sentence type. Marked questions were significantly more often heard as unmarked questions than statements. In the conditions in which at least part of the last word was heard, there was significant interaction between sentence type and tone. In both conditions, statements with high-rising tone combination (‘12’) were the statements which were most often heard as unmarked questions. All kinds of marked questions were heard as unmarked questions more often than all kinds of statements, but in both conditions, the marked questions with rising tone (‘12’ or ‘32’) were seldom classified as unmarked questions. In the condition where the whole last word was heard, marked questions with the tone combinations high-high (‘11’), low-high (‘31’) and low-falling (‘34’) were heard as unmarked questions most often, while in sentences with only half of the word, it was the marked questions with high-low (‘13’) and high-falling (‘14’) tones. For the two conditions where the last word was not complete, there was additional interaction with focus. In both cases, marked questions with
medial focus were most frequently heard as unmarked questions, then those without focus. Statements with final focus were most often identified as unmarked questions, although the difference to the other statement kinds is not significant. Of the marked questions, those with initial focus were least often classified as unmarked questions.
5 Summary and Further Research

In this thesis paper, two kinds of Y/N questions were presented which are syntactically the same as the corresponding statement, except that for one of them a final question particle is added. It was investigated whether there is a prosodic difference between the three sentence types, and at which point it is audible.

5.1 Is there a distinct question intonation?

In literature on question intonation in Mandarin Chinese, the following findings concerning the realization of questions were reported:

- Global raising of the pitch contour
- Gradually rising or level pitch contour
- Increase of pitch range
- Local rise in the last syllable
- Shorter syllables, except for the final one in unmarked questions
- Higher and rising intensity values
- Tonal changes in the last syllables

Almost all of these were also found in our data. The most important cue involving fundamental frequency was the level pitch contour. In nearly all questions, level pitch towards the end was found. Due to the fact that the pitch contour of the statement declines towards the end, the difference increases. The level pitch of questions was found to start on or after a late high peak before the test word, i.e. at the syllable tian1 for short sentences, and at the syllable chao1 for long sentences. In short sentences, raising of the f0 values of earlier peaks occurred in some cases, sometimes even starting from the initial syllable ta1. Similarly, in long questions with medial focus, some prefocal f0 values were raised. In long unmarked questions without focus, the only condition in which the pitch curve was not level, a final rise occurred. This was not found in any other conditions. It was assumed in chapter 3.4 that there was a high pitch target at the end of the sentence in questions, and level pitch was the standard method of reaching this target. When it comes to duration, the only systematic cue which was found was shortening of the final syllable for marked questions. The intensity values were higher at some points, especially in marked questions, and in both kinds of questions, the intensity of the last syllable was significantly larger than in statements. The tones of the final syllable were altered, so that level tones were raised and rising, tones with a rise rose more steeply, and falling tones fell less steeply. Thus, the results of chapter 3.4 are the following:

- In most questions, there is level f0 starting with a late peak
- In short sentences or sentences with medial focus, there is raising in the initial part of the sentence
- In longer sentences, a final rise may occur
- During the last syllables: tone-specific features
Summary and Further Research

- During the last syllable: durational features
- Higher intensity values

In chapter 4, it was established that when the entire last word was heard, but not the *ma* particle in marked questions, 96% of statements were heard as statements, and 90% of questions were heard as questions. This means that if the whole sentence up to the *ma* particle is heard, questions and statements are perceived as being different. Therefore the results for this chapter are:

- The differences between questions and statements are perceivable.

5.2 Are there different kinds of question intonation?

In the literature, little was found on the difference between the two kinds of questions. It was hypothesized that the prosodical marking of syntactically unmarked questions may be stronger due to the fact that these questions are not marked in any other way. This is the Functional Hypothesis (Haan, 2001, p. 56). Evidence that it applies to Mandarin Chinese was found for example by Shen (1990) and Lee (2005, p. 90), but there were also researchers who found evidence to the contrary, for example Chang (1998), whose speakers pronounced marked questions with more prominent prosodic marking, or Tseng (2003), who found evidence that the two kinds of questions are pronounced differently: unmarked questions with global raising of pitch, marked questions with final rise.

In chapter 4, it was established that the two kinds of questions were not significant except for in long questions without focus, in which marked questions were pronounced significantly higher at some points throughout the whole sentence. In all other conditions of different length and focus position there was no significant difference. In visual comparison, no systematic differences were found, either. Marked and unmarked questions seem to differ only with respect to the duration of the last syllable, which is considerably shortened in marked questions compared to unmarked questions, and in the realization of sentence-final tones. In marked questions, the last part of the second, third, or fourth tones are realized on the *ma* particle, therefore the tonal realization of these syllables differs from that of unmarked questions. The Functional Hypothesis was not confirmed for our data. Therefore the results of chapter 3.4 for the difference between the two kinds of Y/N question are the following:

- During the last syllables: tone-specific features
- During the last syllable: durational features

In chapter 4, for the conditions in which the entire last word was heard, but not the *ma* particle in marked questions, it was found that the two kinds of questions showed significant differences in their identification accuracies. 89% of unmarked questions were perceived correctly, in contrast to 72% of marked questions. Marked questions were identified significantly worse than both unmarked questions and statements. Accordingly, the sentences which were wrongly identified as unmarked questions were significantly more often marked questions than statements. Although the accuracy of question identification is relatively high at this point, unmarked questions seem to be
difficult to distinguish from marked questions without the final particle. Nevertheless, an accuracy of 72% is not due to chance, therefore it can be assumed that at least some marked questions were perceived correctly at this point.

- The differences between ma-particle questions and syntactically unmarked questions are perceivable at least in most cases.

5.3 At what point does the distinction start?

In the literature on question intonation, the following possible starting points of question marking were found:

- sentence-initially
- after the predicate
- during the last two syllables, independently of whether they form a lexical unit or not
- after the focused constituent

The difference between the two kinds of questions was expected to begin at one of the two following points:

- sentence-initially
- during the last two syllables

The results of chapter 3.4 indicate that the difference between questions and statements becomes apparent at one of the following points:

- sentence-initially in short sentences or long sentences with medial focus
- after a late peak at which level pitch starts
- during the last syllables in exceptional cases

Focus was found to interact with question marking, but not in the same way as in Liu and Xu’s data (Liu and Xu, 2005). In sentences with medial focus, the difference starts earlier, while in short sentences with initial focus, it starts later. The two kinds of questions start to differ during the last two syllables:

- during the last two syllables: tone-specific features
- during the last syllable: durational features

In chapter 4, two conditions were tested in which the last word was not heard completely. In the first condition, only the first syllable of the test word was heard, and in the second condition, the test word was not heard at all. In the condition in which only the first syllable was heard, 89% of statements were heard as statements, while only 62.5% of questions were heard as questions. In the condition in which the last word was not heard, 85% of statements were identified as statements, while only 45.5% of questions were identified as questions. While the identification accuracy of statements stays relatively high, questions are identified significantly less accurately with each
missing syllable. When both syllables are missing, the accuracy is at chance level. This indicates that the features on the last syllables are relevant to mark questions. It also indicates that statement is the default value for unsure cases. The distinction between questions and statements thus becomes most salient at the following point:

- During the last two syllables

Looking at the differences between questions in these two conditions, it becomes clear that marked questions are again identified significantly worse than unmarked questions. In the condition in which half of the last word was heard, 51% of unmarked questions were identified correctly, but only 20% of marked questions. In the condition in which the last word was not heard, unmarked questions were identified correctly 34% of the time, marked questions only 14%. When looking at the mistakes that were made, the most usual mistake was to classify the questions as statements. This indicates that in most sentences, no question marking was heard. Of those which were classified correctly, in both conditions, unmarked questions with medial focus and marked questions with final focus were identified most accurately. The location where the difference between the two kinds of questions becomes most salient is the following:

- During the last syllable

When looking at the types of mistakes made in the three conditions tested, it becomes clear that the realization of the tone of the last syllable was an important factor in the differentiation of the two kinds of questions. Tones of marked questions which were pronounced very different (for example final rising tone) helped distinguish marked questions from unmarked questions, while tones which were similar (for example final high tone or low-falling tone) lead to confusion between the two kinds of question. Statements with high-rising tone combination were often heard as unmarked questions, due to the fact that the rise is very steep and high. In sentences in which the first syllable of the test word was heard, its f0 height was often a differentiating factor between all three kinds of sentence types. High realization of the first test word syllable was often interpreted as question marking (e.g in tones 13 and 14 in marked questions, which were often heard as unmarked questions). Low realization of this syllable often lead to classification as statement (for example in tones 31 in unmarked questions). In sentences in which the last word was incomplete or missing, focus features were relevant to distinguish between all three kinds of sentence types. Questions with medial focus were recognized as questions most easily, due to global raising. The two kinds of questions were however often confused with each other in sentences with medial focus. Statements with initial or medial focus were most often identified correctly, while statements with final focus were often heard as questions. This is probably due to the fact that post-focal pitch is lowered, while the f0 values of statements with final focus stay high until they drop sharply on the last syllable, which was not heard in these conditions. Marked questions without focus were often heard as unmarked questions, while unmarked questions with final focus were often heard as marked questions. Marked questions without focus showed level pitch after chao1, while this was the only focus condition in which unmarked questions remained at statement level, and were marked by final rising. The fact that hearers expected long unmarked questions to be raised, too, seems to indicate that the results found for these questions are not the usual way these questions are marked. The two kinds of questions with final focus do not show
any differences, though, so this second fact must remain unexplained. The results of chapter 4 thus indicate that the following sentence type distinguishing cues are perceivable at the following locations:

- before the last word: interaction with focus
- before the last syllable: interaction with focus and the f0 height of the first tone of the test word
- during the last syllable: realization of the second tone of the test word (f0 height and duration)

However, sufficient accuracy in the identification of questions is only reached before the last syllable, and sufficient accuracy in the differentiation of the two kinds of questions is only attained when the last syllable is heard.

5.4 Further research

Distinguishing features on the last two syllables In chapter 4 it became clear that question recognition depends highly on the differentiating features on the last syllables. It is however still not clear whether all of these are relevant. It may be interesting for a further production experiment to manipulate sentences to isolate each of the differentiating features, and to test them separately.

Focus The interaction with focus was found to be different than expected. In contrast to the findings of Liu and Xu (2005), the point of focus was not the starting point of the differences between questions and statements. In long sentences with medial focus, questions were indicated by global raising as well as level pitch contour, while in long sentences with initial or final focus, only level pitch was used. Short sentences were only tested with initial and final focus. Those with final focus showed global raising as well as level pitch, like in questions without focus, while for those with initial focus, only level pitch was found.

Test sentences In the previous chapters, it was mentioned that even though the results found here are thought to be valid, the results found in this study were influenced by the characteristics of the test sentence. It was explained that the segmental structure of the sentence made it difficult to label at some points, and that the influence of tones on each other lead to suprasegmental changes in the intonation curves of the sentences which were not triggered by question intonation. It was also mentioned that some researchers found that third tones pull down the question curves to statement level, and that length was found to change the realization of questions. Additionally, the choice of the classifier dian3 might have had a biasing effect in the perception experiment. The results of this study should therefore be tested with different kinds of sentences. At the current stage, it would not be feasible to do corpus studies. But it would decrease the error variance due to the choice of sentence and increase the power of the experiment if not only the subjects were randomized, but sentences, too. The factors which influence the sentences and their perception could be controlled, and deviation due to the sentences used could be eliminated, without having to exclude ‘difficult’ sentences from the study. In constructing these sentences, a more careful choice of segments would facilitate the labeling process.
Speakers  In chapter 3, the figures plotted for visual comparison sometimes differed from the actual findings in the statistical analysis, due to the fact that in the latter, variance due to speaker features were taken into account. It may be that a more careful choice of speakers, i.e. a limitation to a certain dialect subgroup within Mandarin Chinese, would lead to clearer results. The languages of China are so diverse that even in the Mandarin dialect family, as noted in the Ethnologue Report, there are dialects that are not intelligible to speakers of standard Mandarin (Gordon, 2005), and, as Duanmu notes, many speakers of Mandarin dialects may be less willing to change their pronunciation, because their dialect is closer to standard Mandarin than that of the other dialect families (Duanmu, 2000, p. 5). For this reason, education was thought to be more important for finding the right experiment participants than adherence to the Mandarin dialect family. However, Duanmu claims that a dialect-free pronunciation of standard Mandarin is not more socially prestigious than a dialectal one, and that even educated speakers only learn as much standard Mandarin as necessary (Duanmu, 2000, p. 5), and Schack suggested that the differences between her study and previous studies might be due to systematic dialectal differences (Schack, 2000, p. 39). Another possibility would be to record more speakers of different origin within the Mandarin Chinese dialect group, so that variances due to individual speakers would be less important for the analysis.

Functional Hypothesis  In chapter 2, Haan’s Functional Hypothesis was introduced. According to this hypothesis, prosodic question marking is strongest in questions which are syntactically unmarked, and weaker in questions which are syntactically marked (Haan, 2001, p. 56). We have seen in the previous chapter that the Functional Hypothesis could not be corroborated by the present analysis. This may not be meaningful, questions are often marked simultaneously by different means (Ultan, 1978, p. 228), but it may also have a reason. It may be that both kinds of questions are syntactically unmarked. There is another ma particle, which occurs in complementary distribution to the ma question particle (Li, 2006, p. 29), (Shen, 1990, p. 41). In statements it is seen to indicate that the speaker thinks the proposition is obvious or self-evident (Beutel, 1996, p. 167), (Howard, 2002), (Chao, 1968, p. 801). It is distinguished from the ma question particle by pitch height (Li, 2006, p. 28, 67) and length (Chao, 1968, p. 801). According to (Li, 2006), these two ma particles are actually the same particle. This would mean that the ma particle does not provide the interrogative force of the sentence, and therefore the ma particle question is syntactically unmarked. Whether or not this is the case, it would be interesting to compare ma particle questions with their homophonous statements.

Function of the questions  Some studies mention an effect of astonishment or surprise on the intonation of the question contour. As mentioned in chapter 4.5, ma particle questions and unmarked questions can be biased towards a possible answer. Both kinds of questions can carry a notion of surprise. Surprise is correlated with higher final rise (Chang, 1998, p. 68), (Chuang et al., 2007, p. 1263), (Ho, 1977, p. 450), general pitch raising (Chang, 1998, p. 58), (Gussenhoven and Chen, 2000, p. 1), (Ho, 1977, p. 450), and greater pitch range (Chuang et al., 2007, p. 1264), (Lee, 2005, p. 84) (Liu and Xu, 2005, p. 79). Chuang et al. note that a large pitch range in general indicates higher speaker involvement, and tentatively propose, because questions without ma have a

\[16\]Liu and Xu (2005) do not actually write that the pitch range is greater in incredulous questions than in neutral ones, but that their post-focal pitch range is slightly less suppressed.
larger pitch range than *ma* questions in their study, that these two questions may differ in their degree of speaker involvement (Chuang et al., 2007, p. 1265). In contrast, Lee maintains that both the unmarked and the *ma* question may be used as echo-questions to express incredulity (Lee, 2005, p. 56). Her findings indicate that echo questions are marked differently from information-seeking questions, namely with larger pitch range expansion (Lee, 2005, p. 106). Information-seeking questions are indicated mainly by a final rise (Lee, 2005, p. 159). These findings indicate that the function of questions must be taken into account when looking at their realizations.
Appendix A: Test Sentences

1. I often see her.

   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 jiao1 . / ? / ma?
   she like daily buy a-bit banana . / ? / a-particle?

   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 you2 . / ? / ma? (sesame oil)
   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 cao3 . / ? / ma? (leek)
   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 liao4 . / ? / ma? (spices)
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 gao1 . / ? / ma? (baby food)
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 you2 . / ? / ma? (cream)
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 liao4 . / ? / ma? (cheese)

2. not is ME.

   TA1 xíhuan1 mei3tian1 mai3 dian3 xiang1 jiao1 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 xiang1 you2 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 xiang1 cao3 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 xiang1 liao4 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 nai3 gao1 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 nai3 you2 . / ? / ma?
   TA1 xíhuan1 mei3tian1 mai3 dian3 nai3 liao4 . / ? / ma?

3. not is APPLE MILK SUGAR VEGETABLE

   ... /SU1YOU2 /SUAN1NAI3 /HUANG2YOU2.
   CREAM YOGURT BUTTER.

   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 jiao1 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 you2 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 cao3 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 xiang1 liao4 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 gao1 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 you2 . / ? / ma?
   ta1 xíhuan1 mei3tian1 mai3 dian3 nai3 liao4 . / ? / ma?
Appendix A: Test Sentences

4.  wo3 chang2 kan4jian4 ta1.
   I often see her.
   ta1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang ljiao1 . / ? / ma?
she like daily at lihang supermarket buy a-bit banana . / ? / a-particle?

5.  bu4 shi4 WO3.
   not is ME.
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang ljiao1 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang lyou2 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang liao4 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3gao1 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3you2 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3lao4 . / ? / ma?

6.  bu4 shi4 zai4 li4MIN2 chao1shi4.
   not is at LiMIN supermarket
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang ljiao1 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang lyou2 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 xiang liao4 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3gao1 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3you2 . / ? / ma?
   TA1 xi3huan1 mei3tian1 zai4 li4hang2 chao1shi4 mai3 dian3 nai3lao4 . / ? / ma?

3.  bu4 shi4 PING2GUO3 /NIU2NAI3 /BAI2TANG2 /SHU1CAI4
   not is APPLE MILK SUGAR VEGETABLE
   ... /SU1YOU2 /SUAN1NAI3 /HUANG2YOU2.
   CREAM YOGURT BUTTER.
References


References


References


References


Ich erkläre, dass ich die am

28.07.2008

an der Universität Potsdam abgegebene Magisterarbeit mit dem Titel

‘The prosodic realization of two kinds of Y/N-questions in Mandarin Chinese’

selbst verfasst habe, alle nicht von mir stammenden Textpartien als solche gekennzeichnet habe, und keine nicht angegeben Quellen und Hilfsmittel verwendet habe.

Berlin, 20.07.2008

Mira Grubic
Deutsche Zusammenfassung

In der vorliegenden Arbeit geht es um die prosodische Realisierung von zwei Arten von Ja/Nein-Fragen im Mandarinchinesischen. Während die syntaktisch unmarkierte Frageform (UMQ) mit neutralen Aussagen genau homomorph ist, z.B.

ta1 mai3 dian3 xiang1jiao1.
sie kaufen ein-wenig Banane
‘Sie kauft ein paar Bananen.’

ta1 mai3 dian3 xiang1jiao1?
sie kaufen ein-wenig Banane
‘Kauft sie ein paar Bananen?’

wird die ma Frageform (MQ) durch finale Fragepartikel markiert.

ta1 mai3 dian3 xiang1jiao1 ma0?
sie kaufen ein-wenig Banane Fragepartikel?
‘Kauft sie ein paar Bananen?’

Um korrekt und rechtzeitig zu antworten muss der Sprecher also bevor das ma ausgesprochen wird schon wissen, ob es sich um eine Frage oder eine Aussage handelt. Eine Literaturrecherche ergab, dass durchaus vor diesem Punkt ein prosodischer Unterschied zwischen diesen Fragen und Aussagen angenommen wird, jedoch Uneinigkeit herrscht, worin dieser genau besteht. Über die Unterschiede zwischen den beiden Arten von Fragen gab es kaum Literatur.

In einem Produktionsexperiment wurden Sätze zur Analyse erhoben, die sich in ihrer Länge, Fokusposition, und Tonkombination des letzten Wortes unterschieden. Sie wurden von je drei Sprecherinnen als Aussage, ma-Frage und syntaktisch unmarkierte Frage gelesen. Dies ergab 441 Sätze. In der anschliessenden Analyse ergab sich dass die Fragen von den Aussagen dadurch unterschieden, dass (a) ihre Grundfrequenzkontur gegen Ende des Satzes nicht abfiel, (b) ihre Grundfrequenzkontur in kurzen Sätzen und Sätzen mit mittlerem Fokus global angehoben war, (c) sie in einem Fall eine satzfinale steigende Grundfrequenzkontur aufwies, (d) die Töne der letzten zwei Silben unterschiedlich realisiert wurden, (e) die längere Dauer der letzten Silbe in ma Fragen, (f) sie höhere Intensität aufwiesen. Die beiden Arten von Fragen unterschieden sich nur durch Merkmale (d) und (e). Es war also zu erwarten dass Fragen und Aussagen schon vor dem aus den letzten zwei Silben des Satzes bestehenden Testwort unterscheidbar sind, während der Unterschied zwischen den beiden Arten von Fragen erst danach erkennbar wird. Dies wurde im nächsten Schritt getestet.

In einem darauffolgenden Perzeptionsexperiment wurden die langen Sätze einer Sprecherin von zwölf Teilnehmenden in drei Konditionen gehört: (a) bis zum Ende des finalen Testworts (also ohne ma in MQs); (b) bis zur Hälfte des Testworts; (c) bis vor dem Testwort. Dadurch sollte getestet werden ob die Sprecher die Unterschiede überhaupt erkennen, und ab welchem Punkt die Unterschiede erkennbar sind. Das Experiment zeigte dass für die Unterscheidung von Fragen und Aussagen die letzten zwei Silben relevant waren, und für die Unterscheidung der zwei Fragearten die letzte Silbe.