TONE LANGUAGE AND SONG:

AN OPTIMALITY THEORETIC MODEL
OF THE INFLUENCE OF A SUNG MELODY ON THE
INTERPRETATION OF MANDARIN LEXICAL TONES

Maaike Vondenhoff, 0141119
Ma Thesis for the master's in General Linguistics
University of Amsterdam
August 2009

Thesis supervisor: Prof. dr. Paul Boersma
Second reader: dr. Wolfgang kehrein

Acknowledgements

I would like to offer a word of gratitude to the following people, for helping me in the process of finishing this thesis.

Firstly, to my thesis supervisor Paul Boersma, thank you for your academic mentoring and critical advice, and mostly, many thanks for staying so patient.

Also, I would like to thank my mother and stepfather, who have been the best in offering me support and encouragement. It is thanks to them that I succeeded in regularly doing the dishes and eating enough vegetables while working.

Furthermore, to my dear friend David, musician and fellow linguist, whose feisty discussions have been a source of ideas about the relationship between language and music.

To many of my other friends, thank you for listening patiently to all my ranting about the subject of my thesis. Be assured that everything will return back to normality soon.

And last but not least, to the coffee baristas of the Coffee Company on the corner, I would never have made it without your excellent cappuccinos and double espressos to keep me awake and energized in times of hard work.

Abstract

The present study focuses on formalizing musical auditory cue constraints in an Optimality Theory model, in such a way that their hierarchical ranking results in the model generating Mandarin lexical tone perceptions based on a given stimulus melody. This model of a virtual Mandarin listener comes to different constraint rankings in different learning conditions by applying the Gradual Learning Algorithm, each of these conditions reflecting a different grammar set that will be compared to original data obtained from real listeners. The purpose for running these simulations was twofold. Firstly, by doing so I will show that for the perception and identification of lexical tones in a song, Mandarin Chinese listeners are likely to use both the direction of the melody and pitch register equally often. Furthermore, it seems that the pitch and direction of the preceding syllable with respect to the target syllable is of more influence than the succeeding one. However, the results do suggest that a listener's decisions in the experiment were largely based on chance, which would imply that for the understanding of the lyrics in Mandarin Chinese songs, other elements than those analysed by the model are of importance. Secondly, by showing that constraints for the perception of melody can be hierarchically ranked, like constraints for the perception of speech, and generate results comparable to the answers of real language users, I want to add an argument for the position that language and music share similar cognitive underpinnings.

1. Introduction

In the fields of linguistics and music sciences, there have been two main opinions about the relation between language and music. One of these is that language and music share many resemblances. Advocates of this side of the academic quarrel concerning the similarities between the two disciplines draw their evidence from the fields of developmental, structure and prosody studies, as well as neuroimaging research. The opposing opinion, as the word implies, states that language and music do not have much in common at all. People defending this side of the debate base their opinion for a large portion on language and music impairment studies in which patients have lost one ability or the other.

Yet, one area in which music and language unmistakably converge is the phenomenon of tone languages, where fundamental pitch is of importance for understanding the meaning of a word. When learning a tone language, the language learner who is new to the concept of lexical tones could find himself wondering: How do composers write a song in a tone language? Do they keep a strict melodic correspondence with the intonation of the lexical tones, or do they allow themselves musical creativity? To people who are native to a tone language, this question seems trivial or even silly, for they have been talking and singing all their lives.

Several studies of classical and traditional composition of songs sung in tone languages around the world show that composers use various techniques to create a new melody, while at the same time preserving the intelligibility of the words in the song (Chao, 1956; Hsu, 1964; Levis, 1936; List, 1961; Wong and Diehl, 2002; Yung, 1983). But if composers write their songs according to certain strategies, what are the strategies used by listeners? And in particular, what elements of a song melody influence mandarin listeners' perception of lexical tone?

I expected to find that in addition to the direction of the melody, the pitch register is of equal importance. In order to find out what elements of the melody have an influence on the perception of Mandarin lexical tones in the lyrics of a song, I ran a computer simulation of a Mandarin Chinese listener, listening to songs. Based on Optimality Theory (Prince & Smolensky, 1993), I taught the virtual listener to recognize the four lexical tones of the Mandarin language, using the Gradual Learning Algorithm (Boersma & Hayes, 2001). By running the simulation (Praat 5.0.47, Boersma & Weenink, 2009) under different learning conditions, I will be able to see which of the learning conditions, or constraint sets, could generate a set of data that yielded the best resemblance with a set of data obtained in a perception experiment with real Mandarin listeners. Rephrasing the questions posed in the paragraph above yields the following research question: with what cue constraint set will my model of a Mandarin Chinese listener be best able to imitate the answer of real Chinese listeners?

In addition to answering this question, I wish to contribute in the debate about the resemblance between music and language. By formally showing that musical cues can be processed in terms of lexical meanings, I am hoping to add another argument in favor of those who advocate the theory that language and music share an underlying cognitive system If it is indeed the case that constraints for the perception of melody can be hierarchically arranged, in a similar fashion as linguistic constraints, in such a way that their ranking reflects a language user's choice preference for a certain lexical tone given a stimulus input, then it would seem plausible to assume that it is very well possible that music and language share an underlying system that structures manifestation of both disciplines in a similar way.

In chapter 2 the two disciplines of music and language are compared and arguments in favor for as well as against a shared cognitive system are considered. Chapter 3 will be dedicated to tone languages and their characteristics. In the same chapter I also investigate strategies used by composers when writing songs in those tone languages.

In the fourth chapter research about the strategies listeners apply when listening to songs in a tone language is considered and how this relates to the present research. Also, I will discuss aspects of Optimality, after which I will proceed to discuss methods and results of the simulations.

In chapter 5, I will discuss the model design of my research and experimental methods, after which I will proceed with the analysis. Finally, in chapter 6, the conclusion is drawn and I will discuss the results in the light of previous research and my own hypotheses.

2. Music and language

Music and language appear to be very different. No one in their right mind will confuse a Bach sonata and president Obama's inauguration speech. The reason we recognize language and music as two different entities, according to widespread scientific belief, is that music and language are processed in separate modules in our brains. For example, in some cases people with severe brain damage to Broca's or Wernicke's area, causing their language abilities to be severely impaired, could still sing songs or hum a tune (Yamadori, 1977; Brust, 2003). Also, research outside of cognitive impairment studies confirms a separate cognitive processing of language and music input. Besson (1998) indicated with an fMRI study that the melody and words from French operas are processed separately. To many scientists, such cases make a strong point for the theory of brain specialization and modularity of language and music processing. They believe that language and music are two very different phenomena that certainly do not share the same cognitive pathways.

However, research to supply evidence of the opposite is just as plentiful. Nakada (1998) suggested that for the reading musical scores there exists an overlap of locus in the brain with word reading. Also, neuroimaging research has shown that musical syntactic processing activates those areas in the brain that are known to be used for language processing as well (Maess 2001), which suggests an overlap in music-language processing. Furthermore, McMullen & Saffran (2004) did a developmental comparison of the acquisition of knowledge in the domains of language and music throughout childhood, and they found that "there may be similar developmental underpinnings in both domains, suggesting that modularity is emergent rather than present at the beginning of life." (p. 289) In their article, McMullen and Saffran compare the development of linguistic and musical abilities in infants and find that, even without entering the complicated discussion of whether such abilities are innate or not, the development and features of music and language have very much in common on structural, prosodic, grammatical and even, though to a lesser extent, on semantic grounds.

A possible reason for many a scientist to firmly hold on to the belief that music and language do not share many similarities, might primarily be an intuitive one: from the time we were born we have had more occasion to get properly acquainted with language than with music, because of "an exposure to fewer examples of musical phrases than linguistic ones" (McMullen & Saffran, 2004, p. 297). Therefore, we have had more chance to develop fluency for the former, while the latter became relatively neglected. Still, Deutsch et al. (2004) found an "intriguing parallel between the critical periods involved in the acquisition of speech and language on the one hand, and the acquisition of absolute pitch on the other" (p. 342). The phases of language and music acquisition show some striking similarities, particularly in terms of the time frame involved.

It has even been claimed that in principle every form of behaviour is structured in the same way (Liberman, 1975; Gilbers, 1992; Gilbers & Schreuder, 2002; Schreuder, 2006). "If this claim is true, language and music should have much in common, since both disciplines are examples of temporally ordered behaviour" (Schreuder, 2006, p. 5). As we will see in the next sections, it is indeed the case that many resemblances exist between the two disciplines of language and music.

2.1 Prosody in language and music

Of all areas of traditional linguistics, the prosodic features of language and music are perhaps the first one would think of when considering the similarities between the two disciplines. After all, intuitively, melody, stress and rhythm are those elements one would be expected to define music with. As it turns out, music and language show many resemblances in these aspects of prosody.

For instance, it has been found that similar phrase boundary markers can be found in both linguistic and musical utterances. It has been observed that at the end of both musical and linguistic phrases there often is a decline in pitch and a lengthening of the final note or syllable (McMullen & Saffran, 2004; Hayes & MacEachern, 1998; Gilbers & Schreuder, 2002; Jusczyk & Krumhansl, 1993). Related to the final lengthening at phrase boundaries may be a phenomenon that is known as catalexis in the study of prosody, which is the occurrence of a lower number of notes in even-numbered bars. It is most "easily explained with reference to the incompleteness of the underlying segment of the poems" (Lindblom and Sundberg, 1969, p. 70).

Also there seems to be a correlation between musical pitch perception and what dialect or language the individual has been exposed to since childhood: "It has been found that the way the pitch class circle (of the tritone paradox) is oriented with respect to height is related to the language or dialect to which the individual has been exposed and to the pitch range of his or her speaking voice." (Deutsch et al., 2004, p. 340. For a detailed definition of this complicated phenomenon I refer the reader to this article) In other words, there is a connection between the way we perceive pitch and the language experiences we have had. In addition, as we will see in the next chapter as well, Patel and Daniele (2002) found that the influence of rhythm of spoken English and French on musical song composition is significant.

On the cognitive level, research has shown that responses to nonlinguistic human vocal sounds and processing musical pitch are strongest in the right superior temporal area (Belin, Zatorre & Ahad, 2002; Zatorre, 2003), which indicates the "plausibility to accounts of musical and linguistic evolution that emphasize emotional communication through prosody as a primary forebear of both systems" (McMullen & Saffran, 2004, p. 300).

But perhaps the most striking crossover between language and music in the study of prosody is the area of tone languages. These are languages that use pitch as a means of distinguishing between the meanings of words. As it turns out, musicians native to a tone language are more often capable of recognizing pitch heights without reference, a phenomenon known as absolute or perfect pitch (Deutsch et al., 2004). Furthermore, "at the neurological level, there is strong evidence that the brain structures underlying the processing of lexical tone overlap with those underlying the processing of phonemes in speech" (Deutsch et al., 2004, p. 344).

Finally, Schreuder (2006) shows that there is an indication that the mood of emotional prosody in speech is similar to musical modality. "We found a tendency that a sad mood can be expressed by using intervals of three semitones, i.e. minor thirds. Cheerful speech mostly has bigger intervals than thirds, but when thirds are used, these thirds tend to be major thirds" (p.

163). This relates to what Houston (2008) proposes about the relation between musical intervals in speech: “The larger intervallic distance of a major third and the smaller distance of a minor third from a common tonic (C to E v C to E^{b}) appear to be associable with linguistic 'equivalents' within phonetic vowel space and consonantal articulation” (p. 23).

2.2 Phonology in language and music

Language comes to existence by making infinite combinations of a discrete set of sounds. Every culture chooses its own combination of phonemes out of all possibilities that can possibly be produced. This is also the case for music. Of all possible notes, intervallic changes, and use of musical instruments, each culture uses only a finite set to compose infinite possibilities of musical expression. In other words, the structure of sounds of both language and music have a discrete infinite character (McMullen & Saffran, 2004; Houston, 2008).

Related to these discrete sound sets in music and language are the way they are perceived categorically, something that is known in the field of linguistics as *categorical perception*: the perception of two sounds as belonging to two different categories even though they are acoustically very similar, and vice versa: the perception of two sounds as being the same while their auditory properties are very different. A classic example would be the Japanese perception of /l/ and /r/ as belonging to the same phonemic category where an English speaker distinguishes between the two (see for instance Polivanov, 1931). It was previously thought that categorical perception was a characteristic unique to language, but recent research has shown that the perception of nonspeech sounds, like musical signals, is also subject to this phenomenon. For example, the difference between a plucked and a bowed string is perceived categorically by adults and infants (Cutting & Rosner, 1974; Jusczyk, Rosner, Cutting, Foard & Smith, 1977), and musical intervals are also labeled categorically, even by nonmusicians (Smith, Kemler Nelson, Grohskopf & Appleton, 1994). This is generalized in the following way by McMullen and Saffran (2004): “Both (spoken) language and music are generated from a finite set of sounds, (and these) sounds are organized into discrete categories, facilitating representation and memory. Auditory events are subject to categorical perception” (p. 291).

Also, the way we perceive the structure of these sounds and their categories is very similar. Spoken languages are bound to the limits of temporal processing. Speech, like music, is perceived as "frequency spectra, arrayed as pitches (...), organized temporally with the relevant structures unfolding in time" (McMullen & Saffran, 2004, p. 290).

2.3 Structure in language and music

The infinitely combinatorial nature of speech and music, discussed in the previous paragraph, is a direct result of the way these elements are structured: "As important as segmental and suprasegmental cues are for learning both musical and linguistic systems, the real power of these systems comes from their infinitely combinatorial nature" (McMullen & Saffran, 2004, p. 296). This combinatorial nature comes in the form of cultural-specific nuanced rules for well-formed strings.

The recursiveness of language used to be defined as a characteristic unique to human language, (Hauser, Chomsky & Fitch, 2002), even though obviously some structure could be recognized. As a solution to this occurring structure, attempts were made to explain these structures in terms of Markovian or stochastic models (Fucks & Lauter, 1965; Hiller & Isaacson, 1959), when, in fact later, many investigations in the nature of structure of musical composition found that in music the same recursive structures emerge as in language (Lindblom & Sundberg, 1969; Lehrdahl & Jackendoff, 1983).

Also in the fields of development and cognition studies there has been research to support the evidence for a similarity of structure in music and language. For instance, it was found that infants as from the age of 7 months were capable of pattern induction in music. Further similarities were found "(...) in early neural processing of linguistic and musical syntax, both of which nature make use of broca’s area and its right-hemisphere homologue. Response has been demonstrated in nonmusicians, indicating that explicit training is not necessary for this level of implicit knowledge to develop" (McMullen & Saffran, 2004, p. 298).

2.4 Meaning in language and music

Despite the analogies between language and music in the fields of study described above, it is difficult to find a similar analogy for the field of semantics. There are no such things as nouns or verbs in music to convey meaning, and if any meaning is conveyed in the composition of a musical piece, probabilities are high it is a rather abstract one based on moods and emotions.

Still, it is exactly those emotional common bearings that are considered to be a common ground for a shared basis for meaning in music and language. Especially the differences in perception of consonant and dissonant key preference "have long been posited as an important underpinning of emotion in music" (McMullen & Saffran, 2004, p. 298). We have already seen in section 2.1 that speech and music share similar pitch interval distances, and that these pitch distances are an indication of mood. This was also what Cook, Fujisawa & Takami (2004) concluded: utterances perceived as having a positive affect showed a major-like pitch structure, whereas utterances with negative affect were more similar to minor-like pitch structure. Schreuder (2006) investigated this matter as well and hypothesized the results she found were an indication that the mood of emotional prosody of speech and music are similar. In addition, Houston (2008) illustrated that there may be a metaphorical resemblance of two phonetically contrasting stimuli and melodies with contrasting minor/major pitch intervals (see also p. 9 of this thesis). In his experiment, 18 of 20 participants associated *bouba* with the minor and *kiki* with the major melody. "In a phonetic sense, this would imply that they associated voiceless stops/ high front vowels with the major melody, and labials/ low back vowels with the minor melody" (p. 20).

Maybe the most obvious example in which melody carries meaning is the case of tone languages. In these languages, the intonation of a word is responsible for a distinction in meaning. As we will see in chapters 4 and 5, in particular in the context of a song, the musical melody will be shown to be of influence on the lexical perception of listeners native to a tone language.

2.5 Summary

Considering the research discussed above, it appears that it is not unreasonable to assume that music and language share many properties in the areas of prosody, phonology, structure and even semantics. It may even be possible that all manifestations of temporally ordered behavior share a common structure.

Despite the growing body of evidence discussed above, there are still a few matters that remain unanswered. If it is indeed the case that language and music have much in common in terms of a shared cognitive system, then it should be possible to analyze manifestations of both disciplines with one single model. Therefore the remaining chapters are dedicated to contribute one more piece to the puzzle that is the debate of whether music and language share a common basis. As said in the introduction of this thesis, the purpose of the present study is twofold, as well as focusing on the influence of musical melody on the interpretation of Mandarin lexical tones, the purpose is also to contribute supporting evidence for those who advocate a common shared system.

3. Tone languages

In the previous chapter I discussed four areas in which music and language have been said to show many resemblances. Specifically their prosodic features seem to be mutually linked to a great extent, of which the strongest linkage comes from tone languages (Deutsch et al., 2004). In such languages as Mandarin, Cantonese and Thai "words take on arbitrarily different lexical meanings depending on the tones in which they are enunciated" (Deutsch et al., 2004, p. 343). It gets more interesting to see what happens when lexical notes occur within an actual musical environment. It arises questions like: do composers obey the rules of linguistic intelligibility, or do they allow themselves to be musically creative? And what is even more puzzling: to what extent are listeners, in their turn, influenced by the melody of a song when listening to the lyrics?

But before we get to answer these questions, we will first have to consider what exactly a tone language is and understand in what ways these research questions are justified. After that, we will proceed to have a look into the different composing strategies that are found in the world's tone languages. For the purpose of my research I will only discuss Asian tone languages, for these are the focus of the present study.

3.1 Tone languages

In most European languages prosodic features of speech do a great deal to help us understand the speaker's intentions. Rhythm of speech helps us determine which of the utterances are separate constituents, accent and stress help us focus on the most relevant piece of information, and intonational features are strong indicators for structure. By changing the tonal pattern, for example, one can indicate whether an utterance should be interpreted as an declarative or a question.

In tone languages, however, fundamental frequency (F_0) is crucial for understanding the meaning of words. Otherwise similarly pronounced words have utterly different meanings depending on with which tone they are enunciated. Distinguishing between meanings by using pitch in a tone language is of equal importance as the distinction between two different vowels in an English minimal pair like pit-pot. In other words, a tone language is "a language (...) in which almost every morpheme is composed of not only segmental phonemes, but also phonemic pitch pattern. The phonemic burden, or phonological load, of this element of pitch pattern, or tone, is of the same magnitude as that of vowels" (Chao, 1956, p. 52).

Wang (1967) divides the world's tone languages three different groups, the first two of which are American-Indian languages and African languages. He finds that the third group of tone languages, Sino-Tibetan languages, is different from and more complex than the first two in three ways. In the first place, tones in Sino-Tibetan tone languages are almost exclusively used lexically. That is to say, their tones do not serve a syntactical, nor morphological purpose. Secondly, tones of the third group are more complex in the way that they have distinct shapes and that they come in larger numbers than in the American-Indian languages or African languages. Then finally, in Sino-Tibetan tone languages tone sandhi[1] is more complex and dependent on individual tones, whereas tone sandhi in the two other groups is what Wang calls syntagmatic deplacement, i.e. "each syllable receives its tone from its (usually left) neighbor" (p. 94).

A slightly overlapping and dichotomous division has been observed by McCawley (1978). According to him, the world's tonal languages can be divided into pitch-accent languages, like Japanese, and true tonal languages, such as many Sino-Tibetan languages like Mandarin and Cantonese. In a pitch-accent language, like standard Japanese, "the only distinctive melodic characteristic of a phrase is the location of the syllable where the pitch drops" (p. 113). An example of such a pitch-accent distinction is shown in figure 1 below. The apostrophe precedes the syllable on which the pitch drops.

Ka'kiga - oyster
Kaki'ga - fence
Kakiga - persimmon

Figure 1. Example of minimal pitch-accent pairs in Japanese. Taken from McCawley (1978), p.113.

The second type of tone languages McCawley distinguishes are true tonal languages. In a true tonal language it is the melodic *contour* that makes a semantic difference. Mandarin Chinese is said to be a textbook case of such a true tonal language, where a difference in F_0 contour results in the listener hearing a different meaning. An example of such a difference in pitch contour is the following pair of syllables: /yao4/ (要, "to want") and /yao3/ (咬, "to bite")[2]. In both syllables, the same sequence of consonants and vowels is present, but because of their difference in pitch, or lexical tone, they do not have the same meaning.

The lexical tones of true tonal languages can be distinguished by their contours. Abramson (1978) discusses a classification that divides lexical tone movements into dynamic or contour tones, and static or level tones. This convention of classifying lexical tones will prove to be of use in this research, as we will see later on, and "although imprecise, the typological dichotomy is useful" (Abramson, 1978, p. 319). Figures 2 and 3 below illustrate how this classification can indeed be used as a rough identification of the lexical tones of tone languages like Mandarin (figure 2) and Cantonese (figure 3). We can see from these examples that the distribution of dynamic and level tones in tone languages can vary: only half of the Cantonese tones have a dynamic shape (high-rising, low-falling and low-rising), whereas in Mandarin three tones out of four have a dynamic character (the high-level one being the only static tone).

---

[1] The phenomenon known as tone sandhi is defined by the Oxford Concise Dictionary of Linguistics as follows: The phonetic modification of tones in the context of those on preceding of following syllables.

[2] The numbers behind the syllable /yao/ refer to the lexical tone with which each syllable is enunciated. A number 1 stands for the first Mandarin Chinese lexical tone (a high level pitch), a

Tonal categories[3]	Values
1. *High-level*	55
2. *Mid-rising*	35
3. *Falling-rising*	214
4. *High-falling*	51

Figure 2. Mandarin tones

Tonal categories[4]	Values
1. *High-level*	55
2. *High-rising*	35
3 *Mid-level*	33
4. *Low-falling*	21
5. *Low-rising*	23
6. *Low-level*	22

Figure 3. Cantonese tones

The tonal values in the second column of each figure above are numerical descriptions of the movement of pitch of each tone. This method of numerical describing the pitch of lexical tones was first introduced by Chao (1930). Anderson (1978) defines this method as a "five-level system in which a tone 55 begins in the highest level/register and remains level, 35 begins in the middle of the scale and rises to high, and so on. But the actual range of phonetically distinguishable tones is not limited to 5 values" (p. 141). In other words, "tone value is a directional value. The absolute pitch interval is not pertinent to tonal composition" (Xiao-nan, 1989, p. 67).

After all this talk about the nature and importance of lexical tones in a tone language, one must not make the mistake to think that fundamental frequency is the only feature used by speakers and listeners to distinguish between tones. "Thanks to the liberal amount of redundancy that is usually present in all languages, so is Chinese without tones also intelligible, provided that it is otherwise perfect in pronunciation, construction and use of words" (Chao, 1956, p. 53). Although it is known that lexical tones of tone languages, and thus word meanings, are primarily indicated with pitch, other elements are also recognized. Yung (1983) distinguishes long and short duration next to fundamental frequency when describing the lexical tones in

number 2 stands for the second tone (a rising pitch), and so on. All Mandarin Chinese tones and there pronunciations are shown in figure 2.

[3] See Howie (1976) for a detailed description of the nature of Mandarin tones. For an overview of the traditional classification of Mandarin tones as opposed to the one presented here, see for example Mei (1970) p. 104, Yip (1980) and Bao (1999) p. 10.

[4] This classification of Cantonese tones is directly taken from Wong and Diehl (2002), p. 203. Because linguists do not seem to agree on the number of Cantonese tones, I decided to use the classification that was adopted by Wong and Diehl. Others have made different classifications, like Yung (1983), who informs us that Cantonese Chinese possesses nine lexical tones.

Cantonese (p. 29). Furthermore, in Mandarin Chinese, as well as pitch, elements like duration and vocal constriction (i.e. the so-called 'creaky voice') play a role in tone recognition; although considered secondary, they may become important under special conditions, such as whispered speech (Chao, 1956, p. 56).

Considering the complexity of tone systems, especially those of true tonal languages of the Sino-Tibetan group like Mandarin or Cantonese, it becomes clear how important tonal features are for recognizing the meaning of spoken utterances in those languages. This brings us right back to the questions I started this chapter with. In a musical environment, like a song, an intuitive observer would expect these tonal features to be preserved, in order for the lyrics to be still intelligible. Without lexical tones, word intelligibility is reduced after all. So how, I wonder, do song composers maintain lexical intelligibility while at the same time creating new melodies without losing musical creativity?

3.2 Composing strategies

For speakers of a non-tonal language it is often hard to imagine how fundamental frequency can play a role just as important in understanding meaning as the vowels and consonants in their own language. This notion becomes even more complicated if we try to imagine in what ways these tonal elements are intertwined with the melodies of songs created in a society in which a tonal language is spoken. The reason that this seems to boggle our minds is that we are just not used to thinking of melody as being an important factor in identify the meaning of an utterance.

Lehiste and Peterson (1961) propose that "the problem of relating contourlike movements to musical intervals seems to be less relevant for a study of English than for a study of tone languages" (p. 425). However, the intertwinement of linguistic prosodic features with musical composition is not an exotic phenomenon exclusively limited to tone languages. Elements like rhythm and stress play an important role in languages as familiar as English and French and could possibly be just as puzzling for a speaker of a tone language as tonal features for an English speaker. This is why the other way around is probably the case just as well: the problem of relating rhythmic patterns and lexical stress to musical movements seems to be less relevant for a study of tone languages than for a study of English.

Patel and Daniele (2002) performed a quantitative analysis to investigate whether the stress patterns of one's native language influence a composer's style. They found that "English and French musical themes are significantly different in (their) measure of rhythm, which also differentiates the rhythm of spoken English and French. Thus, there is an empirical basis for the claim that spoken prosody leaves an imprint on the music of a culture." (p. B35)

Considering the influence the *rhythmic* prosody of a language has on its music, it is reasonable to assume that the *tonal* prosody of a tone language equally influences the way music is composed by its speakers. Levis (1936) proposes that lexical tones do influence Mandarin composers' musical creations: "The distinctive tonal basis of the Chinese language must have had some influence upon the conscious musical expression of people" (p. VI). But one can imagine the difficulties such a composer faces when creating a song melody based on lyrics and their tonal movements in a true tonal language. Should he preserve the lexical melody for the sake of intelligibility, or will he abandon linguistic pitch and indulge in musical creativity? One can imagine that, in order to sustain a certain level of semantic intelligibility, poetical creativity must be reduced (Wong & Diehl, 2002).

Wong and Diehl (2002) discuss three possible options for composers to deal with this dilemma. "The first is to ignore lexical tones and word meaning and to use pitch exclusively to mark the melody. This preserves musicality at the cost of reduced lyric intelligibility. The second option is just the reverse: to preserve the normal pitch variations of lexical tones while ignoring the melody, sacrificing musicality for intelligibility. (...) The third option is intermediate between the first two: songwriters may attempt to preserve at least partially the pitch contrasts of lexical tones while not unduly restricting the melodic role of F_0 changes" (p. 203).

List (1961) examined the relation between lexical tone and musical melody in reciting and chant in Thai. He found that Thai songwriters find no need to develop new tunes, but instead draw from a common pool of pre-existing material. While creating a new song, the composer simply applies those syllables whose tones match the contour of the corresponding place in the melody. "(...) In chant and song found in the traditional everyday life of the people of Central Thailand, speech melody has played the most prominent role. Song melody has been subservient (...)" (p. 31). Nevertheless, he adds that the degree upon which contour tones coordinate musical melody has tended to diminish under the influence of Western music styles.

In traditional Cantonese opera, musical composers also rely on a stock of already existing melodies. However, unlike in Thai where the relative contours of the lexical tones are maintained, opera writers must assign an absolute tone-melody relation. That is to say, within a musical composition, a certain pitch is associated with a certain lexical tone. According to Yung (1983), "the specific pitches and the melodic contour of the aria type are to a great extent determined by the linguistic tones of the text" (p. 39). Although in certain situations the singer is to some extent allowed to depart from absolute matching.

Modern Cantonese composers, however, apply a different strategy when writing a new song. Wong and Diehl (2002) investigated the role of lexical tone in contemporary Cantonese song compositions and they found that, in contrast with traditional Cantonese opera, "corresponding tone sequences preserve only the *direction* of F_0 change" (p. 204). In the four songs they analyzed, they found that in 91.81% of the cases a mapping occurred between the tonal sequence of the words in the lyrics and the musical sequence of the corresponding notes in the melody, instead of an absolute tone to tune mapping practiced by traditional opera composers.

In traditional Mandarin Chinese singing, there used to be great a dependence on lexical tones, which resulted in somewhat stereotyped forms of melodies: "though the dependence was not so close as to fix the melody unambiguously, it suggested and limited the range of possibilities (...)" (Chao, 1956, p. 57). Although a singer was bound to the tones of the words, he was allowed and expected to introduce grace notes to the melody, with the effect of a clearer diction (Chao, 1956). Musical compositions consisted of "simple tunes of limited range and so designed they could easily be remembered" (Yung, 1983, p. 441).

However, most of contemporary Mandarin songwriters seem to have abandoned the rules of relating musical melody to lexical pitch (Levis, 1936; Chao, 1956) Indeed, Vondenhoff (2007) conducted a musical analysis of three contemporary Mandarin songs and found that less than 40% of the melodic note sequences had the same direction as those of the lexical tones on the corresponding syllables. According to Chao this relaxation of rules governing melody-tone correspondences has to do with the low prestige of Mandarin tones in old-style drama. It is not within the scope of this thesis to consider cultural changes over the past century and their influences on people's behavior in China, but the fact remains that the relationship between lexical tone and musical composition seems to be considerably less strong as once used to be the case.

In short, we have seen that in a tone language, fundamental frequency is an important prosodic element on which listeners rely for understanding speech. Composers of songs in the tone languages discussed above therefore often rely on already existing melodies from which they can choose a tune to match with the song lyrics. Also, linguistic tones may play a large role in the sense that either the melody contours are matched to the lexical tones in an absolute correlation, or only the relative pitch contours, or direction, of the lexical tones are preserved. In the case of modern Mandarin song composition, however, it appears that to a large extent even a relative tone to melody relation is abandoned.

4. Tone languages and song

As we have seen in chapter 3, composers of songs in a tone language are often in the habit of in some way preserving the lexical tones of the words in the lyrics. This is due to the fact that without lexical tones, intelligibility of words in a tone language is reduced (Chao, 1954). In Thailand, composers draw from a corpus of existing melodies (List, 1961), which they apply, to the lyrics of their songs. In *traditional* Cantonese music there is an absolute pitch to tone relation (Yung, 1983), while in *contemporary* Cantonese only a relative pitch to tone relation is maintained (Wong & Diehl, 2002). Finally, *classic* Mandarin composers apply stereotype melodies to the words (Levis, 1936; Hsu, 1964). Also, it has been said that in *modern* Mandarin song compositions, the melody to lexical tone relation is mostly abandoned (Levis, 1936; Chao, 1956; Vondenhoff, 2007).

4.1 Listening strategies of songs in a tone language

While each of these studies address the composers' strategies of preserving musical creativity as well as the intelligibility of the song, the questions that naturally arise are the following: what do listeners do with this information that composers put in the song for them? How do speakers of a tone language listen to the words of a song and to the corresponding musical melody? In what way does their perception of the melody influence their interpretation of the lexical tones?

This last question in particular was the focus of Wong and Diehl's article *How can the lyrics of a song in a tone language be understood?* (2002). They investigated how native listeners of Cantonese Chinese extract the lexical meaning of a sung text in a non-contextual and therefore tonal ambiguous environment. They found that, without the help of a semantic context, listeners applied an ordinal mapping rule when interpreting the lexical tones. That is, a mapping "between musical note and tone group occurs such that the direction of pitch change in two consecutive musical notes is the same as in the two consecutive tone groups attached to them" (Wong & Diehl, 2002, p. 204). In other words, while in carefully spoken Cantonese, lexical tone sequences have an absolute and constant percentage of F_0 change (Chao, 1947; Wong & Diehl, 2002), listeners to contemporary Cantonese songs use only the *direction* of pitch change to identify the lexical tones of the lyrics of a song.

Although Cantonese has six lexical tones, three of which have a dynamic or contour character and three a static or level character (see chapter 3, figure 3 for more detail), for their experiment Wong and Diehl used only the three level tones. They presented their subjects with two different sets of three melodies, in which only the tone of the last syllable was subject to change. As the distances between the penultimate and the target tone changed by two, five or nine semitones, participants in the experiment were expected to choose a different answer, the tones of it correlating with the relative pitch differences in the melody. These melodies were all sung with the neutral sentence “Ha6 yat1 go3 zi6 hai6 si#[5]”. The responses of the eight participants showed that over 95% of the cases was as predicted by an ordinal mapping rule. Moreover, “listeners’ tone assignments are most generally characterized by an ordinal mapping rule similar to that used by composers of contemporary Cantonese songs” (p. 208. See chapter 3, § 3.2 of this thesis for a short discussion of their contemporary song analysis).

Inspired by Wong and Diehl's findings, I conducted a similar experiment focusing on Mandarin Chinese melody perception and its influence on lexical tone interpretation (Vondenhoff, 2007). As can be seen in figure 2 on page 17, only one of the Mandarin Chinese tones has a level character and three of them can be characterized as dynamic tones. Since dynamic tones are more complex than level tones (Wang, 1967), I took a slightly different approach in the perception experiment than Wong and Diehl did. Not only were listeners expected to "identify the target tone by comparing the musical on which it was sung to the immediately preceding tone and its musical note" (Wong and Diehl, 2002, p. 206), they were also expected to compare the target tone to the immediately succeeding tone.

[5] “The next word is si#”, in which the meaning of the final syllable was expected to be interpreted as "teacher" (/si1/), "to try" (/si3/), or "yes" (/si6/) depending on the note corresponding to it. The numbers do not refer to the tonal values (second column of figure 3), but to the 'names' of the tones (first column of figure 3).

Ten participants were presented with the target sentence "Xia4 yi2 ge ci2 shi4 yu#, zhe4 ge ci2 hen2 hao3"[6]. The melody of this sung sentence was adapted in such a way that the target syllable *yu#* altered between having a high F_0, an F_0 of medium height, and a low F_0. The preceding syllable *shi4* and following syllable *zhe4* were adapted in a similar fashion, thus combining 3^3 pitch levels into 27 stimulus melodies differing only slightly on the target syllable and its direct musical environment (see Vondenhoff, 2007 p. 31-33 for an overview of all the stimulus melodies). The participants were asked to listen to these 27 stimulus melodies three times in a random order, and to write down the character of the word they heard on the target syllable /yu#/ (see Vondenhoff, 2007, chapter 5 for a more detailed description of the perception experiment). The data acquired in the experiment is presented in table 1 below.

Unlike their Cantonese colleagues, the Mandarin listeners participating in the perception experiment perceived a lexical tone corresponding with the direction of the melody in only 40% of the cases (see table 1 for results). I hypothesized that there could be two main reasons for the discrepancy between these results and those obtained by Wong and Diehl (2002). Firstly, Wong and Diehl argued, "reliance on F_0 is probably greater in tone languages with more tonal distinctions" (p. 204). Since Mandarin Chinese has two tones less than Cantonese Chinese, this could be a possible explanation for the unexpected low correlation between the direction of the melody and the corresponding lexical tone.

A second reason could simply be that Mandarin listeners rely on strategies other than mapping the direction of the melody to that of the interpreted lexical tone. After all, an ordinal mapping strategy found by Wong and Diehl has only been shown to be the case for *level* lexical tones; the interpretation of *dynamic* lexical tones may very well be dependent on other elements of the melody, such pitch *register* in addition to the *direction* of F_0 change, as well as an interaction between preceding and succeeding syllables and their corresponding notes relative to the target syllable.

[6] "The next word is yu#, that is a correct word", in which the meaning of the ambiguous syllable *yu#* was expected to be interpreted as "roundabout" (/yu1/), "fish" (/yu2/), "rain" (/yu3/), or "bath" (/yu4/), depending on the melodic context. The numbers do not refer to the tonal values (second column of figure 2), but to the 'names' of the tones (first column of figure 2).

This brings us to the core of the present study: if not direction alone, which elements of the melody of a song do play a role in the recognition of lexical tones?

	yu1	**yu2**	**yu3**	**yu4**	
HHH	11	4	6	6	*27*
HHM	11	7	4	5	*27*
HHL	12	9	1	5	*27*
HMH	5	10	9	3	*27*
HMM	10	7	7	3	*27*
HML	8	8	10	1	*27*
HLH	1	6	17	3	*27*
HLM	4	2	19	2	*27*
HLL	4	4	16	3	*27*
MHH	14	7	1	5	*27*
MHM	11	5	4	7	*27*
MHL	14	6	1	6	*27*
MMH	11	6	8	2	*27*
MMM	14	4	7	2	*27*
MML	9	5	7	6	*27*
MLH	6	7	13	1	*27*
MLM	8	4	14	1	*27*
MLL	6	8	12	1	*27*
LHH	13	6	2	6	*27*
LHM	13	8	1	5	*27*
LHL	8	8	3	8	*27*
LMH	20	3	1	3	*27*
LMM	17	2	4	4	*27*
LML	19	3	2	3	*27*
LLH	2	3	19	3	*27*
LLM	4	5	13	5	*27*
LLL	7	4	12	4	*27*
Total	*262*	*151*	*213*	*103*	**729**

Table 1. Subjects' responses per melody (taken from Vondenhoff, 2007, p. 21). Grey cells contain the expected results.

4.2 *Language and music in Optimality Theory*

Before getting to try and find out more about the matter of how listeners of a song in a tone language identify the lexical tones of the lyrics, we will first have a look at the model it is going to be investigated with: Optimality Theory, and in particular the Gradual Learning Algorithm.

In short, what Optimality Theory (OT) entails is "a means for precisely determining which analysis of an input *best satisfies* (or least violates) a set of conflicting conditions. (...) The means that a grammar uses to resolve conflicts is to rank constraints in a *strict dominance hierarchy*. Each constraint has absolute priority over all the constraints lower in the hierarchy" (Prince & Smolensky, 1993, p. 2). The Gradual Learning Algorithm (GLA) "makes the assumption that selection points for natural language constraints are distributed normally, within the mean [μ] of the distribution occurring at the ranking value" (Boersma & Hayes, 2001, p. 49).

While the next chapter is dedicated in detail to *how* an OT/GLA model will help to find out more about what happens inside a listener's decision mechanism when listening to words in a melody, in this chapter I will shortly discuss *why* I think this model will yield the best insights. There are three reasons why OT and especially the GLA are very useful tools in the present research.

Firstly, and most importantly, the GLA has been shown to be a reliable tool in linguistic research. "If the language learner has access to an appropriate inventory of constraints, then a complete grammar can be derived, provided an algorithm is available that can rank the constraints on the basis of the input data" (Boersma and Hayes, 2001, p. 45), the algorithm in question of course being the Gradual Learning Algorithm. I will not discuss the GLA and how it works in great detail here, for now it suffices to say that Boersma and Hayes showed that it successfully deals with speech errors, intermediate well-formedness, free variation and matching corpus frequencies (Boersma & Hayes, 2001, p. 78). As we will see in chapter 5 especially these last two assets are of value in the case of the present study.

Secondly, this theory, although first introduced in the field of phonology, has already been used as a research tool to investigate the grey area that overlaps both language and music. In fact, it is not the work of phonologists that provided significant conceptual antecedents for OT (Prince & Smolensky, 1993, p. 2), it owes much to people like Wertheimer (1923), who considered the perception of structured grouping of input signals in terms of Gestalt Theory, and Lehrdahl and Jackendoff (1983), concerned with a formal generative theory of tonal music.

One instance of research surpassing the border between linguistics and music science is done by Hayes and MacEachern (1998), who formalized the assessment of the goodness of verse quatrains in English folk verse, finding "principles of how grouping and rhythmic structure can be cued with phonological material" (p. 505). Furthermore, Gilbers and Schreuder (2002; 2003) showed the resemblances between music and language, especially their rhythmic qualities by an analysis of rhythmic variability in OT (see chapter 2 as well). They regarded their conclusions "in relation to the study of temporally ordered behaviour" (Gilbers & Schreuder, 2002, p. 2). In their view, "the observation that language and music show so many similarities strengthens the hypothesis that the same structures and principles hold for all temporally ordered behaviour. (...) It is the way our brain works: our cognitive system structures the world surrounding us in a particular way in order to understand everything in the best way" (p. 22).

This brings me to the third and final reason for using OT as an analysis method: by formally showing that musical cues can be processed in terms of lexical meanings, I am hoping to add another argument in the debate about the similarities between language and music. After all, if it is true that constraints about the perception of melody can be hierarchically arranged, similar to their linguistic counterparts, in such a way that their ranking reflects a language user's choice preference for a certain lexical tone given a stimulus input (as we will see is the case in chapter 5), then it would seem plausible to assume a position with those who favor the theory that language and music share an underlying cognitive system.

5. Model design and experimental methods

So far we have seen that song composers in different tone languages follow certain strategies to allow themselves to be creative in their musical compositions, while at the same time preserving a fair level of intelligibility of the lexical tones that are characteristic of their language. Some of these strategies include using stereotypical melodies from a corpus of existing melodies, and thus limiting musical creativity for optimal intelligibility, or using the speech melody of the lyrics as the basis of the song melody, in this way allowing more freedom of composition. The latter strategy was also described by Wong and Diehl (2002, see chapter 3 for more detail), who called it ordinal mapping, or the preservation of (in their case Cantonese) lexical tones by using a relative, ordinal melody scale for song compositions in which the direction of the lexical tones is maintained.

Wong and Diehl also showed that Cantonese *listeners* use a similar strategy: when asked to interpret level tones in a sung melody, they were influenced by the direction of the melody, or the relative tonal distance of the preceding note with respect to the note on the target syllable. They showed that in 95% of the cases, a mapping was found between the direction of the melody and the lexical tone of the target syllable (see chapter 4).

Inspired by Wong and Diehl's research, Vondenhoff (2007, see chapter 4 for more detail) investigated whether such a strong correlation between the direction of the musical melody and the direction of lexical tones could also be found for Mandarin, in other words whether Mandarin listeners also use an ordinal mapping strategy similar to their Cantonese colleagues. Even though the results of the experiment indicated that the participants were indeed influenced by the melody when interpreting the lyrics of the songs, ordinal mapping occurred in less than 40% of the cases. The same figure was also found in a musical analysis of three contemporary Mandarin songs. I hypothesized that in addition to the direction of the melody, other musical cues might be of importance for the identification of lexical tone, both dynamic and level, in a Mandarin Chinese song.

But if the direction of the melody is not the only factor in Mandarin songs to help people understand the lyrics, what other cues can be found in the music that will do the trick? As we have seen in chapter 3, speakers and listeners of a complicated tone language like Mandarin Chinese are dependent on pitch to understand what is being said (or sung), after all (Wang, 1967; Chao, 1956; Yip, 1980; Wong & Diehl, 2004).

As I hypothesized in reference to the results of the perception experiment discussed in the previous chapter, intuitively one would expect a listener not only to map the direction of the melody when identifying the words of a song, but also for her to listen for matching tone levels. That is to say, when hearing a melody with a general low pitch, a listener would hear a lexical tone in the low register, and for a rather high melody a tone in the high register would mostly be picked as the optimal choice. For Mandarin Chinese, this would translate to a third tone and a first tone respectively. Additionally, when dropping the melody from a high note to a low note, it makes sense for a listener to hear a fourth tone syllable more often than in the case of the melody dropping only half the distance, say from a medium high note to a low note.

These intuitions are confirmed by Wang (1967) and Liu (2008). In his article *Phonological features of tone,* Wang created a set of features in which contour tones are distinguished from each other by specific binary features. This framework includes tone features for both tone movement and pitch (See Wang 1967, Table I on p. 97 for a complete overview of his features of tone). Using the features proposed in his article, Mandarin Chinese tones would be specified as shown in Figure 4 below. Again, Chao's tone letter notation (1933; see Chapter 5 of this thesis for a short description) proves useful for indicating the four Mandarin tones.

According to Wang, it is both the *direction* of the intonation (i.e. a rising contour or a concave contour) and the *pitch* of a syllable that are of importance for recognizing a lexical tone. Rephrasing this in the words of Liu (2008): "(There are) two major characteristics of Asian tonal systems: the contour and the register" (p. 150). This, in short, is what I hypothesize to be the answer to the question of how listeners recognize the lexical tones of the lyrics of a Mandarin Chinese song: while it has been shown that for the identification of *level* tones the direction of the melody is sufficient, for the identification of level tones as well as their dynamic counterparts, both the *contour* of the song melody and its *register* are important. Furthermore, I am curious to see whether the direction and register of the preceding syllable with respect to the target syllable are more important than the direction and pitch of the succeeding syllable (or vice versa).

	55	35	315	51
High	+	+	-	+
Central	-	-	-	-
Mid	-	-	-	-
Rising	-	+	+	-
Falling	-	-	+	+
Contour[7]	-	-	+	-
Convex	-	-	-	-

Figure 4. Mandarin tones in terms of Wang's features of tone.

5.1 Modeling the perception of Mandarin tones

As said, the purpose of this research is to investigate whether it is indeed true that Mandarin listeners rely on both the direction and the pitch of the melody in order to identify the lexical tones of words in a song. In order to test this hypothesis, we will have a look inside the grammar of a virtual Mandarin Chinese listener, using an Optimality Theoretic framework (Prince and Smolensky, 1993) modeled after Boersma and Hayes' Gradual Learning Algorithm (2001). I discussed this algorithm and its purposes and advantages in the previous chapter, where it was seen that this device is an excellent tool for analyzing frequency based corpuses of language data and that it is capable of handling optionality.

In order to find out what auditory cues play the most important role in the perception and interpretation of Mandarin lexical tones in a musical environment, said virtual listener will learn to perceive and interpret the four tones under seven different learning conditions, each condition being a different set of grammar rules or cue constraints. These learning conditions

[7] For two reasons I will henceforth call the Contour tone feature *Concave*: in the first place, I wish to avoid ambiguity as I referred in Chapter 2 to contour tones as dynamic lexical tones as opposed to static or level lexical tones. Secondly, because the word concave describes the movement of descending/rising melody in a better way.

will be discussed in the next section of this chapter. The idea is that the better the constraint set represents the data from the perception experiment, the less differences there will be between the virtual listener's simulated perception of the lexical tones and the perceptions of the real listeners who participated in the experiment.

For this purpose, I translated Wang's tone features (1967, see previous section for more detail) into twelve cue constraints for each of the four lexical tones. This resulted in a set of 48 cue constraints describing the melodic movements of the 27 experiment stimuli. In addition to Wang's original features I added the feature "*Low*" describing the low melody contours that can be found in the stimuli as well. As I said in chapter 4, the stimulus melodies used in the perception experiment consist of three pitch levels. These pitch levels are represented in the grammar by constraints such as *[High] /3/ ("A high pitch is not a third tone"), violated in those cases in which the target syllable has a high pitch, *[Mid] /1/ ("A medium high pitch is not a first tone"), violated in those cases in which the target syllable has a medium pitch and *[Low] /2/ ("A low pitch is not a second tone"), violated in those cases in which the target syllable has a low pitch.

Even though in Wong and Diehl's research only the correlation between the preceding note and the target syllable was taken into account, in the Mandarin perception experiment the correlation of both preceding and succeeding melody pitches were compared to the target syllable, because contour tones "are sequences of level tones in languages with a suprasegmental representation of tone" (Leben, 1980, p. 39). In other words, it seems plausible to regard "a falling tone as a sequence of HL, a rising tone as a sequence of LH, and so forth" (Leben, 1980, p. 38). As three of the four the Mandarin lexical tones are contour tones, and therefore are more complex than the level tones investigated in Wong and Diehl's research, it seems plausible that the melody movements of both directly neighboring syllables are involved, instead of only the preceding one. For instance, a tone may seem high pitched relative to the preceding note, but if followed by a higher note, it would be perceived as an overall rising melody in comparison.

Therefore, in addition to those constraints based on Wang's tone features, I added cue constraints for preceding and succeeding tonal movements to the grammar of the virtual Chinese listener. Taking all these considerations into account, the following auditory cues were used for the simulation: High, Mid, Low, Level, Convex, Concave[8], Preceding level, Preceding rising, Preceding falling, Succeeding level, Succeeding rising, Succeeding falling. An overview of all cue constraints can be found below in table 2.

Pitch register constraints			
*[High] /1/	*[High] /2/	*[High] /3/	*[High] /4/
*[Mid] /1/	*[Mid] /2/	*[Mid] /3/	*[Mid] /4/
*[Low] /1/	*[Low] /2/	*[Low] /3/	*[Low] /4/
Direction constraints			
*[Level] /1/	*[Level] /2/	*[Level] /3/	*[Level] /4/
*[↗↘] /1/	*[↗↘] /2/	*[↗↘] /3/	*[↗↘] /4/
*[↘↗] /1/	*[↘↗] /2/	*[↘↗] /3/	*[↘↗] /4/
Preceding constraints (all direction constraints)			
*[Lev, prec] /1/	*[Lev, prec] /2/	*[Lev, prec] /3/	*[Lev, prec] /4/
*[↗, prec] /1/	*[↗, prec] /2/	*[↗, prec] /3/	*[↗, prec] /4/
*[↘, prec] /1/	*[↘, prec] /2/	*[↘, prec] /3/	*[↘, prec] /4/
Succeeding constraints (all direction constraints)			
*[Lev, succ] /1/	*[Lev, succ] /2/	*[Lev, succ] /3/	*[Lev, succ] /4/
*[↗, succ] /1/	*↗, succ] /2/	*[↗, succ] /3/	*[↗, succ] /4/
*[↘, succ] /1/	*[↘, succ] /2/	*[↘, succ] /3/	*[↘, succ] /4/

Table 2. Cue constraints used in the simulation.

As is shown in table 2, these cue constraints can be divided into two general types: direction constraints that govern the movement and direction of melodic contour of the song, and pitch register constraints, which are of use for recognizing height and pitch level. Examples of direction constraints

[8] A convex tone feature translates as a “rising and falling contour” and a concave tone feature refers to a “falling and rising contour”. For the sake of brevity, I used the respective icons ↗↘ and ↘↗ to refer to these types of tone features.

are *[↗↘] /4/ (“A syllable with a fourth tone does not have a convex melody”), *[Lev] /3/ (“A syllable with a third tone does not have a level melody”) and *[↘, prec] /1/ (“A syllable with a first tone is not preceded by a syllable on a higher pitch”). An example of a pitch constraint is *[High] /3/ (“A syllable with a third tone does not have a high pitch”). The cases in which these constraints are violated are shown in tableau 1 below.

	*[High] /yu1/	*[Mid] /yu1/	*[Low] /yu1/	*[↗↘] /yu1/	*[↘↗] /yu1/	*[Lev] /yu1/	*[↗, prec] /yu1/	*[↘, prec] /yu1/	*[lev, prec] /yu1/	*[↗, succ] /yu1/	*[↘, succ] /yu1/	*[lev, succ] /yu1/
HHH	*[9]					*						*
HHM	*										*	
HHL	*										**	
MHH	*						*					*
MHM	*			*			*				*	
MHL	*			*			*				**	
LHH	*						**					*
LHM	*			*			**				*	
LHL	*			*			**				**	
HMH		*			*			*		*		
HMM		*						*				*
HML		*						*			*	
MMH		*							*	*		
MMM		*				*			*			*
MML		*							*		*	
LMH		*					*			*		
LMM		*					*					*
LML		*		*			*				*	
HLH			*		*			**		**		
HLM			*		*			**		*		
HLL			*					**				*
MLH			*		*			*		**		
MLM			*		*			*		*		
MLL			*					*				*
LLH			*						*	**		
LLM			*						*	*		
LLL			*			*			*			*

Tableau 1. Constraint violations per stimulus melody.

[9] One star means the constraint is violated once, two stars means it is violated twice. Double violation of a constraint happens in those cases in which the pitch drops or rises two levels instead of one.

Another division can be made between preceding constraints and succeeding constraints. Preceding constraints would be needed to recognize tone movement of the preceding note with respect to the target, like *[↗, prec] /4/ ("a syllable with a fourth tone is not preceded by a syllable on a lower pitch"). Similarly, succeeding constraints are there to recognize tone movement of the following note with respect to the target, for example *[Lev, succ] /3/ ("a syllable with a third tone is not succeeded by a syllable on the same pitch"). Again, those cases in which these constraints are violated can be found in tableau 1.

To put it in a more formal way, let's suppose that the input melodies have the following structure $<_{-1}\sigma\ ,\ _{0}\sigma\ ,\ _{+1}\sigma\ >$, where $_{0}\sigma$ is the target syllable and $_{-1}\sigma$ and $_{+1}\sigma$ are the preceding and the succeeding syllable respectively. If we then consider the structure mentioned above, then we can define each of the constraints as follows.

*[High] /1,2,3,4/
tone 1 (2, 3, 4)[10] = 0 ← $(F_0\ {}_0\sigma = H)$

*[Mid] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_0\sigma = M)$

*[Low] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_0\sigma = L)$

*[Level] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_{-1}\sigma = F_0\ {}_0\sigma = F_0\ {}_{+1}\sigma)$

*[↗↘] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_{-1}\sigma < F_0\ {}_0\sigma) \wedge (F_0\ {}_0\sigma > F_0\ {}_{+1}\sigma)$

*[↘↗] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_{-1}\sigma > F_0\ {}_0\sigma) \wedge (F_0\ {}_0\sigma < F_0\ {}_{+1}\sigma)$

*[↗, prec] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_{-1}\sigma < F_0\ {}_0\sigma)$

*[↘, prec] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← $(F_0\ {}_{-1}\sigma > F_0\ {}_0\sigma)$

[10] The statement is true for lexical tone 1, lexical tone 2, lexical tone 3 or lexical tone 4.

*[Lev, prec] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← ($F_{0\ -1}\sigma$ = $F_{0\ 0}\sigma$)

*[↗, succ] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← ($F_{0\ +1}\sigma$ < $F_{0\ 0}\sigma$)

*[↘, succ] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← ($F_{0\ +1}\sigma$ > $F_{0\ 0}\sigma$)

*[Lev, succ] /1,2,3,4/
tone 1 (2, 3, 4) = 0 ← ($F_{0\ +1}\sigma$ = $F_{0\ 0}\sigma$)

Even though Wang argues that "(positing) the three features contour, falling and rising (would be redundant and) no paradigm would ever use all three of these", all three of them are included in the present study nevertheless. I agree with Wang that this creates a certain level of redundancy; after all, a convex feature can easily be represented with a preceding rising and a succeeding falling constraint together. Still, having these extra constraints will enable me to investigate their influence on the perception of lexical tone in different circumstances. The way this is done will be described in the next section.

5.2 Learning conditions and results

In the previous section I hypothesized that interpreting the lexical tones of a Mandarin Chinese song depends on four factors: the direction of the melody, the pitch of the melody, the preceding melody direction and pitch relative to the target syllable, and the succeeding melody direction relative to the target syllable. To find out what the influences are of each of these factors, the virtual listener will recognize the four lexical tones of the Mandarin language by learning the ranking of the constraints in its grammar by means of the Gradual Learning Algorithm. For this purpose I devised six different learning conditions, or cue constraint sets, as well as one condition in which there will not be any ranking. By an absence of a ranking, I mean that the virtual listener will not be able to pick an optimal candidate on the basis of constraint ranking, in other words, its perception of Mandarin tones will depend on chance alone.

1. No constraint ranking (no learning)
2. All constraints included
3. Direction constraints
4. Pitch constraints
5. Preceding constraints
6. Succeeding constraints
7. Preceding and succeeding constraints together

Having divided the 48 constraints into these seven sets, I will be able to find under which condition my virtual listener is best capable of imitating the data of the perception experiment most accurately. Depending on how well each set represents the original data obtained in the perception experiment, the generated set of input/output pairs will be an accurate imitation of the original set. Comparing the generated distribution of melody/lexical tone pairs from the virtual listener with the data obtained from real Chinese listeners participating in the perception experiment, I will be able to see which set of cue constraints yields the least difference. The smaller the difference, the better the set of constraints represents the data, or in short: the better the rules, the better the copy.

But before my virtual listener could even start to recognize the lexical tones in the melodies presented to it (see chapter 4 for more detail about the stimulus melodies), it was first necessary to teach it how to recognize those melodies. Before initiating the learning phase, all constraints were ranked equally high; in this case I assigned a ranking value of 100.0 to all of them. The actual ranking value does not matter, what does matter is the notion that the constraints will gradually shift along the scale when the ranking values change.

See Tableau 2 as an example of what the initial constraint ranking of a simplified version of the OT grammar looks like. For the sake of simplicity I will not be showing a tableau with all 48 constraints, only the ones for recognizing low pitch, falling preceding melody contour and rising succeeding melody contour. As initial ranking, all constraints start at the same value, in this case at ranking value 100.0. Since the model's decisions are based on a hierarchical constraint ranking, having all constraints at the same ranking results in the virtual listener not being able to choose the right candidate. More specifically from a perception point of view: it is not able to recognize any lexical tones based on the melodic input.

[MLH]	*[Low] /1/	*[Low] /2/	*[Low] /3/	*[Low] /4/	*[↘, prec] /1/	*[↘, prec] /2/	*[↘, prec] /3/	*[↘, prec] /4/	*[↗, succ] /1/	*[↗, succ] /2/	*[↗, succ] /3/	*[↗, succ] /4/
☞ /yu1/	*				*				*			
☞ /yu2/		*				*				*		
☞ /yu3/			*				*				*	
☞ /yu4/				*				*				*

Tableau 2. Initial constraint ranking.

The next step in the learning phase consists of presenting melody-tone distribution pairs, the learning data, obtained in the perception experiment (see table 1 in chapter 4 for the experiment data). The learning procedure I used was the stochastic Gradual Learning Algorithm, which in short implies that for each evaluation of an input-output pair, a small amount of noise is temporarily added to the constraint rankings. Each time an auditory input signal (i.e. a stimulus melody) is presented, the grammar decides which output candidate (i.e. which of the four tones) is the best choice. Since the grammar is error driven, it will notice when the candidate generated according to the current ranking is not the same candidate as the one presented to it. This is what we can see happening in tableau 3 below. When presented with stimulus melody MLH , according to the ranking of its constraints, /yu1/ is picked as the optimal candidate of the four options (indicated by the pointing finger), whereas the learning datum told it to go for /yu3/ (indicated by the check sign).

[MLH]	*[Low] /2/	*[Low] /3/	*[↗, succ] /4/	*[↗, succ] /3/	*[↘, prec] /3/	*[Low] /1/	*[↘, prec] /2/	*[Low] /4/	*[↗, succ] /2/	*[↗, succ] /1/	*[↘, prec] /1/	*[↘, prec] /4/
☞ /yu1/						*				**	*	
/yu2/	*!						*		**			
✓ /yu3/		*!		**	*							
/yu4/			*!*					*				*

Tableau 3. The current ranking generates the wrong output candidate.

Being error driven, the algorithm will take its mistake as an indication that the grammar needs to change in such a way that in the future the right candidate is more likely to be generated. It will take measures to move each of the offending constraints a small step along the hierarchic scale. Since it is likely that those constraints for which the *correct* candidate suffers violation are ranked too high, and those constraints for which the learner's actual *output* suffers violation are ranked too low, the grammar will move the former a small step down the ranking scale and the latter a small step higher up the ranking scale (see for more detail Boersma and Hayes, 2001, p. 52-53). This small step is called the plasticity of the constraints, which Boersma and Hayes define as "the numerical value by which the algorithm adjusts the constraints' ranking value at any given time" (p. 52). The values of the plasticity in the simulation are shown in table 3 below. The learning simulation was run in accordance with this schedule, based on Boersma and Hayes' (2001, p. 79-80), but slightly adapted for the experiment.

	Plasticity	Evaluation noise
First 100,000 runs	2.0	2.0
Second 100,000 runs	0.2	2.0
Third 100,000 runs	0.02	2.0

Table 3. Learning Schedule.

In tableau 4 below is shown what happens when the algorithm decides to move the offending constraints. The arrows indicate in which direction the constraints are moved.

[MLH]	*[Low] /2/	*[Low] /3/	*[↗, succ] /4/	*[↗, succ] /3/	*[↘, prec] /3/	*[Low] /1/	*[↘, prec] /2/	*[Low] /4/	*[↗, succ] /2/	*[↗, succ] /1/	*[↘, prec] /1/	*[↘, prec] /4/
☞ /yu1/						←*				←**	←*	
/yu2/	*!						*		**			
✓ /yu3/		*! →		** →	*→							
/yu4/			*!*					*				*

Tableau 4. Adjustment of the constraint ranking.

After running the learning simulation a few more times and repeating the algorithm cycles of generation and adjustment described above, the grammar will have moved its constraints considerably along the ranking scale. We can see in tableau 5 below that the constraints preventing /yu3/ to be generated as the optimal candidate - *[Low] /3/, *[↗,succ] /3/ and *[↘,prec] /3/ - have gained lower ranking and have therefore become less offensive than previously was the case. Still, the correct candidate cannot be perceived according to this constraint ranking, which is why the grammar will continue the process of evaluation and reranking, shown in said tableau 5. Eventually though, the system will obtain a constraint ranking that allows the correct candidate to be chosen (tableau 6). The algorithm will notice no mismatch between the learning datum and the generated output and no action is taken.

[MLH]	*[Low] /2/	*[Low] /1/	*[Low] /3/	*[↗, succ] /4/	*[↗, succ] /3/	*[↘, prec] /3/	*[↘, prec] /2/	*[Low] /4/	*[↗, succ] /2/	*[↗, succ] /1/	*[↘, prec] /1/	*[↘, prec] /4/
/yu1/		*!								**	*	
/yu2/	*!						*		**			
✓ /yu3/			*!→		**→	*→						
☞ /yu4/				←**				←*				←*

Tableau 5. Further adjustment of the constraint ranking

[MLH]	*[Low] /2/	*[Low] /1/	*[↗, succ] /4/	*[↗, succ] /1/	*[↘, prec] /1/	*[Low] /3/	*[↘, prec] /2/	*[Low] /4/	*[↗, succ] /2/	*[↗, succ] /3/	*[↘, prec] /3/	*[↘, prec] /4/
/yu1/		*!		*	**							
/yu2/	*!						*		**			
☞ /yu3/						*				**	*	
/yu4/			*!*					*				*

Tableau 6. Learning datum and generated output are the same: the system has reached an optimal constraint ranking for this auditory input.

Now consider what happens if the grammar is presented with a different learning datum, one which it has not encountered yet and for which it has not had the opportunity to evaluate its constraint ranking. This is of course something that happens frequently, as the data consists of 27 stimulus melodies with four different possible candidates. If that is case, once more, the algorithm will cycle through the steps discussed above and make adjustments to the constraint ranking values as soon as a mismatch is encountered (tableau 7).

[HLL]	*[Low] /2/	*[Low] /1/	*[↗, succ] /4/	*[↗, succ] /1/	*[↘, prec] /1/	*[Low] /3/	*[↘, prec] /2/	*[Low] /4/	*[↗, succ] /2/	*[↗, succ] /3/	*[↘, prec] /3/	*[↘, prec] /4/
/yu1/		*!			**							
/yu2/	*!						**					
✓ /yu3/						*! →					**→	
☞ /yu4/								←*				←**

Tableau 7. Adjustment of the constraint ranking after evaluation of a new learning datum.

Eventually, after repeatedly going through these steps of generation, comparison and adjustment, the grammar "stabilizes with ranking values that yield a distribution of generated outputs that mimics the distribution of forms in the learning data" (Boersma and Hayes, 2001, p. 54), thus including relative frequency of free variants.

After completing the learning process for each of the learning conditions, the grammar generated a new set of input and output form distributions, based on the constraint rankings in each of the different grammars corresponding to the seven learning conditions. As I said at the beginning of this section, the better the set of cue constraints, the less of a difference there will be between the output generated by the algorithm and the answers given by the real Chinese listeners who participated in the perception experiment.

These differences were calculated as follows. After each set of 100,000 runs of presenting the learning data, the positive differences between the learning data and the generated data was calculated. For instance, after 300,000 runs of the data in the All Constraints Included condition, of the 100,000 cases the input HHH was presented to the virtual listener, /yu1/ was generated 45,428 times (see first row, first column of table 5 in the Appendix). Of the 27 cases HHH was presented to real listeners, /yu1/ was heard 11 times (see first row, first column of table 1). These numbers yield the fractions 0.454 and 0.407 respectively, which results in a positive difference of 0.047. This positive difference was computed after each of the 100,000 runs, resulting in three numbers for each of the input/output pairs. An average was calculated of these three, after which the sum of the average positive differences per input was multiplied by 100 (to turn them into percentages) and divided by two (to rule out double differences). The results of these computations are shown in table 4. Each of the next few sections will be dedicated to the results of the learning conditions individually.

% Av. Δ per stimulus	No Learning	All	Direction	Pitch	Prec	Succ	Prec+succ
HHH	15.74	10.43	8.26	13.22	13.39	9.14	11.57
HHL	27.90	10.51	12.68	9.23	24.67	9.76	9.06
HHM	16.90	7.14	1.60	6.45	13.67	2.87	7.26
HLH	37.94	12.74	15.96	13.44	17.02	21.18	14.44
HLL	34.31	7.14	15.99	5.35	16.27	31.64	18.89
HLM	31.50	6.77	17.63	14.74	20.15	23.70	16.17
HMH	20.19	11.79	21.15	28.47	18.89	19.76	22.21
HML	21.22	9.95	12.90	24.87	14.50	28.84	12.40
HMM	13.87	10.26	21.23	1.80	20.08	7.46	17.00
LHH	23.09	5.48	11.12	4.81	6.97	20.31	9.95
LHL	14.04	12.60	12.55	14.41	23.28	16.05	22.05
LHM	27.74	10.40	10.30	9.05	2.29	10.54	7.75
LLH	45.26	12.01	22.58	16.65	34.82	25.32	27.83
LLL	20.10	6.14	14.50	14.33	12.66	17.43	12.97
LLM	23.05	10.50	4.76	9.84	17.36	10.55	13.62
LMH	48.73	7.22	8.04	13.59	21.75	50.81	23.68
LML	45.56	5.56	9.20	23.61	17.80	27.15	19.87
LMM	37.81	8.48	12.02	19.69	15.25	24.21	12.22
MHH	27.71	4.86	10.56	9.02	7.56	24.19	10.10
MHL	26.98	9.72	9.58	8.49	4.65	11.91	2.79
MHM	16.58	8.10	8.12	9.40	14.06	8.16	14.55
MLH	24.24	12.30	8.37	13.04	6.02	8.49	9.87
MLL	24.18	15.16	6.19	16.78	7.58	27.68	7.51
MLM	31.32	13.14	10.07	12.44	11.88	9.78	11.56
MMH	20.31	10.18	25.06	9.75	12.98	19.31	18.01
MMM	27.96	7.07	11.81	8.56	17.85	12.32	16.08
% Av. Δ All Stimuli	26.43	9.50	12.43	12.79	14.85	18.30	13.96

Table 4. Percentage difference per individual stimulus for each of the different learning condition. The grey cells are the "anomalies" whose differences were 20% or more.

5.2.1 *No constraint ranking*

The first learning condition does not really qualify as a learning condition, as it does not involve any learning at all. By not presenting any data to the grammar, the learning algorithm will not get a chance to go through the learning cycles that cause it to adjust its constraints' rankings (described above). Because of this, each constraint is still ranked at its initial value of 100.0. Since it is the hierarchic constraint ranking that determines which candidate is to be generated, an absence of such a constraint ranking implies that the virtual listener cannot take proper interpreting decisions. This is a very important notion, for as we will see, comparing the results of the no learning condition with the data from the experiment shows just how much the real listeners' decisions are based on chance.

Without any constraint ranking, I found a difference of 26.43% between the original data and the input/output pairs generated by the grammar in the no constraint condition (table 4). This number is almost exactly the same as the differences calculated after distributing the four lexical tones equally per input melody (i.e. if each candidate were chosen once every four times). For an equal distribution of 25,000 times per 100,000, a difference of 25.96% could be calculated. In other words, almost 75% of the participants' answers could be predicted by a machine picking randomly from the four candidates each time an input melody was presented to it (see table 4 in the appendix for the data generated under this condition). Admittedly, the grammar without learning performs worse than any of the other grammars in almost all of the cases, but such a high result is nevertheless rather sobering. It suggests that Mandarin might not be influenced by melodic cues after all.

As Vondenhoff (2007) suggested (see chapter 3), even though the traditional rules of musical composition in China used to consist of a rich system of rules governing the relationship between linguistic tone and melody, these rules have mostly been abandoned by contemporary songwriters. This was earlier proposed by Chao (1956) and Levis (1936). Because of this relaxation of the traditional Chinese art of song composition, listeners may not be used anymore to interpreting the musical information possibly hiding inside melodies in terms of lexical tones. But before drawing any conclusions, let us first have a look at the grammar's performance in the other six learning conditions.

5.2.2. Direction constraints

For this learning condition, the nine constraints on the bottom rows of Table 2 were used by the grammar of the virtual listener. These included those constraints that are responsible for recognizing level, convex, concave, rising and falling features of the stimulus melodies. This was also the condition investigated by Wong and Diehl in their research for Cantonese listening strategies. As said in the beginning of this Chapter, Wong and Diehl found an almost 95% correlation between the mapping of the direction of the song melody and the direction of the lexical tone. When reproducing their experiment, I did not find such a strong correlation (Vondenhoff, 2007). Still, with a mapping between direction of melody and lexical tones of 40%, I expected that this set containing direction constraints will perform fairly well.

However, looking at the numbers, we see that compared to the original data set, the positive difference after 100,000 runs of presenting the data yields an average percentage of 12.4% difference, or to say it differently 87.6% correct (table 3). At first sight, this appears to be pretty accurate. But when compared to the condition in which the grammar was not allowed to learn a hierarchic constraint ranking (§ 5.2.1.), 87.6% correct seems not so impressive next to a 75% correct of the no learning condition. Still, this part of the answers that could not be "predicted" on the basis of chance could be predicted by a grammar consisting of direction constraints. In other words, it seems that listeners of Mandarin Chinese songs do use the direction of the melody to some extent when interpreting the lexical tones of the lyrics.

If we then proceed to look more closely at the results, a rather surprising pattern emerges. Looking at the generated distribution pairs for which the model performed worst, in other words the anomalies of the results that have a difference of over 20%, we can see that the pairs for which the model performed worst are those involving lexical tones 1 and 3. More specifically,

[LLH] /3/, [MMH] /1/, [HMH] /3/ and [HMM] /3/ were the ones that had the highest percentages (individual results can be found in table 6 in the appendix). In other words, these are the instances in which the model had difficulty imitating the data from the perception experiment. Incidentally, one of these (LLH) is also one of the cases in which real listeners found the least trouble in interpreting the lexical tone of the target syllable (see table 1).

5.2.3. Pitch constraints

Comparing the results of the "pitch constraints" condition of the simulation with the data acquired in the perception experiment, a similar pattern can be seen as the one described in the previous section. Those instances in which the model perceived a lexical tone with a difference of 20% or higher (in this set only three instances of such anomalies occurred), the lexical tones to be interpreted were tones 1 or 3. Additionally, the general performance of the pitch constraints condition with an overall difference of around 12.8% (or 87.2% correct), turned out to be only very slightly less accurate than the "direction constraints" set. This is in accordance with my hypothesis stating that the pitch of the melody is more or less equally important for Mandarin listeners as direction of that melody. In other words, the auditory input of the fundamental frequency of a syllable alone seems to be as sufficient a cue for interpreting the lexical tone as the direction.

The two sets together, however, should complement each other and the virtual listener is expected to perform better in the "all constraints included" condition. This is indeed the case, as we shall see in § 5.2.4. This intuition is also reflected by the complementation of anomalies occurring in the two conditions: in three of the six cases in which a difference of 20% or higher per stimulus melody was found in the "direction" set, these melodies in the "pitch" set had less than a 20% difference.

Let us go back at the similarities between the "pitch constraints" set and the "direction constraints" set mentioned at the beginning of this section. Looking at the distributions of the individual input/output pairs after running the simulation with the "pitch" constraints only[11], once more we can see that those cases in which the model had trouble generating the input-output pairs of the perception experiment, all of which involved lexical tones 1 and 3.

5.2.4. All constraints included

As previously mentioned in § 5.2.2, the "all constraints included" condition yielded the best results. With an average percentage difference of 9.5% this data achieves over 90% accuracy. This is to be expected, for it implies that even though each separate set has their own "specialty" and generates their own "gaps" or anomalies, these anomalies are covered by combining the sets together. After running the simulation under the "all included" condition, not one anomaly was found. Seeing that of all the constraint sets, the "all included" set seems to best represent the experiment data, the hypothesis is that Mandarin use both the direction and the pitch register of the melody of a song to understand the lyrics.

5.2.5. Preceding or Succeeding constraints

An interesting outcome of the simulation can be seen if we look at the results of the "preceding constraints" condition compared to the "succeeding constraints" condition. As stated at the beginning of this chapter, I was curious to find whether the influence of succeeding melody movement is just as strong as the movement of the melody of the syllable previous to the target syllable, stronger, or even less strong. But as can be seen in table 4, the difference between the original input-output distributions and the computer generated ones is much greater for the "succeeding constraints" set (with 18.3% difference almost nearing the percentage yielded in the "no ranking" constraint set) than for the 'preceding constraints set" (15.5%). Obviously, the latter did a better job at representing the data, which is also reflected by the fact that the two sets together hardly performed any better, as we shall see in the next paragraph.

Once more, when looking at the anomalies (20% difference or more) at the level of individual input-output pairs, we can see the same curious

[11] To avoid having to impose tedious lists of input-output pairs upon the reader, I refer to the appendix for an overview of all the input-output pair distributions.

pattern emerging as previously described for the "direction" and "pitch" simulation conditions: each of the cases in which the model had difficulty imitating the original data involved lexical tones 1 and 3.

5.2.6. Preceding and Succeeding constraints

Even though taken separately the "preceding constraints" set had a lower percentage difference than the "succeeding constraints" set, one would still expect them to perform better when combined in one big set. This intuition is confirmed when comparing anomalies in the "preceding" column with those in the "succeeding" column of table 3: in three of the six cases of anomalies in the "preceding" column, there is not an anomaly in the "succeeding" column. This suggestion of complementation notwithstanding, together they succeed in only very slightly improving their accuracy in re-generating the pair distributions from the perception experiment. While the virtual listener generates a set of data with a 14.85% difference when presented with only preceding constraints, when presented with both preceding and succeeding constraints it re-generates a data set that differs 13.96% from the original data. This 0.89% can hardly be called an improvement, which means that adding constraints that govern the recognition of succeeding melody contours does not make that much of a difference.

In addition, the same curious pattern can be found as I described in the previous paragraphs and which by now it should not come as a surprise: two instances of the four anomalies were ones where real listeners had a strong preference for either tone 1 or 3 and this preference was not shared by the virtual listener.

5.*3* ***Summary***

In short, what we have seen in the past few sections, is that the "no constraint ranking" set turned out to represent the experiment data with an accuracy of 75%. This suggests that the Mandarin listeners participating in the experiment based a large portion of their decisions on guessing, instead of using the melody direction or pitch as a signal for lexical tones.

Furthermore, the "direction" condition yielded better results than the "no ranking" condition, suggesting that those answers that were not random guesses have been based on the direction of the melody. In addition, the "pitch" condition yielded very similar results, which would also suggest that some portion of the answers given in the experiment were based on pitch as well. The "all included" condition, however, yielded the highest accuracy of all the constraint sets. This would strengthen the assumption that for Mandarin Chinese listeners to identify the lexical tones of a Mandarin song, the direction of the melody and the pitch register are of equal importance.

Finally, the results of the simulations suggest that the pitch register and melodic contour of the preceding syllable with respect to the target syllable is of more influence than those of the succeeding syllable. This is supported by the fact that "preceding" and "succeeding" constraints combined do not much improve their accuracy in representing the experiment data.

6. Conclusion and discussion

With this research I intended to focus on two main goals. The first one, as stated in the introduction, was to participate in the debate about whether language and music share common cognitive pathways. Research has shown that in many areas of linguistics that used to be previously thought to be unique for language, music shares similar characteristics, and vice versa. In the fields of semantics, prosody, syntax and phonology many striking similarities were found in previous research.

By modeling the perception and processing of music and language simultaneously, I hoped to add another argument to the growing body of evidence in favor of a shared system for both language and music. If both can be modeled with one algorithm, maybe language is not so special after all.

The second and main goal of this research was to investigate which elements of Mandarin Chinese songs are of importance when interpreting the lexical tones of the individual words of the lyrics. Unlike Wong and Diehl's findings for the Cantonese Ordinal Mapping, I hypothesized to find that pitch is just as influential as the direction of the melody. In addition, I was curious to see whether Mandarin Chinese listeners prefer using the preceding syllable as a reference for interpreting the lexical tone of a specific syllable, or rather the melodic characteristics of the syllable that followed.

For this purpose, I created a grammar model of a Mandarin Chinese listener, using Optimality Theory and the Gradual Learning Algorithm. The grammar was my virtual Mandarin Chinese listener the data acquired in the perception experiment I applied six different grammar conditions. For each grammar set, or condition, I computed the differences with the original set of data. The idea was: the bigger the difference, the worse the set of cue constraints was able to represent the data from real listeners.

What I found was that, when forced to rely on melodic auditory cues alone and in the absence of context, listeners of Mandarin Chinese songs seem to base their decisions for a large portion on chance. However, even if the characteristics of the melody are not as important for Mandarin listeners as for their Cantonese colleagues, when deprived of context, for the interpretation of the lexical tones of words in a sung environment both the direction of the melody and pitch register appear to be of equal importance.

Furthermore, the results suggest that the contour of the melody directly preceding a syllable is more important than the characteristics of the syllable directly following it. Even when the sets of preceding and succeeding constraints are united in one grammar, the difference between the generated data and the original experiment data does not become much smaller.

Looking at these conclusions, I cannot but put forward an argument that is twofold. Firstly, the notion that the real listeners seemed to have based their lexical tone interpretations on chance may be inherent to an internally faulty experiment. Because the singer of the stimulus melodies that were used in the perception experiment is not native to the Chinese language[12], the melodies may have confused the participants in the experiment. As discussed in chapter xx, traditional Mandarin singers were expected to insert grace notes for a clearer diction during singing (Chao, 1956), grace notes that, for obvious reasons, were not incorporated in the melodies presented to the real listeners. Their confusing may have resulted in them picking more or less randomly from the four lexical tones each time a stimulus was presented to them and abandoning their assumed strategy of using the contour and register of the melody.

But secondly, and perhaps more reasonably than confusion among the population participating in the experiment, would be to assume that listeners of contemporary Mandarin Chinese songs are not in the habit of using melody contour and pitch register of the melody when interpreting the lyrics of those songs. If, as suggested (Levis, 1936; Chao, 1956; Vondenhoff, 2007), contemporary song composers do not incorporate the features of lexical tones in the melody of a song anymore, then there is no use for listeners of those songs to try and find musical elements to help them indentify those lexical tones. In addition, as was earlier suggested in chapter

[12] Unfortunately, I was unsuccessful in getting any of the Chinese singers I approached to sing the stimulus melodies for me, so I was forced to sing them myself. Naturally, I asked a Chinese teacher to check whether my pronunciation was correct and as natural possible. More details can be found in Vondenhoff, 2007.

4, there may be less need to depend on fundamental frequency in a language with a smaller inventory of lexical tones (This would of course be the primary reason for the relaxation of those rules that govern the relation between melody and lexical tone, in the first place). Instead, they may rely on other factors, like the context of the song, other prosodic cues such as duration, voice quality such as creaky or breathy voice, or perhaps their expectation of what songs are supposed to be about.

Still, even when deprived of such a context to aid the listeners' expectations, the direction of the melody and its pitch seem to get a good shared second place.

Appendix

The output tables below contain pairs of generated melody-tone distributions of only the third run of presenting 100,000 learning data. The reason is that it is plausible that the data acquired after 300,000 runs is most accurate.

Output distributions without constraint ranking				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.250	0.250	0.251	0.249
HHL	0.248	0.251	0.251	0.250
HHM	0.248	0.249	0.252	0.251
HLH	0.252	0.249	0.250	0.249
HLL	0.249	0.252	0.249	0.249
HLM	0.248	0.251	0.250	0.251
HMH	0.249	0.251	0.251	0.249
HML	0.252	0.251	0.248	0.249
HMM	0.252	0.251	0.248	0.250
LHH	0.251	0.250	0.250	0.249
LHL	0.251	0.247	0.252	0.250
LHM	0.250	0.250	0.249	0.250
LLH	0.251	0.248	0.251	0.250
LLL	0.252	0.248	0.251	0.249
LLM	0.248	0.250	0.251	0.250
LMH	0.253	0.248	0.250	0.249
LML	0.248	0.250	0.249	0.253
LMM	0.251	0.252	0.249	0.247
MHH	0.250	0.250	0.249	0.250
MHL	0.249	0.250	0.251	0.250
MHM	0.251	0.250	0.249	0.250
MLH	0.251	0.250	0.248	0.251
MLL	0.250	0.251	0.248	0.252
MLM	0.251	0.249	0.251	0.249
MMH	0.249	0.251	0.251	0.248
MML	0.251	0.252	0.248	0.249
MMM	0.247	0.252	0.251	0.250

Table 5. Output distribution for the "no constraint ranking" condition.

Output distributions All constraints included				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.454	0.203	0.176	0.167
HHL	0.368	0.302	0.116	0.214
HHM	0.371	0.302	0.116	0.211
HLH	0.168	0.196	0.563	0.073
HLL	0.152	0.215	0.552	0.081
HLM	0.166	0.195	0.568	0.072
HMH	0.267	0.278	0.369	0.086
HML	0.403	0.275	0.290	0.032
HMM	0.284	0.331	0.286	0.098
LHH	0.529	0.215	0.037	0.219
LHL	0.415	0.255	0.086	0.244
LHM	0.417	0.255	0.083	0.245
LLH	0.135	0.141	0.599	0.125
LLL	0.324	0.105	0.446	0.125
LLM	0.137	0.140	0.596	0.127
LMH	0.736	0.138	0.076	0.050
LML	0.641	0.113	0.109	0.138
LMM	0.671	0.108	0.061	0.160
MHH	0.529	0.214	0.038	0.219
MHL	0.414	0.255	0.084	0.246
MHM	0.414	0.256	0.086	0.244
MLH	0.169	0.195	0.564	0.071
MLL	0.152	0.215	0.553	0.080
MLM	0.168	0.194	0.565	0.073
MMH	0.296	0.238	0.301	0.165
MML	0.448	0.161	0.230	0.161
MMM	0.467	0.129	0.278	0.126

Table 6. Output distributions for the "all constraints included" condition.

Output distributions Direction				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.409	0.143	0.301	0.147
HHL	0.405	0.256	0.159	0.180
HHM	0.406	0.254	0.159	0.181
HLH	0.200	0.199	0.531	0.070
HLL	0.178	0.276	0.456	0.090
HLM	0.201	0.202	0.527	0.070
HMH	0.199	0.200	0.532	0.068
HML	0.432	0.213	0.322	0.034
HMM	0.180	0.273	0.456	0.092
LHH	0.585	0.163	0.059	0.194
LHL	0.413	0.256	0.091	0.239
LHM	0.415	0.254	0.090	0.241
LLH	0.179	0.188	0.486	0.147
LLL	0.409	0.143	0.302	0.146
LLM	0.180	0.190	0.486	0.144
LMH	0.716	0.114	0.114	0.055
LML	0.610	0.176	0.077	0.137
LMM	0.584	0.163	0.059	0.194
MHH	0.584	0.163	0.059	0.193
MHL	0.415	0.256	0.090	0.239
MHM	0.412	0.255	0.091	0.242
MLH	0.201	0.201	0.530	0.068
MLL	0.179	0.276	0.456	0.089
MLM	0.198	0.200	0.532	0.070
MMH	0.180	0.190	0.485	0.144
MML	0.406	0.256	0.160	0.178
MMM	0.408	0.144	0.301	0.146

Table 7. Output distributions for the "direction constraints" condition.

Output distributions Pitch constraints				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.442	0.248	0.095	0.215
HHL	0.441	0.245	0.095	0.219
HHM	0.443	0.246	0.095	0.217
HLH	0.171	0.178	0.560	0.092
HLL	0.169	0.179	0.561	0.091
HLM	0.170	0.178	0.562	0.090
HMH	0.471	0.192	0.225	0.113
HML	0.470	0.193	0.225	0.112
HMM	0.472	0.192	0.226	0.110
LHH	0.442	0.245	0.097	0.217
LHL	0.444	0.243	0.096	0.217
LHM	0.441	0.246	0.096	0.216
LLH	0.169	0.179	0.560	0.092
LLL	0.170	0.179	0.562	0.090
LLM	0.173	0.179	0.557	0.092
LMH	0.475	0.192	0.222	0.111
LML	0.473	0.193	0.224	0.110
LMM	0.475	0.191	0.225	0.109
MHH	0.443	0.246	0.095	0.215
MHL	0.440	0.245	0.096	0.218
MHM	0.441	0.246	0.098	0.214
MLH	0.172	0.178	0.560	0.091
MLL	0.168	0.178	0.563	0.091
MLM	0.170	0.180	0.559	0.090
MMH	0.472	0.193	0.225	0.109
MML	0.473	0.190	0.227	0.110
MMM	0.475	0.192	0.225	0.109

Table 8. Output distributions for the "pitch constraints" condition.

Output distributions Preceding constraints				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.330	0.194	0.323	0.153
HHL	0.328	0.194	0.321	0.156
HHM	0.330	0.194	0.323	0.153
HLH	0.210	0.231	0.484	0.075
HLL	0.209	0.231	0.482	0.078
HLM	0.211	0.227	0.487	0.075
HMH	0.208	0.230	0.485	0.077
HML	0.209	0.230	0.485	0.076
HMM	0.211	0.228	0.484	0.077
LHH	0.536	0.198	0.076	0.190
LHL	0.534	0.197	0.075	0.194
LHM	0.534	0.196	0.076	0.194
LLH	0.330	0.193	0.323	0.154
LLL	0.332	0.193	0.323	0.152
LLM	0.330	0.194	0.323	0.153
LMH	0.534	0.198	0.073	0.194
LML	0.534	0.197	0.074	0.195
LMM	0.536	0.196	0.074	0.195
MHH	0.533	0.197	0.075	0.194
MHL	0.536	0.198	0.075	0.191
MHM	0.536	0.198	0.074	0.193
MLH	0.210	0.229	0.485	0.075
MLL	0.212	0.227	0.484	0.076
MLM	0.212	0.230	0.482	0.076
MMH	0.330	0.193	0.324	0.153
MML	0.329	0.194	0.322	0.155
MMM	0.331	0.194	0.323	0.152

Table 9. Output distributions for the "preceding constraints" condition.

Output distributions Succeeding constraints				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.396	0.186	0.274	0.144
HHL	0.431	0.238	0.136	0.195
HHM	0.426	0.241	0.140	0.194
HLH	0.250	0.188	0.467	0.095
HLL	0.399	0.186	0.274	0.140
HLM	0.250	0.188	0.465	0.097
HMH	0.250	0.192	0.462	0.096
HML	0.428	0.237	0.139	0.195
HMM	0.398	0.185	0.276	0.141
LHH	0.397	0.187	0.276	0.141
LHL	0.428	0.238	0.138	0.196
LHM	0.427	0.239	0.137	0.196
LLH	0.252	0.189	0.464	0.095
LLL	0.399	0.185	0.277	0.140
LLM	0.252	0.191	0.462	0.095
LMH	0.253	0.187	0.464	0.096
LML	0.429	0.237	0.139	0.195
LMM	0.399	0.186	0.277	0.139
MHH	0.396	0.185	0.277	0.142
MHL	0.430	0.238	0.139	0.193
MHM	0.429	0.237	0.141	0.193
MLH	0.253	0.188	0.463	0.096
MLL	0.397	0.185	0.277	0.141
MLM	0.252	0.188	0.464	0.096
MMH	0.249	0.189	0.466	0.097
MML	0.429	0.239	0.138	0.194
MMM	0.397	0.185	0.278	0.140

Table 10. Output distributions for the "succeeding constraints" condition.

Output distributions Preceding and Succeeding constraints				
	/yu1/	/yu2/	/yu3/	/yu4/
HHH	0.358	0.165	0.316	0.161
HHL	0.357	0.238	0.207	0.197
HHM	0.356	0.235	0.210	0.199
HLH	0.181	0.208	0.551	0.061
HLL	0.253	0.227	0.435	0.085
HLM	0.179	0.209	0.552	0.060
HMH	0.178	0.208	0.554	0.060
HML	0.259	0.348	0.285	0.108
HMM	0.250	0.229	0.436	0.085
LHH	0.564	0.167	0.093	0.176
LHL	0.509	0.225	0.063	0.203
LHM	0.510	0.222	0.062	0.206
LLH	0.289	0.169	0.416	0.127
LLL	0.361	0.166	0.312	0.161
LLM	0.287	0.168	0.417	0.128
LMH	0.493	0.203	0.136	0.168
LML	0.510	0.225	0.062	0.204
LMM	0.564	0.166	0.093	0.177
MHH	0.565	0.167	0.092	0.176
MHL	0.510	0.225	0.062	0.204
MHM	0.511	0.224	0.060	0.205
MLH	0.180	0.207	0.553	0.060
MLL	0.252	0.229	0.432	0.086
MLM	0.179	0.208	0.554	0.060
MMH	0.288	0.168	0.417	0.127
MML	0.354	0.238	0.209	0.199
MMM	0.363	0.166	0.311	0.160

Table 11. Output distributions for the "preceding and succeeding constraints" condition.

References

Abramson, Arthur S. (1978). Static and dynamic acoustic cues in distinctive tones. In: *Language and Speech* **21**, 319-325.

Anderson, Stephen R. (1978). Tone Features. In: *Tone: a Linguistic Survey,* edited by Victoria A. Fromkin. New York: Academic Press.

Bao, Zhiming (1999). *The Structure of Tone.* New York Oxford: Oxford University Press.

Belin, P., R.J. Zatorre & P. Ahad (2002). Human temporal-lobe response to vocal sounds. In: *Cognitive Brain Research* **13**, 17-26.

Besson, M. (1998). Singing in the brain: Independence of lyrics and tunes. In: *Psychological Science* **9**, 494-498.

Boersma, Paul and B. Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**-I, 45-86

Boersma, Paul P.G. and David Weenink (1992-2007). *Praat*, version 4.5.19.

Brust, J. (2003). Music and the neurologist: A historical perspective. In: I. Peretz & R. Zatorre, eds, *The cognitive neuroscience of music*, 181-191. Ofxord University Press, New York.

Chao, R.C. (1947). *Cantonese Primer.* Cambridge, M.A.: Harvard University Press.

Chao, Yuen Ren (1930/1980). A System of Tone-Letters. *La Maitre Phonetique* **45**, 24-47.

Chao, Yuen Ren (1956). Tone, intonations, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In: *For Roman Jakobson : Essays on the Occasion of His Sixtieth Birthday, 11th October 1956.* Edited by M. Halle, H.G. Lunt, H. McLean and C.H. Van Schooneveld. The Hague: Monton & Co.

Cook, N.D., T. Fujisawa & K. Tamaki (2004). Application of a psycho-acoustical model of harmony to speech prosody. In: B. Bel and I. Marlien, eds. *Proceedings of Speech Prosody* , 147-150. Nara, Japan.

Cutting, J. & B. Rosner (1974). Categories and boundaries in speech and music. In: *Perception and Psychophysics* **16**, 564-570.

Deutsch et al. (2004). Absolute pitch, speech, and tone language: some experiments and a proposed framework. In: *Music perception*, **21**-3, 339-356.

Deutsch, Diana (1997). The tritone paradox: a link between music and speech. In: *Current Directions In Psychological Science* **6**, 174-180.

Fucks, W. & J. Lauter (1965). *Exaktwissenschaftliche Musikanalyze*. Westdeutscher Verlag, Opladen.

Gandour, Jackson T. (1978). The Perception of Tone. In: *Tone: a Linguistic Survey,* edited by Victoria A. Fromkin. New York: Academic Press.

Gilbers, D. (1992). *Phonological Networks, A Theory of Segment Representation*, PhD thesis, Grodil 3, Groningen.

Gilbers, D. & M. Schreuder (2002). Language and music in Optimality Theory, In: *ROA* 571-0103. URL: http://roa.rutgers.edu/view.php3?roa=571

Hauser, M.D., N. Chomsky & W.T. Fitch (2002). The faculty of language: What is it, who has it, and how does it evolve? In: *Science* **298**, 1569-1579.

Hayes, B. P. and M. McEachern (1998). Quatrain form in English folk verse. *Language* **74**-III, 473-507.

Hiller, L. & L. Isaacson (1959). *Experimental music: Composition with an electronic computer.* McGraw-Hill, New York.

Houston, David (2008). *Music and Language: The Case for Music in Linguistic Curricula and Research.* Saarbrücken: VDM Verlag Dr. Müller (MA thesis).

Howie, John Marshall (1976). *Acoustical Studies of Mandarin Vowels and Tones.* Cambridge: Cambridge University Press.

Hsu, Dolores Menstell (1964). Musical elements of Chinese opera. In: *Musical Quarterly* **50**, 439-451.

Jusczyk, P., B. Rosner, J. Cutting, C.F. Foard & L.B. Smith (1977). Categorical perception of non-speech sounds by two-month-old infants. In: *Perception and Psychophysics* **21**, 50-54.

Jusczyk, P.W. & C.L. Krumhansl (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. In: *Journal of Experimental Psychology: Human perception and performance* **19**, 627-640.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch.* Oxford University Press, New York.

Leben, W. (1980). *Suprasegmental phonology.* Doctoral Thesis, MIT, Cambridge. [Garland Press, New York 1980]

Lehiste, J. and G.E. Peterson (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustic Society of America* **33**, 419-425.

Lehrdahl, F. and R. Jackendoff (1983). *A generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Levis, John Hazedel (1936). *Foundations of Chinese Musical Art*. Shanghai: North-China Daily News.

Liberman, M. Y. (1975). *The International System of English.* Bloomington: Indiana University Linguistics Club.

Lindblom, B. and J. Sundberg (1970). Towards a generative theory of melody. *Swedish Journal of Musicology* **52**, 71-88

List, George (1961). Speech Melody and Song Melody in Central Thailand. *Ethnomusicology* **5**, 16-32.

Liu, T.H. (2008). A fresh look at the paradoxal nature of Chinese contour tones. In: *Proceedings of ConSOLE* **XV**, 149-163.

Maess, B. (2001). Musical syntax is processed in broca's area: An meg study. In: *Nature Neuroscience* **4**, 540-545.

Matthews, P.H. (1997). *Oxford Concise Dictionary of Linguistics.* Oxford, N.Y.: Oxford University Press.

McCawley, James D. (1978). What is a Tone Language? In: *Tone: a Linguistic Survey,* edited by Victoria A. Fromkin. New York: Academic Press.

McMullen, Erin and Jenny R. Saffran (2004), Music and Language: A Developmental Comparison. *Music Perception*, **21-3**, 289-311

Mei, Tsu-Lin (1970). Tones and Prosody in Middle Chinese. *Harvard Journal of Asiatic Studies* **30**, 86-110.

Nakada, T. (1998). 'Musical brain' revealed high-field (3 esla) functional mri. In: *Neuroreport* **9**(17), 3853-3856.

Patel, Aniruddh D. and Joseph R. Daniele (2003). An empirical comparison of rhythm in language and music. In: *Cognition* **87**, B35-B45.

Polivanov, E.D. (1931). La perception des sons d'une langue étrangère. In: *Traveaux du Cercle Linguistique de Prague* **4**, 79-96. [English translation: The subjective nature of language sounds. In: E.D. Polivanov (1974):

Selected works: articles on general linguistics, 223-237. Mouton, The Hague]

Prince, A. & P. Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar.* Technical Report TR-*2*, Rutgers University Center for Cognitive Science [published in 2004 by Blackwell, Malden].

Schreuder, M. (2006), *Prosodic Processes in Language and Music.* PhD Dissertation, Groningen Dissertations in Linguistics (GRODIL) **60**, University of Groningen.

Vondenhoff, M. (2007). *Meaning in Melody, The correlation between lexical tone and musical pitch in Mandarin Chinese*. BA Thesis, University of Amsterdam, Amsterdam [unpublished].

Wang, W. (1967). The phonological features of tone. *International Journal of American Linguistics* **33**, 93-105

Wertheimer, M. (1923). First published as: Untersuchungen zur Lehre von der Gestalt II. In: *Psycologische Forschung* **4**, 301-350. Translation published in Ellis, W. (1938). *A source book of Gestalt psychology* (p. 71-88). Routledge & Kegan Paul, London.

Wong, Patrick C.M. and Randy L. Diehl (2002). How can the Lyrics of a Song in a Tone Language Be Understood? *Psychology of Music* **30**, 202-209.

Xiao-nan, Susan Shen (1989). *The Prosody of Mandarin Chinese.* Los Angeles: University of California Press.

Yamadori, A. (1977). Preservation of singing in broca's aphasia. In: *Journal of Neurology, Neurosurgery and Psychiatry* **40,** 221-224.

Yip, Moira (1980). *The Tonal Phonology of Chinese*. Ph.D. dissertation, MIT.

Yung, Bell (1983). Creative Process in Cantonese Opera I: The Role of Linguistic Tones. *Ethnomusicology* **27**, 29-47.

Zatorre, R.J. (2003). Neural specialization for tonal processing. In: I. Peretz & R.J. Zatorre, eds., *The cognitive neuroscience of music*, 231-246. Oxford University Press, Oxford.

Front cover illustration: Tom Johnston (1974). Syncopated Texture. In: *Imaginary Music*, Editions **75**, Paris.