Experiments on the modular nature of word and sentence phonology in Chinese Broca's patients

Liang, J.

Experiments on the modular nature of word and sentence phonology in Chinese Broca's patients

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6000
e-mail: lot@let.uu.nl
http://wwwlot.let.uu.nl/

Cover illustration: Response box used in the experiments described in this dissertation.

ISBN-10: 90-78328-00-2
ISBN-13: 978-90-78328-00-1

NUR 632

Experiments on the modular nature of word and sentence phonology in Chinese Broca's patients

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 10 mei 2006
klokke 14.15 uur

door

LIANG JIE

geboren te Xinjiang, China
in 1960

Promotiecommissie

promotor:	prof. dr. V.J.J.P. van Heuven
referent:	prof. dr. Y.R.M. Bastiaanse, RU Groningen
overige leden:	dr. J. Caspers
	prof. dr. C.H.M. Gussenhoven, Radboud Universiteit Nijmegen
	prof. dr. R.A.C. Roos
	dr. J.M. van de Weijer

The first year (2000/01) of the research reported in this dissertation was financially supported by a grant from the China Scholarship Council (12-months stay at the LUCL phonetics laboratory).

Contents

Preface

The present journey started in September 1999 when I first went to Prof. Zhang Benshu, a neurologist in the Neurology Department of Tianjin General Hospital with my very ambitious proposal: Tones and Intonation of Chinese Broca's patients. My proposal with a long question list did keep her busy for quite a long time, hunting for patients who would meet my requirements. She called me early in the morning or late at night whenever there was a potential patient for my research. I travelled back and forth between the patients' homes and hospitals, making recordings. Sponsored by the Chinese Scholarship Council, I spent my first year abroad at Leiden University, the Netherlands, during which I did acoustic analysis and visited the international conference, Science of Aphasia. In 2002 I defended my thesis in Chinese, which focused on the speech production by aphasic patients in Nankai University, P.R. China. In 2003, supported by Xinjiang Normal University, my PhD thesis, *Research on tones and intonation of Chinese Broca's aphasic speech* was published in Chinese. However, the number of questions on my list had increased rather than decreased. Still, please allow me to express my heart felt thanks to my promotor at Nankai University, Prof. Shi Feng, Prof. Zhang Benshu in Tianjin General Hospital and all my patients for their generous help.

Thanks to Leiden University, I was able to continue my journey to work on an international PhD thesis with focus on the speech perception by Chinese Broca's patients in terms of lexical tones and sentence intonation. I wish to thank Jack Gandour, Diana van Lancker, Roelien Bastiaanse, Marina Nespor and Carlos Gussenhoven for their insightful input in my research. Many thanks go to Lisa and Rint for their great encouragement. I am grateful to Esterella, Ineke, Jeroen and many people in LUCL for their generous help.

I have been alone in the Netherlands but I never got the chance to feel lonely. My first year in the Netherlands was in Mrs Sees Somer's house with a German neighbour called Elke. Many more times than I can remember, when I returned home cold and wet, my 'Dutch mother' was waiting for me with hot Dutch farmer's soup and 'broodjes'. My second home in the Netherlands was in Max Henninger's house, where I have been spoiled by Max and his super-professional Indonesian kitchen. My third home is the Pacilly family. I am so proud of being a member of the family. Jos, Josée, Sietse and Feike, I remember every moment we spent together such as the wild water canoe ride in Drievliet, the quiet Christmas Eve of 2004, the Sunday sailing among the beautiful islands of the Grevelingenmeer in Zeeland. I must say that I have become much braver ever since I came to know Sietse. My colleagues, Rob, Ellen, Maarten, Jos, Ruben, Petra, and Johanneke in the phonetic lab, are my angels. The morning coffee and afternoon tea made it impossible for me to become homesick. I have become very much attached to Rob, Maarten and Johanneke for whenever I needed help I knew they were always there.

Finally, I wish to take the opportunity to apologize to my parents, my daughter and my husband for having been away for such a long time. If any progress has been made in my career, they can never claim more than they deserve for all the patience, support and love they have for me.

Chapter One

Introduction[1]

Summary

This book investigates effects of brain lesions on prosodic processing in Chinese Broca's patients with respect to the supra-segmental aspects pertaining to (i) lexical pitch features, i.e. lexical tones in Chinese; (ii) post-lexical pitch features, i.e. sentence intonation. The basic overall question is: Where and how is word-prosody (lexical tone) represented in the mental architecture? Is it separate from segmental structure and is it separate from sentence-prosody?

It has been known since the time of Broca (1861) that persons with damage in the region of the middle cerebral artery within the left hemisphere of the brain (LH) frequently suffer from speech and language deficits. However, the nature of the deficits is still a matter of debate. Specifically it is still an open question whether the language impairment is primarily due to processing limitation or a structural deficit.

While verbal memory and semantic processes are complex integrated aspects of language functions, language processing requires perception of phonemes at the earliest stage. This process requires categorization of the simplest unit of speech sounds on the basis to their acoustic features. Comparison of phoneme judgment by patients with Broca's aphasia with that by healthy subjects may help to decide whether a deficit is caused by structural defect or by limited processing in aphasic patients.

There is a second long-standing debate concerning the neural basis of prosodic processing, i.e. whether the production and perception of linguistic prosody (especially of word tone and sentence intonation) is lateralized to the left hemisphere and that of non-linguistic prosody (e.g. the signalling of emotion) to the right hemisphere.

In a tone language like Chinese, pitch is used to signal linguistic contrasts at both word-level and sentence-level. A study of Chinese aphasic speech, therefore,

---

[1] This chapter is based on a literature review published in Chinese as J. Liang (2004) Review on Chinese aphasic studies, *Contemporary Linguistics,* 2, 154-158.

provides us with an opportunity for a better understanding of the deficit nature in Broca's patients and testing the hypothesis concerning the role of the left hemisphere of the brain in the control of prosodic aspects of language.

Motivated by the above two issues, the present investigation focuses on lexical tone and sentence intonation as well as the interaction between the two in the speech of Beijing Broca's patients. It has been repeatedly reported in the literature that such patients suffer defect in both the production and perception of word tones in their language. It is, however, unclear to what extent the abnormalities are linguistic in nature and to what extent they are caused by poor control of the neuro-physiological mechanisms associated with fundamental frequency (F0) variation.

1. Brain and language

The human brain is a paired organ; it is composed of two halves (called cerebral hemispheres) that look rather alike. However, the fact is that the two halves of the human brain are not exactly alike, which fact is referred to as 'brain lateralization'. Each hemisphere has functional specializations, i.e. has functions whose neural mechanisms are localized primarily in that half of the brain.

Most humans (but not all) have left hemisphere specialization for language abilities. The only direct tests for speech lateralization are too invasive to use on healthy people, so most of what we know in this area comes from clinical reports of people with brain injuries or diseases.

1.1. Broca's area

Aphasia is an impairment of the ability to use or comprehend words (in sentence context), usually acquired as a result of a stroke or other brain injury. The lesion causing aphasia was first described by the French neurologist Paul Broca (1861), who located it in the third frontal convolution (both the gyrus and the sulcus) of the left frontal lobe, after which publication this location was called **Broca's area**. A photograph of a lateral view of the brain of Mr. Leborgne (Broca's patient at the time) with lesion to the area is presented in Figure 1.

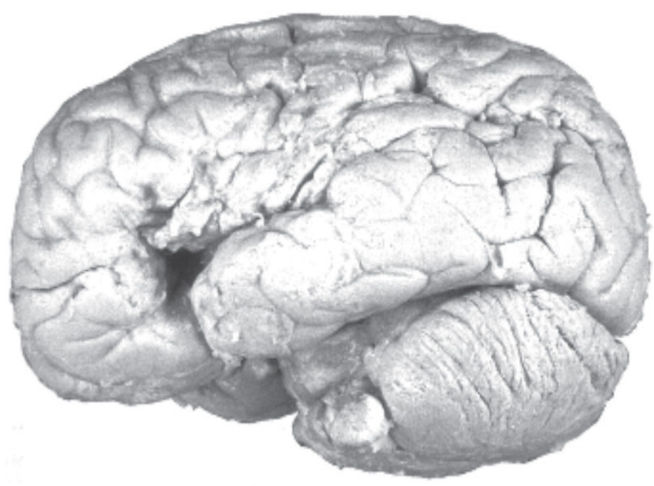


Figure 1: Photograph of the brain of Mr. Leborgne *(lateral view) showing Broca's area. (downloaded from Amunts 2004)*

Broca's discovery starts a scientific theory of localization of cortical functions, which has been a popular research topic for more than a century. Following classical works by Broca, the main principle of the clinical anatomical method is to relate lesion sites to specific linguistic (and non-linguistic) deficits.

Broca's area was the first brain region to which a circumscribed function, i.e. language, was linked. Broca's area is located in the opercular and triangular sections of the inferior frontal gyrus. In most people, Broca's area is in the lower part of the frontal lobe in the left hemisphere of the brain. Broca's area corresponds to areas 44 and 45 in Brodmann's classification system (see Figure 2). It contains the motor speech area and controls movements of tongue, lips, and vocal cords (and other physiological and anatomical structures involved in the production of speech).

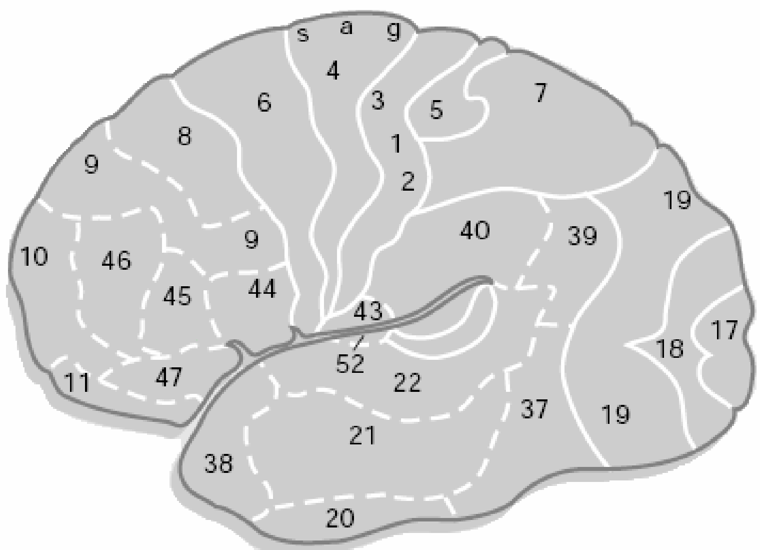


Figure 2: Topography of the left hemisphere of the brain in Brodmann's system.

1.2. Broca's aphasia

Broca's area is the section of the brain which is involved in speech production. People suffering from neuro-physiological damage to this area can usually exhibit a grammatical disorder which is called Broca's aphasia or agrammatism, characterized by the relative omission of grammatical words in sentence production. Broca's aphasia is often accompanied by difficulty in performing the motor or output aspects of speech (dysarthria or verbal apraxia).

Ideally, the role of a given region in language functions would thus be inferred from a single cognitive deficit associated with damage to the particular site. However, such a one-to-one relationship is hardly observed in reality due to many factors that account for the complexity of brain/language relationships. For instance, aphasic patients may exhibit various language disorders and each of them may correspond to several cognitive mechanisms such as their abilities to perceive others or to express themselves. Additionally, adaptive mechanisms, or strategies, taking place in undamaged parts of brain may also contribute to the complexity of symptoms. For instance, considerable recovery from an initial severe aphasia is usually observed over weeks and months after the abrupt lesion onset. It is still under debate whether agrammatism is a syndrome which is defined as a set of symptoms that co-occur often enough to suggest a single cause.

It is now uncontroversial that the classical model described by Broca, is problematic in that it cannot account for the range of aphasic syndromes anatomically (e.g., Petrides and Pandya, 1994; Amunts, Schleicher, Bürgel, Mohlberg, Uylings and Zilles, 1999) as well as linguistically (e.g., Zatorre, Meyer, Gjedde and Evans, 1996; Koski, Wohlschlager, Bekkering, Woods, Dubeau, Mazziotta and Iacoboni, 2002). As knowledge in the biological and linguistic domains accumulated, the concept of Broca's area and its role in language has changed from a center of language production and articulation to a center which is involved in the processing (production as well as perception) of syntax, semantics, phonology, and even in non-language related tasks.

Nevertheless, the study of Broca's aphasia remains a fascinating research area for hypotheses that link the conceptual and technical apparatus of linguistics with the machinery of neuroscience, i.e. linking hypotheses that bridge concepts such as 'distinctive feature' with the mechanisms of neurobiology (e.g. concepts such as 'receptive field'). Experimental research on Broca's aphasia has contributed much to a new functional anatomy of language (Poeppel and Hickok, 2004).

Traditionally, Broca's aphasia is characterized by non-fluent speech, few words, short sentences, and many pauses. Thus, it is often referred to as a 'non-fluent aphasia' because of the halting and effortful quality of speech, but there are other names for this type of language disorder such as 'expressive' Also, it has been called verbal aphasia (Head, 1926), motor aphasia (Goldstein, 1933), and efferent motor aphasia (Luria, 1964).

Depending on the size of the lesion to Broca's area, the symptoms can range from the mildest type (cortical dysarthria) – with intact comprehension and the ability to communicate through writing – to a complete loss of speaking out loud. In most cases, the speech consists almost entirely of content words. The words that patients can produce come with great effort and often sound distorted. The speech melody (intonation) is flat or monotonous. This gives the speech the general appearance of a telegraphic nature, because of the deletion of functional words and disturbances in word order. Interestingly, although the ability to repeat words and phrases after a speaker is impaired, auditory comprehension in conversational speech is relatively intact (Goodglass 1976).

However, since 1980s, Broca's aphasia has been reliably associated with syntactic comprehension deficits (Caplan and Hildebrandt, 1988; Caramazza and Zurif, 1976; Grodzinsky, 1989; Zurif, Swinney, Prather, Solomon and Bushell, 1993; van der Meulen, 2004). More recent studies have shown that these patients may have lexical-semantic comprehension deficits as well (Milberg, Blumstein and Dworetzky, 1987; Milberg, Blumstein, Katz, Gershberg and Brown, 1995).

The nature of the deficits that underlie comprehension problems in Broca's aphasia is still a matter of debate. A number of studies have reported results that are suggestive of processing impairments, rather than a loss of knowledge (Bates, Friederici and Wulfeck, 1987; Brown, Hagoort and Swaab, 1996; Friederici and Kilborn, 1989; Haarmann and Kolk, 1994; Hagoort, 1993, 1997; Hagoort, Brown and Swaab, 1996; Kolk and Weijts, 1996; Milberg, Blumstein and Dworetzky, 1987; Milberg, Blumstein, Katz, Gershberg and Brown, 1995; Swaab, 1997; Tyler, Ostrin, Cooke and Moss, 1995; Zurif, Swinney, Prather, Solomon and Bushell, 1993).

Also, there is broad consensus that, at least in speech perception, the right temporal lobe plays an important role, and, more generally, one of the main consequences of imaging research has been to highlight the extensive activation of the right hemisphere in language tasks.

1.3. Other major terms used in aphasic research

In addition to Broca's discovery, a German neurologist and psychiatrist, Carl Wernicke, discovered, in 1874, that damage to the left posterior section of the superior temporal gyrus, posterior to the primary auditory cortex, on the temporo-parietal junction, i.e. the posterior part of Brodmann's area 22 (see Figure 2), could cause another type of aphasia ─ Wernicke's aphasia.

People with Wernicke's aphasia are traditionally described as having fluent but often meaningless speech containing well-formed sentences with lots of grammatical elements. The patients have normal prosody but poor comprehension. They suffer word-finding difficulties and often make substitutions and neologisms or jargonisms (nonsense words). Unlike agrammatic patients, people with Wernicke's aphasia have relatively normal syntactic abilities.

Closely related to Wernicke's aphasia, anomic aphasia is characterized by naming or word finding problems. According to Goodglass and Kaplan (1983), anomia can be localized with the least reliability of any of the aphasic syndromes. The lesion is found most often in the temporal parietal area.

Therefore, the traditional Broca's patient is non-fluent with spared comprehension whilst the typical Wernicke's patient is fluent with impaired comprehension. Table 1 is a rough traditional classification of the main terms found in recent literature.

In summary, the brain has two roughly identical halves – the left and the right hemispheres, and these differences may form the basis for the first major brain specialization for language – lateralization of language to the left hemisphere. The second major brain specialization for language is within the left hemisphere. Obviously, research on the nature of Broca's aphasia will better our understanding of language grammar and its brain mechanisms.

Table 1: Classification of the main features found in aphasic research

Broca's aphasia	**Wernicke's aphasia**
Non-Fluent	Fluent
Good comprehension	Poor comprehension
Expressive	Receptive
Motor	Sensory
Agrammatic	Neologistic/ Jargonistic

2. Prosody and its acoustic correlates

Individual speech sounds such as vowels and consonants are called 'segments'. Speech involves more than just stringing together individual segments in a sequence. Other properties of speech are laid on top of groups of segments. In linguistics and phonetics the term prosody, literally 'that which accompanies the song', i.e. the music with the words, is most often used to refer to those properties of speech which cannot be derived from the segmental sequence of phonemes underlying human utterances. On the perceptual level those properties lead amongst other things to perceive pattern of relative pitches and syllable prominences, coded in perceived melodical and rhythmical aspects of speech (Nooteboom, 1997).

Prosody can be approached from a variety of view points. A main distinction can be drawn between the more abstract, linguistic, phonological description of prosody and the more concrete, acoustic-phonetic analyses of prosody.

2.1. Linguistic approach to prosody

Linguistic prosody is used – among other things – to disambiguate or to mark the internal organization of sentence constituents (phrasing), to highlight important words within such phrases (accentuation) or to (help) convey the overall communicative import of a sentence (such as question, statement, command, exclamation). Moreover, prosody is used to convey lexical meaning as well as post-lexical meaning.

At the word (or 'lexical') level, for example, in a tone language like Chinese, each syllable is marked with one of the four distinctive melodical patterns, which are called lexical tones. Replacing one tone patterns by another may have the consequence of changing the lexical meaning of a word (see §§ 2.2, 3.2).

At the sentence level, there are at least two different basic syntactic sentence types in a language: statements and questions. The different sentence types are marked by different melodical patterns as well. Prosodic signalling of sentence type is common, especially if no other linguistic elements such as question particles, word order, or verb form are present to facilitate identification. Falling intonation usually indicates a statement, whereas a rising intonation declares a question.

Linguistic prosody such as lexical tone and sentence intonation both draw on the same phonetic features, which are loudness, pitch, and speech tempo (acceleration and deceleration relative to a mean speaking rate). These three main prosodic parameters have their phonetic correlates in intensity, frequency, and duration.

2.2. Acoustic-phonetic approach to prosody

Prosodic features are physically realized in the speech chain in terms of variations of a set of acoustic parameters. Acoustic-phonetic analyses identify the following 'phonetic correlates of prosody': fundamental frequency (F0), length changes in segmental duration, pauses, loudness and voice quality. Such acoustic modulations are used by human speakers to express a variety of linguistic or paralinguistic features, from lexical to post-lexical meaning.

Among the phonetic correlates mentioned above, the primary acoustic correlate of speech melody is the (variation in the) fundamental frequency (F0) of the waveform. A distinction can be drawn between two different phonological uses of F0 on the basis of the linguistic domain over which the F0 patterns extend (Laver, 1994). In a tone language like Chinese, F0 is used to differentiate both lexical tone (at the syllable or word level) and linguistic intonation (at the phrase or sentence level); in a non-tone language, it is only used to identify linguistic entities at levels higher than the word, i.e. at the phrase and sentence level.

Pitch basically refers to the tonal height of a sound object, e.g. a musical tone or the human voice. The term often refers to the fundamental frequency (oscillation frequency) of the glottal oscillation (vibration of the vocal folds) in speech processing while to an auditory (subjective) attribute in psychoacoustics. Pitch has been considered an important component of prosody since the 1950s (Fry, 1955; Fry, 1958; Bolinger, 1958; Lieberman, 1960; Hadding-Koch, 1961).

Prosody, as expressed in pitch, carries linguistic (e.g. lexical meaning or question intonation) and paralinguistic information (e.g. emotional intonation). Pitch movement at word level (local pitch excursion) is the characteristic of lexical tone (linguistic). The four Chinese lexical tones can be expressed in different pitch patterns (see Figure 3) named Tone 1, Tone 2, Tone 3 and Tone 4. A syllable /ma/ marked with one of the four lexical tones, /ma1/, /ma2/, /ma3/ and /ma4/ indicates a different lexical meaning, i.e., 'mother', 'hemp', 'horse' and 'scold', respectively. In addition to local pitch excursion near the edge of the sentence (boundary tones), the pitch setting at post-word level (global pitch excursion) is typical of intonation (linguistic or paralinguistic).

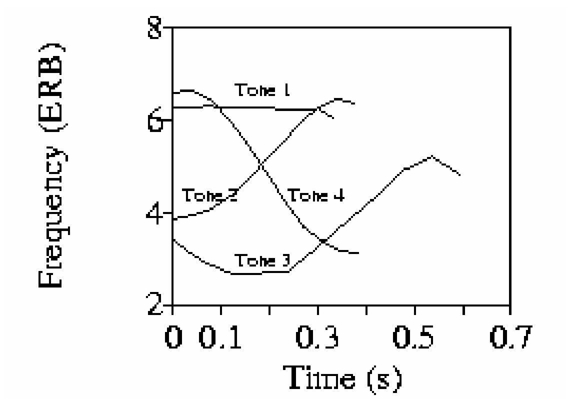


Figure 3: Fundamental frequency (ERB) as a function of time (s) for each of four lexical tones, for a male speaker of the Beijing dialect.

All languages are believed to exhibit intonation systems regardless of whether F0 is or is not also used to signal lexical contrasts at the word level. Therefore, in a tone language like Chinese, the interaction between tone and intonation is of theoretical interest because both lexical tone and intonation are signalled primarily by changes in the same acoustic parameter, F0. The changes in F0 provide us with an excellent opportunity for studying the role of the left hemisphere in the control of prosodic aspects of language and the possible interaction between prosody at word level and at sentence level.

3. Beijing dialect

Beijing dialect usually refers to the dialect spoken in the urban area of Beijing with its own local pronunciations, e.g. much rhotacization, as well as separate vocabulary and slang. Beijing dialect has been adopted as the official language for all of China. Some linguists have given a broader definition of Beijing dialect which also includes some dialects extremely akin to that of Beijing. In the rest of the thesis, I will use the broader definition of Beijing dialect.

3.1. Beijing and other Chinese dialects

The variations in spoken Chinese are traditionally divided into seven to ten groups. The seven main groups are Mandarin, Wu, Xiang, Gan, Hakka, Cantonese (Yue), and Min. The later six are usually referred to as southern dialects. Linguists who distinguish ten instead of seven major groups, would then separate Jin from Mandarin, Pinghua from Cantonese, and Hui from Wu. Beijing dialect is a representative of the Mandarin group, which is also called northern Chinese illustrated in a tree diagram in Figure 4.

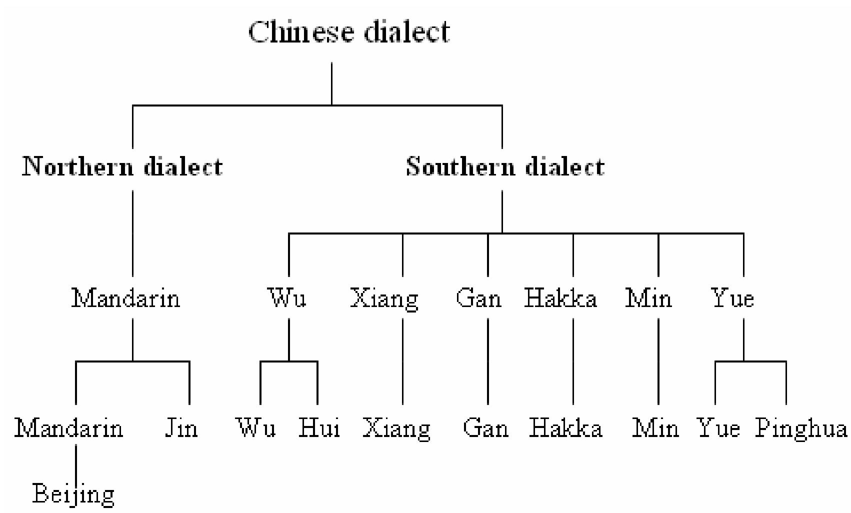


Figure 4: Tree diagram of Chinese dialects

Many dialects of Mandarin are mutually intelligible to some degree. However, the southern (non-Mandarin) dialects are not mutually intelligible with Mandarin; people of southern dialects have to learn the Beijing dialect as if it were a foreign language. For example, a speaker of Nantong dialect (one of the Mandarin varieties) may find the Beijing dialect easier to understand than a speaker of Changsha (one of the Xiang dialects) does.

This impressionistic distinction has been supported by studies on distance in terms of mutual phonological intelligibility between dialects. As an example of the intelligibility computation based on phonology, over 2,700 syllable-words (Beijing University 1962, 1989) were used to calculate mutual intelligibility scores for all the pairs of the 17 dialects contained therein. The degrees of intelligibility for all the dialect pairs were used in cluster analysis to establish a grouping of dialects. Taking into account the differences in initials (onsets of syllables), finals (syllable rhymes) and tones, researchers computed that the mean degree of similarity between

Mandarin dialects is 92%, but is 61, 69, 68, 62, 83 and 53% between Mandarin and Wu, Xiang, Gan, Min, Hakka (Kejia) or Cantonese, respectively (glossika.com). According to Cheng (1994) the Beijing and the Changsha dialect are not mutually intelligible to some degree as the mutual intelligibility index between the two dialects is around 0.61 out of 1.00.[2]

Historically, research suggests that Old Chinese was a toneless language. Most likely, tones arose between Old Chinese and Early Middle Chinese (that is between 500 BCE and 500 CE) as a result of the loss of final laryngeals (Sagart, 1999).3 Most modern Chinese dialects have between three to ten contrastive tones. There exists a rough cline between the northern and southern dialects, which generally have three to five tones, and the south-eastern dialects which have around six to ten. From this point of view, Chinese dialects provide a typological transition between the non-tonal Tungusic, Mongolian and Turkic languages in the north and the tone-rich languages of the Mao-Yao and Kam-Tai groups in the south.

One may hypothesize that it is more difficult for a Changsha-dialect speaker to learn the Beijing dialect than for a Nantong-dialect speaker, and that it is even more difficult for a non-tone language speaker, such as a Uygur speaker, to learn the Beijing dialect than Nantong and Changsha speakers, since the Uygur language is a Turkic language belonging to the Altaic family, while Nantong and Changsha are tone languages within the Sinitic family. Alternatively, one may expect strong interference (negative transfer from one tone language to another); in that case it may be better to learn a tone language from a toneless background. The sites for the Beijing, Uygur, Nantong and Changsha dialects are drawn in Figure 5.

[2] It should be pointed out here that the work cited is a structural linguistic exercise, and does not reflect the actual mutual intelligibility between pairs of Chinese dialects. As far as we know, there is no experimental research that has ever attempted to actually measure intelligibility (as has recently been done for Norwegian dialects by Gooskens and Heeringa, 2004 and for some closely related West-Germanic language varieties by van Bezooijen and Gooskens (2005).

[3] CE stands for 'Common Era'. It is a relatively new term that is experiencing increased usage and is eventually expected to replace AD. The latter is an abbreviation for 'Anno Domini' in Latin or 'the year of the Lord' in English. The latter refers to the approximate birth year of Yeshua ben Nazareth (a.k.a. Jesus Christ). CE and AD have the same and value; 2004 CE = 2004 AD. BCE stands for 'Before the common era'. It is eventually expected to replace BC, which means 'Before Christ'. BC and BCE are also identical in value. Most theologians and religious historians believe that the approximate birth date of Yeshua of Nazareth was in the fall, sometime between 7 and 4 BCE, although we have seen estimates as late as 4 CE and as early as the second century BCE. Of course, one can always interpret the letter 'C' in CE and BCE as referring to 'Christian' or 'Christ's'. The Abbreviations Dictionary does exactly this. [text adapted from http://www.religioustolerance.org/ce.htm]

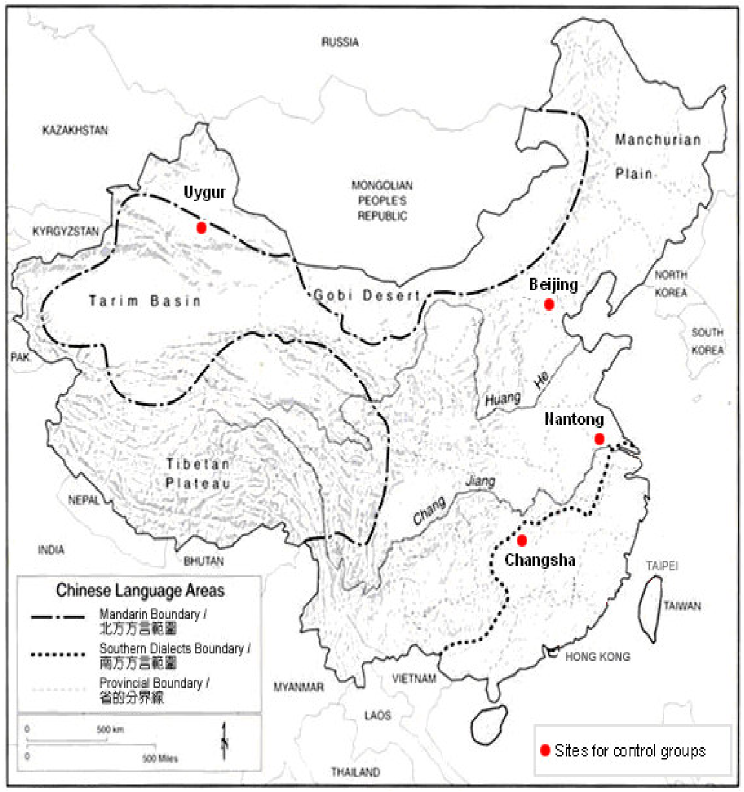


Figure 5: Chinese language areas and physical features (adapted from: Zhouguo Da Bai Ke Quanshu, Beijing, 1988).

3.2. Tone inventories of Beijing, Nantong and Changsha dialects

Beijing dialect has four lexical tones, described as 'high level', 'rising', 'low dipping', and 'falling', which are traditionally also referred to as Tone 1, Tone 2, Tone 3, and Tone 4, respectively.

Five-level descriptive system. In Chinese linguistics, tonal systems are generally described in a so-called five-level system (Chao, 1930), in which the number stands for the relative pitch and each single digit stands for a relatively short tone, e.g. [55] is a high tone and lasts longer than [5], which is also high but takes up roughly half the duration. The tonal systems of Beijing, Nantong and Changsha dialects are presented in Table 2.

Table 2: Tone inventories of Beijing, Changsha and Nantong dialects transcribed in Chao's (1930) five-level system.

Register	**Beijing Dialect**	**Nantong Dialect**	**Changsha dialect**
High	55	55	55
		5	
		4	
Mid	35	35	24
	51	42	41
			33
Low	21(4)	21(3)	13 (1)

Table 2 shows that the three dialects have different numbers of tones but all the tones are from a subset of one and the same system. Therefore, tonal systems of various dialects in Chinese are often described in an ancient Chinese tonal system from which all modern dialects evolved.

Ancient Chinese tonal system. The ancient Chinese tone system arguably comprised four tone categories, called Ping 'level', Shang 'rising', Qu 'departing' and Ru 'entering'. Each tone, in turn, is split into a higher and a lower one called Yin and Yang, respectively. This division is historically conditioned by the voice feature of the onset consonant, i.e., Yin developed from unvoiced onsets and Yang from voiced ones. Normally, words starting with voiced consonants all dropped to the lower tone. Accordingly, the tonal systems of Beijing, Nantong and Changsha dialects are presented in Table 3.

Table 3: Tone systems of Beijing, Nantong and Changsha dialects (the voice feature refers to the onset consonant conditioning the register split).

Tone	Level (Ping)		Rising (Shang)		Departing (Qu)		Entering (Ru)	
Register (voice)	−	+	−	+	−	+	−	+
Beijing	55	35	214		51			
Nantong	21	35	55		42	213	4	5
Changsha	33	13	41		55	21	24	

There is a poem about the ancient Chinese tones as found in the Kangxi Dictionary (康熙字典)

平聲平道莫低昂，'**Ping** tone is level but never low,'
上聲高呼猛烈强，'**Shang** tone rises yet holds high,'

去聲分明哀遠道， ‘**Qu** tone blues low with length,’
入聲短促急收藏。 ‘**Ru** tone clipps short and hide.’

The poem described the Ping tone as a mid-flat tone, the Shang tone as a high flat tone, the Qu tone as a long and low tone, and Ru tone as having a short, clipped sound, about half the length of a Ping tone. According to the poem, four tones were distinguished in ancient Chinese tonal system, i.e. mid-flat, high flat, long-low and short. It is commonly accepted that the Ping tone (mid-flat) described in the ancient poem is best represented in the Changsha tonal system by [33] in modern dialects of Chinese.

Development of Beijing tones. According to Sagart (1999), the main mechanism for tone splits in Chinese dialects from Early Middle Chinese into modern systems is tone lowering after voiced initial consonants. The Middle Chinese level tone (Ping) split into a high Tone 1 (after a voiceless onset) and a rising Tone 2 (after voiced onset – the voiced consonant depressed the first part of the originally high tone). Such splits were less regular, or even absent, in the historical development of other Beijing tones. The development of Beijing dialect is summarized in Table 4.

Table 4: The development of Beijing tones (from Sagart, 1999).

MC initials MC tones	Level	Rising	Departing	Entering
-voice unaspirated obstruents	Tone 1	Tone 3	Tone 4	Tone 1,2,3,4
-voice aspirated stops	Tone 1	Tone 3	Tone 4	Tone 1,2,3,4
+voice obstruents	Tone 2	Tone 4	Tone 4	Tone 2
+Voice sonorants	Tone 2	Tone 3	Tone 4	Tone 4

Table 4 shows that the traditional level tones split into Tone 1 (high level) and Tone 2 (rising), three-quarters of the traditional rising tones developed into Tone 3 (low dipping) and in Beijing dialect the departing tones developed into Tone 4 only.

The four tones in Modern Beijing dialect are perceptually confused by native listeners to different degrees (Howie, 1976). The most frequent confusion is symmetrically between Tone 2 and Tone 3. This is explained by the fact that both tones are essentially rising, the early fall portion in Tone 3 being optional. The second-most frequent perceptual confusion is between Tone 1 and Tone 2.

However, this confusion is asymmetrical: Tone 1 is misheard as Tone 2 but not vice versa (since Tone 2 is misheard as Tone 3). The confusion between Tone 1 and Tone 2 may be accounted for by the fact that they originate from the same historical source, and the difference between them has remained relatively small, i.e. a local depression of the pitch after the onset in Tone 2, which then quickly rises to the same high pitch as in Tone 1. Tone 4 is the least confused tone in Mandarin (Beijing in particular).

4. Literature review on aphasia

The review of the literature on aphasia will be divided in two parts. In the first, the major findings of selected experimental studies concerning Chinese aphasia will be summarized. The second part will concentrate on work carried out on neuro-anatomical correlates of language in general and of prosody in particular.

4.1. Research on Chinese aphasia

Historically, analysis of aphasic speech has been largely restricted to western European languages. In recent years, however, there has been an upsurge of interest in how aphasic speech deficits are manifest in Chinese. By investigating the breakdown of speech due to aphasia in Chinese, it is possible to offer a new perspective on aphasia, thereby increasing our understanding of potential universals in aphasia theory. Since the phonological, orthographic, lexical, and grammatical structures of Chinese differ substantially from those of Indo-European languages, on which major theories of linguistics and psycholinguistics are based, the studies of Chinese aphasia also presents a major challenge to psycholinguists who attempt to understand the dynamics of language processing and language acquisition, and offer new windows on cognitive processes and new opportunities for psycholinguistic studies as well.

In studies of Chinese aphasia dealing with speech production and perception, so far, there have been important findings that typically address the following issues: (1) lexical tone production deficits, (2) the independence of phonological tiers, (3) fluent versus non-fluent word order deficits (4) 'Graded' morphological deficit and (5) lack of sentence embedding in non-fluent speech (Hughes, Chan and Su 1983; Packard, 1986, 1993; Yiu and Fok, 1995).

Packard (1993), perhaps, is the first detailed linguistic analysis of a large body of Chinese natural speech data. In the book four case studies of Chinese speakers are presented. The book is the first attempt to determine how aphasic speech production deficits are manifest in Chinese. It consists of four chapters covering typology-dependent theories, issues in aphasia theory, reconsidering language structure, background information on Chinese, a detailed linguistic description and analysis of deficits in the speech of four aphasic speakers of Chinese, a set of speech characteristics helpful for Chinese aphasia diagnosis, and proposes an explanation of aphasia syndromes based on the theory of linguistic modularity. The complete corpora of four aphasic Chinese speakers, including interlinear and free translations, are presented in an Appendix.

Packard characterizes agrammatism in Chinese as the simplification of syntactic structure, the omission and underemployment of free grammatical morphemes, and the omission and otherwise improper use of morphemes used in forming complex words. Packard finds that complex Chinese words might be analyzed within a theory of word structure (Lexical Phonology and Morphology), which predicts their likelihood of misuse by patients non-fluent aphasics. Packard argues that this word formation deficit extends to other languages as well, making it a universal feature of agrammatism.

In syntax, Packard finds that the functional-word omission and syntactic simplification seen in agrammatism in other languages is also a characteristic of Chinese agrammatism. He claims that qualitative differences between non-fluent and fluent agrammatic speakers in the nature of their word order errors and a lack of complex embeddings in non-fluent speech are found, despite the absence of overt complementizers in Chinese. With regard to the identity of 'co-verb', Packard's answer is 'these data provide neurolinguistic support for the view that the words that are called co-verbs may be verb-like or preposition-like, depending on usage. The evidence is that in non-fluent speech, co-verbs are generally retained when used as verbs (content words) and generally omitted when used as prepositions (functional words)'.

Following the Chinese data, an explanation for the major aphasia syndromes is offered based on the modularity hypothesis developed in cognitive science. The theory posits that Broca's aphasia is the result of computational deficits that occur within linguistic components, while Wernicke's aphasia is the result of deficits that occur in the transfer of information between components.

However, many theories of agrammatism have been challenged, e.g. by the discovery that agrammatic patients can make above-chance judgments of grammaticality. Over the past ten years, more work has been done and more evidence has been found in other languages, which enables us to have a better understanding of the nature of agrammatism. In an attempt to account for the nature of the deficit in agrammatism, two main approaches have been developed, i.e. the structural and the processing one. The former postulates that agrammatic deficits are caused by damage to syntactic representations (Caplan, 1983; Hagiwara, 1995; Friedmann and Grodzinsky, 1997). According to the latter, the deficit in agrammatism cannot be described in terms of a representational deficit, but it is rather caused by processing limitations (Friederici and Frazier, 1992; Kolk, 1995). There are at least three aspects that are worthwhile mentioning.

First, it is crucial to notice that more studies suggest that the performance of aphasic patients on some functional categories may vary across tasks. For instance, there are numerous studies (Hagiwara, 1995; Friedmann and Grodzinsky, 1997), in which it has been argued that in a hierarchical structure of a sentence the lower the position of the functional head and its projection, the more accessible they are to an agrammatic speakers. The studies of agrammatic speakers are based on empirical research including spontaneous speech and acceptability judgment experiments by agrammatic patients in many languages, such as French, Italian, Japanese, and Hebrew.

Secondly, speech production in Broca's aphasia is characterized as agrammatic: function words and grammatical endings are omitted or replaced and speech is reduced to mainly content words, while it has also repeatedly been reported that patients with Broca's aphasia have problems with verb retrieval, both in action naming and in spontaneous speech. For instance, Bastiaanse, Rispens Ruigendijk, Rabadán and Thomspon (2002) conducted several cross-linguistic experiments to show that patients with Broca's aphasia have problems retrieving verbs from the lexicon and with the syntactic operation of Verb Movement. With Dutch Broca's patients, production of finite verbs in the matrix clause that have been moved from their base-generated position, is significantly more impaired than production of finite, non-moved verbs in the embedded clause. This discrepancy between matrix and embedded clause is not found for English lexical verbs, which are never moved overtly, but is similar to the discrepancy observed for English moved and non-moved auxiliaries.

Thirdly, results obtained from the study on the similarity between agrammatic and normal speech have made a significant contribution to the understanding of the nature of agrammatism (Hofstede, 1992; Kolk and Heeschen, 1992; de Roo, 1999).

Chinese is well-known for its impoverished system of grammatical morphology. Li, Bates and Macwhinney (1993) examined how, in the absence of inflections, Chinese speakers employ other types of cues in real-time sentence interpretation. A reaction time technique was designed to tap into the role of word order, noun animacy, the object marker *ba*, the passive marker *bei*, and the indefinite marker *yi*. Results show the following hierarchy of cue strengths in Chinese: passive marker bei > animacy > word order > object marker *ba* > indefinite marker *yi*. The fact that the semi-morphological markers (*ba* and *bei*) are separated by semantic (noun animacy) and syntactic (word order) cues in this strength hierarchy shows that cues are not necessarily grouped together by linguistic type (e.g., morphology > order vs. order > morphology). Complex interactions among cue types were observed in both the decision and the reaction time data.

These findings are compatible with interactive activation models of sentence processing, e.g. the Competition Model claiming that language processing is considered essentially as a competition of multiple cues (Bates and MacWhiney, 1982, 1989). However, they pose problems for models that assume a modular architecture in which morphological, semantic, and syntactic sources of information are insulated from one another at various points in parsing and interpretation. Finally, reaction time data reveal aspects of processing that are often not available in results from choice response measures, attesting to the usefulness of reaction time studies at the sentence level.

Chinese poses an interesting test of the above models, because its grammar is so austere, with few obligatory features. Lu, Bates, Li, Tzeng, Hung, Tsai, Lee and Chung (2001) conducted an on-line grammaticality judgment task with healthy and agrammatic speakers of Chinese, using the small set of constructions that do permit judgments of grammaticality in this language. Patients with Broca's and Wernicke's aphasia showed similar patterns, with above-chance discrimination between grammatical and ungrammatical forms, suggesting once again that patients with

Broca's aphasia are not unique in the degree of sparing or impairment that they show in receptive grammar. It is argued that there is some sensitivity to grammatical well-formedness in Chinese patients with aphasia, but the effect is fragile for patients and probabilistic for healthy speakers.

Law and Leung's (2000) paper reported the performances of two Cantonese speakers with aphasia on tasks examining their sentence processing deficits. The data on sentence comprehension show that thematically non-canonical sentences, full passives, and subject-gap sentences present greater difficulty to these patients than canonical sentences, truncated passives, and object-gap sentences, respectively. These patterns are consistent with previous observations on Chinese speakers with aphasia and are expected given the structural differences between Chinese and English. In a Cantonese grammaticality judgment test, a set of structures are identified that can elicit clear judgments from normal subjects and agrammatic subjects, contrary to the claim that grammaticality judgments in Chinese are probabilistic and fragile. Most interestingly, the patients' overall performance patterns reveal a double dissociation between sentence comprehension and judgment of sentence well-formedness, i.e. sentence comprehension impaired while judgment of sentence well-formedness intact or vice versa. The results suggest that the two tasks are supported by independent processing mechanisms.

Studies in English and Italian have shown that non-fluent patients (patients with Broca's aphasia) find it more difficult to produce verbs than nouns, while some fluent patients (including patients with Wernicke's aphasia and anomic aphasia) show the opposite profile, which provides an evidence for a double dissociation. Many explanations have been offered for this double dissociation. Grammatical accounts claim that verb deficits reflect differences in morphological and/or syntactic complexity. Semantic-conceptual accounts propose that verbs are based on action meanings, which are stored in anterior motor regions (areas left of BA 4 in Figure 2); nouns are based on object meanings, which are stored in sensory cortex (BA 1 to 3 in Figure 2). Lexical accounts assume that verbs and nouns are stored in separate regions of the brain, independent of their semantic content.

Again Chinese offers a new explanation for the above issue for in Chinese many words are compounds with a complex internal structure, including VN compound verbs like "LOOK-BOOK" ("read") and VN compound nouns like "STAND-GOOSE" ("penguin"). Hence words may be nouns at the lexical level, but they contain verbal elements at the sublexical level, providing a challenge to existing explanations for the noun-verb dissociation. An object- and action-naming study was conducted with Chinese patients with Broca's and Wernicke's aphasia, designed to elicit several different compound types (VN nouns, VN verbs, VNN nouns, NNN nouns and NN nouns). Chen and Bates (1998) replicated the noun-verb double dissociation at the whole-word level, and provided further evidence for a double dissociation at the sublexical level: Broca's errors more often on the verb morpheme within VN nouns as well as VN verbs; Wernicke's errors more often on noun morphemes, and they often produce verb morphemes where none are required (e.g., substituting VV for NN words). Hence explanations for the noun-verb dissociation must apply at both the lexical and the sublexical level, a problem for all current accounts.

The lesion models derived from lexical tone production deficits in Chinese aphasia have added insight into the organization of the healthy brain and into the variety of processes that take place during the comprehension and production of prosody. We shall talk about this part of research in detail in the next section.

Recently, researchers have used Chinese as a crucial test case, using neuro-imaging techniques to approach the fundamental question whether shared neural mechanisms at higher cortical levels are engaged for pitch perception of linguistic and non-linguistic auditory stimuli. The results show that Chinese subjects show increased activity in the left premotor cortex, pars opercularis, and pars triangularis across four identification tasks (consonant, vowel, tone and pitch) while English subjects show increased activity in the left inferior frontal gyrus regions only in the vowel task and in right inferior frontal gyrus regions in the pitch task (Gandour, Wong, Hsieh, Weinzapfel, van Lancker and Hutchins, 2000; Hsieh, Gandour, Wong and Hutchins, 2001).

As we can see from the brief account above, the studies of Chinese aphasia have greatly broadened our perspectives on the organization of language in the brain. We have reasons to believe that as the research progresses, we shall benefit more.

4.2. Research on neuro-anatomical correlates of linguistic prosody

The contributions of the left and right cerebral hemispheres to the processing of linguistic prosody have been the goal of a great deal of study. Despite the increased interest in this issue, the precise nature of such deficits remains unclear (Baum and Pell, 1999, for a review).

4.2.1. Dichotic listening studies

When different melodies are played at the same time into the right and left ears, which melody does the person hear? The person always recognizes the melody that he heard in his left ear better. This is how we know that the right brain hemisphere, corresponding primarily to the left ear, is better at music. Similarly, if different words are spoken simultaneously into the right and left ears, the right ear, i.e. the left hemisphere, has better recognition.

To investigate lateralization of prosodic processing, a dichotic listening task, involving the simultaneous presentation of a pair of different stimuli to the right and left ear, is often used. Dichotic research has regularly shown a right-ear advantage (i.e. a left-hemisphere dominance) for linguistic stimuli (e.g. Bryden and Murray, 1985; Cutting, 1974; Dwyer, Blumstein and Ryalls, 1982), attributable to the right ear being more directly connected to the language-dominant left hemisphere and the left ear to the non-dominant right hemisphere through the primary contralateral pathways (e.g. Kimura, 1967).

Van Lancker and Fromkin (1973) compared ear preferences in tone-language speakers for three sets of stimuli: (i) pitch differences within language stimuli (lexical tone in the tone language Thai); (ii) language stimuli without pitch differences (consonant-vowel words pronounced on a mid-tone); and (iii) pitch differences alone (hums). Results from 22 native Thai listeners demonstrate that lexical tones and consonants are better heard at the right ear corresponding to the left hemisphere, while the hums show no ear preference. Van Lancker concluded that pitch discrimination is lateralized to the left hemisphere when the pitch differences are linguistically processed.

4.2.2. Neuro-imaging studies

Neuro-imaging studies, however, showed that speech prosody is processed in neither a single region nor a specific hemisphere, but engages multiple areas comprising a large-scale spatially distributed network in both hemispheres (Gandour, Tong, Wong, Talavage, Dzemidzic, X, Li and Lowe, 2004). Based on the studies of Chinese tone and intonation with healthy subjects, Gandour et al. proposed that lateralization is influenced by language experience that shapes the internal prosodic representation of an external auditory sign, and speech prosody perception is mediated primarily by the right hemisphere, but is left-lateralized to task-dependent regions when language processing is required beyond the auditory analysis of the complex sound.

In a literature review, Friederici and Alter (2004) put forward a different proposal, according to which the degree to which prosodic features are segmentally bound determines lateralization (Shipley-Brown, Dingwall, Berlin, Yeni-Komshian and Gordon-Salant, 1988). Prosodic expression that extends over a longer phrase or sentence (e.g. intonation) is thought to be lateralized to the right hemisphere, whereas prosodic expression that is associated with a syllable (e.g. tone) is thought to be ‘drawn’ toward the left hemisphere.

Imaging studies tell us what areas participate in a particular (linguistic) function, and lesion studies tell us what areas are necessary for normal functioning (Gandour, 2000: 76). This statement motivated the present study.

4.2.3. Lesion studies

Investigations on the brain-damaged speakers, i.e. lesion studies, have yielded quite inconsistent results. For example, investigations of the control of F0 in brain-damaged speakers have yielded mixed results, sometimes demonstrating deficits on the part of LH-damaged speakers (e.g. Cooper, Soares, Nicol, Michelow and Goloskie, 1984; Danly and Shapiro, 1982; Danly, Cooper and Shapiro, 1983; Ryalls,

1982) and sometimes showing deficits for RH-damage speakers (e.g. Baum and Pell, 1997; Behrens, 1989; Bryan, 1989; Shapiro and Danly, 1985; Weintraub, Mesulam and Kramer, 1981).

Contrarily, studies of agrammatic speakers of tone languages such as Chinese (Naeser and Chan, 1980; Packard, 1986), Thai (Gandour and Dardarananda, 1983; Gandour, Holasuit Petty, and Dardarananda, 1988), and Norwegian (Ryalls and Reinvang, 1986) have yielded quite consistent results, indicating that for tone language patients with aphasia, damage to the left hemisphere (rather than to the right hemisphere) impairs both tone production and perception.

However, the existing literature on lesion in the language-dominant hemisphere has provided neither a detailed description of nor a satisfactory explanation for the impaired prosodic components with respect to tone and intonation of Chinese aphasic speech.

4.2.4. Hypothesis

Several hypotheses have been advanced concerning the neural basis of prosodic processing, including right-hemisphere dominance for all aspects of prosody (Werintraub, Mesulam and Kramer, 1981), right-hemisphere specialization for emotional (paralinguistic) prosody but left-hemisphere specialization for linguistic prosody (Van Lancker, 1980). Even further, a lexicalization hypothesis has been put forward to account for the apparent left-hemisphere specialization of tones as opposed to the more bilateral representation of stress and intonation (Packard, 1986). It assumes that the functional lateralization of a prosodic feature is determined by whether the feature is specified in the mental lexicon, i.e. only lexically specified prosodic features are lateralized to the left hemisphere; these would at least include the features of lexical tone and lexical stress.

Van Lancker and Fromkin (1973) compared ear preferences in tone language speakers for three sets of stimuli (See §§ 4.2.1. for detail) and results from 22 native Thai listeners demonstrate that lexical tones and consonants are better heard at the right ear corresponding to the left hemisphere, while the hums show no ear preference. Van Lancker concluded that pitch discrimination is lateralized to the left hemisphere when the pitch differences are linguistically processed.

As described in the previous section, pitch cues play different roles in a language system. Van Lancker (1980) referred to the several roles of pitch in language as occurring at six levels in a functional hierarchy.

(1) The first level in the functional hierarchy of pitch in speech – and the least linguistically structured one – is that of voice quality.
(2) The second level of functional use of pitch in speech is often called 'para-linguistic', signalling personality traits and emotional states in the pitch contour.

This ‘affective’ function of tone has been seen to reflect individual psychology more than to represent features of the linguistic system.

(3) The third level is closely related to the affective level, but often analyzed as linguistic, in the sense of having patterns or structure; it is the distinguishing of attitudes by intonation, including attitudes toward the speaker himself, toward the remark being spoken, or toward the listener.
(4) The fourth level, syntactic use of intonation, refers to contrasts between types of sentence, such as question and statement, or types of clauses, such as appositive vs. restrictive.
(5) The fifth level is the most discretely structured use of pitch that lies in the phonological and lexical domains. Use of pitch to distinguish minimal pairs of lexical items has been called a “phonological” function of tone.
(6) The sixth level is the segmental level, e.g. Pitch contrasts also help listeners to tell voiced from voiceless counterparts of consonants at the same place of articulation.

Van Lancker (1980) concluded that linguistic analyses of tone and intonation, as well as experimental and clinical studies of pitch in the speech signal, indicate that pitch cues play various roles in language behaviour. These different functions, occurring at different levels of the grammar, are located along a continuum from most structured (e.g. tones) to least structured (e.g. voice quality) in linguistic terms. Here the term voice quality refers to the habitual spectral composition of the speaker’s sounds (as determined by the average shape of the speaker’s vocal tract and by the characteristics of the vocal cord movements). Hemispheric laterality studies show that highly structured pitch contrasts are associated with left cerebral processing, whereas least linguistically structured pitch cues are specialized to the right hemisphere. Intermediate functional roles of pitch, those conveyed on intonation contours, are correspondingly ambiguous with respect to laterality. Van Lancker formulated her conclusion as follows: ‘pitch in the acoustic signal is processed in the brain according to its functional context, properties of which may be specialized in either hemisphere.’ (Van Lancker, 1980: 201) This is often referred to as the functional lateralization hypothesis. This hypothesis is taken as point of departure in the present investigation.

4.3. Linguistic experience in the hemispheric processing of prosody

Recent neuro-imaging studies have demonstrated that processing lexical tones of Chinese (Klein, Zatorrre, Milner and Zhao, 2001; Wang, Sereno, Jongman and Hirsch, 2003) and Thai (Gandour, Wong, Hsieh, Weinzapfel, Van Lancker and Hutchins, 2000) by native speakers of these languages predominantly involves the left-hemisphere functional regions, whereas such hemispheric specification is not characteristic of non-native speakers.

Results have been also reported from dichotic listening studies, comparing four groups of listeners (Chinese, English–Chinese bilinguals, Norwegian, and

American) in their processing of Chinese lexical tones (Wang, Behne, Jongman.and Sereno, 2004) and show that for proficient early bilinguals, who have acquired the functional use of the Chinese pitch contrasts, hemispheric processing of Chinese lexical tones becomes native-like. They propose a dynamic pattern in the processing of Chinese tones by non-native speakers, from no hemispheric lateralization to native-like left-hemisphere dominance.

These results consistently show that native speakers of tone languages process tonal contrasts predominantly in the left hemisphere, whereas linguistically irrelevant pitch information such as hums yielded no significant ear effects. For non-native speakers whose native languages do not have tonal distinctions, tones as well as hums were not processed as language and were thus not lateralized in the left hemisphere. Taken together, these studies suggest that left-hemisphere specialization for tones occurs only when they are part of the speaker's linguistic system.

Two subsequent questions arise given the above-mentioned patterns of tone processing by native and naive (or lack of lexical tone experience) speakers.

First, what is the hemispheric processing pattern of lexical tones for agrammatic speakers of Chinese with brain damage in the left hemisphere? Do they process lexical tones in left-hemisphere language-sensitive regions as was exhibited by native speakers or do they behave like non-native speakers of Chinese, i.e. with no hemispheric lateralization? If there is deficit in the linguistic system of the patients with aphasia, we may predict that the patients would behave like non-native speakers.

The second question concerns the hemispheric specialization of intonation. Do Chinese speakers process intonation in the same way as they do lexical tones? Do agrammatic speakers of Chinese with LH damage process intonation native-like or non-native-like?

5. Overview of the thesis

The goal of this study is to find out whether linguistic prosody at the word-level and sentence-level are processed equally well by Chinese patients with Broca's aphasia relative to normal subjects and second language (L2) learners of Chinese. Laboratory-controlled experiments were carried out with Beijing patients with aphasia who had a unilateral damage in the left hemisphere of the brain. We aimed to test how and to what extent they process Chinese lexical tones and intonation relative to Beijing healthy speakers and Beijing learners with and without tone language as their native language.

5.1. Primary goal

Phonology plays a crucial role in language processing. It is the medium by which sound information maps onto higher levels of language processing (e.g. words). There has been a strong convergence of results from the neuropsychological literature and the neuro-imaging literature to suggest that the neural basis of phonological processing is lateralized to the left hemisphere, and encompasses a distributed neural system that includes posterior brain structures (superior temporal gyrus) and anterior brain structures (inferior frontal gyrus) (Blumstein, 1998; Pugh, Shaywitz, Shaywitz, Constable, Skudlarski, Fulbright, Bronen, Shankweiler, Katz, Fletcher and Gore, 1996; Fiez, Raichle, Miezen, Petersen, Tallal, and Katz, 1995; Paulesu, Frith and Frackowiak, 1993; Paulesu, Frith, Snowling, Gallagher, Morton, Frackowiak and Frith, 1996; Démonet, Chollet, Ramsay, Cardebat, Nespoulous, Wise, Rascol and Frackowiak, 1992; Démonet, Price, Wise and Frackowiak, 1994; Démonet, Fiez, Paulesu, Petersen and Zatorre, 1996; Sergent, Zuck, Lévesque and MacDonald, 1992; Zatorre, Evans, Meyer and Gjedde, 1992; Zatorre, Meyer, Gjedde, and Evans, 1996, but cf. Poeppel, 1996).

The primary goal of the present study is to study the effect of brain lesion on the processing of linguistic prosody by Broca's patients in terms of lexical tone and sentence intonation. We are interested in the modular structure of the speech/ language apparatus. As tone languages permit rigorous control over the phonological influence on F0 patterns and as it has been suggested that detailed, acoustic-phonetic studies of intonation should be carried out on tone languages whose tonal phonology is well understood (Ladd, 1981), Beijing dialect has been chosen in the present study. That is to say, we shall examine how well Beijing Broca's patients process tones and intonation and to what extent the deficit may be caused by a structural deficit and/or processing limitations.

5.2. Hypotheses

We rest our investigation on the following three hypotheses. We started out from the assumption that language users may employ different mechanisms when processing speech in a tone language, depending on their native language background. We investigated the degree of difficulty in language processing, i.e. the impairment of lexical tones and intonation by patients with Broca's aphasia, and compared this with the degree of processing difficulty experienced in the acquisition of Beijing as a tone language by non-native but healthy learners from varying linguistic backgrounds. The idea is that native language speakers and second language (L2) learners have different mental lexicons and that this difference will be reflected in their speech behaviour. Among L2 learners, difference in L1 – whether tone language or not – will also lead to a different mental lexicon and the difference in turn will be observed in their speech behaviour.

Hypothesis One: Research has demonstrated that language acquisition begins at a very early age and proceeds at an amazingly fast pace. Early exposure to a language has a lasting impact on speech processing routines in adults, i.e., listeners use a processing apparatus specially tuned to their mother tongue. Consequently, they have difficulty in dealing with sound structures that are alien to the language they heard since infancy. That is to say, speakers have a structural deficit to the phonological contrasts which do not occur in the native language. In other words, L2 speakers display a phonological 'deafness' to the language. Research suggests that the phonological 'deafness' is robust in that it is resistant to learning a second language, and even to specific training (e.g. Dupoux and Peperkamp, 2002).

Our hypothesis is that the degree of the phonological 'deafness' increases as phonological similarity between the first language (L1) and the second language (L2) decreases. For instance, L2 speakers of a tone-less language L1 will display a complete deafness to tone contrasts in comparison with L2 speakers of tone-language L1. L2 speakers of a closely related tone-language L1 will have less 'deafness' than those L2 speakers of a less closely related tone-language L1.

If Beijing Broca's patients have a structural deficit, they would display some phonological 'deafness', which may bear some resemblance to the performance of non-native speakers of Beijing dialect. If the patients' processing resources are limited, they would still behave like their native healthy controls when they are provided enough time.

We predict that all the L2 speakers will show a structural deficit in the perception of lexical tones in Beijing dialect. Uygur speakers will show a severe structural deficit in lexical-tone processing in comparison with the two tone-language groups as the Uygur language is a tone-less language, belonging to the Altaic family while Beijing, Nantong and Changsha are tone languages, belonging to the Sinitic family. Within the tone-language groups, we predict that it is more difficult for Changsha speakers than for Nantong speakers to process lexical tones and sentence intonation in Beijing dialect as Nantong is more closely related to Beijing dialect than Changsha dialect is.

Additionally, we would like to test whether or not unique, neural mechanisms in the left hemisphere are recruited for the processing of linguistic prosody. More specifically, we want to find out how patients with lexical tone impairment categorize pitch variation. Based on the group-comparative analysis, similarity or difference will be found and discussed in terms of pitch variation (lexical tone versus sentence intonation) and response patterns (native speakers versus non-native ones).

Hypothesis Two: If the locations in left and right hemisphere are of paramount importance in determining hemispheric specialization, then unilateral lesions should result in uniform behavioural deficits regardless of the linguistic status of the prosodic elements. If, on the other hand, different linguistic functions result in hemispheric specialization, then we may expect behavioural deficits to vary depending on the specific role of prosodic elements in language. Another possibility is that both hemispheres are engaged in prosody, but attend to different aspects of the speech signal. By focusing on F0, we can obtain a clear picture of abstract linguistic elements in relation to their phonetic and neural correlates. Most studies that investigated intonation in unilaterally damaged patients of tone languages employed sentence-length stimuli in emotional contexts without comparable stimuli in non-emotional contexts.

Hypothesis Three: If, as suggested language abilities are functions of the left hemisphere, in tone languages, lexical tone, which is a phoneme and can distinguish among lexemes, should play the same linguistic role as segments (e.g. vowels). Therefore, like vowels, lexical tone is also subject to deficit following damage to the left hemisphere.

More generally, the hypotheses to be tested are whether lexical tone and linguistic intonation are separate functions with separate locations in the brain and to what extent the lesion affects linguistic prosody at word level and sentence level. Although the importance of speech melody is generally well understood, relatively little attention has been paid to prosodic deficits following focal brain damage.

5.3. Approach

Following the above hypotheses, we carried out a comparative study on lexical tone and sentence intonation by Broca' patients of Beijing dialect and two types of healthy speakers, viz. speakers from L1 speakers (Beijing) as well as L2 Beijing speakers with tone-language L1 (Nantong and Changsha) and a tone-less L1 (Uygur).

Methods. We tested our hypotheses by an experimental approach to lexical tones and intonation with Broca's patients as well as healthy speakers. The experimental approach primarily depends on acoustic measurements of speech production and perception: (i) fundamental frequency in lexical tone production, and (ii) subjects' responses and their reaction time to stimuli manipulated systematically in terms of their acoustic make-up.

We ran two types of experiments, (i) production and (ii) linguistic perception of the pitch variation at word level and sentence level. As there is no norm for sentence intonation as there is for lexical tones in terms of correctness, and production of patients involves much more factors than perception, we confined our production experiment only to lexical tones.

We first ran a lexical production experiment, whereby we determined whether the subjects correctly produced the words, either the tone pattern or the vowel as it was intended by the speaker of the utterance, and compared the results of the two. Secondly, we carried out a series of perception experiments on linguistic prosody (lexical tone and sentence intonation), with natural stimuli and systematic resysthesized stimuli, to study how well the patients perceived and characterized lexical tones and sentence intonation relative to healthy controls as well as to non-native speakers. Lastly, we probed into the issue how time pressure affected subjects in characterizing lexical tones.

Subjects. According to our hypotheses, we recruited a group of Broca's patients who speak the Beijing dialect as their native language and had suffered unilateral brain damage in the left hemisphere, and three groups of Beijing learners as our control groups. Taking together, the four groups are, (i) native Beijing (L1) speakers, and (ii) Beijing learners (L2), who are sub-grouped into L2 speakers with tone language L1 (Nantong and Changsha speakers), and L2 speakers with a tone-less L1 (Uygur speakers). The sites where the four groups of speakers were recruited are illustrated in Figure 5 on page 12.

Fourteen Broca's patients as diagnosed by Professor Zhan Benshu from Tianjin General Hospital were recruited in our experiments. These patients are native speakers of Beijing dialect and all of them had single unilateral damage in the left hemisphere (detailed information is available in the following chapters where necessary). Thirty Beijing native speakers made up a healthy controlled group and fifty-one Chinese non-Beijing students constituted a second-language learner (L2) group. The L2 group is further divided into two types based on whether or not their mother tongue is a tone language, i.e. twenty-one speakers whose native language is a tone language (Nantong or Changsha) and thirty native speakers of a tone-less language, Uygur. All the subjects took part in the experiments during the years of 1999 to 2003.

5.4. Main issues

Our primary concern is the effect of brain lesion on the processing of prosody in Chinese aphasic speech, i.e., whether or not there is a structural deficit in Beijing speakers with aphasia. More specifically, we shall address the following five main questions:

(1) To what extent does brain lesion affect lexical tones in the speech production of Broca's patients? Are all lexical tones in Beijing dialect equally impaired in speech production? What happens to the vowels that are the primary tone-bearing segments for lexical tone? (Chapter 2)
(2) To what extent does lesion affect linguistic prosody in speech perception? Are lexical tone and sentence intonation equally impaired in the perception of Chinese patients with aphasia, because pitch variation is used both at the word and sentence level in Chinese? (Chapter 3)
(3) To what extent does phonological 'deafness' affect linguistic prosody in speech perception? Are lexical tone and sentence intonation equally impaired in the perception of L2 speakers of Beijing dialect? (Chapter 4)
(4) What are the characteristics of lexical tone categorization by Chinese patients with aphasia relative to healthy controls as well as non-native speakers? (Chapter 5)

(5) In line with Question 4, what role does time pressure play in lexical tone categorization? Is it true that the use of prosody by patients with Broca's aphasia deteriorates progressively when they are allowed less time to perform language processing tasks? This would allow us to conclude whether lexical tone deficit is due to processing limitation or a structural deficit. (As we only found a strong effect of time pressure in one of five continua, we reported the preliminary result in Appendix I)

5.5. Expected findings

As has been stated, we propose to evaluate the deficit displayed by agrammatic speakers by testing whether the linguistic knowledge the agrammatic speakers acquired at an early age is affected. If we find that agrammatic patients' performance is intermediate between that of native Beijing speakers and non-native Beijing learners, i.e. showing an ear which is phonologically half normal and half 'deaf', we will claim that the linguistic knowledge which is acquired at early age is affected in agrammatic patients.

5.6. Outline of the remainder of the dissertation

The thesis comprises six chapters plus one appendix. Part of Chapter One, Chapter Two to Five and Appendix I come from six papers, published in, accepted by, or submitted to professional journals or specialized volumes. Chapters Two to Five and the Appendix address the issues (1) to (5) presented above, respectively; Chapter Six summarizes some of the possible consequences of the present study and suggestions for further research.

The remainder of the thesis is then organized follows. In Chapter Two, we reported a case study on production of a brain-damaged Chinese speaker, providing evidence for separate tonal and segmental tiers in the lexical specification of words. Chapter Three examines the perception of pitch variation at the word level and sentence level by patients with LH damage. We did a parallel study with native and non-native speakers of Beijing dialect in Chapter Four in order to evaluate the deficit (if there is any). Chapter Five is devoted to the perceptual representation of lexical tones in Chinese listeners (with aphasia) relative to native and non-native healthy controls. We further did an example study of time pressure in Appendix I. Chapter Six presents a conclusion chapter summarizing the findings we obtained in this investigation, and placing them in a more general framework.

We hope our investigation will shed light on the relationship between language and mechanisms of neurobiology, and also offer new windows on cognitive mechanisms for psycholinguists who attempt to understand the dynamics of language processing and language acquisition.

Chapter Two

Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker[4]

Abstract

We present an acoustic study of segmental and prosodic properties of words produced by a female speaker of Chinese with left-hemisphere brain damage. We measured the location of the point vowels /a, e, ə, i, y, o, u/ and determined their separation in the vowel plane, and their perceptual distinctivity. Similarly, the acoustic properties of the four lexical tones were measured in the F0 × time space. The data for our brain-damaged speaker were compared with those of a healthy control speaker. Results show that the patient's vowels hardly suffered from her lesion (relative to the vowel dispersion in the healthy control speaker), but that the identifiability of the four lexical tones was greatly compromised. These findings show that the tonal errors in aphasic speech behave independently of the segmental errors, even though both serve to maintain lexical contrasts in Chinese, and are therefore part of the lexical specification of Chinese words. The present study suggests that the specification of segmental and tonal aspects of lexical entries in Chinese, and in tone languages in general, are located or processed separately in the brain.

---

4 This chapter has been published as J. Liang and V.J. van Heuven (2004) Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language,* 91, 282-293.

1. Introduction[5]

Studying the breakdown of segmental and tonal aspects of word production in acquired language aphasia can provide valuable insight into the nature of normal speech production. In Chinese both segmental and tonal structure are used to differentiate individual words in the lexicon. Packard (1986) demonstrated that left-damaged non-fluent aphasic speakers of Chinese suffered from a tonal production deficit that was both quantitatively and qualitatively equivalent to the deficit experienced by these speakers in the production of consonants. This was taken as evidence that, in tone languages, the lexical specification of the tone contour resides in the left hemisphere, along with consonantal information. Note that no mention is made of vowel production in Packard's study. However, all the available evidence to date points to the fact that the vowels, and any voiced coda segment that may follow them, carry the tonal information (e.g. Howie, 1976). Clearly, then, vowel production is closely linked to tonal production, to the extent that one would predict that vowel deficits and loss of tonal contrasts should always co-occur in aphasic speech. On the basis of these results it seems obvious that the lexical specification of words in a tone language such as Chinese is holistic: tones and segments reside in the same module of the brain, and in the case of a lesion either both are affected or both remain unaffected.

The present research again targets the question whether the segmental and tonal specification of lexical items in Chinese may or may not be independent of each other in aphasic speech. Our study is motivated by the consideration that the past literature on patients with anterior lesions in the language-dominant hemisphere has provided neither a detailed description of, nor a satisfactory explanation for, the impaired lexical prosody relative to the loss of phonemic quality in Chinese aphasic speech.

In Packard's (1986) study, the aphasic speech was perceptually scored such that deviations had to be identified between a correct model token (produced by a healthy speaker) and the aphasic's repetition of the model. Listeners were not required to perceptually identify the tones or consonants in the aphasics' speech. Consequently there were no functional measures of tonal or segmental loss. Moreover, the perceptual comparisons were made by only three listeners, and no acoustic analyses – let alone automatic classification – of the aphasics' tokens were attempted.

---

[5] This article is the full version of a much shorter conference paper presented at the EURESCO conference "The Science of Aphasia", in Giens, France, 14-19 September 2001. The authors thank the Leiden University Foundation (LUF) for subsidizing the first author's trip to this conference. The first author also gratefully acknowledges a scholarship granted her by the China Scholarship Council, which funded her stay at ULCL-HIL in 2000/01. We thank Robert S. Kirsner of UCLA for checking and correcting our English.

Gandour, Holasuit Petty and Dardarananda (1988) and Gandour, Ponglorpisit, Khunadorn, Dechongki, Boongird, Boonklam and Potisuk (1992) did a series of acoustic phonetic studies on tone loss in Thai (a language with a five-member lexical tone contrast), showing that almost all tonal substitutions resulted in mid, low, or falling tones. Of the five Thai tones, the falling tone proved most resistant to disruption. In cases of severe non-fluent aphasia, this tone was often the only one the patient could produce successfully.

The F0 range of the tone space for brain-damaged speakers is not compressed. Gandour and co-workers claimed that the extent to which a left hemisphere non-fluent aphasic speaker may experience difficulty in producing Thai tonal contrasts, cannot be attributed to a narrowing of the F0 range and that it is only in longer utterances that non-fluent aphasic speakers exhibit flatter F0 patterns. Gandour at el. further concluded that F0 disturbance in aphasia is due primarily to a timing deficit over larger-sized linguistic units. Consequently, Gandour et al.'s results seem to contradict Packard's (1986) finding that non-fluent aphasic speakers of Chinese exhibited a far more extensive deficit in tone production.

Recently, a cross-linguistic functional neuroimaging study has indicated that healthy Chinese subjects showed increased activity in the left precentral gyrus/premotor cortex (PRG), as well as in the pars opercularis (F3o) and pars triangularis (F3t) of the left inferior frontal gyrus (IFG) for each of four discrimination tasks involving consonants, vowels, tones and tone-like non-speech pitch contours, respectively (Hsieh, Gandour, Wong and Hutchins, 2001). This finding suggests that all the phonological properties of Chinese words, both segmental and tonal, are located in the same parts of the brain, from which the prediction follows that all phonological properties should be affected to the same degree when a native speaker of Chinese has a brain lesion in these parts. Unfortunately, the literature provides neither a detailed account of loss of tone contrasts across the full set of lexical tones in the language, nor does it contain an indication of the severity of the tonal deficit relative to the concurrent loss of segmental contrasts. Our present neuropsychological study aims to fill in this gap.

1.1. Short introduction to the Standard Chinese vowel and tone system

Before turning to our experiment, it may be useful to first present a brief description of the vowel and tone inventory of Standard Chinese.

Vowel phonemes in Standard Chinese

Much has been written on the nature of the vowel system of Standard Chinese (e.g. Hartman, 1944; Hocket, 1947; Howie, 1976; Wiese, 1997). A detailed discussion of the vowel phonemes is beyond the scope of the present study. All we need here is a simple statement that may serve as a minimal linguistic framework for an acoustical classification of vowel tokens. Therefore, we have adopted the seven-category vowel system as displayed in (1).

(1)

	front	central	back
high	i	y	u
mid	e	ə	o
low		a	

Several authorities group /i/ and /y/ together in the high-front category. However, since lip rounding, which would then be the sole feature differentiating rounded /y/ from unrounded /i/, ceteris paribus, results in greater acoustic backness (i.e. lower F2 formant values), we felt justified in reclassifying /y/ as high-central, as indicated in the table. Also, the literature frequently mentions low allophones of mid /e/ and /o/. We therefore expect considerable variation between the mid and low categories for these vowels acoustically along the F1 formant dimension), but did not set up separate mid and low categories for these vowels.

Standard Chinese lexical tones in isolation

Chinese has four lexical tones: 'high level', 'rising', 'low dipping', and 'falling', which are traditionally referred to as Tone 1, Tone 2, Tone 3, and Tone 4, respectively. The four tones of Standard Chinese are distinguished primarily by the F0 contour during the vowel (Howie, 1976). Despite the overriding importance of F0 movement as an acoustic correlate of tonal category, there are other, secondary correlates in the phonetic signal. Specifically, duration is moderately well correlated with tone category (Howie, 1974; Ho, 1976) as are differences in the amplitude contours (Howie, 1976; Whalen and Xu, 1992).

There is a long-standing debate in the linguistic literature whether contour tones should be considered single contour units or sequences of two level units, e.g., a low followed by a high or a high followed by a low. Abramson (1978), Clark (1978), Pike (1948) and Wang (1967) treat contour tones found in languages such as Thai and Mandarin as single units, while Duanmu (1994), Gandour (1974), Leben (1973), Woo (1969) and Yip (1991) analyze these contour tones as sequences of high and low targets. For instance, Tone 2 can be analyzed as a single rise in the former, 'unitary' approach, or as a bi-tonal sequence of a level low tone followed by a level high tone. In the latter, 'compositional' approach, any Chinese tone is viewed as a sequence of two subsyllabic timing units, or 'morae', each of which functions as a tone-bearing unit. We will not enter into this debate at this time, and for expository simplicity assume the unitary solution, as exemplified in (2):

(2)

	Static tone	Dynamic tone
High	Tone 1 (‘high level’)	Tone 2 (‘rising’)
Low	Tone 3 (‘low dipping’)	Tone 4 (‘falling’)

1.2. Structure of the study

In our experiment we employed an isolated (monosyllabic) word-reading task with materials in which the seven vowels and four lexical tones were systematically varied. Speech materials were collected from one left-hemisphere damaged female speaker of Chinese (see below), as well as from a healthy control speaker, such that the range of tonemic and segmental lexical contrasts in the language was adequately sampled. The recordings were then analyzed in two different ways. First, tones and vowels were submitted for identification to a group of native Chinese listeners, in order to determine the extent of the loss of contrast in both the tonemic and segmental aspects of the aphasic speech. Second, the materials were acoustically measured and analyzed in terms of their physical distinctiveness relative to that of the healthy control subject’s speech.

2. Methods

2.1. Primary recordings

Speaker. PYF, a 38 year old, right-handed female technician, participated in this study in April 2000, one month after the onset of her left-hemisphere damage. After suffering her lesion, PYF’s spontaneous speech was cumbersome, characterized by word-finding difficulties, incomplete syntactic constructions, and errors in sound production that rendered her speech virtually unintelligible. Her non-verbal communication was still effective, and apart from her aphasia, she was able to carry out the activities of everyday life without difficulty. CT brain imagining then indicated atrophy in the left precentral gyrus (PRG), and pars triangularis (F3t). PYF was clinically diagnosed as a Broca’s aphasic by Prof. Zhang Benshu in Tianjin General Hospital. PYF was without concomitant dysarthria as reported by the referring neurologist, and as judged by the first author from the patient’s performance on the verbal agility subtest of Chinese Diagnostic Aphasia Examination (Gao, 1993), which is a Chinese adaptation of the Boston Diagnostic Aphasia Examination. PYF had completed twelve years of formal education and spoke a regional accent of Standard Chinese typical of Jixian area in Tianjin, China.

Recording procedure. Our aphasic speaker produced single words in citation form by reading out a word list in a quiet a room in the hospital. The list was taken from the general list for ‘tones’, ‘initials’ and ‘finals’ of the (revised) Word List for Dialect Investigation (Institute of Linguistics, Chinese Academy of Social Sciences, 1999), which was designed as a descriptive tool to determine the phonotactics (including the distribution of lexical tone) of any Chinese dialect. The general word list consists of three parts: (i) 99 words for tone types, (ii) 115 words for syllable onset types (i.e. initial consonants), and (iii) 110 words for rhyme types (i.e. vowels plus postvocalic consonants), comprising a set of 324 monosyllabic words.

During the recording session the speaker was seated at a desk, and read the stimulus words from sheets of paper. The recordings were made onto a cassette deck (Sony WM-GX614) using a high-quality omni-directional microphone clipped onto the speaker's lapel, about 10 cm below the mouth.

If the speaker failed to read out a word, she was prompted immediately by the experimenter to repeat the attempt. If she failed on the second trial, no further attempts were undertaken to get the speaker to produce the item. In all, PYF managed to record 217 tokens. In 107 cases, more or less evenly distributed over the stimulus list, the speaker failed to respond with the required item, even after the second attempt. Clearly, the assignment was difficult for the speaker; yet, we know both from the clinical tests run beforehand and from the comments of the speaker herself that the problems did not arise during the reading of the visual stimulus, but were entirely the result of her seriously impaired speech production.

The normal control speaker was a male professional radio broadcaster, aged 34, who volunteered for the job and received no remuneration for his services. These recordings were made in a radio studio, using high-quality equipment. The control speaker read the 324 items from paper one by one, and required no second trials.

2.2. Acoustic analysis

The recordings were digitized (22 KHz, 16 bit) and then acoustically analyzed using the speech signal processing package Praat (Boersma and Weenink, 1996).

We decided to use the full seven-member inventory of Chinese vowels, namely /a, e, ə, i, y, o, u/, as this would allow us to best determine the size and orientation of the patient's vowel space relative to that of the healthy control speaker. The center frequencies of the lowest four vowel formants (F1, F2, F3, F4) were estimated by an LPC formant tracking algorithm (Split-Levinson algorithm, 512-points Hamming window, 10 ms window shift).

Twelve tokens of each vowel type were selected, both for the patient and for the normal speaker. Twelve was the largest number of tokens available for all vowel types in the word list. It was felt that no purpose was served by having unequal numbers of tokens across types. When more than twelve tokens were available for a given vowel type, only the first twelve were included in the selection.[6] The formants

[6] In fact, the maximum number of tokens available in the word list for any vowel type was 15 (for /a/).

were sampled at the point in time where F1 reached its maximum, i.e., when the mouth was maximally open. This moment is generally accepted as the most representative point in the acoustic specification of vowel quality. Perceptually realistic scaling of vowel quality was then obtained by expressing vowel openness in terms of F1 in Barks (number of critical bands). Vowel backness was captured by the second formant, F2, plotted in Barks as well (see Bladon and Lindblom, 1981; Hayward, 2000).

In order to acoustically measure the tones, fundamental frequency (F0) was determined by the autocorrelation-based pitch tracking algorithm implemented in Praat. Automatic pitch period detection was supplemented by manual correction of the F0 values. The tokens were more or less evenly distributed over the four lexical tones (although Tone 3 remained somewhat underrepresented). F0 was calculated every 10 ms. As the (sonorant portions of the) onset consonants in the syllable have been shown to carry no distinctive information with respect to lexical tone in Chinese (Howie, 1976), the relevant time portion of the tone pattern included only the vowel and the sonorant coda consonant following it (if at all present). The F0 pattern for each token was then time-normalized by computing the F0 at 11 equidistant points in time from beginning to offset of the tonal stretch defined above. F0 measurement points were then psychophysically scaled in ERBs (Equivalent Rectangular Bandwidths), which is currently held to be the most adequate perceptual representation of pitch and pitch intervals in speech (Hermes and van Gestel, 1991; Ladd and Terken, 1995; van Heuven and Haan, 2000). The ERB scale is in between a linear, physical frequency scale (in hertz) and a representation in terms of a musical scale (in semitones).

2.3. Perceptual identification

Thirty listeners, between 20 and 22 years of age, equally divided between males and females, all of whom were university students and native speakers of Standard Chinese, wrote down the 217 monosyllabic words spoken by PYF in an open-set format. The listeners were told that only ordinary, everyday Chinese words would be contained on the tape. Each word was played twice with a 5-second pause between repetitions. After every tenth stimulus pair a short beep was inserted to orient the listeners and ensure that they would not get lost and skip one or more stimuli.

The stimulus tape was presented to the listener over headphones in a language laboratory. Responses were entered in the appropriate spaces on the answer sheets either in Chinese characters or in Pinyin (alphabetic transliteration with diacritics marking tone) depending on the individual listener's preference.

The 324 items produced by the control speaker were not offered for perceptual identification. We assumed that both tones and segments would be identified (close to) 100 percent correct. The results of the acoustic measurements and – especially – of the subsequent automatic classification procedure indicate that this assumption was basically correct.

3. Results

We now turn to the results of the listening tasks. We will first examine the perceptual distinctiveness of the segmental and tonal contrasts in PYF's speech and then consider the acoustic measurements. Our goal here was to see to what extent the seven vowel types and the four lexical tones can be identified from their acoustic make-up automatically, by means of a classification algorithm.

3.1. Perception of segmental and tonal contrasts

The results of the listening experiment are presented in Table 1 (for tone identific-ation) and Table 2 (for vowel identification).

Table 1: Absolute and relative (in %) frequencies of tones identified by 30 listeners crosstabulated against stimulus tone as intended by the (aphasic) speaker. Correct identifications are on the main diagonal in bold face against a shaded background.

Intended Tone (down)	Perceived tone (across)						
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Unid.	Total
Tone 1 (55)	**1575 (87.5)**	35 (1.9)	70 (3.9)	82 (4.6)	38 (2.1)		1800 (100)
Tone 2 (35)	1442 (78.8)	**160 (8.7)**	106 (5.8)	73 (4.0)	49 (2.7)		1830 (100)
Tone 3 (214)	811 (67.6)	74 (6.2)	**241 (20.1)**	48 (4.0)	26 (2.2)		1200 (100)
Tone 4 (51)	1233 (73.4)	32 (1.9)	69 (4.1)	**276 (16.4)**	68 (4.0)	2 (0.1)	1680 (100)
Total	5061 (77.7)	301 (4.6)	486 (7.5)	479 (7.4)	181 (2.8)	2 (0.0)	6510 (100)

Listeners failed to produce an interpretable response in only two cases (out of the total of 30 x 217 = 6510 responses). In 181 cases the response contained a word with the neutral tone (which was not elicited in the production task). The large majority of the responses were constrained to the target set of monosyllabic words with full tone. On average, only 35.6% of the tones were identified correctly, but with great differences between the rates for individual tones: Tone 1 (high level, 87.5% correct identification) < Tone 3 (low dipping, 20.1%) < Tone 4 (falling, 16.4%) < Tone 2 (rising, 8.7%). The effect of stimulus tone on percent correct identification is significant by a one-way analysis of variance (ANOVA) with tone as a fixed factor (disregarding neutral-tone responses), $F(3, 213) = 300$ ($p < 0.001$). A subsequent post-hoc analysis (Scheffé procedure, $\alpha = 0.05$) indicates that Tone 1 differs from all other tones, but that tones 2 and 4 do not differ from each other, as is also the case

for tones 3 and 4.
As is predictable from the high percentage of correct responses for Tone 1, there is a strong bias in the confusion pattern of the identifications favouring Tone 1 responses. It is by far the most frequent response alternative for each of the four stimulus tones: 78.8%, 67.6% and 73.3% for tones 2, 3 and 4, respectively. It is impossible to determine, at this point in the data presentation, what might be the cause of this asymmetry in the confusion pattern: do all the tones assume properties that make them sound like Tone 1, or is tone not specified at all in PYF's speech and is Tone 1 chosen as a default (possibly because it is the most frequent tone in the Chinese lexicon)?

Table 2: Absolute and relative (in %) frequencies of vowels identified by 30 listeners crosstabulated against stimulus vowel as intended by the (aphasic) speaker. Correct identifications are on the main diagonal in bold face against a shaded background.

vowel (down)	Perceived vowel (across)								
	a	e	ə	i	y	o	u	unid.	Total
a	**356 (98.9)**	1 (0.3)	1 (0.3)			2 (0.6)			360 (100)
e		**356 (98.9)**	2 (0.6)	2 (0.6)					360 (100)
ə			**354 (98.3)**	1 (0.3)		2 (0.6)	3 (0.8)		360 (100)
i		1 (0.3)	13 (3.6)	**337 (93.6)**				9 (2.5)	360 (100)
y	1 (0.3)	8 (2.2)	18 (5.0)	24 (6.7)	**293 (81.4)**	11 (3.1)	5 (1.4)		360 (100)
o	2 (0.6)	24 (6.7)	3 (0.8)		2 (0.6)	**318 (88.3)**	11 (3.1)		360 (100)
u		1 (0.3)	11 (3.1)			47 (13.1)	**301 (83.6)**		360 (100)
Total	359 (14.2)	391 (15.5)	402 (16.0)	364 (14.4)	295 (11.7)	380 (15.1)	320 (12.7)	9 (0.4)	2520 (100)

The identifiability of the seven vowel types is much less compromised (91.9% correct), and does not differ much across vowel type: 98.9%, 98.9%, 98.3%, 93.6%, 81.4%, 88.3%, and 83.6% correct for /a, e, ə, i, y, o, u/, respectively, as shown in Table 2.

Nine out of the total of 30 (listeners) x 7 (types) x 12 (tokens) = 2,520 responses, or 0.4%, were missing; all missing data were instances of the vowel /i/. Perceptual confusion is restricted to the high round vowels: /u/ is confused in 17 percent of the tokens, typically with /o/ (13%), while /y/ has 19 percent error identification, including 6% confusion with /i/.

The effect of vowel type on the error percentage is significant, $F (6, 77) = 4.6$ ($p < 0.001$) but none of the seven vowel types differ from one of the others (Scheffé procedure, $\alpha = 0.05$). Clearly, then, vowel identification in PYF's speech is superior to the identification of her lexical tones.

3.2. Acoustic analysis

The results from the perception experiment show that the aphasic speaker's tones were severely impaired while her vowels were relatively well preserved. In order to investigate their physical distinctiveness, we will now present the acoustical analysis of the 217 tokens of lexical tones and the twelve tokens of seven vowels /a, e, ə, i, y, o, u/ relative to those of the control speaker.

We first determined the mean **duration** of the four tones of the patient and of the healthy professional speaker and compared our results with the mean duration of monosyllabic words of eight non-professional speakers (Xu, 1997) shown in Figure 1.

First of all, Figure 1 shows that our professional speaker observes a very deliberate speaking style, with much longer durations of each of the four tones than is found for the non-professional speakers in the reference sample collected by Xu (1997). Also, the figure reveals that the tones differ substantially in their duration. Specifically, Tone 3 is considerably longer than the other three tones, which do not differ very much from each other. This durational pattern is found both for our control speaker and for Xu's reference sample. Therefore, we provisionally conclude that the relative timing behaviour of our professional speaker is representative of the normal speaker sample at large.

In contradistinction, the durations in the aphasic speaker do not differ across the four lexical tones. A two-way ANOVA with speaker (aphasic versus professional) and lexical tone as fixed effects shows that both main effects and their interaction were significant, $F (1, 426) = 13.5$ ($p < 0.001$) for speaker, $F (3, 426) = 49.4$ ($p < 0.001$) for tone, and $F (3, 426) = 29.4$ ($p < 0.001$) for the interaction. Subsequent one-way ANOVAs indicate that the effect of lexical tone is highly significant for the professional control speaker, $F (3, 213) = 82.1$ ($p < 0.001$). Scheffé post hoc tests for contrast ($\alpha = 0.05$) indicate that Tone 1 is significantly shorter (380 ms) and Tone 3 significantly longer (585 ms) than the other two tones, which do not differ from one another (424 and 439 ms for Tones 2 and 4, respectively). The effect of tone is much smaller – though still significant – for our aphasic speaker, $F (3, 213) = 7.3$ ($p < 0.001$). Post hoc tests reveal that Tone 4 is significantly shorter (390 ms) than the other three tones, the durations of which do not differ from each other (440, 431 and 462 ms for tones 1, 2, and 3, respectively).

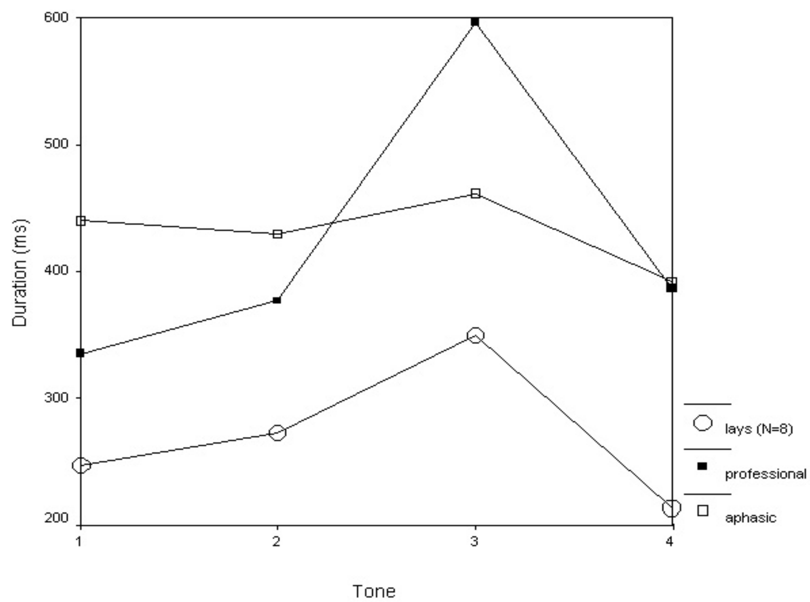


Figure 1: Duration (ms) of four lexical tones plotted separately for one aphasic speaker, one professional control speaker, and a group of eight non-professional normal speakers as reference.

Clearly, then, these results show that the tonal contrast in the speech of our aphasic speaker is severely compromised in the temporal domain. The durational pattern deviates substantially from that of the normal speakers (both our professional speaker and the reference sample). These results are compatible with the finding that the lexical tones in the aphasic's speech were very poorly identified by our listeners.

Figure 2 plots the **fundamental frequency** (F0) traces of the 217 tokens of lexical tones produced by the aphasic speaker (left-hand panels) and for the control speaker (right-hand panels) separately for each of the four lexical tones (tones 1, 2, 3 and 4 from top to bottom). Each of the eight panels contains between 40 and 61 time-normalized F0 traces. Time normalization was achieved by reducing the raw F0 curve for each token to eleven equidistant points along the time-axis, which affords adequate time resolution for the F0 contours.

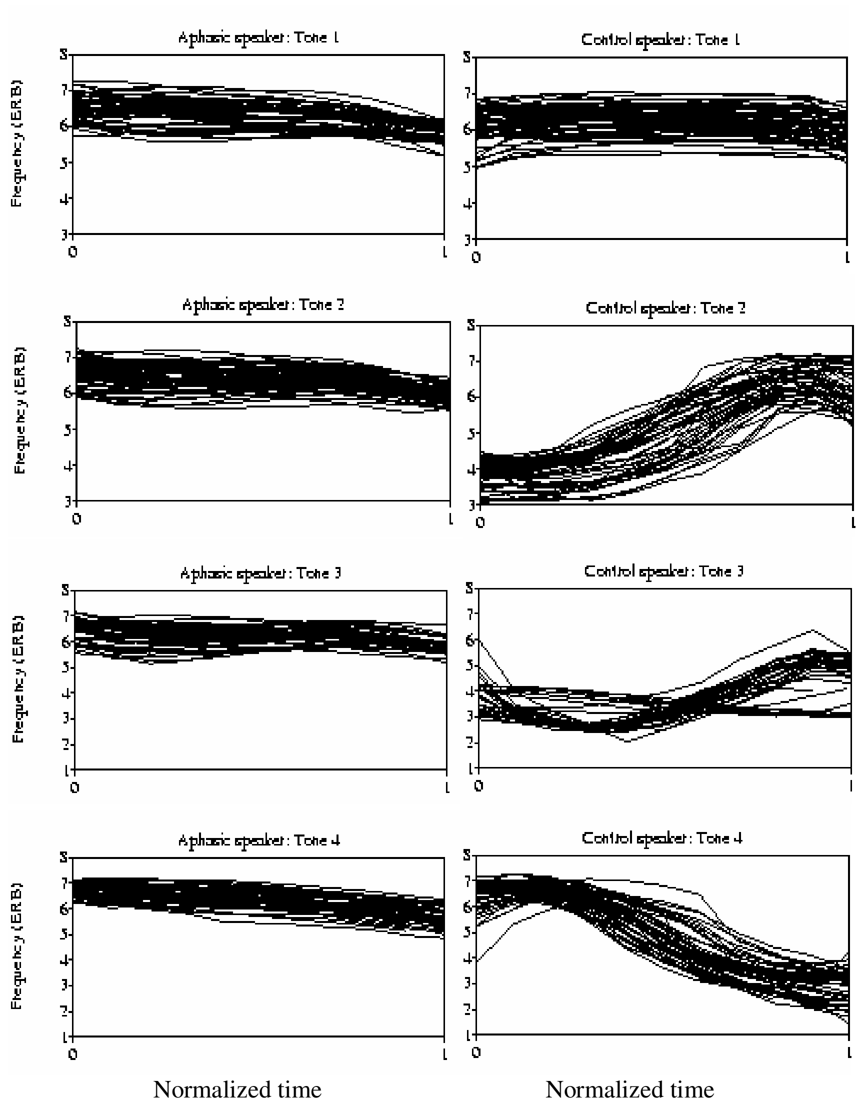


Figure 2: F0 traces plotted as a function of normalized time of the 217 tokens of lexical tones produced by the aphasic speaker (left-hand panels) and for the control speaker (right-hand panels) broken down by lexical tone (tones 1, 2, 3 and 4 from top to bottom).

Visual inspection of the data reveals that the four lexical tones are clearly grouped into four tightly clustered, distinct patterns for the normal speaker. The average patterns correspond closely to what has been found in the literature. The high level Tone 1 starts in the speaker's high pitch range and remains high without drastic pitch movement except for a slight dip in the middle of the vowel and a slight rise toward the end of the syllable. The rising Tone 2 begins in the speaker's mid pitch range, remains level or drops slightly during the first half of the vowel, and rises up to high at the end. The low Tone 3 is phonetically a low falling tone. It starts in the speaker's mid range and falls to the low range, usually accompanied by laryngealization (creaky voice) during the second half of the syllable. The falling Tone 4 usually peaks around the vowel onset, and then falls to the low pitch range at the end.

It has been found that the beginning and end points of all tones cluster at three distinct levels rather than spread evenly across a continuum, i.e., Tone 1 and Tone 4 both start high, very close to where Tone 1 and Tone 2 end, Tone 2 and Tone3 both begin in the mid range, while Tone 3 and Tone 4 both fall to the low pitch range (Shih, 1988). This is exactly the pattern that we observe in our control speaker. Again, we conclude that our professional speaker is representative of the population of Standard Chinese speakers at large. The qualitative description offered here is brought out even more clearly if we abstract from token-individual variation, and just plot the mean F0 traces for each of the four lexical tones as a function of (non-normalized) time, as is done in Figure 3 (right-hand panel for the control speaker).

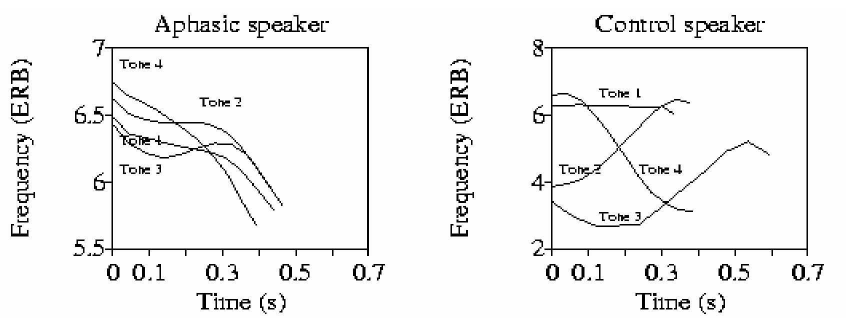


Figure 3: Fundamental frequency (ERB) as a function of time (s) for each of four lexical tones, for one aphasic speaker (left) and one control speaker (right). The traces are averaged across 40 to 61 tokens per tone type per speaker.

The corresponding token-individual traces (Figure 2 left-hand panels) for the aphasic speaker are hard to distinguish from one another. Nevertheless, a MANOVA procedure (using Wilk's Lambda statistic) on the eleven F0 points and the duration of the tone run for the aphasic speaker reveals a small but significant effect of tone, $F(36, 597.6) = 4.7$ ($p < 0.001$). The same procedure reveals a much larger effect for the healthy control speaker, $F(36, 597.6) = 85.3$ ($p < 0.001$). The small overall differences among the four tones produced by the aphasic speaker are displayed in Figure 3 (left-hand panel), where it was necessary to magnify the F0 scale; note that

PYF's tone range is compressed between 5.5 and 7 ERB, while for the control speaker the range runs between 2 and 8 ERB.

In PYF's production, all four tones start in the high pitch range, and end in a comparatively low pitch range, which, however, is not low enough for Tone 4, let alone for Tone 3. Yet, during the time interval between 0.1 and 0.3 seconds after the tone onset, we can see some tonal differentiation:

- Tone 4 stands out clearly: it begins higher, has shorter duration and drops more steeply and more steadily than any other tone.
- Tone 2 remains high for the greater proportion of its duration; this seems to reflect an attempt on the part of our speaker to maintain a high target towards the end of the configuration.
- Tone 3 goes down to an F0 minimum, and then rises. This modulation is slight but unmistakable. The pattern then corresponds to the high-low-high tone sequence that is characteristic of the low dipping tone produced in isolation.
- Tone 1, finally, runs in the middle of the parameter space defined by the other three tones. It seems to behave as the unmarked – or default – pattern.

After the initial 0.3 s of the tone, the patient does not differentiate the tones any longer. She seems unable to reach or maintain any high F0 target when more than 0.3 s has elapsed. After this point in time, F0 drops down for each of the tones by the same interval with similar steepness.[7] Shen and Lin (1991) describe the contrast between Tone 2 and Tone 3 in terms of the relative timing of the low turning point in the configuration. Tone 2 reaches its F0 minimum quite soon after the onset, whilst the minimum is not reached within the first 100 ms for Tone 3.

If our qualitative description of PYF's tone patterns is accepted, our data suggest that the onset and offset parts of our patient's tones are severely damaged but that some measure of differentiation between the four tones remains in the middle parts. The information seems to be present in the acoustic signal, but it is not picked up by the human listeners, simply because the contrast is too subtle – probably below threshold – and is drowned by inter-token variability.

Let us now turn to the acoustic analysis of the vowel data. Figure 4 presents the **vowel tokens** of the two speakers as points in the F1-by-F2 plane. Since the data are plotted along Bark scales, the distances in the visual plane should correspond well with the perceived magnitude of differences in vowel quality, both for the female (aphasic) and for the male (control) speakers.

---

[7] Interestingly, Gårding, Kratochvil, Svantesson and Zhang (1986) describe the difference between Tone 4 and Tone 3 in terms of an early downshift of F0 in Tone 4 relative to Tone 3. The F0 in Tone 3 remains high until the middle of the syllable, and then drops, whilst the F0 in Tone 4 begins its fall at roughly 25% percent of the syllable's duration and reaches its bottom value at roughly 75%. In our aphasic speaker Tone 4 falls right from the beginning and at no point during its time course does its slope level off.

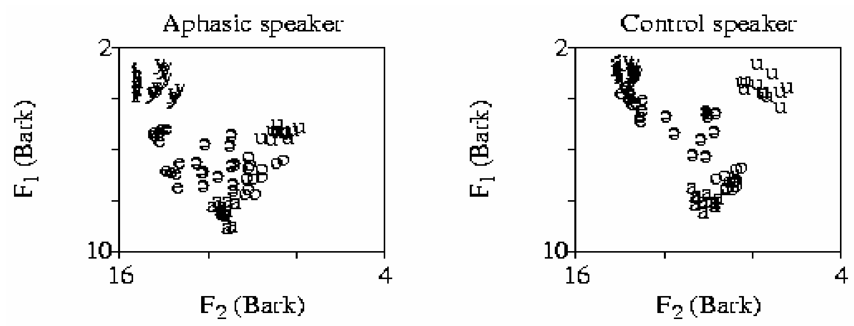


Figure 4: F1 (as a correlate of vowel height) plotted against F2 (as a correlate of vowel backness), both in Barks, for tokens of 7 vowel types for an aphasic speaker (left panel) and a normal control speaker (right panel).

Visual inspection of the results indicates that the vowels produced by the control speaker can easily be sorted into seven discrete types. It is quite feasible to draw a polygon around each group of vowel tokens such that the polygons do not overlap, i.e. there is perfect separation between the seven categories. The vowel space of our aphasic speaker is somewhat distorted relative to that of the control speaker. Specifically, PYF's high back vowel /u/ is substantially lowered, and comes rather close to the /o/-tokens. However, for PYF's vowels, too, perfect separation of the tokens into seven types is afforded by drawing non-overlapping polygons. Deferring statistics until the next section, the data in Figure 4 allow us to understand why our listeners were quite able to identify the vowels in the speech of the aphasic speaker.

3.3. Automatic acoustic classification

Although it is relatively simple to devise an algorithm that identifies the vowels of either the aphasic or the control speaker, for instance based on the polygons in the F1-by-F2 planes, such a classification algorithm would be much more problematic in the case of the lexical tones. In order to quantify the relative degree of success with which tones and vowels are differentiated in the speech of our aphasic speaker as opposed to the control speaker, we need one single classification procedure that can be applied to both sets of phenomena. Linear Discriminant Analysis (LDA) is such a procedure. The LDA was performed stepwise such that all acoustic parameters were entered initially, and eliminated in backward fashion if the algorithm's performance did not deteriorate by more than a criterial amount. Thus, the final solutions presented here, are maximally economical: they afford the highest classification performance in spite of severe data reduction (Klecka, 1980).

The full set of lexical tone tokens were submitted to LDA with the eleven equidistant F0 measurements (in ERB) as predictors for the aphasic and the control tokens separately. Similarly, LDAs were run on the vowel data (with F1, F2, F3, and F4 at the intensity maximum as predictors).

The results of the tone classification are shown in Table 3 as a confusion matrix. The tone type predicted (identified) by the LDA is crosstabulated against the actual tone type – i.e., as intended by the speaker – in the same way as was done in Table 1 for the human perception data.

Table 3: Tone Classification by Linear Discriminant Analysis. Identification scores are in percentages. Correct decisions are indicated along the main diagonal, in bold face against a shaded background.

Speaker	Correctly classified	Actual Tone	Identified by LDA (%)				Total
			Tone 1	Tone 2	Tone 3	Tone 4	
Aphasic	53.5	Tone 1	**50**	25	5	20	100
		Tone 2	16	**59**	12	13	100
		Tone 3	20	40	**30**	10	100
		Tone 4	5	27	0	**68**	100
Control	96.8	Tone 1	**98**	2	0	0	100
		Tone 2	0	**92**	8	0	100
		Tone 3	0	2	**98**	0	100
		Tone 4	0	0	0	**100**	100

The control speaker's tones are classified between 92 and 100 percent correct, with an average of 97%. This classification performance was afforded on the basis of just two out of the total of twelve predictors (eleven F0 points and the duration): one point rather early in the time course of the F0 trace (point 2) and one rather towards the end of the contour (point 8).

Clearly, the identification by LDA is substantially poorer in the case of the aphasic speaker's tones. These are classified at a modest 55 percent correct, with a lot of confusion among them. Tone 4 is correctly identified most often (68%) while Tone 3 is poorest (30%). Still, all the tones are identified at better-than-chance level (= 25%), and significantly so for Tones 1, 2, and 4. This indicates that useful acoustic information remains in the aphasic speaker's realizations which is not picked up by human listeners, who have not learnt to use these cues.

Table 4 presents the results of LDAs on the vowel data of the aphasic speaker and the control. Not surprisingly, given the earlier findings in terms of dispersion of the vowel tokens in the formant space, percent correct classification is close to 100 for both speakers: 96 % correct for the aphasic speaker against 98% for the control speaker. These results are fully compatible with the earlier finding that 90% of the aphasic speaker's vowel tokens were identified correctly.

Table 4: Vowel Classification by Linear Discriminant Analysis. Identification scores are in percentages. Correct decisions are indicated along the main diagonal, in bold face against a shaded background.

Speaker	Correctly classified	Vowel	Identified by LDA (%)							Total
			a	e	ə	i	y	o	u	
Aphasic	96.4	a	**100**	0	0	0	0	0	0	100
		e	0	**100**	0	0	0	0	0	100
		ə	8.3	0	**91.7**	0	0	0	0	100
		i	0	0	0	**100**	0	0	0	100
		y	0	0	0	8.3	**91.7**	0	0	100
		o	0	0	0	0	0	**91.7**	8.3	100
		u	0	0	0	0	0	0	**100**	100
Control	97.6	a	**100**	0	0	0	0	0	0	100
		e	0	**91.7**	0	8.3	0	0	0	100
		ə	0	8.3	**91.7**	0	0	0	0	100
		i	0	0	0	**100**	0	0	0	100
		y	0	0	0	0	**100**	0	0	100
		o	0	0	0	0	0	**100**	0	100
		u	0	0	0	0	0	0	**100**	100

On aggregate, the results in this section have shown that the aphasic speaker suffers severe loss of tones, but retains all of the vowel phoneme contrasts (about as well as the control speaker).

4. Discussion

4.1. Task-dependent laterality during language processing

It is generally acknowledged that language is subserved by left-hemisphere (LH) mechanisms. It is not clear at this stage whether the LH is specialized for speech, for linguistic processing in all modalities (including, for instance, writing and sign language), or for even more general cognitive processing. The cue-dependent hypothesis advocates that the same basic neural mechanisms underlie the processing of complex auditory stimuli regardless of their linguistic relevance (Ivry and Lebby, 1993, Pöppel, 1997; Warren, 1999). The task-dependent hypothesis states that specialized neural mechanisms exist when activated by speech only (Imaizumi, Mori, Kiritani, Hosoi and Tonoike, 1998).

Our findings do not support the cue-dependent hypothesis that left inferior frontal regions (in our patient: PRG, F3t) are specialized for rapid temporal processing of complex acoustic stimuli (cf. Schwarts and Tallal, 1980; Tallal, Miller and Fitch, 1993). Chinese tones are of relatively long durations (250–350 ms) and yet they are lateralized to the LH instead of RH, as is evidenced by the severe loss of tone contrast in our LH-damaged patient PYF.

On the other hand, our findings are in accord with known effects of LH brain lesions, as found in loss of contrast in tone production and/or perception (Gandour and Dardarananda, 1983; Eng, Obler, Harris and Abramson, 1996; Moen and Sundet, 1996), in dichotic listening (Van Lancker and Fromkin, 1973; Ip and Hoosain, 1993; Moen, 1993), and in functional neuro-imaging (Gandour, 1998; Gandour, Wong, Hsieh, Weinzapfel, Van Lancker and Hutchins, 2000; Hsieh et al., 2001). Both these and our own findings support the task-dependent hypothesis of cerebral lateralization whereby LH mechanisms selectively mediate the processing of linguistic information irrespective of acoustic cues or type of phonological unit (i.e., segmental or suprasegmental). However, our findings are not compatible with earlier studies reporting that, in isolated monosyllables, even non-fluent LH aphasic speakers are able to control relative differences in duration associated with the five lexical tones in Thai (Gandour et al., 1992).

Under the task-dependent hypothesis of cerebral lateralization, viz., that pitch discrimination is lateralized to the LH when the pitch differences are linguistically processed (Van Lancker and Fromkin, 1973), our findings further suggest that the left PRG and F3t in the IFG regions are specialized for lexical tone rather than for vowel production in Chinese.

4.2. System in the loss of tones?

We have seen in the results obtained for our patient PYF that the contrasts among the four lexical tones of Chinese were severely reduced. Moreover, the four tones did not suffer equally; some tones were more affected by PYF's aphasia than certain others. The least compromised single tone was Tone 1, with 88 percent of the tokens being correctly identified. Human identification was second best for Tone 3 (the low dipping tone, 20% correct). This tone type, then, is much more difficult for our patient to produce correctly than Tone 1. Tone 1 and Tone 3 share one important characteristic: they can be analyzed as static tones: the high level tone, and the low level tone, respectively. Dynamic tones, involving a movement of pitch from a low to a high target (a rise) or vice versa (a fall) are more difficult than any static (level) tone. Tone 4, the dynamic falling tone, is the second most difficult tone for PYF to produce (16% correct), to be followed only by the dynamic rising Tone 2 with a mere 9% correct.

In human vocal pitch production, the pitch has a tendency to drop over the course of an utterance, among other things as a consequence of the reduction of transglottal air pressure as lung air is used up during phonation. This tendency has been called downdrift (Ladd, 1996) or declination ('t Hart, Cohen and Collier, 1990). It would make sense to find, therefore, that implementing a fall is less problematic for PYF than producing a rising pitch, which would be a dynamic tone change conflicting with the natural tendency of vocal pitch to fall.

In light of the above considerations, we suggest the following account of PYF's compromised tone production in isolated monosyllabic words:

(i) maintaining a level pitch is easier than executing a pitch change,
(ii) producing high pitch is easier than producing a low pitch, and
(iii) producing a change from high to low is easier than a change from low to high

There are several questions, however, that should be answered before we can reach any satisfactory conclusions with respect to the loss of tone contrasts in our patient's aphasia.

First of all, we need to know to what extent the tonal impairment is only an aspect of PYF's speech production or a more pervasive feature of her linguistic competence. In the latter case she should be unable to identify tones in the speech of fellow Chinese speakers. Given PYF's satisfactory performance on the diagnostic test battery it would seem reasonable to assume that only speech production is affected by her aphasia, but a more definitive test procedure would be called for to corroborate this preliminary conclusion.

Second, we must establish the extent to which PYF is able to control her sentence melody in abstraction of the lexical tones. If, for example, she were still able to make her pitch go up or down at the end of the utterance so as to mark the sentence as a statement or a question (for example, by forcing her to refrain from using the dedicated Chinese question particle /ma/ to signal the difference in sentence type), we would infer that the tonal deficit is not a matter of vocal pitch control but rather resides in the (more abstract) specification of the lexical tones.

Importantly, however, we have shown that, whatever the source of the problem with PYF's tone production, it is not the case that the specification or implementation of the lexical tones interacts with the specification or production of the vowel types. From a physiological point of view such an interaction would be unlikely, since different and largely independently operated sets of muscles are involved in the laryngeal structures responsible for generating vocal pitch and pitch change, and the supralaryngeal structures that are engaged when different vowels are produced. Assuming for the moment that the problems in PYF's tone production are not a matter of muscle control (as her sentence melody sounds relatively unaffected), we must assume that the problem exists in a more central module, probably at the level of the lexicon. Given, then, that PYF's vowels are still intact but her tones are compromised, we may provisionally conclude that, at least for this single individual, the specification of lexical tone is not located in the same place in the structure of the brain as the lexical specification of vowel type.

5. Conclusion

Our data suggest that segmental structure and lexical tone are separate functions with separate locations in the brain. If in the future these findings are replicated with larger groups of patients, we would be forced to conclude that the phonological or phonetic components of the grammar are more diffusely represented within the language-dominant left hemisphere than heretofore believed. Such a structure would be consistent with the non-linear representation in autosegmental phonology, in which tone and segments are represented on separate tiers (Goldsmith, 1990).

Chapter Three

Separate representations for lexical tone and sentence intonation? A perception study of Chinese aphasic patients[8]

Abstract

Background: Previous research reports lexical tone deficit with left-hemisphere (LH) damaged patients speaking a tone language. Imaging studies with healthy tone-language listeners show that prosodic functions predominantly activate the left hemisphere for lexical tone and the right hemisphere for intonation. There is no report available yet on intonation perception by tone language listeners after unilateral damage in the left hemisphere.

Aims: We addressed two questions: (1) How well do Chinese aphasic patients with LH damage perceive pitch variations at word-level and sentence-level? (2) Are Chinese aphasic patients with LH damage differentially sensitive to linguistic levels (lexical tone versus intonation) as predicted by task-dependent hypothesis or to some particular physical property, i.e. temporal factor, as predicted by cue-dependent hypothesis?

Methods and Procedures: Fourteen chronic Chinese-speaking (Beijing dialect) aphasic subjects with unilateral LH damage diagnosed as Broca's aphasia participated in two laboratory- controlled perception experiments, on lexical tone and sentence intonation (statement vs. question), in which we manipulated the intonation contour of the whole utterance as well as the final syllable. As controls, there were thirty healthy native speakers of Beijing dialect.

Outcome and results: Results revealed that relative to their healthy control listeners the aphasic patients were significantly poorer and slower in lexical tone identi-

---

[8] This chapter will appear as J. Liang and V.J. van Heuven (under revision) Separate representations for lexical tone and sentence intonation? A perception study of Chinese aphasic patients. *Aphasiology*. It is an extended version of a paper first presented at the conference 'The Science of Aphasia', in Trieste, Italy, 22-27 August 2003. We gratefully acknowledge travel grants awarded by the conference organization and by the Leiden University Fund (LUF) for subsidizing the first author's trip to the conference.

fication but were not significantly different in intonation identification (statement vs. question) and as equally fast in terms of decision time.

Conclusions: These findings show that lexical tone is perceived independently of intonation in Chinese, even though both of them primarily depend on pitch variation for contrasts, and therefore, the results suggest that lexical tone and sentential intonation in Chinese, and in tone languages in general, are located or processed separately in the brain.

1. Introduction

Chinese is a lexical tone language. Words in such languages do not only differ in the sequence of vowels and consonants ('segments') but also by word melody ('tone'). In the linguistic literature on Chinese languages tones are customarily described as relative pitches within the speaker's tonal range. The lowest relative pitch at or near the bottom of the speakers pitch range is called '1' and the highest pitch level '5' (Chao, 1930). Standard Chinese (Mandarin, Beijing dialect) has four lexical tones: Tone 1 (high level, 55), Tone 2 (rising, 35), Tone 3 (dipping, 214) and Tone 4 (falling, 51). An often-cited minimal quartet, i.e. the same segments making up a syllable with different word meanings depending on the tone, in Beijing Chinese is /ma1, ma2, ma3, ma4/ 'mother, hemp, horse, scold', respectively.

Moreover, when certain tones occur in sequence the tone on an earlier word may change its phonetic shape (coarticulation) or even its phonological type (assimilation) under the influence of an adjacent tone. We will refer to these changes in contexted tones as tonal sandhi, side-stepping the linguistic debate whether the change is phonetic or phonological. An example of tone sandhi is seen in a sequence of Tone 3 immediately followed by another Tone 3: here the first (canonically dipping) tone changes into a rising tone (which may or may not be identical to a Tone 2).

Pitch, then, is used in Chinese both at the word level to differentiate between four (or more) lexical tones and at the sentence level to signal – among other things – differences in clause type such as statement versus question. Cross-linguistically there is a clear tendency for languages to make a choice for either word tones or sentence intonation. That is to say, if a language such as Chinese, has a substantial inventory of word tones, it will exploit the pitch dimension less for the signalling of contrasts at the sentence level. As a corollary it has often been noted that Chinese does not normally use sentence melody to signal the difference between declarative and interrogative sentences but uses a dedicated question particle to express the latter function. This, then, is a functional view that claims that if some phonetic dimension is exploited in one area of the phonology, it will not be used to the same extent in another part of the phonology (see, for instance, Ross, Edmondson and Seibert, 1986; Gandour, Larsen, Dechongkit, Ponglorpisit and Khunadorn, 1995; Seddoh, 2002).

Imaging studies on pitch perception of healthy speakers show that pitch processing is influenced by language experience (Gandour, Wong, Hsieh, Weinzapfel, Van Lancker and Hutchins, 2000; Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong and Lurito, 2002; Gandour, Wong, Dzemidzic, Lowe, Tong and Li, 2003; Gandour, Tong, Wong, Talavage, Dzemidzic, Xu, Li and Lowe, 2004; Hsieh, Gandour, Wong and Hutchins, 2001). Their findings show that for Chinese listeners prosodic functions at word level predominantly engage the left hemisphere (LH); prosodic functions at the sentence level elicit bilateral activation, with a right hemisphere (RH) preference. In contrast, listening to the same Chinese stimuli, English listeners show bilateral activation for lexical tone but predominantly right-sided activation for intonation.

Several hypotheses have been forwarded concerning the neurobiological mechanism underlying speech perception. The hypotheses involve either acoustics or linguistics, or both, as the source of hemispheric lateralization. Hypotheses involving an acoustically driven mechanism are called cue-dependent. These claim that speech processing is subserved by neurobiological mechanisms specialized for particular aspects of the acoustic signal, e.g. the absolute size of temporal domain to account for the hemispheric specialization of local and global processing, irrespective of communicative or linguistic relevance as suggested by the cue-dependent hypothesis (e.g. Behrens, 1989; Poeppel, 2001). Hypotheses focusing on linguistic mechanisms are called task dependent. These claim that unique, neural mechanisms are recruited for the speech domain as suggested by the task-dependent or domain-specific effects (e.g. Van Lancker, 1980; Ross and Mesulam, 1979). Hybrid views engaging both linguistic and acoustic processing claim the existence of dynamic interactions between the two hemispheres, i.e., during spoken-language comprehension processes in the left and the right hemispheres are assumed to interact dynamically. There are two models especially worth mentioning in this context, (i) a comprehensive model of speech perception that is mediated primarily by RH regions for complex-sound analysis but which is lateralized to task-dependent regions in the LH when language processing is required (Gandour et al., 2004), and (ii) a dynamic dual-pathway model of auditory language comprehension that claims that segmental, lexical (i.e., tone) and syntactic information is processed in the LH, suprasegmental sentence-level information (i.e. intonation) in the RH; this model additionally claims that the involvement of the LH increases as the linguistic nature of either the stimulus or the task gets larger (Friederici and Alter, 2004).

Imaging studies tell us what areas participate in a particular (linguistic) function; lesion studies tell us what areas are necessary for normal functioning (Gandour, 2000: 76). Therefore, in the present study we aim to provide a lesion study on pitch perception by speakers of a tone language. So far we have only reports on the perception of lexical tones but not on intonation perception. For instance, one case study of Chinese aphasic patients with an LH intracerebral bleed reported that lexical tone perception was impaired (Naeser and Chan, 1980; Eng, Obler, Harris and Abramson, 1996), and another report showed that lexical tone identification of four monolingual Thai aphasic patients differed from that of normal controls (Gandour and Dardarananda, 1983).

Our questions are as follows:

(1) How well do Chinese aphasic patients with LH damage perceive pitch variations at word-level and at the sentence-level?

(2) Are Chinese aphasic patients differentially sensitive to linguistic levels (lexical tone versus intonation) as predicted by a task-dependent hypothesis or to some particular physical property, i.e. temporal factor, as predicted by a cue-dependent hypothesis?

For each of the three types of hypotheses formulated above we expect the following results.

- If speech processing is subserved by **cue-dependent** neurobiological mechanisms specialized for the time course of speech prosody, irrespective of communicative or linguistic relevance, we should find equal impairments for both lexical tones and sentence intonation, at least when they are of the same duration.
- If, however, unique, **task-dependent** neural mechanisms are recruited for the speech domain, we may have two possibilities: (i) equal impairment for lexical tones and sentence intonation since both belong to the speech domain, or (ii) differential impairment since word tones and sentence intonation operate at different levels in the linguistic hierarchy. In both possibilities there will be competition for resources between lexical tone perception and sentence intonation since these processes are controlled by speech-specific neural mechanisms. With equal impairment of tone and intonation perception competition for resources will be the same as in a healthy listener, giving equal weight to both tasks; with differential impairment one task will be executed at the expense of the other. In the latter case, however, the unimpaired task may actually be performed better by a patient than by a healthy listener, since competition is reduced or even absent.
- In the case of **hybrid** neurobiological mechanisms specialized for both linguistics and acoustics, for instance the left hemisphere for linguistics and the right hemisphere for acoustics, we would expect patients with LH damage to display a deficit in language processing but to retain sensitivity to non-linguistic sounds.

In our study, we set up two laboratory-controlled perception experiments on lexical tone and sentence intonation (statement vs. question) identification, respectively, in which we manipulated the intonation contour of the whole utterance as well as the boundary tone in the sentence-final syllable. The duration of the latter is comparable to that of lexical tones, so that the cue-dependent effect can be tested.

Our prediction is that Chinese aphasic patients with LH damage first of all, have a lexical tone deficit. The tone deficit may be accompanied by an intonation deficit (equal impairment), showing that pitch perception at both levels involves a single shared module in the LH). Alternatively, intonation perception may be less impaired (relative to the healthy controls), showing one of two things: (i) intonation may be processed by a separate module in the brain, possibly located in the RH or (ii) intonation is processed in the same module as lexical tones but wins out in the competition (greater impairment of tones would then be beneficial to intonation perception).

2. Methods

We set up a three-factor experiment with listener type, stimulus type and task condition as the principal independent variables in the construction of the stimulus materials.

2.1. Listener type

Fourteen right-handed Beijing aphasic listeners, native Beijing speakers from Tianjin, P. R. China with twelve years of formal education at least, aged 31–80, average 55, were diagnosed by Professor Zhang Banshu from Tianjin General Hospital as non-fluent patients with Broca's aphasia characterized by word-finding difficulties, incomplete syntactic constructions, and errors in sound production.

Table 1: MRI or CT scan findings of the lesion site in the left hemisphere for individual patients. Judgments made by Prof. Zhang Banshu and radiologists at Tianjin General Hospital)

	Name	PYF	ZZL	XWC	CRY	HSQ	QCF	YSY	CH	ZHR	YJJ	FZJ	ZLJ	TDH	LKM
Patients	Age	39	50	68	43	47	63	69	80	31	54	50	69	56	52
	Sex	f	m	f	M	m	f	f	f	m	m	m	m	m	m
	Time after onset (months)	23	25	5	6	6	4	2	12	2	6	12	4	35	39
	Years of formal education	12	14	17	16	16	17	17	13	16	15	15	17	13	13
Frontal lobe	Precentral gyrus lower part	●			●	●				●				●	●
	Superior f. gyrus posterior								●						
	Inferior gyrus pars triangularis	●	●			●					●		●	●	
	Broca's area pars operculalis						●	●					●		●
Parietal lobe	Postcentral gyrus lower part	●	●	●	●							●	●		
	Supramarginal anterior							●	●						
	Gyrus posterior						●								
	Posterior						●								
	Superior lobule										●				
Cingulate	Gyrus posterior			●					●						

Production studies on the tones and vowels of one of the patients, the severest case, showed that lexical tones were seriously damaged while the vowels were preserved (Liang and van Heuven, 2004). The patients' non-verbal communication was still effective, and apart from their aphasia, they were able to carry out the activities of everyday life without difficulty. All of them suffered from unilateral damage in the left frontal and/or parietal lobe (detailed information presented in Table 1) and showed normal hearing sensitivities at 0.5, 1, and 2 kHz following a pure-tone air-conduction screening.

None of the participants had been diagnosed with neurological or psychiatric illness prior to experiments, apart from a single-event cerebro-vascular accident (CVA) with damage in the LH of the brain. Patients participated in this experiment during January, 2002 to January, 2003.

Thirty healthy Beijing controls were native speakers of Beijing dialect, aged between 21 and 70, average 40, 17 male and 13 female. All of them were right-handed with normal hearing and at least twelve years of formal education. They took part in the experiments in September 2002.

2.2. Stimuli

Recordings used in the experiments were from a male native speaker of Beijing dialect (Standard Chinese), experienced in sound recording. All the recordings were made in a sound-insulated booth at the Phonetics Laboratory, Leiden University, on digital audio tape (DAT) using a Sennheiser MKH-416 unidirectional microphone, then transferred to computer memory and downsampled to 16 kHz (16 bit amplitude resolution) using the Praat speech processing package (Boersma and Weenink, 1996).

For the lexical part, four frequent lexical words were used, covering the Beijing four-tone system, /ma1 ma2 ma3 ma4/ (see introduction). In order to study the effect of context, each of the four lexical-tone stimuli was read aloud five times in isolation and five times in medial position in a fixed carrier sentence: [wo3 ʂuo1 … dʐə4 kə0 dzi4] 'I say … this word', which took 1.5 seconds. In all there were 40 stimuli for lexical tone identification, i.e. 4 (tone) × 5 (repetition) × 2 (in a carrier sentence and isolation).

For the post-lexical (intonation) part, a single question utterance was used: [kai1 sun1 jiŋ1 kʰai1 fei1 tɕi1] 'must Sun Ying fly flying machine?' meaning 'Is it Sun Ying's turn to fly the plane?' The utterance contains high level tones (Tone 1) only, in order to reduce the lexical tonal influence. The sentence can be interpreted as either a statement or a question depending on the intonation. The speaker was instructed to say the sentence as a statement and a question as naturally as possible; again, five tokens of each were recorded. Before recording the utterances the speaker was given the context within which the target utterance was to be interpreted, such that there was narrow focus on the subject 'Sun Ying'.

Studies have shown that intonation is perceived in terms of specific fundamental frequency (F0) patterns (Gårding and Abramson, 1965; 't Hart, Collier and Cohen, 1990), which involve the same acoustic parameter that is employed in the signalling of lexical tone (Pike, 1948; Wang, 1972; Howie, 1976). Previous acoustic and perception studies on question intonation of Standard Chinese have shown that besides the use of question particles, the main acoustic correlate of the statement ~question contrast is the F0 (Howie, 1976; Tseng, 1981; Coster and Kratochvil, 1984). Since our sentences did not contain any lexical (including particles) and/or syntactic question markers, they are optimally suited for studying the role of intonation in signalling the difference in sentence type between statement and question.

Using the Praat software the pitch contour of one question utterance was stylized with seven pivot points connected by straight lines in a log-frequency (semitone, ST) by linear time representation. It is generally accepted that the pitch variation in the final syllable signals the statement ~ question contrast in Standard Chinese (Tsao, 1967; Gårding, 1985, 1987; Kratochvil, 1998; Yuan, Shih and Kochanski, 2002) but there is disagreement on the issue whether the overall pitch level of the utterance provides a subsidiary cue to the contrast (De Francis, 1963; Kratochvil, 1998) or not (e.g. Tsao, 1967). Kratochvil (1998), in fact, described the question intonation of Beijing Dialect as accompanied by a raising of the overall pitch level by about 50 Hz. We will perceptually test the relative contributions of the overall pitch and of the final fall/rise to the statement ~question contrast.

Therefore, a two-dimensional stimulus continuum was generated (through PSOLA resynthesis, cf. e.g. Moulines and Verhelst, 1995) from the question utterance, by systematically combining seven overall pitch levels and seven different boundary tones.[9] The overall pitch levels were decremented by 1-ST steps, which adequately covered the range we established in the production of question and statement tokens by our speaker.

The seven boundary tones were generated by decrementing the terminal pitch of the question token in six steps of 0.5 ST. Again, the 3.5-ST difference between the highest and lowest terminal frequencies in the boundary tone continuum corresponds closely to what we found in our speaker's production. By manipulating only the frequency values while leaving the time coordinates unaltered, 49 different pitch patterns were generated according to the following schema (Figure 1).

[9] In order to keep the experiment within manageable bounds pitch manipulations were introduced to the question base only. We informally observed that the quality of the stimuli was better overall when the question base was used. The statement base was never used; however, it did provide an estimate of what the F0 fall should be like to generate convincing statement melodies on the question base.

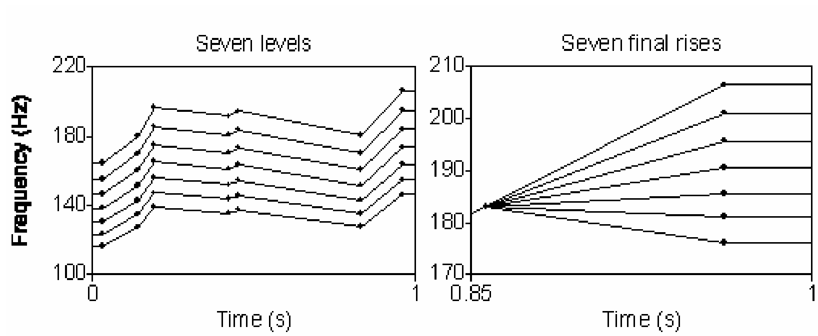


Figure 1: Overview of the manipulations of pitch level (7 steps) and boundary tone (7 steps). Orthogonal variation of level and boundary tone yields 49 different stimulus types

2.3. Task conditions and procedure

Considering that lexical tone stimuli contained naturally produced tokens only, we ran the two experiments in a fixed order, such that the lexical-tone experiment always preceded the intonation test (with potentially ambiguous, hybrid exemplars). The stimuli were randomized and presented by computer binaurally over headphones (Sony MDR-V3) at a comfortable listening level. Listeners were tested one at a time in a quiet room, in the case of patients often in the subject's own home. A specially designed keyboard was used with either four buttons marked with the corresponding Chinese characters (in font 72) for the tone identification, or two buttons marked with '。' and '?', i.e., punctuation marks that are used in Chinese for statement and question.[10] Before each part of the experiment, specific instructions were given both orally and in writing. Each experimental task was preceded by a short practice session, with four practice trials for the tone identification task and seven for the intonation task. Each experiment lasted about 20 minutes. Decisions made and reaction times (measured from the offset of the stimulus, precision 1 ms) were stored in the computer. Stimulus presentation and response collection were controlled by E-prime software.[11]

In the **lexical tone experiment**, the subjects decided which of the four words they had heard by pressing one of the four buttons on the response keyboard with their right hand each time they heard a stimulus. They were asked to avoid errors in the first stage of the experiment (no time pressure). In the second stage the subjects were instructed not only to avoid errors but also to perform the task as quickly as they could manage (time pressure). The four buttons were evenly spaced across the

[10] The keyboard was designed and built by J.J.A. Pacilly at the Universiteit Leiden Phonetics Laboratory.
[11] The E-prime script for stimulus presentation and response collection was written by J.J.A. Pacilly.

top row on the response box; a black key for continue/start was located in the center of the bottom row. The 40 lexical-tone stimuli, 20 in a carrier sentence and 20 in isolation were presented to the listeners twice in two blocks (once without time pressure and a second time with time pressure).

When there was no time pressure, there was a fixed 3000-ms (isolated targets) or 5000-ms (targets in carrier) inter-stimulus interval (ISI) after the offset of the stimulus, irrespective of the reaction time. When no response was given within the ISI, the subject timed out, and the next stimulus was presented. In the sections with time pressure imposed on the listener, the next stimulus started 1000 ms (in isolation) or 2000 ms (in a carrier sentence) after the response; the same time-out constraint applied as above. This partly self-paced stimulus presentation in the time-pressure condition prompted the subjects to speed up their reaction time. In the debriefing after the experiment, subjects confirmed that they had felt pressured; the effects of pressure are also apparent from the results: not only were reaction times much faster, but also the number of timed-out responses was higher.

In the intonation experiment, the listeners were required to choose between statement and question by pressing with their right hand one of the two buttons marked with '。' or '?' each time they heard a stimulus. As before, the subjects performed their task without and with time pressure. The 49 stimuli were presented twice in two blocks in random order, i.e. once without and once with time pressure. Every stimulus was preceded by a fixed reference sound, which was a single token of the neutral vowel (schwa) spoken by the same speaker who recorded the stimuli, and PSOLA resynthesized at 119 Hz, which is the lowest pitch used in the experiment, i.e. the onset pitch of the lowest resynthesized version.[12] The reference precursor preceded the test sentence by 500 ms (offset to onset). In the absence of time pressure, there were fixed 6000-ms ISIs between stimuli (offset to onset), irrespective of the reaction time. When time pressure was imposed, the next stimulus was presented 3000 ms after the previous response.

In the lexical-tone identification task subjects were exposed to prototypical exemplars of the categories only. In the sentence-intonation identification task the extremes of the continuum may be taken as the category prototypes (steps 1 and 7) but the intermediate steps are non-prototypical hybrids. As a result of this difference between the two sets of stimuli, a direct comparison of the results from these two different tasks should proceed with some caution.

In the next section, we report the results obtained with the types of listener for lexical tone identification and post-lexical identification of sentence type. We will first analyze the data collected for the perception of the lexical tone contrast. In the second part we will complement this with the corresponding analysis of the sentence

[12] The reference signal was a sequence of fourteen identical periods sampled from the middle of a natural schwa token. The artificial intensity contour comprised a 40-ms rise time, 50-ms steady state, and a 40-ms decay time.

intonation contrast. It should be noted beforehand that these two parts are rather different conceptually. In the matter of the lexical tones, we can determine whether the listener correctly identified the tone pattern as it was intended by the speaker of the utterance. Also, we may determine how much time it took the listener, whether patient or healthy, to correctly decide on the tonal category. Since the stimuli were presented for identification once with and once without emphasis on speed of response (time pressure), we predict that listeners traded accuracy for speed when time pressure was on, i.e., were prepared to gamble in the case of ambiguous stimuli in order to gain speed. In the identification of sentence melody (question ~ statement), all the stimuli were derived through parametric manipulation from a single question utterance. There is no pre-given norm here of what would be a statement and what counts as a question. So in this section of the data we will just analyze the percentage of 'question' responses and the decision latencies; we expect longer latencies as the choice between the response alternatives is more evenly balanced (which would be a sign of ambiguity in the stimulus). It is unclear, in the identification of sentence melody, how time pressure should affect the quality of the response.

3. Results

3.1. Lexical tone identification

Percent correct lexical tone identification scores were computed and submitted to a repeated-measures style analysis of variance (RM-ANOVA) with listener type (controls, patients) as a between-subject factor, and with stimulus tone, context (carrier sentence, citation form) and time pressure (on, off) as within-subject factors. The ANOVA shows that, overall, percent correct was significantly different for type of listener [$F (1, 42) = 14.6$ ($p < 0.001$)], and for time pressure [$F (1, 42) = 41.8$ ($p < 0.001$)]. Significance was also found for the interaction between listener type and time pressure [$F (1, 42) = 29.9$ ($p < 0.001$)]. There were several other significant (higher-order) interactions, which we do not report here as their significance disappears when the raw data are corrected for response bias (see d-prime analysis below).

A similar RM-ANOVA was carried out on reaction time for the lexical-tone responses. It shows that, overall, reaction time was significantly different for listener type [$F (1, 42) = 47.2$ ($p < 0.001$)], context [$F (1, 42) = 9$ ($p = 0.005$)], and time pressure [$F (1, 42) = 154.9$ ($p < 0.001$)]. Significant two-way interactions were found between tone type and context [$F (2.4, 99.1) = 17.9$ ($p < 0.001$)], and between context and time pressure [$F (1, 42) = 35$ ($p < 0.001$)].[13] Also, there were two more marginally significant two-way interactions for listener type × time pressure [$F (1, 42) = 4.5$ ($p = 0.041$)] and for lexical tone × time pressure [$F (2.7, 113.2) = 3.6$ ($p =$

[13] Degrees of freedom were Huyhn-Feldt corrected for all F-ratios reported in this paper. For main effects the correction has no consequences but for many interaction effects the degrees of freedom are reduced, and are no longer integers.

0.019)]. Additionally, three three-way interactions as well as the four-way interaction proved significant; however, we will not analyze the higher-order interactions.

The correct tone identification (percent) and the associated reaction time (ms) as a function of tone type broken down by the listener type are presented in Figure 2.

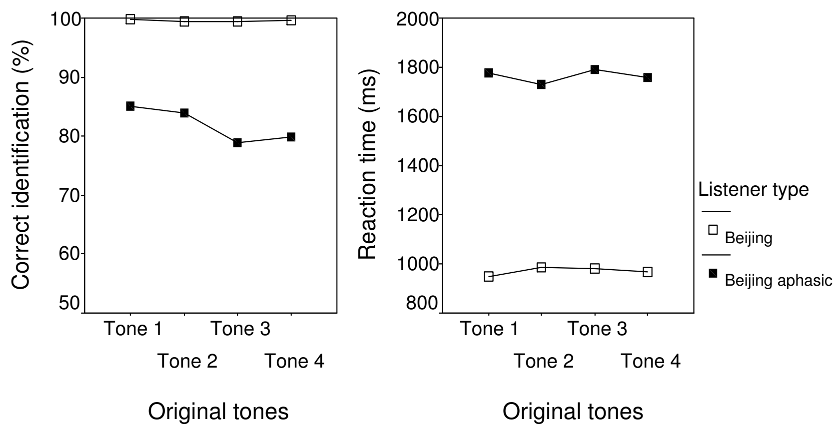


Figure 2: Correct tone identification (percent) and associated reaction time (ms) as a function of tone type broken down by listener type (Beijing vs. Beijing aphasics).

Figure 2 shows that the aphasic group was much poorer in tone identification and much slower in the associated reaction time relative to the control group.

In order to have a clear picture of the distribution of these data, we also plotted percent correct by the listener groups against the associated reaction time broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation in Figure 3.

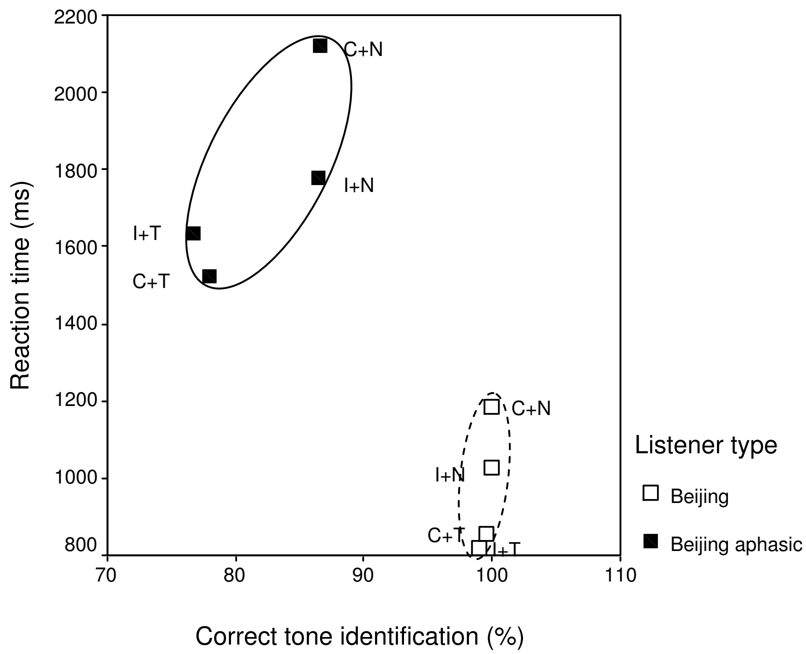


Figure 3: Percent correct identification of lexical tone plotted against reaction time (ms) for listener type (control vs. aphasic) in two time-pressure conditions (no time pressure 'N' versus time pressure 'T') and two contexts (in carrier sentence 'C' versus isolation 'I'). Spreading ellipses were drawn around the centroids by hand.

Figure 3 shows the following effects:

(i) **Listener type**. The aphasic listeners took almost twice as long as the control listeners in identifying lexical tones. The ellipse drawn around the four centroids for the aphasic group (solid line) does not overlap with the ellipse drawn around the centroids for the control group (dashed).

(ii) **Time pressure**. Time pressure leads to considerably shorter reaction time but has little effect on percent correct for the healthy listeners. In contrast, time pressure reduces percent correct as well as the reaction time for the aphasic group.

(iii) **Context**. Both types of listeners took longer time when identifying tones in the carrier sentence than in isolation. Importantly, however, embedding the target in a carrier did not affect percent correct for both groups.

From the above findings, we conclude that our aphasic listeners still possess an essential characteristic of tone-language speakers: they performed as accurately with contexted targets as with isolated targets.

As a last stage in the analysis of lexical tone identification let us break down the results by the four separate tones. Figure 4 plots percent correct identification for each of the four lexical tones broken down by listener type. The presentation is broken down further into four panels, one for each combination of context (carrier sentence, isolation) and time pressure (with, without).

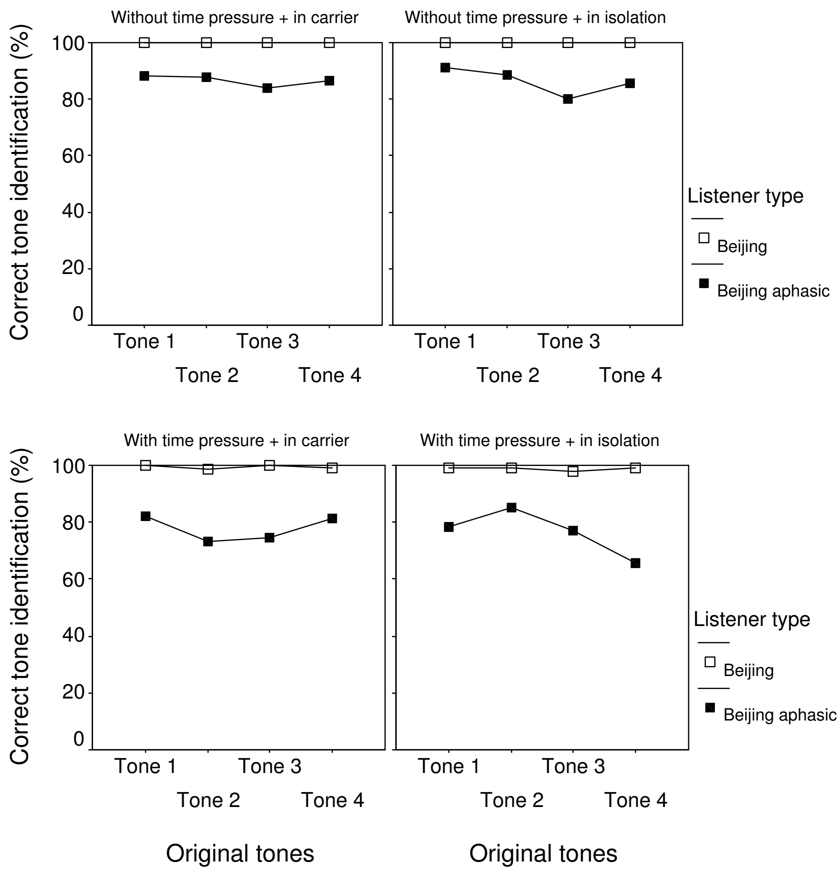


Figure 4: Correct tone identification (percent) as a function of tone type broken down by listener type: one for each combination of context (in carrier sentences, isolation) and time pressure (with, without).

From Figure 4 we can see the following:

(i) Beijing listeners identified the tones almost perfectly with no less than 99.3 percent correct regardless time pressure.

(ii) The aphasic listeners performed poorly with correct identification ranging from 65 to 91%. Time pressure affected the percent of lexical tone

> identification, especially in the case of Tone 2 and Tone 3 while context effect does not affect Tone-1 identification so much as it does Tone 2, Tone 3 and Tone 4.

Clearly, the tone identification task is easy for the healthy listeners. There is no point in analyzing the perceptual confusion structure for this group, as their level of performance is near-perfect. However, a tone-confusion analysis makes sense for the aphasic listeners. Therefore, Table 2 presents the lexical-tone confusion for Beijing aphasic listeners broken down by context (carrier vs. isolation) and time pressure (presence vs. absence of time pressure).

Table 2: Confusion matrices of tone identification by Beijing aphasic listeners broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation. Bold black numbers on shaded background are correct responses, off-diagonal cells identify confusions (≥ 10%).

Listener type	**Time pressure**	**Tones**	**Responses**							
			in carrier sentence				in isolation			
			Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
Beijing aphasic	No	Tone 1	**88.1**	7.5	3.0	1.5	**91.3**	1.4	7.2	
		Tone 2	4.5	**87.9**	4.5	3.0	1.4	**88.6**	**10.0**	
		Tone 3	8.8	1.5	**83.8**	5.9	8.6	7.1	**80.0**	4.3
		Tone 4	4.5	4.5	4.5	**86.6**	8.6	5.7		**85.7**
	Yes	Tone 1	**82.4**	8.8	7.4	1.5	**78.5**	4.6	9.2	7.7
		Tone 2	7.8	**73.4**	7.8	**10.9**	7.4	**85.3**	5.9	1.5
		Tone 3	9.0	**10.4**	**74.6**	6.0	**12.1**	4.5	**77.3**	6.1
		Tone 4	**12.5**	6.3		**81.3**	7.5	**10.4**	**16.4**	**65.7**

We observe that Beijing aphasic listeners have a fairly even spread of the tone confusions across the four types. Only in seven cases are tones confused in more than 10%. When time pressure was off, this happened only once for tones in isolation (Tone 2 → Tone 3) and never at all for tones in a carrier sentence. When time pressure was on, confusion was a little more widespread; it affected the perception of tones 2, 3 and 4 but never of Tone 1.

In order to come to an overall characterization of the performance of the two listener groups, we determined Receiver Operating Characteristic (ROC) curves for each individual listener in each of two groups (patients versus controls) with and without time pressure, and broken down further for presence versus absence of context. Since ROC analysis applies to dichotomies only, the responses were analyzed in terms of four binary oppositions: Tone-1 vs. not Tone-2, Tone-2 vs. not Tone-2, Tone-3 vs. non-Tone-3, and Tone 4 vs. not Tone-4. We then computed d'-values as a measure of detectability of each lexical tone among its competitors (with standard correction for perfect scores and zero false alarms). Figure 5 plots the d'-scores averaged over the four tones, but broken down by context (carrier sentence, isolation), time pressure (on, off) and by context (in carrier sentence: left panel; in isolation: right panel).

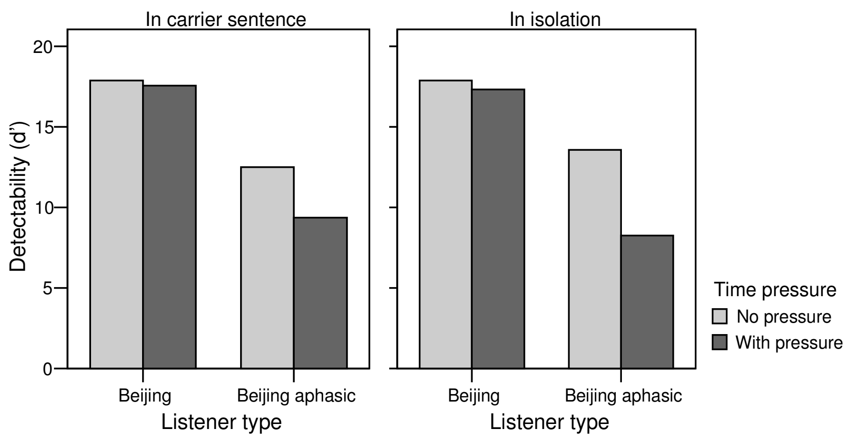


Figure 5: Detectability measure (d') for listener type broken down by presence versus absence of time pressure.

The d'-values were submitted to a similar RM-ANOVA as described above. Significant effects were still found for time pressure [F (1, 42) = 62.4 (p < 0.001)] and for listener type [F (1, 42) = 47.8 (p < 0.001)]. Additionally, only the second-order interaction between time pressure and listener type [F (1, 42) = 40.9 (p < 0.001)] and the third-order interaction, viz. time pressure × listener type × context [F (1, 42) = 5.8 (p = 0.02)], remain significant.

To summarize, the results show that the control listeners were significantly more sensitive to the lexical tones than the aphasic listeners overall, even though the patients discriminated between the four tones much better than chance. When time pressure was applied the patients' performance deteriorated much more than that of the controls. The patients came closest to the controls when listening to citation forms without time pressure; performance was intermediate for contexted tones without time pressure, and poorest when time pressure was applied.

3.2. Post-lexical identification (statement vs. question intonation)

In this section, we will analyze question responses to the two pitch manipulations in detail, i.e. the final rises and the overall pitch levels, in order to investigate the perception of question intonation across the two groups of listeners. When we analyze the effect of manipulation of the boundary tone, we average the data over the seven overall pitch levels. Conversely, when analyzing the effect of overall pitch level, we average over the seven final rises. This procedure reduces the number of data points by a factor 7. This procedure was adopted after we ascertained that the effects of boundary tone and overall pitch level were essentially independent. Although a 4-way RM-ANOVA – with final rise, overall level and time pressure as within-subject factors and listener type as a between-subjects factor – indicated a small interaction between rise and level, $F(24.1, 1013) = 2.5$ ($p < 0.001$), it accounts for a mere 4 percent of the total variability in perceived percent questions (as opposed to 70% and 23% for the main effects of final rise and overall pitch level, respectively). Moreover, visual inspection (see appendix 1) reveals that the weak interaction is only found with the healthy controls responding without time pressure. In this condition only, there is a tendency for the effect of overall pitch level to be smaller for the lowest boundary tones than elsewhere along the continuum but this does not seem to invalidate our assumption of independence for the two stimulus factors.

3.2.1. Question responses to manipulation of final pitch movement

Figure 6 presents the percentage of question responses, and the associated reaction time, for listener type as a function of the pitch manipulation of the final rise. Positive X-values denote a final rise, negative values are falls; '0 ST' indicates that the final syllable is spoken at a monotone.

Percent question decisions. A RM-ANOVA with final rise (7 values) and time pressure (on, off) as within-subject factors and listener type as a between-subject factor was carried out on the clause-type responses (statement = 0, question = 1). Significant effects were found for final rise [$F(2.5, 103.1) = 95.5$ ($p < 0.001$)], for the two-way interaction of final rise × time pressure [$F(5, 208.7) = 5.1$ ($p < 0.001$)] and for the three-way interaction, i.e. final rise × time pressure × listener type [$F(5, 42) = 3.6$ ($p = 0.004$)]. Crucially, no significant effect at all was found for listener type.

The effect of final rise is due to well-defined cross-overs from question to statement for the two groups of listeners. Typically, rises are perceived as question markers, whereas falls and the final monotone yield a clear majority of statement judgments. It is only the small-sized rises (between 0.5 and 1.0 ST) that are ambiguous between question and statement. No significant interaction between final rise and listener group is found, which indicates that the psychometric function for the aphasic listeners is as steep as that for the healthy listeners.

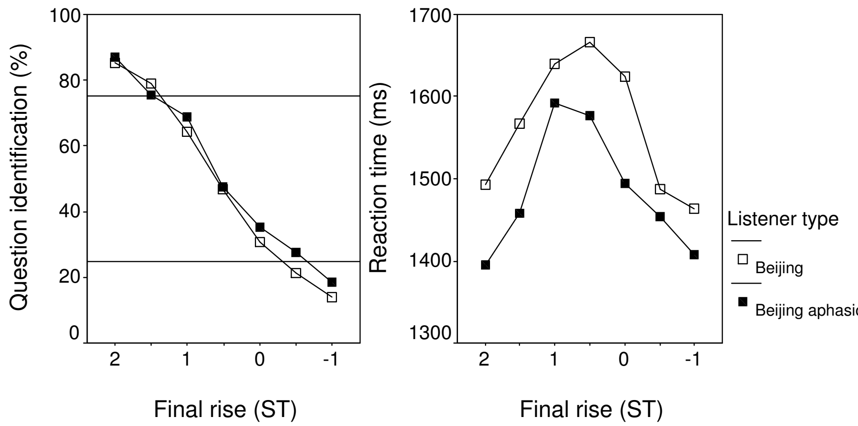


Figure 6: Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of final boundary tone (semitones) for Beijing healthy and aphasic listeners.

Decision time. A similar two-way RM-ANOVA was then carried out on the decision time. We observed just two large and significant main effects, i.e. for time pressure [F (1, 42) = 113.5 (p < 0.001)], and for final rise [F (4.6, 193.6) = 10.6 (p < 0.001)]. Neither the effect of listener type not any interaction reached significance.

The effect of final rise seems to follow in a straightforward fashion from the ambiguity of the stimulus as shown in the left-handed panel of Figure 6. For both groups of listeners, the two most extreme rise values are unambiguous exemplars of questions, yielding fast decisions. Likewise, the two final falls (negative rise values) are unambiguous tokens of statements, with fast decision times. Only the three middle values are ambiguous, with question responses between 25 and 75%. Here the decisions are delayed by some 100 ms. Aphasic listeners are overall faster than the controls by ca. 55 ms; this effect, however, is insignificant but at least the aphasics are not slower to respond than the controls, which is unusual in itself.

Figure 7 is a further breakdown by time pressure of the right-hand panel of Figure 6. The effect of time pressure as such is, predictably, that responses are a lot faster (over 400 ms on average) when time pressure is applied. With time pressure on, the aphasic patients are somewhat faster than the healthy listeners (50 to 100 ms) but the time pressure × listener type interaction was not significant. We note, again, that the aphasic patients perform the task at least as proficiently as the health controls, both in terms of choice and reaction time, and irrespective of time pressure.

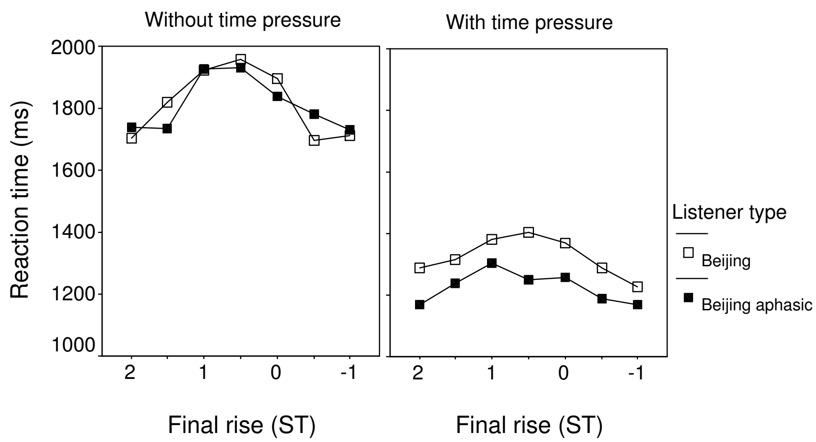


Figure 7: Reaction time (ms) for question identification without time pressure (left panel) and with time pressure (right panel) as a function of excursion size of final boundary tone (semitones) for Beijing aphasics and healthy listeners.

3.2.2. Question responses to manipulation of overall pitch level

Now let us consider the clause-type responses and their associated decision time by the two groups of listeners as a function of the manipulation of pitch level (but averaged over the seven boundary tone values). In the same way as we did with the final rise, we ran a three-way RM-ANOVA on the percentage of question responses to manipulation of overall pitch level. A significant main effect was revealed for pitch level only [$F (3.8, 161.4) = 12.5$ ($p < 0.001$)]. Significance was also found for the interaction between pitch level × listener type [$F (3.8, 161.4) = 3.7$ ($p = 0.008$)] and for the three-way interaction [$F (5.8, 241.5) = 3$ ($p = 0.008$)].

Figure 8 presents the percentage of question responses, and the associated decision time, for listener type as a function of the manipulation of the overall pitch level. The X-value indicates the number of 1-ST steps by which the overall pitch level was decremented relative to the original question token.

The significant effect of pitch level is as would be expected: higher overall pitch levels yield more question responses. However, in contradistinction to manipulating the final rise, the manipulation of overall level fails to produce well-defined cross-overs (left panel). Moreover, an approximation to a cross-over is seen only for the control subjects; the aphasic listeners do not seem to use the overall pitch level at all, hence the significant listener type × pitch level interaction.

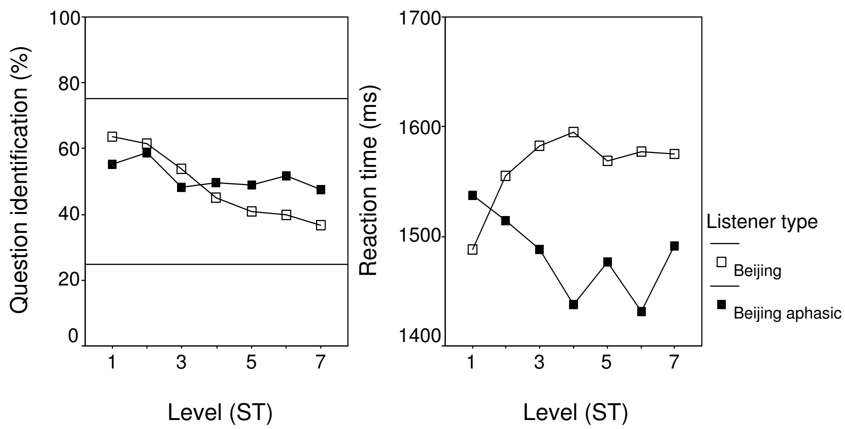


Figure 8: Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of pitch level (semitones) for Beijing aphasics and healthy listeners.

These are the same data as before but with one factor (final rise) replaced by another (pitch level). Therefore, the ANOVA on the decision times showed, predictably, the same significant effect of time pressure [$F(1, 42) = 112.7$ ($p < 0.001$)]. Significance was also found for the interaction between overall pitch level and listener type [$F(5.7, 238.7) = 2.3$ ($p = 0.036$)]. No other effects or interactions reached significance. In order to understand the interaction, a further breakdown of the results is presented in Figure 9, which plots percent question responses (upper panels) and the associated reaction times (lower panels) separately for absence (left) and presence (right) of time pressure.

Without time pressure the healthy listeners come closer to realizing a cross-over than when performing under time pressure. Also, in the absence of time pressure, the relationship between question responses and reaction time is as predicted (and also as found in the analysis above). Again, when the stimulus is ambiguous between question and statement, the reaction time is slow, but when the categorization is (reasonably) clear, the decisions are faster. This predicted relationship is found for the Beijing healthy group. With time pressure applied, the curves are shallow or completely level, and the local maximum in the decision-time curves is largely absent. The aphasics have a similar flat response curve, and their decision-time curve has no local maximum regardless of time pressure.

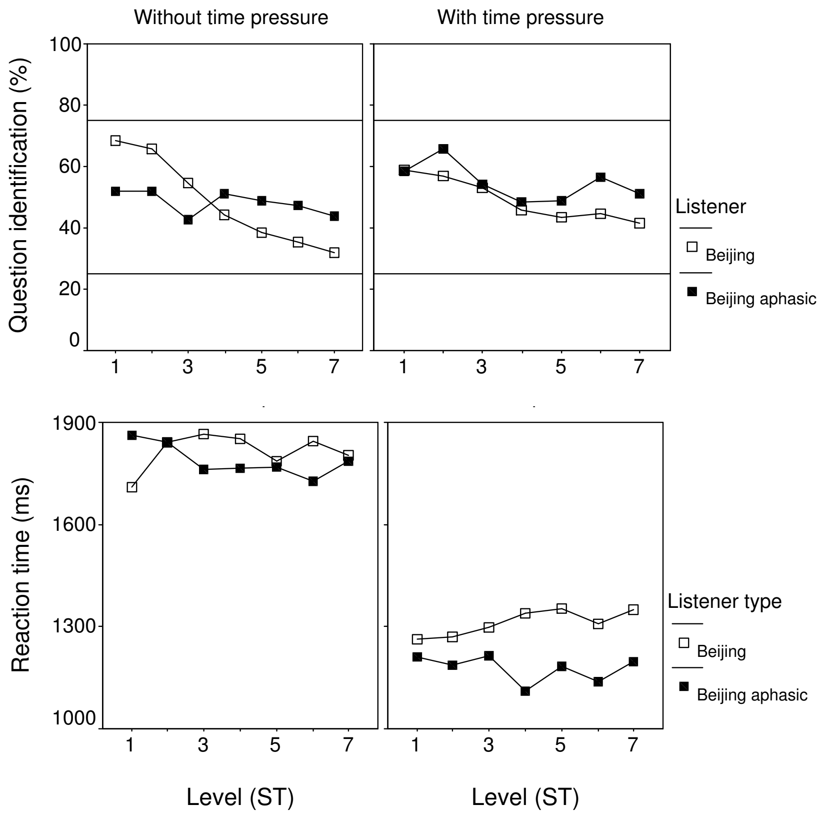


Figure 9: Question identification (percent, upper panels) and their associated reaction time (ms, lower panels) as a function of pitch level broken down by listener type, with and without time pressure (left and right panels, respectively).

In summary: in terms of question responses, patients and controls use the final rise in the same way. However, the patients do not use the overall pitch-level cue; the controls show a tendency towards a cross-over for the level cue (but clearly a lesser cue than the final rise) which is more pronounced if there is no time pressure. There is a large effect of time pressure on the overall speed of the task performance. Patients are somewhat faster than controls overall but the difference is not significant. Patients are faster for lower overall pitch levels, whilst controls are slower there; the reason for this interaction is unclear.

4. Conclusions and discussion

4.1. Summary of main findings

Before we proceed to draw any conclusions, let us first summarize, in Table 3, what we have found in our experiments. In this table an asterisk indicates that the healthy Beijing control listeners displayed a significantly better task performance than the aphasic listeners in terms of tone discrimination (d'), and speed thereof (ms) or in sensitivity to manipulation of pitch (boundary rise, overall pitch) in the identification of clause type. Cells are empty when no significant difference could be established by the ANOVA.

Table 3: Summary of findings for perception of lexical tone and sentence intonation: task performance of Beijing controls versus aphasic listeners. An asterisk indicated significantly better task performance for the healthy controls.

Stimulus type		Findings: controls - aphasics	
Lexical tone	Carrier	Correct tone discrimination	*
		Decision time	*
	Isolation	Correct tone discrimination	*
		Decision time	*
Sentence intonation	Final rise	Sensitivity	
		Decision time	
	Level	Sensitivity	*
		Decision time	

Relative to Beijing listeners, the aphasic listeners have poorer identification of lexical tones, both in terms of percent correct and decision time. At the same time the aphasics are as sensitive to pitch manipulations at the sentence level as are the Beijing controls, and equally fast (in fact a little faster but not significantly so), when determining clause type from the shape of the final boundary tone. However, the aphasics do not use the overall pitch-level cue here, which is a weak, secondary cue for the control subjects. Next, we will discuss the implications of these results in the light of the issues raised in our introduction.

4.2. Word-level versus sentence-level pitch variation

Our first research question asked how well Chinese aphasic patients with LH damage perceive pitch variations at word-level and at the sentence-level. Our results show that the Broca's patients involved in our study suffer from a considerable deficit in lexical tone perception. The control listeners had (near-) perfect tone-identification scores, whether or not obtained under time pressure, and irrespective of the tones being presented in citation form or in context, and they took their decisions relatively fast. The patients, in contrast to this, often identified the tones incorrectly (10 to 25% error) and needed more time for their decisions. Moreover, their performance deteriorated substantially when they had to perform the identification under time pressure. The fact that the patients' lexical tone identification was clearly more successful when no time pressure was applied shows that the tone deficit is at least partly due to a processing limitation. The patients' performance improved as time constraints were removed. However, it is not easy to determine whether the removal of all processing limitations would have resulted in perfect task performance (i.e. at the level of the control group). We would argue that the basic task of lexical tone identification – given the prototypical exemplars used in our experiment – was extremely simple for native listeners of Chinese, and made hardly any demands on the subjects. Yet, in spite of this, the patients' error rate was never less than 10%, even in the easiest condition, i.e. in citation forms with no time pressure. This would indicate to us that the patients did not only suffer from a processing limitation but that their mental representation of the four lexical tones was also compromised to some extent. Given that our patients sustained brain lesion, it comes as no surprise that their decision times, whether or not time pressure was applied, were slower overall than those obtained for the control subjects; we point out, nevertheless, that the effect of the lesion is very large indeed: the patients took almost twice as long over their lexical tone decisions (1800 ms on average) as the control subjects (1000 ms).

The perception of pitch-related linguistic contrasts at the sentence level, as determined in our experiment, did not differ markedly between patients and healthy control listeners. On the whole, both groups of listeners used the manipulations of the boundary tones in very much the same way and with equal proficiency when having to decide whether a speech utterance sounded like a statement or like a question. Not only did the patients respond to the manipulations of the boundary pitch with the same precision as the control listeners, reaching complete cross-overs from question to statement, they also performed their task – quite surprisingly – even faster than the controls (although not statistically significantly so). There is, however, one aspect in which the controls' perception of clause type (statement, question) differs from that of the patients. This is the use of the overall pitch-level cue. The patients with Broca's aphasia did not use the manipulation of overall pitch level at all when deciding whether the utterance was a question or a statement; their decision was entirely based on the properties of the final boundary tone. When the boundary tone was a rise, a question was perceived, when it was level or fell, the percept was that of a statement. For the control listeners, however, the manipulation of the overall pitch level provided a secondary, weaker cue, to the clause type, such that higher overall pitch added to the percept of a question.

Given that the patients with Broca's aphasia performed much poorer than the controls on the lexical-tone identification task but were equal to the controls in terms of identifying statements and questions, we conclude that the representation and/or processing of word-level prosody (such as involved in the recognition of lexical tones in Beijing Chinese) is separate from the processing of pitch at the sentence level.

4.3. Cue-dependent versus task-dependent mechanism

Now that we have ascertained that lexical tone perception was much more compromised than the perception of sentence intonation, we can address the issue of the mechanism that may be involved. The cue-dependent hypothesis (see introduction), predicts equal impairment of lexical tone and sentence intonation, at least when the stimuli have the same temporal properties. In our materials the relevant cues that enable the choice between the lexical tones or clause types, were confined to the temporal domain of one syllable. Equal impairment of tone and intonation, however, was not found, so that the cue-dependent hypothesis must be rejected. The task-dependent hypothesis is compatible with two possible outcomes (see introduction): (i) equal impairment for lexical tones and sentence intonation, since both belong to the speech domain, or (ii) differential impairment, since word tones and sentence intonation operate at different levels in the linguistic hierarchy. Given that no equal impairment was obtained, possibility (i) is ruled out. However, differential impairment, which was found, is compatible with the second configuration of results that would support the task-dependent hypothesis.

So, the results of our study support the task-dependent view over the cue-dependent hypothesis. Moreover, it was predicted that, in the case of unequal impairment of lexical tone and sentence intonation, there would be competition for resources. Now that the patients' lexical tone perception is compromised, more resources can be made available to the perception of sentence intonation. We observed that our patients were at least as fast in performing the intonation task as the control subjects and were still able to clearly differentiate between question and statement percepts. This finding, then, can be interpreted as reduced competition for resources, as was predicted by the task-dependent hypothesis.

We may then go on and ask if there is any need for entertaining a hybrid mechanism. It should be clear, at the outset of this discussion, that if our results are compatible with one of the non-hybrid mechanisms, in our case with the task-dependent hypothesis, they are a fortiori compatible with a hybrid hypothesis. So in order to abandon the task-dependent hypothesis in favour of a hybrid mechanism, we would have to look for additional evidence.

4.4. Use of context-sensitive sandhi rules

For both groups of listeners, whether aphasic or healthy, identifying tones presented in context took considerably longer than identifying the same tones in isolation. Interestingly, context slowed our listeners down but did not lead to more errors. The canonical shapes of the four tones of Beijing change considerably when preceded and/or followed by other tones in context. We assume that healthy listeners apply some phonological rule in order to relate the surface tone to its canonical form. Apparently, applying this rule is a more or less error-free process but it takes time. For aphasic listeners whose lexical tone impairment is caused by a structural deficit, the application of the tone-sandhi rules is not automatic and error-free, leading not only to delay but also to errors.

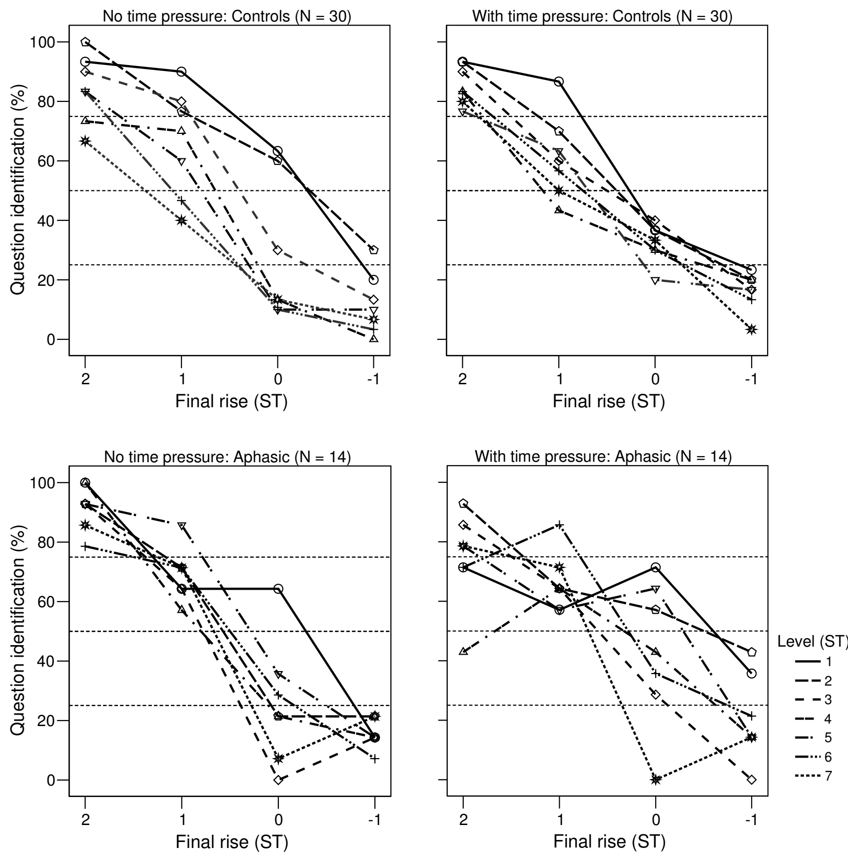
Addressing the issue of how tone sandhi words are mentally represented, Zhou and Marlsen-Wilson (1997) studied the effect of auditory priming on lexical decision by healthy Beijing listeners. They found that lexical decision to words affected by tone sandhi was more error-prone and took longer. Their hypothesis was that word forms with tone sandhi are not represented in the lexicon. Only canonical forms are stored in the lexicon, so that citation forms can be directly matched to the lexical representations; when words are presented in context, the effects of tone sandhi have to be undone by phonological rules. Our data are partially consistent with these findings in that words presented in a carrier sentence only took longer to recognize but did not lead to more errors, whether the listeners were healthy or aphasic.

The aphasic overall percent correct tone recognition is substantially poorer and takes much more time than that of the control group. These findings are consistent with previous lesion studies (Naeser and Chan, 1980; Eng et al., 1996). However, poor and slow as they were, the patients still possess an essential characteristic of healthy speakers: they performed as accurately with contexted targets as with isolated targets, i.e., they were still able to apply their implicit knowledge of tone-sandhi processes. The finding is consistent with previous studies on Chinese lexical tone identification in context with healthy subjects (Xu, 1994). Even if their processing capacity is limited, the aphasic listeners could still decode tones in connected speech as well (but not as fast) as they did in citation forms.

Additionally, the aphasic listeners have a fairly even spread of the tone confusions among the four types in the absence of time pressure. When time pressure was on, there was more confusion, typically affecting the perception of Tones 2, 3 and 4 but never of Tone 1. We claim, therefore, that Tone 1 has special status in Chinese, which would be a language-specific effect. This claim is also consistent with an earlier finding in the tone production of a single Beijing aphasic speaker, who preserved Tone 1 much better than any of the other tones (Liang and van Heuven, 2004).

In conclusion, we replicated earlier findings to the effect that aphasic listeners of a tone language such as Thai or Chinese have a significant loss of lexical tone identification (Gandour et al., 1983; Gandour et al. 2000; Hsieh et al., 2001; Hughes, Chan and Su, 1983; Klein, Zatorre, Milner and Zhao, 2001; Yiu and Fok, 1995). Since in all these studies, including our own, the patients had a lesion in the LH, it must be the case that the lexical tones are primarily lateralized to the left hemisphere. What is new in our result is that the brain damage is not sensitive to the difference between canonical shapes of tones (isolated citation forms) and assimilated forms in context.

Appendix: Full breakdown of perception of sentence intonation



Percent question responses as a function of final boundary tone and of overall pitch level, broken down by type of listener (aphasic, bottom versus control, top) and by time pressure (absent, left versus present, right).

Chapter Four

Chinese tone and intonation perceived by L1 and L2 listeners[14]

Abstract

Pitch is used in Chinese both at the word level to differentiate between four (or more) lexical tones and at the sentence level. A functional view claims that if some phonetic dimension is exploited in one area of the grammar (including the phonology) it will not be used to the same extent in another part of the phonology. We addressed the issue by asking how well learners of Chinese as a second language (L2) perceive pitch variations at word and sentence-levels. We assume that if there is competition between these two uses of pitch, we should expect L2 speakers whose native language (L1) is a tone language to be more sensitive to lexical tones and less sensitive to sentence intonation than L2 speakers whose L1 is a non-tone language.

Eighty-one Chinese-speaking (Beijing dialect) subjects participated in two laboratory-controlled perception experiments, on lexical tone and sentence intonation (statement vs. question), in which we manipulated the intonation contour of an entire utterance as well as that of the final syllable. We recruited one group of L1 listeners (native Beijing speakers) and three groups of L2 listeners, i.e. two groups with a tone-language L1 (Nantong and Changsha dialect) and one group with non-tone language L1 (Uygur, an Altaic language).

Results revealed that in perceiving tones and intonation in Beijing dialect, relative to L1 performance, L2 listeners with a non-tone-language L1 (Uygur) proved less sensitive to lexical tones but were at the same time more sensitive to intonation than L2 listeners with a tone-language L1. These results support the functional view, i.e., if some phonetic dimension is exploited in one area of the grammar, e.g. lexical tones, it will not be used to the same extent in another part of the phonology, e.g. sentence intonation.

[14] This chapter will appear as J. Liang and V.J. van Heuven (accepted) Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven and T. Riad (eds.), *Tone and Intonation*. Berlin: Mouton de Gruyter. It is an elaboration of a poster presented at the International Conference on Tone and Intonation, in Santorini, Greece, 9-11 September 2004. We gratefully acknowledge a travel grant by the Leiden University Fund (LUF) for subsidizing the first author's trip to the conference.

1. Introduction

Pitch is used in Chinese both at the word level to differentiate between four (or more) lexical tones and at the sentence level to signal – among other things – differences in sentence type such as statement versus question. Cross-linguistically there is a clear tendency for languages to make a choice for either word tones or sentence intonation. That is to say, if a language such as Chinese has a substantial inventory of word tones, it will exploit the pitch dimension less for the signalling of contrasts at the sentence level. In this connection, it is interesting that it has often been observed that Chinese possesses a sentence-final question particle (ma), which may signal interrogativity along with any pitch differences. This, then, is a functional view that claims that if some phonetic dimension is exploited in one area of the grammar (e.g. the phonology), it will not be used to the same extent in another part of the grammar (see, for instance, Ross, Edmondson and Seibert, 1986; Gandour, Larsen, Dechongkit, Ponglorpisit and Khunadorn, 1995; Seddoh, 2002).

There is a substantial body of experimental results on the question how well speakers of non-tone languages, such as English and Dutch, learn to speak and recognize lexical tones, mostly in Mandarin – which is the most widely spoken (and taught) tone language in the world. Results typically indicate that non-tone learners have poorly defined cross-overs in tone identification tasks, and show considerable inter-speaker variability in toneme boundaries (e.g. Leather 1987). There have been virtually no studies addressing the question of native-language interference in the acquisition of lexical tones by speakers of another tone language. A very recent literature survey (Wang, Jongman and Sereno, 2005) mentions two studies that touch upon this question, but only tangentially. Gandour (1983) studied perceptual similarity between pairs of natural tones in Mandarin, Cantonese, Taiwanese and Thai, but the task did not involve the perceptual identification of tones. Similarily, Vakoch and Wurm (1996) presented natural tokens of Mandarin and Cantonese tones to Mandarin, Cantonese and English listeners in a discrimination task. Their results indicate that English listeners discriminate the tones less successfully than listeners who speak a tone language. We are not aware of studies addressing tone identification by non-native learners with or without a tone-language L1.

Our questions are as follows:

(1) How well do learners of Chinese as a second language (L2) perceive pitch variations at the word level and at the sentence-level in comparison with L1 listeners of Chinese?
(2) Are L2 learners of Chinese differentially sensitive to linguistic levels (lexical tone versus intonation)?
(3) Are L2 listeners with a non-tone L1 less sensitive to lexical tones but more sensitive to sentence intonation than L2 listeners with a tone L1?

In our study, we set up two laboratory-controlled perception experiments on lexical tone and sentence intonation (statement vs. question), respectively, in which we manipulated the intonation contour of the whole utterance as well as the (non-lexical) boundary tone in the sentence-final syllable. As for subjects, we recruited 30

Beijing dialect speakers as the control sample, and compared their performance with that of four groups of non-native learners of Beijing dialect (Mandarin). The L2 groups differed in terms of native-language type (tone vs. non-tone language). Moreover, within the group of learners with a tone-language background a further division was made depending on the complexity of the native tone system (four, six or seven-tone inventory). This set-up will allow us to distinguish between several possibilities. We may find that all listeners show the same behavioral pattern irrespective of their L1 background. Alternatively, results may differ depending on the L1 type (tone versus non-tone) and/or the complexity of the tone system. If the latter results should obtain, the hypothesis that lexical tone and intonation perception compete for resources may be tested. We predict that, relative to L1 listeners, learners with a tone-language L1 should be more sensitive to lexical tones and less to intonation than non-tone L1 listeners.

2. Methods

We set up an experiment with native language, stimulus type and task as the principal factors. Each factor will be explained in a subsection.

2.1. Native language

Our study involved four groups of listeners, who all had the Chinese nationality but differed in their native language.

(1) Listeners of the **Beijing** dialect, which is very similar to the Standard language (Mandarin). The 30 Beijing native speakers were aged between 21 and 70, average 40, 17 male and 13 female. All of them had at least twelve years of normal education.

(2) Listeners speaking the seven-tone **Nantong** dialect of Chinese. The eight Nantong dialect listeners were construction workers, aged 22–48, who had just spent their first year in Beijing and learned Beijing dialect through natural exposure to the language.[15]

(3) Listeners speaking the six-tone **Changsha** dialect of Chinese. The thirteen Changsha listeners were first-year students at Hunan University, aged 19–21, who also intended to take the HSK,[16] which test is required for teacher certification in China. They had learned Standard Chinese either through formal training or self-study.

---

[15] In fact, some passive knowledge of Beijing dialect may have been acquired through watching television in Mandarin before the Nantong and Changsha speakers moved to the capital.

[16] HSK stands for China's Hanyu Shuiping Kaoshi 'Chinese Proficiency Test', a standardized test at the state level designed and developed by the HSK Center to measure the Chinese proficiency of foreigners, overseas Chinese and students from Chinese national minorities (http://www.hsk.org.cn).

(4) Listeners living in the People's Republic of China speaking **Uygur**, a Tungusic (Altaic) non-tone language which is not related to Chinese. The thirty Uygur listeners were students of Xinjiang Normal University, aged 18–20[17], who had had just finished a one-year intensive training course in Standard Chinese (Beijing dialect) and intended to pass the HSK, which is obligatory for them if they wish to continue their higher education in China.

In total, eighty-one Chinese volunteers participated. All of them were right-handed and reported no speech or hearing disorders. Composition of the listener groups is summarized in Table 1.[18]

Table 1: Listener types.

Language family		Language	Number of word tones	Number of subjects
Sinitic	Mandarin	Beijing	4	30
		Nantong	7	8
	Xiang	Changsha	6	13
Altaic	Turkic	Uygur	0	30

2.2. Stimuli

Two types of stimuli were used in the experiments for the lexical (word tone) as well as for the post-lexical (intonation) part. Recordings used in the experiments were from a male native speaker of Beijing dialect (Standard Chinese), experienced in sound recording. All the recordings were made in a sound-insulated booth at the Phonetics Laboratory, Leiden University, on digital audio tape (DAT) using a Sennheiser MKH-416 unidirectional microphone, then transferred to computer memory and downsampled to 16 kHz (16 bit amplitude resolution) using the Praat speech processing package (Boersma and Weenink, 1996).

For the lexical part, four frequent lexical words were used, covering the Beijing four-tone system, /ma^1 ma^2 ma^3 ma^4/ 'mother, hemp, horse, scold', respectively. In connected speech, some of the lexical tones may change their tonal category in

[17] The Xinjiang Uygur Autonomous Region is an area where bilingual language policy is in practice. Television programs and news papers, for instance, are issued in both Standard Chinese and in Uygur. The majority of the Uygur children receive their pre-university education in Uygur and take the university entrance examination in the Uygur language. When they are accepted by universities, they receive a one-year compulsory language training in Standard Chinese and they are required to pass Band 5 of the HSK before being allowed to continue their university education as all the courses at the university level are offered in Standard Chinese.

[18] The authors thank Prof. Ning Chunyan (in Changsha), Prof. Li Bing (in Changsha), Associate Prof. Guo Weidong (in Ulumqi), and Mr. Li Gang (in Beijing) for their kind help in recruiting subjects. We also thank all our subjects for participating in the experiments.

context. For example, the low tone in Beijing dialect changes into a rising tone when followed by another low tone. When that happens, the derived rising tone is perceptually indistinguishable from the lexical rising tone (Wang and Li, 1967). This kind of categorical tone shift due to tonal context is usually referred to as ‘tone sandhi’. In order to control the context effect, the lexical-tone stimuli were recorded by reading aloud a list of the four words in a carrier sentence and in isolation respectively, four times for each. In all there were 40 stimuli for lexical tone identification, i.e. 4 (tone) × 5 (repetition) × 2 (in a carrier sentence and isolation).

For the post-lexical (intonation) part, a single question utterance was used: /kai^1 sun^1 jiŋ^1 kʰai^1 fei^1 tɕi^1/, ‘must Sun Ying fly flying machine?’ meaning ‘Is it Sun Ying’s turn to fly the plane?’ The basic utterance contains high level tones (Tone 1) only, in order to reduce the lexical tonal influence. The sentence can be interpreted as either a statement or a question depending on the intonation. Only tokens with question intonation were elicited. The speaker was instructed to ‘say’ the sentence written with a question mark as naturally as possible.

Studies have shown that intonation is perceived in terms of specific fundamental frequency (F0) patterns (Gårding and Abramson, 1965; ’t Hart, Collier and Cohen, 1990), which involve the same acoustic parameter that is employed in the signalling of lexical tone (Pike, 1948; Wang, 1972; Howie, 1976). Previous acoustic and perception studies on question intonation of Standard Chinese have shown that besides the use of question particles, the main acoustic correlate of the statement~question contrast is the fundamental frequency (Howie, 1976; Tseng, 1981; Coster and Kratochvil, 1984). Since our sentences did not contain any lexical (including particles) and/or syntactic question markers, they are optimally suited for studying the role of intonation in signalling the difference in sentence type between statement and question.

Using the Praat software the pitch contour of one question utterance was stylized with seven pivot points connected by straight lines in a log-frequency (semitone, ST) by linear time representation. It is generally accepted that the pitch variation in the final syllable signals the statement~question contrast in Standard Chinese (Tsao, 1967; Gårding, 1985, 1987; Kratochvil, 1998; Yuan, Shih and Kochanski, 2002) but there is disagreement on the issue whether the overall pitch level of the utterance provides a subsidiary cue to the contrast (De Francis, 1963; Kratochvil, 1998) or not (e.g. Tsao, 1967). Kratochvil (1998), in fact, described the question intonation of Beijing Dialect as accompanied by a raising of the overall pitch level by about 50 Hz. We perceptually tested the relative contributions of the overall pitch and of the final fall/rise to the statement~question contrast. To this effect, a two-dimensional stimulus continuum was generated (through PSOLA resynthesis, cf. e.g. Moulines, and Verhelst, 1995) from the question utterance, by systematically combining seven overall pitch levels and seven different boundary tones. The overall pitch levels were decremented by 1-ST steps, which adequately covered the range we established in the production of question and statement tokens by our speaker. The seven boundary tones were generated by decrementing the terminal pitch of the question token in six steps of 0.5 ST. Again, the 3.5-ST difference between the highest and lowest terminal frequencies in the boundary tone continuum corresponds closely to what we found in our speaker's production. By manipulating only the frequency values while leaving the time coordinates unaltered, 49 different pitch patterns were generated according to the following schema (Figure 1).

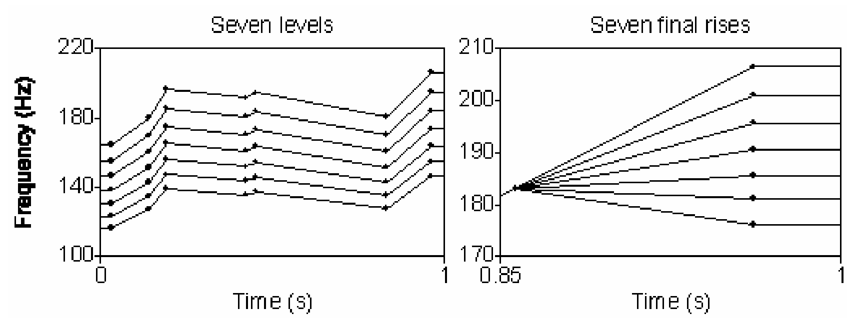


Figure 1: Overview of the manipulations of pitch level (7 steps) and boundary tone (7 steps). Orthogonal variation of level and boundary tone yields 49 different stimulus types

2.3. Tasks and procedure

Considering that lexical tone stimuli contained naturally produced tokens only, we ran the two experiments in a fixed order, such that the lexical-tone experiment always preceded the intonation test (with potentially ambiguous, hybrid exemplars). The stimuli were randomized and presented by computer binaurally over headphones (Sony MDR-V3) at a comfortable listening level. Listeners were tested one at a time in a quiet room. A specially designed keyboard was used with either four buttons marked with the corresponding Chinese characters (in font 72) for the tone identification, or two buttons marked with '。' and '?', i.e., punctuation marks that are used in Chinese dialects for statement and question.[19] Before each part of the experiment specific instructions were given both orally and in writing. Each experimental task was preceded by a short practice session, with four practice trials for the tone identification task and seven for the intonation task. Each experiment lasted about 20 minutes. Decisions made and reaction times (measured from the offset of the stimulus, precision 1 ms) were stored in the computer. Stimulus presentation and response collection were controlled by E-prime software.[20]

In the **lexical tone experiment**, the subjects decided which of four words they had heard by pressing one of the four buttons on the response keyboard each time they

[19] The keyboard was designed and built by J.J.A. Pacilly at the Universiteit Leiden Phonetics Laboratory.
[20] The E-prime script for stimulus presentation and response collection was written by J.J.A. Pacilly

heard a stimulus. In the first stage of the experiment they were asked to avoid errors (no time pressure). In the second stage the subjects were instructed not only to avoid errors but also to perform the task as quickly as they could manage (time pressure). The four buttons were evenly spaced across the top row on the response box; a black key for continue/start was located in the center of the bottom row. The 40 lexical-tone stimuli, 20 in a carrier sentence and 20 in isolation were presented to the listeners twice in two blocks (once without time pressure and a second time with time pressure).

When there was no time pressure, there was a fixed 3000-ms (isolated targets) or 5000-ms (targets in carrier) inter-stimulus interval (ISI) after the offset of the stimulus, irrespective of the reaction time. When no response was given within the ISI, the subject timed out, and the next stimulus was presented. In the sections with time pressure imposed on the listener, the next stimulus started 1000 ms (in isolation) or 2000 ms (in a carrier sentence) after the response. The shorter ISI in the time-pressure condition prompted the subjects to speed up their reaction time. In the debriefing after the experiment, subjects confirmed that they had felt pressured; the effects of pressure are also apparent from the results: not only were reaction times much faster, but also was the number of timed-out responses higher (even though the ISI had not been reduced).

In the **intonation experiment**, the listeners were required to choose between statement and question by pressing one of the two buttons marked with '□' or '?' each time they heard a stimulus. As before, the subjects performed their task without and with time pressure. The 49 stimuli were presented twice in two blocks in random order, i.e. once without and once with time pressure. Every stimulus was preceded by a fixed reference sound, which was a single token of the neutral vowel (schwa) spoken by the same speaker who recorded the stimuli, and PSOLA resynthesized at 119 Hz, which is the lowest pitch used in the experiment, i.e. the onset pitch of the lowest resynthesized version.[21] The reference precursor preceded the test sentence by 500 ms (offset to onset). In the absence of time pressure, there were fixed 6000-ms ISIs between stimuli (offset to onset), irrespective of the reaction time. When time pressure was imposed, the next stimulus was presented 3000 ms after the previous response.

[21] The reference signal was a sequence of fourteen identical periods sampled from the middle of a natural schwa token. The artificial intensity contour comprised a 40-ms rise time, 50-ms steady state, and a 40-ms decay time.

3. Results

In this section, we report the results obtained with the types of listener for lexical tone identification and post-lexical identification of sentence type. We will first analyze the data collected for the perception of the lexical tone contrast. In the second part we will complement this with the corresponding analysis of the sentence intonation contrast. It should be noted beforehand that these two parts are rather different conceptually. In the matter of the lexical tones, we can determine whether the listener correctly identified the tone pattern as it was intended by the speaker of the utterance.

Also, we may determine how much time it took the listener to correctly decide on the tonal category. Since the stimuli were presented for identification once with and once without emphasis on speed of response (time pressure), we predict that listeners traded accuracy for speed when time pressure was on, i.e., were prepared to gamble in the case of ambiguous stimuli in order to gain speed. In the identification of sentence melody (question ~ statement), all the stimuli were derived through parametric manipulation from a single question utterance.

There is no pre-given norm here of what would be a statement and what counts as a question. So in this section of the data we will just analyze the percentage of 'question' responses and the decision latencies; we expect longer latencies as the choice between the response alternatives is more evenly balanced (which would be a sign of ambiguity in the stimulus). It is unclear, in the identification of sentence melody, how time pressure should affect the quality of the response.

3.1. Lexical tone identification

A four-way repeated measures analysis of variance (RM-ANOVA) with type of listener (Beijing, Nantong, Changsha and Uygur) as a between-subject factor, and with stimulus tone (Tone 1 to Tone 4), context (in carrier sentence vs. in isolation) and time pressure (on vs. off) as within-subject factors was carried out on the lexical tone responses.

It shows that, overall, percent correct identification was significantly different for the four types of listener [$F(3, 77) = 65.4$], for context [$F(1, 77) = 24.1$], and for tone [$F(3, 223) = 7.9$] ($p < 0.001$ in all cases). Significance was also found for the interactions of type of listener × context [$F(3, 77) = 40.5$ ($p < 0.001$)], and of listener type × tone [$F(8.7, 223) = 12.5$ ($p < 0.001$)].[22] Scheffé post-hoc tests divided the listener type into two groups, i.e. (i) Uygur, and (ii) Nantong, Beijing and Changsha.

[22] Degrees of freedom were always Huyhn-Feldt corrected, which may yield non-integer values for interaction terms.

A similar four-way RM ANOVA was carried out on the reaction time of the lexical-tone responses. It shows that, overall, reaction time was significantly different for the various types of listener [F (3, 77) = 38.5 (p < 0.001)], context [F (1, 77) = 57.1 (p < 0.001)], time pressure [F (1, 77) = 416.3 (p < 0.001)] and for tone [F (2.9, 225.8) = 4.2 (p = 0.006)].

Significant effects were found for second-order interactions of listener type × context [F (3, 77) = 19.4 (p < 0.001)], listener type × time pressure [F (3, 77) = 24.9 (p < 0.001)], context × time pressure [F (1, 77) = 28.9 (p < 0.001)], context × tone [F (3, 231) = 9 (p < 0.001)], time pressure × tone [F (2.9, 226.6) = 4.5 (p = 0.005)], and listener type × tone [F (8.8, 225.8) = 2.4 (p = 0.014)]. Also, three-way interactions proved significant for listener type × context × time pressure [F (3, 77) = 13.7 (p < 0.001)]. Scheffé post-hoc tests divided the listener type into two groups: (i) Changsha, Beijing and Nantong; (ii) Uygur.

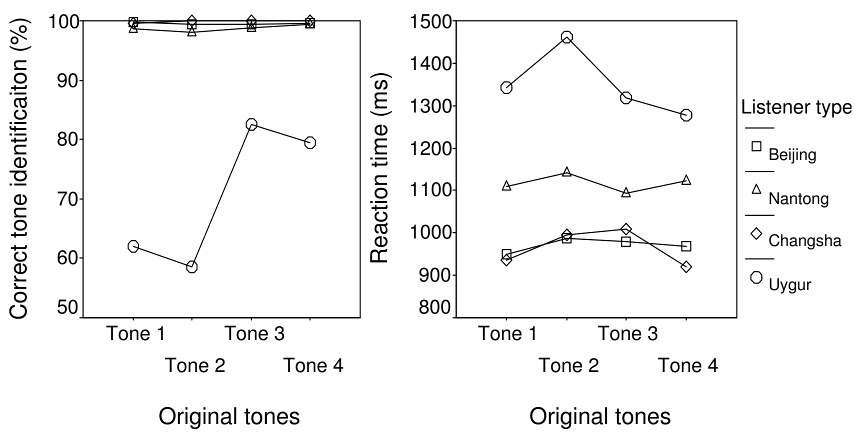


Figure 2: Correct tone identification (percent) and associated reaction time (ms) as a function of tone type broken down by four types of listener.

Correct tone identification (percent) and the associated reaction time (ms) are presented in Figure 2 as a function of tone type broken down by the four types of listener.

Figure 2 shows that the Uygur group was much poorer in tone identification, and much slower in their reaction time relative to the tone-language groups. In order to have a clear picture of the distribution of these data, we also plotted percent correct by the four speaker groups against the associated reaction time broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation in Figure 3.

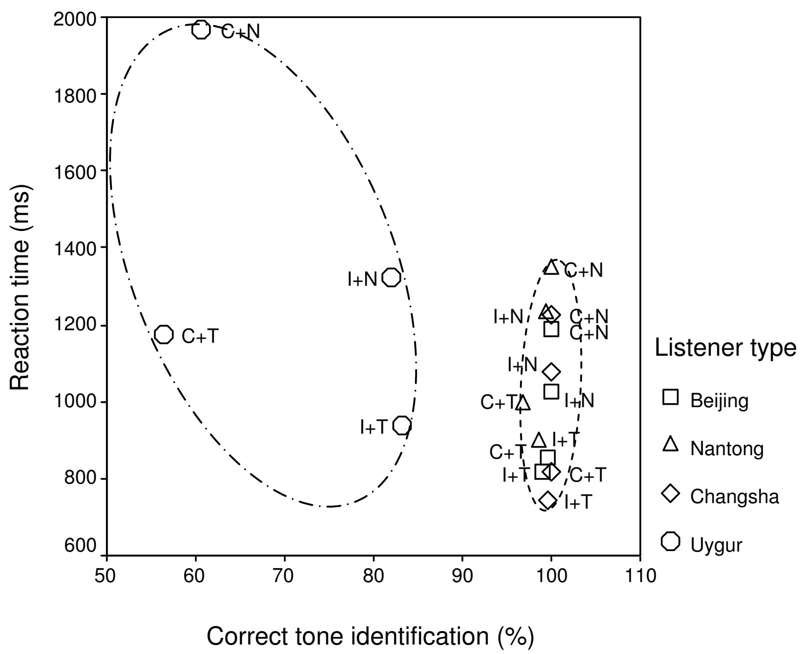


Figure 3: Percent correct identification of lexical tone plotted against reaction time (ms) for four types of listener (differing in native language background) in two time-pressure conditions (no time pressure 'N' versus time pressure 'T') and two contexts (in carrier sentence 'C' versus isolation 'I'). Ellipses were drawn around the centroids by hand.

Figure 3 shows the following effects:

(i) **Language background**. Uygur non-tone-language listeners (dash-dotted line) are separate from the tone-language groups.

(ii) **Time pressure**. In each type of listener time pressure leads to considerably shorter reaction time but has little effect on percent correct. Post-hoc tests above show that the Uygur groups were significantly different from the tone-language groups, while the three tone-language groups were not significantly different from each other.

(iii) **Context**. All listeners took longer when identifying tones in the carrier sentence than in isolation. Importantly, however, embedding the target in a carrier did not affect percent correct for any tone-language group, but it clearly reduced percent correct for the non-tone-language Uygur listeners.

From the above findings, we conclude that the contexted targets created a problem for Uygur listeners, such that their overall percent correct tone recognition is substantially poorer and takes much more time than that of the tone-language listeners.

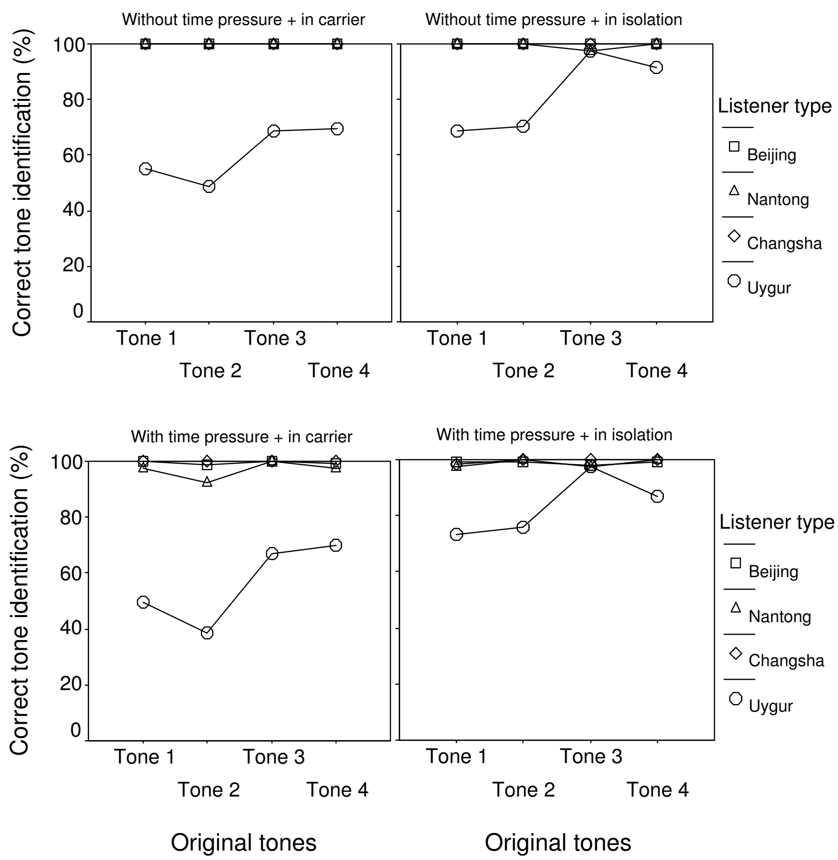


Figure 4: Correct tone identification (percent) as a function of tone type broken down by four types of listener: one for each combination of context (in carrier sentences, isolation) and time pressure (with, without).

As a last stage in the analysis of lexical tone identification let us break down the results by the four separate tones. Figure 4 plots percent correct identification for each of the four tones broken down by listener type. The presentation is broken down further into four panels, one for each combination of context (carrier sentence, isolation) and time pressure (with, without).

From Figure 4 we can see the following:

(i) Beijing and Changsha listeners identified the tones almost perfectly with no less than 98.5 percent correct.

(ii) The Nantong group was less perfect, with 92.5% correct Tone-2 identification.

(iii) The Uygur group was the poorest, with percent correct ranging from 40 to 98.

Clearly, then, being a native speaker of some other Chinese tone language affords a listener to perform as well in the Beijing tone identification task as the native listeners of the Beijing dialect themselves. There is no point in analyzing the perceptual confusion structure for these three groups of listeners, as their level of performance is nearly perfect. However, a tone-confusion analysis makes sense for the Uygur listeners. Therefore, Table 2 presents the lexical-tone confusion for Uygur listeners broken down by context (carrier vs. isolation) and time pressure.

Table 2: Confusion matrix of tone identification by Uygur listeners broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation. Bold numbers against a shaded background are correct responses.

Listener type	Time pressure	Tones	Responses							
			in carrier sentence				in isolation			
			Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
Uygur	No	Tone 1	**55.2**	24.8	4.8	15.2	**68.5**	21.5	5.4	4.7
		Tone 2	23.3	**48.6**	17.1	11.0	22.1	**70.5**	1.3	6.0
		Tone 3	16.0	12.0	**68.7**	3.3		2.0	**97.3**	0.7
		Tone 4	5.5	17.2	7.6	**69.7**	2.7	4.7	1.3	**91.3**
	Yes	Tone 1	**49.6**	30.2	12.2	7.9	**73.3**	16.7		10.0
		Tone 2	38.6	**38.6**	7.6	15.2	18.1	**75.8**		6.0
		Tone 3	8.1	21.5	**67.1**	3.4	1.3		**97.3**	1.3
		Tone 4	6.0	10.7	13.4	**69.8**	4.7	8.7		**86.7**

There is considerable confusion for Uygur listeners, especially when the tones were presented in carrier sentences. The largest confusion occurred between Tone 1 and Tone 2: Tone 1 tokens were misidentified as Tone 2 in 30% and 25% with and without time pressure, respectively, while 39% and 23% of Tone 2 were misidentified as Tone1.

In order to come to an overall characterization of the performance of the four listener groups, we determined Receiver Operating Characteristic (ROC) curves for each individual listener in each of the four groups (Beijing, Nantong, Changsha and Uygur) with and without time pressure, and broken down further for presence versus absence of context.

Since ROC analysis applies to dichotomies only, the responses were analyzed in terms of four binary oppositions: Tone-1 vs. not Tone-2, Tone-2 vs. not Tone-2, Tone-3 vs. non-Tone-3, and Tone 4 vs. not Tone-4. We then computed d'-values as a measure of detectability of each lexical tone among its competitors (with standard correction for perfect scores and zero false alarms).

Figure 5 plots the d'-scores averaged over the four tones, but broken down by context (carrier sentence, isolation), time pressure (on, off) and by context (in carrier sentence: left panel; in isolation: right panel).

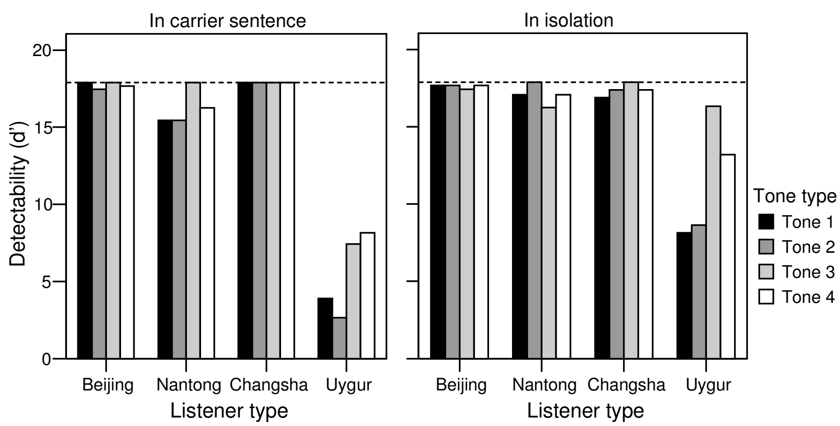


Figure 5: Detectability measure (d') for listener type broken down by tone type (Tone 1, Tone 2, Tone 3 and Tone 4). Maximum d' = 17.89

The d'-values were submitted to a similar RM-ANOVA as described above. Significant effects were still found for listener type [F (3, 77) = 92.3 (p < 0.001)], time pressure [F (1, 77) = 8.8 (p = 0.004)], context [F (1, 77) = 22.8 (p < 0.001)] and tone [F (2.9, 222.9) = 10 (p < 0.001)].

Significant effects were also found for two second-order interactions, i.e. context × listener type [F (3, 77) = 37.4 (p < 0.001)] and tone × listener type [F (8.7, 222.9) = 12.6 (p < 0.001)]. One third-order interaction, viz. context × tone type × listener type [F (8.4, 216.5) = 2.4 (p = 0.016)] remains significant as well. Scheffé post-hoc tests separated the Uygur listeners from the tone-language groups, who did not differ from each other (α = 0.01).

To summarize, the results show that the tone listeners were significantly more sensitive to the lexical tones than the non-tone listeners (Uygurs) overall, even though the Uygurs discriminated between the four tones much better than chance. When the tones were presented in context, the Uygurs' performance deteriorated much more than that of the tone-language groups'. The Uygur listeners came closest to the tone-language groups when listening to citation forms, especially Tone 3; performance was intermediate for Tone 4 in isolation, and poorest with contexted Tone 2 and Tone 1.

3.2. Post-lexical identification (statement vs. question intonation)

In this section, we will analyze question responses to the two pitch manipulations in detail, i.e. the final rises and the overall pitch levels, in order to investigate the perception of question intonation across the two groups of listeners. When we analyze the effect of manipulation of the boundary tone, we average the data over the seven overall pitch levels. Conversely, when analyzing the effect of overall pitch level, we average over the seven final rises.

This procedure reduces the number of data points by a factor 7. This procedure was adopted after we ascertained that the effects of boundary tone and overall pitch level were essentially independent. Although a four-way RM-ANOVA – with final rise, overall level and time pressure as within-subject factors and listener type as a between-subjects factor – indicated a small interaction between rise and level [$F(28.6, 2205) = 4$ ($p < 0.001$)], it accounts for a mere 4 percent of the total variability in perceived percent questions (as opposed to 70% and 23% for the main effects of final rise and overall pitch level, respectively).

Moreover, visual inspection (see appendix) reveals that the weak interaction is only found with the native speakers (the Beijing listeners) responding without time pressure. In this condition only, there is a tendency for the effect of overall pitch level to be smaller for the lowest boundary tones than elsewhere along the continuum but this does not seem to invalidate our assumption of independence for the two stimulus factors.

3.2.1. Question responses to manipulation of final pitch movement

Figure 6 presents the percentage of question responses, and the associated reaction time, for the four types of listener as a function of the pitch manipulation of the final rise. Positive X-values denote a final rise, negative values are falls; '0 ST' indicates that the final syllable is spoken at a monotone.

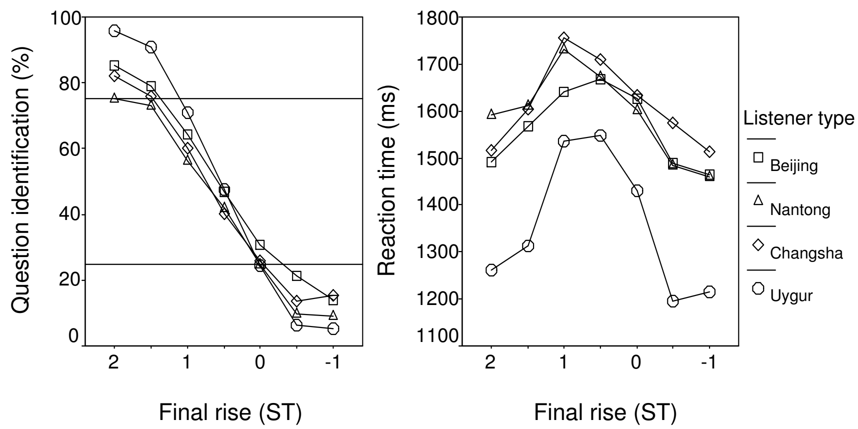


Figure 6: Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of final boundary tone (semitones) for listeners from four languages spoken in China.

Percent question decisions. A three-way RM-ANOVA with listener type as a between-subject factor, and final rise and time pressure as within-subject factors was carried out on the sentence-type responses (statement = 0, question = 1). A large significant effect was obtained for final rise [F (2.9, 227) = 289 (p < 0.001)], but no significance was found for either listener type [F (3, 77) = 1.3 (p = 0.27)] or time pressure [F (1, 77) = 0.047 (p = 0.83)]. The effect of final rise is due to well-defined cross-overs from question to statement for all four groups of listeners. Typically, final rises are perceived as question markers, falls and the final monotone yield a clear majority of statement judgments. It is only the small-sized rises (between 0.5 and 1.0 ST) that are ambiguous between question and statement. The effect of listener group is not interesting to our analytical problem, as it might only reveal slight differences in response bias.

Significance was also found for the interaction of listener type × final rise [F (8.8, 227) = 3.6 (p < 0.001)] and final rise × time pressure [F (4.8, 366) = 2.8 (p = 0.02)] but not for listener type × time pressure [F (3, 77) = 0.22 (p = 0.89)]. There was no significant third-order interaction. The significant interaction of listener type × final rise is crucial, since it reveals differences in sensitivity, i.e., the psychometric function is clearly steeper for Uygurs than, for instance, for Changsha listeners.

We tested this by first computing for each listener a boundary width, which was defined as the distance in step-size units separating the 25% and 75% cross-over points as determined by probit interpolation. The steepness (slope coefficient) of the cross-over was then calculated by dividing the boundary width into 50. Thus, a slope coefficient of 10 would indicate that increasing the terminal frequency of the final rise by 1 semitone (i.e. 2 steps) yields an increment of the number of 'question' responses by 10 percentage points. More generally, the higher the slope coefficient, the steeper the cross-over and the more sensitive the listener. The results are listed in Table 3.

Table 3: Slope measure (N, mean, SD) for manipulation of terminal boundary pitch for four groups of listeners broken down by task performance with and without time pressure.

Listener type	*N*	*Without time pressure*		*With time pressure*	
		Mean	*SD*	*Mean*	*SD*
Beijing	30	20.9	12.9	23.1	20.9
Nantong	8	27.1	15.4	16.4	10.2
Changsha	13	16.4	9.5	22.9	16.9
Uygur	30	42.1	29.1	32.8	23.9

From Table 3 we can see that Uygur listeners have the steepest slopes (42.1), which indicates that these listeners are more sensitive to the changes in the final rise than the three tone-language groups. A two-way RM-ANOVA on the slope measures with listener type as a between-subject factor and time pressure as a within-subject factor show significant effects for listener type [F (3, 77) = 5 (p = 0.003)] and not for time pressure [F (1, 77) = 1.2 (p = 0.28)]. However, the second-order interaction between listener type and time pressure was found significant [F (3, 77) = 3 (p = 0.034)]. Scheffé post-hoc tests show that the Uygur group was significantly different from the other tone-language groups, which did not significantly differ among themselves ($\alpha = 0.01$).

Decision time. A similar analysis was carried out on the reaction time for sentence-type responses. Here significance was found for listener type [F (3, 77) = 6.7 (p < 0.001)], final rise [F (4.4, 340) = 27.4 (p < 0.001)] and for time pressure [F (1, 77) = 165.2 (p < 0.001)]. Significance was also found for the second-order interaction final rise × time pressure [F (4.8, 367) = 5.8 (p < 0.001)], but not for the interactions of listener type × final rise [F (13.2, 340) = 1.5 (p = 0.103)] and listener type × time pressure [F (3, 77) = 1.6 (p = 0.206)]. Scheffé post-hoc tests show that the Uygur group was significantly different from the other tone-language groups (Beijing, Nantong and Changsha), which did not significantly differ among themselves ($\alpha = 0.01$), i.e. 1357 vs. 1570, 1595 and 1624 ms respectively ($\alpha = 0.01$).

The effect of final rise seems to follow in a straightforward fashion from the ambiguity of the stimulus. The clearest example is seen in the case of Uygur listeners. The two most extreme rise values are unambiguous exemplars of questions, yielding fast decisions. Likewise, the two final falls (negative rise values) are unambiguous tokens of statements, with fast decision times. Only the three middle values are ambiguous, with question responses between 25 and 75%. Here the decisions are delayed by some 100 to 200 ms. The same effect, though less extreme, is seen for the other listener groups.

Figure 7 is a further breakdown (by time pressure) of the right-hand panel of Figure 6. The effect of time pressure as such is, predictably, that responses are a lot faster (over 400 ms on average) when time pressure is applied. However, without time pressure the listeners are especially slow when having to decide on ambiguous stimuli. When the need arises to respond quickly, they save relatively more time on ambiguous stimuli than on unambiguous ones.

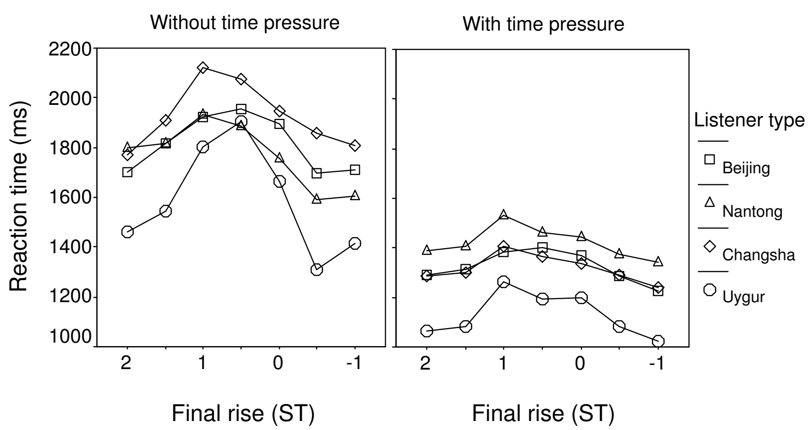


Figure 7: Reaction time (ms) for question identification without time pressure (left panel) and with time pressure (right panel) as a function of excursion size of final boundary tone (semitones) for listeners from four languages spoken in China.

There is a further complication in the data. When there is no time pressure, the difference in reaction time between slow and fast listener groups is ca. 440 ms. When time pressure is on, the difference between the extreme groups is reduced to 300 ms. Consequently, we would expect there to be greater overlap between the four groups such that the groups are less distinct when performing under time pressure. However, the reverse proves to be the case.

The statistical separation between the four listener groups is clearly better when subjects perform under time pressure than when not. Also, the order among the four groups changes considerably. Uygurs are always the fastest group, regardless of time pressure. Without time pressure the three groups of tone-language listeners are relatively close together and only Changsha listeners are significantly slower than the others.

3.2.2. Question responses to manipulation of overall pitch level

Now let us consider the sentence-type responses and their associated reaction time from the four types of listener as a function of the manipulation of pitch level. The data are presented in Figure 8.

We ran a similar three-way RM-ANOVA on the sentence-type responses as we did with final manipulation, with listener type as a between-subject factor and with overall pitch level and time pressure as within-subject factors. A significant effect was revealed for pitch level [F (3.5, 272.9) = 54.18 (p < 0.001)], but not for listener type [F (3, 77) = 1.35 (p = 0.27)] nor for time pressure [F (1, 77) = 0.05 (p = 0.83)]. The interactions of listener type × pitch level [F (10, 272.9) = 4.30 (p < 0.001)] and of time pressure × pitch level [F (6, 462) = 7.41 (p < 0.001)], however, were significant.

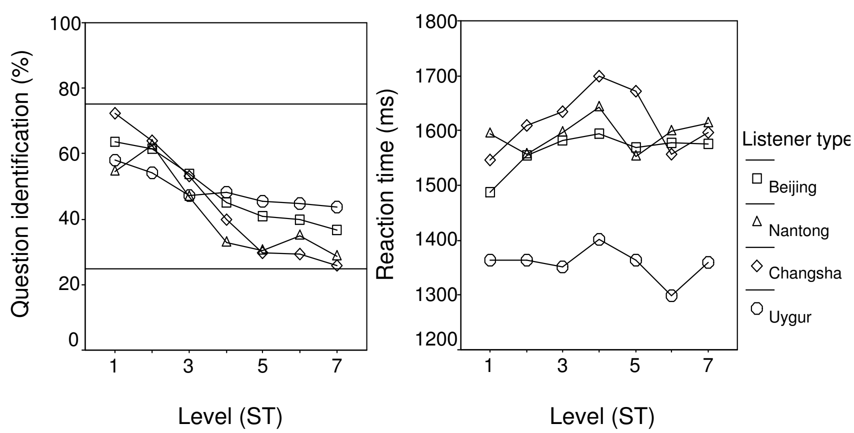


Figure 8: Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of pitch level (semitones) for listeners from four languages spoken in China.

Again, the main effect of listener type is irrelevant (see above). The effect of pitch level is as would be expected: higher overall pitch levels yield more question responses. However, in contradistinction to manipulating the final rise, the manipulation of overall level fails to effect a well-defined cross-over, at least when the responses are averaged across all four listener groups (left panel). In fact, a full cross-over (from 78 to 20% question responses) is obtained only for Changsha listeners, and only when responding without time pressure.

In Table 4, we present slope measures for manipulation of overall pitch level for the four groups of listeners with and without time pressure.

Table 4: Slope measure (N, mean, SD) for manipulation of overall pitch level for four groups of listeners broken down by task performance with and without time pressure.

Listener type	*N*	*Without time pressure*		*With time pressure*	
		Mean	*SD*	*Mean*	*SD*
Beijing	30	7.7	6.6	2.0	9.9
Nantong	8	9.8	12.1	5.6	4.7
Changsha	13	13.5	9.9	8.9	12.1
Uygur	30	2.7	2.4	1.7	2.5

Table 4 shows that the steepest slope coefficient (13.5) is obtained, when no time pressure is applied, by the Changsha group, closely followed by the Nantong and Beijing groups, while the Uygur listeners have the flattest curves. A two-way RM-ANOVA on the slope measures with listener type as a between-subject factor and time pressure as a within-subject factor shows significant effects for listener type [F $(3, 77) = 6.6$ ($p < 0.001$)] and time pressure [F $(1, 77) = 14.9$ ($p < 0.005$)]. No significant effect was found for the second-order interaction. Scheffé post-hoc tests divided listener type into three subsets, (i) Uygur, Beijing, (ii) Beijing, Nantong, and (iii) Nantong, Changsha ($\alpha = 0.01$).

Decision times. These are the same data as before but with one factor (final rise) replaced by another (pitch level). Therefore, the RM-ANOVA on the reaction times showed, predictably, significant effects for listener type [F $(3, 77) = 6.75$ ($p < 0.001$)], for time pressure [F $(1, 77) = 164.74$ ($p < 0.001$)] and for pitch level [F $(5.8, 443) = 3.44$ ($p = 0.003$)]. Again, Scheffé post-hoc tests divided the Uygur listeners from the rest of the other three tone-language groups.

In order to better understand the interaction between listener group and time pressure, a further breakdown of the results is presented in Figure 9, which plots percent question responses (upper panels) and the associated reaction times (lower panels) separately for absence (left) and presence (right) of time pressure.

Crucially, without time pressure the listener groups come closer to realizing a cross-over than when performing under time pressure. Also, in the absence of time pressure, the relationship between sentence-type responses and reaction time is as predicted (and also as found in the analysis above). Again, when the stimulus is ambiguous between question and statement, the reaction time is slow, but when the categorization is (reasonably) clear, the decisions are faster. This relationship is found for all listener groups. With time pressure applied, however, all response curves are shallow or completely level, and local maxima in the reaction-time curves are largely absent.

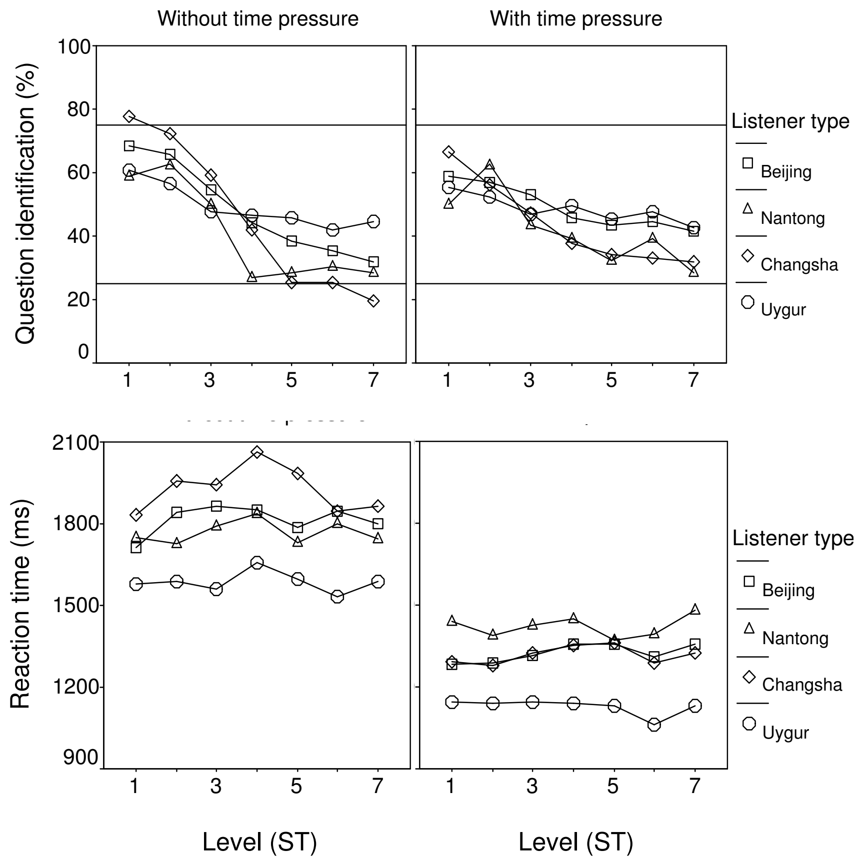


Figure 9: Question identification (percent, upper panels) and their associated reaction time (ms, lower panels) as a function of pitch level broken down by four types of listener, with and without time pressure (left and right panels, respectively).

In summary, we plotted slope values of the four groups of listeners separated by manipulation type in two panels (right: final rise; left: overall pitch level) and broken down by time pressure (yes vs. no) in Figure 10.

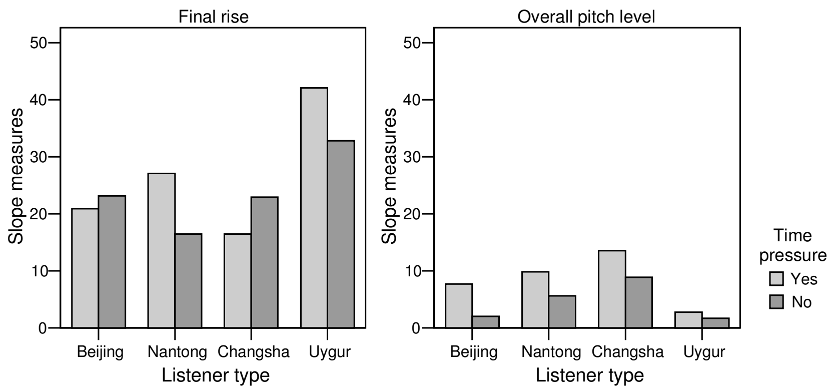


Figure 10: Slope measures as a function of listener type separated by manipulation type (final rise, right; overall pitch level, left) and broken down by time pressure (On/Yes vs. Off/No).

A three-way RM-ANOVA with listener type as a between-subject factor and manipulation type (final vs. level) and time pressure (on vs. off) as a within-subjects factors was carried out on the slope values and yielded significant effects for listener type [$F(3, 77) = 3.1$ ($p = 0.031$)], manipulation type [$F(1, 77) = 45.8$ ($p < 0.001$)] and time pressure [$F(1, 77) = 8.7$ ($p = 0.004$)]. Significance was also found for the second-order interaction of listener type × manipulation type [$F(3, 77) = 6.5$ ($p = 0.001$)] and for the third-order interaction listener type × time pressure × manipulation [$F(3, 77) = 3.1$ ($p = 0.03$)]. Scheffé post-hoc tests show that the Beijing group significantly differed from the Uygur group only ($\alpha = 0.01$).

From Figure 10, we can see that manipulating the final rise has a much stronger effect on the perception of sentence type than manipulation of overall pitch level, regardless of time pressure. Among the four groups, the Uygurs were most sensitive to the final rise and least to overall pitch level, with or without time pressure. The tone-L2 listeners are more sensitive to pitch level as a question cue than the native Beijing listeners, both with and without time pressure.

3.3. Comparison of identification in lexical and sentence intonation

In the preceding sections we presented the results for perception of lexical tone and of sentence intonation separately. We will now attempt to evaluate the differences between Beijing L1 and L2 listeners by taking into account the results from both types of test simultaneously. As we explained before, the two perception tests cannot be compared directly due to the fact that one (lexical tone) can be evaluated in terms of correct or incorrect, but the other (intonation) has no pre-given norm. This difference should be borne in mind when considering the following results.

In Figure 11 we have plotted the four listener groups performing with and without time pressure in a two-dimensional plane defined by their sensitivity to the final rise as a question cue (vertically) and the detectability of the lexical tones (d'-values, horizontally).

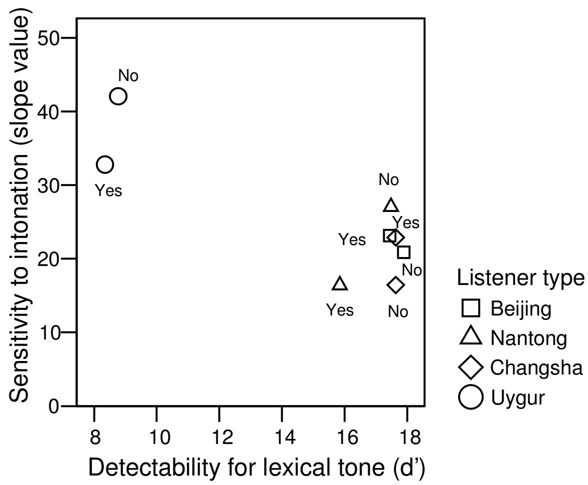


Figure 11: Sensitivity to final rise as cue for question intonation (slope coefficient) plotted against detectability based on percent of lexical tone identification (d prime) broken down by listener type and time pressure (No: no time pressure; Yes: with time pressure).

Figure 11 shows a negative correlation between sensitivity to intonation and successful lexical-tone identification. The Uygur group with a non-tone language background has relatively poor lexical-tone identification but is highly sensitive to manipulation of the final rise as a question cue; for the tone-language listeners the situation is the reverse: these are much more successful at lexical-tone identification but rather insensitive to the final boundary pitch as a question cue at the sentence level.

4. General discussion

4.1. Summary of main findings

Before we proceed to drawing any conclusions, let us first summarize, in Table 5, what we have found in our experiments. In this table a plus or minus sign indicates that the L2 listeners displayed a significantly better or poorer task performance, respectively, than the Beijing controls, in terms of tone discrimination (d'), and speed thereof (ms) or in sensitivity to manipulation of pitch (boundary rise, overall pitch) in the identification of sentence type. Cells are empty when no significant difference could be established by the RM-ANOVA.

Table 5: Summary of findings for perception of lexical tone and sentence intonation: task performance of L1 listeners (Beijing) versus L2 listeners (Nantong, Changsha, Uygur). A plus or minus sign indicates significantly better or poorer task performance, respectively, by L2 listeners than by Beijing controls.

Stimulus type			L2 listeners re. Beijing		
			Nantong	Changsha	Uygur
Lexical tone	In carrier	Detectability (d′)			–
		Decision time			–
	Citation	Detectability (d′)			–
		Decision time			–
Sentence intonation	Final rise	Sensitivity (slope)			+
		Decision time			+
	Level	Sensitivity (slope)		+	
		Decision time			+

The tone-language listener groups correctly identify lexical tones in Beijing dialect as often as Beijing listeners themselves, and they are as insensitive to pitch manipulations at the sentence level as the Beijing group, although their task performance may be slower. Additionally, a further analysis of tone perception by the Uygurs shows a significant difference between the two conditions, in carrier sentences vs. in isolation; this difference is not manifest in Table 5, but was found earlier in Figure 3. Next, we will discuss the results concerning the perception of lexical tone and sentence intonation by Chinese L1 and L2 speakers.

4.2. Lexical tone deficit exhibited by L2 speakers

For all four groups of listeners, whether having a native language with or without lexical tone phenomena, identifying tones presented in context took considerably more time than identifying the tones in citation forms. For the non-tone Uygur listeners, however, there was also a substantial increase in identification errors for tones presented in context; this effect simply did not arise with any of the tone-language groups: context slowed them down but did not lead to more errors. The canonical shapes of the four tones of Beijing change considerably when preceded and/or followed by other tones in context. We assume that listeners with a tone-language background apply some phonological rule in order to relate the surface tone to its canonical form. Applying this rule is a more or less error-free process but it takes time. For the non-tone Uygur listeners the application of the tone-sandhi rules is not automatic and error-free, leading not only to delay but also to errors.

The finding that the listeners are able to factor out the effect of tonal context is consistent with previous studies on Chinese lexical tone identification in context (Xu 1994). For Uygur listeners, however, undoing the effect of tone sandhi is a problem (as it was shown to be for American listeners, cf. Broselow, Hurtig, and Ringen, 1987).

Typically, Uygur listeners learned the tones in citation forms first (in supervised learning) but had to acquire the tone sandhi rules later through natural exposure outside the classroom. Of course, the tone listeners simultaneously acquired the tones both in their sandhi form as well as in citation form. Even if they are L2 speakers of Beijing dialect, they could still decode tones in connected speech as well as they did in citation forms. In conclusion, we have provided evidence for lexical tone deficit in tone-language learning that is only experienced by L2 speakers with a non-tone L1 and the positive transfer of tones, both in citation forms and in context, with L2 learners with a tone-language background. At the moment, we cannot determine the main factor that governs the positive transfer exhibited by our L2 subjects, i.e. Nantong and Changsha speakers. It may be due to the exposure to radio or television in Standard Chinese (which resembles the Beijing dialect) or to their tonal experience in their mother tongue. Further investigation is needed to separate the two factors.[23]

4.3. Intonation perception by Chinese speakers

In non-tone-languages such as English or Dutch, questions typically end on a rising pitch and statements have falling terminal pitch. In Chinese, however, the difference between statement and question intonation is more complicated because of the interaction of tone and intonation. For instance, question intonation with Tone 2 (rising tone) in sentence-final position has a rising shape but it has falling pitch with Tone 4 (falling tone).

Since the lexical tonal modification in Chinese employs the same dimension (i.e. pitch) as intonation does, it tends to blur the visibility of the latter to the point that the very existence of intonation in Beijing dialect (and other Chinese dialects) has been questioned (Kratochvil, 1998).

Our findings suggest that F0 is primarily perceived at the lexical level for listeners of a tone language but at the sentence level for listeners of a non-tone language.[24] Our results are in line with Seddoh's hypothesis (2002), i.e. the linguistic typological differences between tone and non-tone languages suggest that native speakers of tone languages may be less able, relative to native speakers of intonation

---

[23] We thank Yi Xu for comments and valuable suggestions.

[24] This conclusion is speculative to some extent. It may be the case that the processing of pitch at the sentence level may still get priority if the full range of linguistic functions is tested, i.e. not just the signalling of sentence type but also the marking of focus and phrase boundaries (cf. Xu and Xu, 2005).

languages, to use intonation to signal different sentence types. Our perceptual results are in line with cross-linguistic production data from tone languages (e.g. Ewe, Chinese and Thai) and intonation languages (e.g. English, French and Russian) (cf. Gandour et al., 1995; Ross et al., 1986).

Gandour et al. (2003) investigated the identification of statement versus question intonation in Chinese with Chinese and (non-native) English listeners, using natural tokens only. The native Chinese listeners were more accurate in identifying the sentence type than the English non-native listeners, but curiously enough, the English listeners were significantly faster. Table 6 combines the findings by Gandour et al. and our own. It contrasts the performance of Chinese listeners with that of non-tone listeners (English, Uygur); the better language type is indicated by a '+'-sign, the counterpart is given a '–'.

Table 6: A comparison of the performance between Chinese and English / Uygur listeners at judging clause-type from Chinese intonation with natural speech or manipulated one.

Type of stimuli in Chinese	Task/measure	Tone-language Chinese	Non-tone language English / Uygur
Natural speech, four utterances	Accuracy	+	–
	Reaction time	–	+
Manipulated speech, one basic utterance	Sensitivity	–	+
	Reaction time	–	+

From Table 6, we see that the non-tone-language listeners are always faster than the tone-language listeners regardless of the type of stimuli (natural or manipulated), and they are more sensitive to intonation. We suggest that listeners with a non-tone language background give processing priority to sentence melody – leading to fast decision times. Speakers of a tone language have learnt to divide their resources between two competing tasks: listening for pitch cues bearing on word meanings (lexical tone) as well as monitoring pitch at the sentence level.

This would account for the longer processing times as well as the diminished sensitivity to pitch cues at the sentence level for Chinese listeners. Chinese listeners, of course, are still better at identifying the sentence melody as intended by a fellow-Chinese speaker, simply because they know the system better than foreign listeners do.

More than likely the Uygur listeners used the intonational system of the Uygur language when deciding whether the target utterance was a statement or a question. We are not aware of any published studies on the melodic marking of sentence type in Uygur but it seems safe to assume that this language uses the semi-universal association of high terminal pitch for question and low pitch for statement (see also Gussenhoven 2004).

The exact implementation of the categorical distinction, however, differs from one language to the next (Bolinger 1978; Ladd 1981), which would account for the fact that the category boundary between statement and question is not the same for our Uygur listeners as for the Beijing listeners. For all this, the finding that the cross-overs from statement to question are steeper, and the identifications are faster for the Uygur listeners, indicates that they were less bothered by the lexical tones than the tone-language listeners were. This finding concurs with recent cross-linguistic perceptual studies by Gussenhoven and Chen (2000) and Chen (2005). These studies, too, show that Chinese listeners use differences in sentence melody less effectively than listeners of non-tone languages (Dutch and Hungarian) when having to interpret nonsense stimuli (meaningless CVCVCV sequences) as either a statement or a question.

In line with this, we conclude that, once free of lexical-tone interference, listeners become more sensitive to intonation. This is tantamount to saying that lexical-tone interference is the primary factor in reducing the sensitivity to pitch cues at the sentence level in listeners with a tone-language background.

4.4. Lateralization of lexical tone and sentence intonation

Imaging studies on pitch perception of healthy speakers show that pitch processing is influenced by language (Gandour, Wong, Hsieh, Weinzapfel, Van Lancker, and Hutchins, 2000; Gandour, Wong, Lowe, Dzemidzic, Satthamnuwong, Tong, and Lurito, 2002; Gandour, Wong, Dzemidzic, Lowe, Tong, and Li, 2003; Gandour, Tong, Wong, Talavage, Dzemidzic, Xu, Li and Lowe, 2004; Hsieh, Gandour, Wong, and Hutchins, 2001). Their findings show that for Chinese listeners prosodic functions at the word level predominantly engage the left hemisphere (LH); prosodic functions at the sentence level elicit bilateral activation, with a right hemisphere (RH) preference. In contrast, listening to the same Chinese stimuli, English listeners show bilateral activation for lexical tone perception but predominantly right-sided activation for intonation.

Several hypotheses have been forwarded concerning the neurobiological mechanism underlying speech perception. The hypotheses involve either acoustics or linguistics, or both, as the source of hemispheric lateralization. Hypotheses involving an acoustically driven mechanism are called cue dependent. These claim that speech processing is subserved by neurobiological mechanisms specialized for particular aspects of the acoustic signal (e.g. the absolute size of a temporal domain) to account for the hemispheric specialization of local and global processing, irrespective of communicative or linguistic relevance as suggested by the cue-dependent hypothesis (e.g. Behrens, 1989; Poeppel, 2001).

Hypotheses focusing on linguistic mechanisms are called task dependent. These claim that unique, neural mechanisms are recruited for the speech domain as suggested by the task-dependent or domain-specific effects (e.g. Van Lancker, 1980; Ross and Mesulam, 1979).

Hybrid views engaging both linguistic and acoustic processing claim the existence of dynamic interaction between the two hemispheres, i.e., during spoken-language comprehension processes in the left and the right hemispheres are assumed to interact dynamically. There are two models especially worth mentioning in this context, (i) a comprehensive model of speech perception that is mediated primarily by RH regions for complex-sound analysis but which is lateralized to task-dependent regions in the LH when language processing is required (Gandour et al., 2004), and (ii) a dynamic dual-pathway model of auditory language comprehension that claims that segmental, lexical (i.e. tone) and syntactic information is processed in the LH, but suprasegmental sentence-level information (i.e. intonation) in the RH; this model additionally claims that the involvement of the LH increases as either the stimulus or the task is more linguistic in nature (Friederici and Alter, 2004).

First, our findings cannot be explained by the cue-dependent hypothesis, which depends crucially on the absolute size of temporal domain of the speech, e.g., the left hemisphere preferentially extracts information from short 20−50 ms temporal-integration windows while the right hemisphere homologues preferentially extract information from long 150−250 ms integration windows (Hickok and Poeppel 2000; Poeppel 2001). In our stimuli the time domain of the manipulation of sentence intonation was either the syllable (word) in the case of the final rise (150 ms) or the entire sentence (six syllables, roughly 1000 ms) in the case of overall pitch level. Nevertheless, Uygur listeners only showed lexical tone deficit but their perception did not differ from that of the tone-language groups in intonation identification, and in fact they were significantly faster – regardless of the temporal difference.[25]

Second, our findings cannot be explained by a task-dependent hypothesis, either, because the selective impairment that occurred in lexical-tone perception indicates that the areas necessary for lexical tone processing are not the same; therefore the processing of lexical tone and sentence intonation does not happen in one single module.

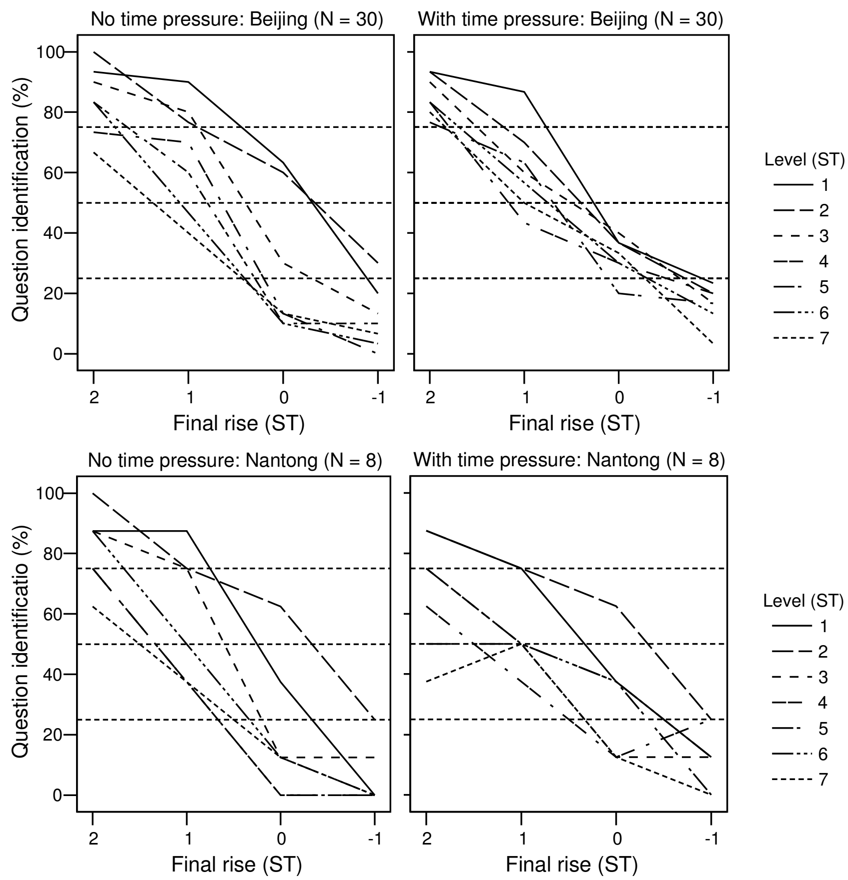
However, our findings are compatible with hybrid hypotheses claiming the existence of dynamic interaction between the two hemispheres. The differential patterns between tone-language and non-tone-language groups (i.e. greater sensitivity to lexical tone but reduced sensitivity to intonation) can be accounted for if the perception of speech prosody is mediated by different areas or in separate modules (Gandour et al., 2004).

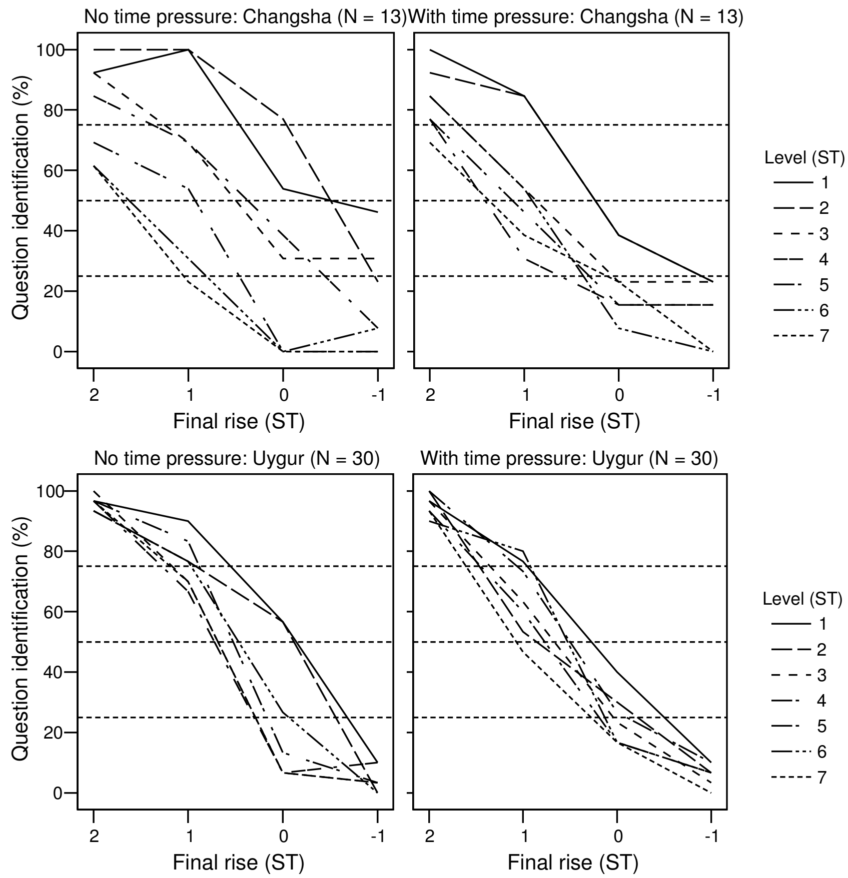
---

[25] It may be objected that the temporal domain of signalling the sentence type spans the entire duration of the utterance, since it was found, for instance for Danish (Thorsen, 1980) and Dutch (van Heuven and Haan 2000) that questions are globally characterized by a gradually rising pitch from beginning of the sentence onwards. However, we know from other research where such global pitch cues for question (vs. statement) were pitted against the local cue in the final rise, the latter cue is decisive. No combination of global cues could ever reverse the percept that was evoked by the final boundary cue (van Heuven and Haan 2002).

In conclusion, then, this study demonstrates that only L2 speakers with a non-tone L1 exhibit lexical tone impairment in learning a tone language, but the same speakers show better sensitivity to sentence intonation than L2 speakers with a tone L1. Our results, finally, support the functional view, i.e., if some phonetic dimension is exploited in one area of the grammar, e.g. lexical tones, it will not be used to the same extent in another part of the phonology, e.g. sentence intonation (Ross et al. 1986; Gandour et al., 1995; Seddoh, 2002).

Appendix: Full breakdown of perception of sentence intonation




No time pressure: Changsha (N = 13)
With time pressure: Changsha (N = 13)
No time pressure: Uygur (N = 30)
With time pressure: Uygur (N = 30)
Question identification (%)
0
20
40
60
80
100
Final rise (ST)
2
1
0
-1
Level (ST)
1
2
3
4
5
6
7

Chapter Five

Perceptual representation of lexical tones in Chinese (aphasic) listeners[26]

Abstract

Introduction. Previous research reports lexical tone deficits in the perception of speech by left-hemisphere (LH) damaged patients speaking a tone language. However, the nature of the tone deficit remains unknown, i.e., whether the impairment is caused by a structural deficit or by processing limitations. We ask by what mechanism Standard Chinese (i.e. Beijing dialect) aphasic patients with LH damage perceive their lexical tones. Do they perceive tones as if they are learners of Beijing dialect (L2 speakers) with (i) a non-tone L1 or (ii) some tone-language L1? Our assumption is that early exposure to a language has a lasting impact on speech processing routines in adults, i.e., listeners use a processing apparatus especially tuned to their mother tongue. Consequently, they have substantial difficulty in dealing with sound structures that are alien to the language they heard as infants. That is to say, L2 speakers are structurally deficient when processing phonological contrasts which do not occur in their native language.

Our hypothesis is that the degree of phonological 'deafness' appears inversely with the phonological similarity between L1 and L2. For instance, L2 speakers of Chinese with a non-tone language L1 will display more deafness to tone contrasts than learners with a tone-language L1; similarly, L2 speakers of Standard Chinese with of a closely related tone-language L1 will have less tone deafness than learners with a more distantly related tone-language L1. Crucially, however, L2 speakers do not have obvious processing limitations. We test our hypotheses by a comparative study on lexical tone identification by five groups of listeners, (i) Beijing aphasic patients, (ii) Beijing L1 speakers, (iii) Beijing L2 speakers with a closely-related tone-language L1 (Nantong), (iv) Beijing L2 speakers with a distantly-related tone-language L1 (Changsha), and (v) Beijing L2 speakers with a non-tone language L1 (Uygur). We predict that all the L2 speakers will show a structural deficit in their perception of lexical tones in Beijing dialect. The Uygur speakers will be the

[26] This chapter was submitted to *Lingua* as J. Liang and V.J. van Heuven Perceptual representation of lexical tones in Chinese (aphasic) listeners.

severest case since their native tongue is a non-tone language in the Altaic language family. Within the two tone-language groups, although both belong to Sinitic family, we predict that tone perception is more difficult for Changsha speakers than for Nantong speakers as Nantong is more closely related to Beijing dialect than Changsha dialect is. By the same criterion which differentiates L2 speakers from L1 speakers perceiving lexical tone representations in Beijing dialect, we evaluate the nature of lexical tone impairment of Beijing aphasic speakers. If the impairment is due to processing limitation, the patients should display a similar identification pattern to that of Beijing L1 speakers when there is enough time. If the impairment is caused by a structural deficit, the patients would behave like L2 speakers, showing some measure of deafness.

Methods and Procedures: Fourteen chronic aphasic subjects (Beijing dialect) with unilateral LH damage (diagnosed as Broca's aphasia) participated in an experiment on lexical tone identification, in which pitch parameters were varied in ten steps along five continua. Control subjects belonged to three tone-language groups (30 Beijing, 8 Nantong, 13 Changsha) and one Chinese non-tone language group (30 Uygur).

Predictions: We expect that the Beijing speakers will display more well-defined cross-overs and less randomness in their responses than the L2 speakers, and the L2 speakers with tone-language L1 will do better than those with non-tone L1 speakers (Uygur). Two predictions are possible for the aphasic patients: (i) if their processing ability is limited, their responses will resemble those of native Beijing listeners except that their response will be slower or will show a high degree of randomness as they are unable to access the underlying tone system; but (ii) if their perceptual categories are compromised, their responses will have less clearly defined cross-overs and more randomness, which is characteristic of L2 speakers.

Results and conclusions: A large amount of data was condensed into two parameters: (i) response entropy, representing the degree of randomness in the response behaviour, and (ii) transmitted information, representing the quality of the perceptual categories established along the tone continua. Results reveal that the patients not only exhibited less well-defined cross-overs and more randomness in their responses than L1 speakers do, but also performed more poorly than L2 speakers with tone-language L1. However, the patients were still as sensitive to pitch variation as the healthy native speakers were.

Our data have theoretical implications regarding hemispheric specialization in the perception of speech prosody, supporting the hypothesis of dynamic interactions between the two hemispheres, i.e., the perception of speech prosody is mediated primarily by RH regions for the analysis of complex non-linguistic sounds but is lateralized to task-dependent regions in the LH when language processing is required.

1. Introduction

Lexical tone. Chinese is a so-called lexical tone language. Words in such languages do not only differ in the sequence of vowels and consonants ('segments') but also by word melody ('tones'). Standard Chinese (Mandarin, Beijing dialect) has four lexical tones: Tone 1 (high level), Tone 2 (rising), Tone 3 (dipping) and Tone 4 (falling). For a fuller account of the tones we refer to Chapter 2, § 1.2.

What happens when a Chinese speaker suffers from a brain lesion (e.g. as the result of a stroke) in the left hemisphere (LH), the dominant brain half for language processing? Previous studies report that tone-language listeners with such damage typically show a deficit in their perception of the lexical tones. For instance, Naeser and Chan (1980) show that a monolingual Chinese aphasic who had suffered a LH intracerebral bleed, exhibited lexical tone impairment. Gandour and Dardarananda (1983) examined four monolingual Thai aphasic patients and found that their lexical-tone identification was significantly poorer than that of normal controls.

In a lexical-tone identification experiment (Liang and van Heuven, 2004a), Chinese aphasics were found to share characteristics with listeners who speak a non-tone language, i.e. overall tone recognition was substantially poorer and took considerably longer than that of tone-language control listeners. However, unlike the non-tone language listeners, no significant context effect was found for the aphasic listeners, i.e., they performed as accurately with contexted targets as with isolated targets. This indicates that the Chinese aphasics were still able to apply their implicit native knowledge of the language to the task of perceptually undoing language-specific effects of tone-sandhi in order to identify lexical tones.

Van Lancker (1980) proposed a scale of hemispheric specialization associated with different domains of pitch contrast with lexical tone being the "most linguistically structured" and affect and voice quality being the "least linguistically structured." When pitch contrasts are more linguistic, they are lateralized to the left hemisphere; when they are less linguistic, they are lateralized to the right. We will refer to Van Lancker's view as the functional account.

Gandour and Dardarananda (1983) asked left-hemisphere damaged (LHD) subjects to identify words that differed minimally in the five Thai lexical tones. They found that LHD patients were less competent at identifying Thai tones relative to one normal and one RHD subject. Eng, Obler and Harris (1996) found that LHD patients could not perceive tones as accurately as normal listeners. Mandarin tones were also investigated in dichotic-listening studies, e.g. by Baudoin-Chial (1986) and by Wang, Jongman, Sereno (2001), who found a right-ear advantage for tone identification only with the Mandarin-speaking subjects but not with the English control listeners. Taken together, the results indicate that Mandarin tones are predominantly processed in the left hemisphere by native Mandarin speakers.

Let us assume, therefore, that lateralization to the left hemisphere in the processing of some physical/acoustic cue depends on its linguistic function. Therefore, acoustic cues will be used less effectively in linguistic processing by patients with unilateral damage in the LH. We hypothesize, then, that Chinese aphasic speakers with LH damage will fail to perceive tones in terms of a well-defined and finite set of linguistic categories.

Hallé, Chang and Best (2004) conducted a cross-linguistic study comparing Taiwan Mandarin and French listeners on discrimination and identification tasks using three different tone continua derived from natural Mandarin utterances. In their research, the French listeners showed no increased sensitivity near category boundaries for the Mandarin Chinese tone contrasts but Mandarin listeners did. This study shows that there is an effect of tone categories on the performance of native speakers of Mandarin. In other words, the Taiwanese listeners' tone perception is rather categorical and linguistic whereas the French listeners perceive these tones as non-linguistic and psychophysical phenomena. The study further suggests that tone categories cover a large range of variation around a prototype (see also Kuhl, 1991; Polka, 1995). The crucial argument supporting categorical tone perception is the cross-linguistic difference observed between the Mandarin and French listeners. Mandarin listeners outperformed French listeners in both identification accuracy and between-category (but not within-category) discrimination. French listeners' discrimination is not influenced by tone categories and is instead determined by psychophysical factors.

To test our hypothesis that Chinese listeners with LH damage will not be able to identify the lexical tones, we ran a lexical tone identification task with Beijing aphasic speakers, Beijing healthy speakers and three groups of non-native speakers of Beijing Chinese. We aim to determine to what extent the LHD patients resemble the healthy Beijing controls or are more alike one of the non-native control groups, with native languages that differ slightly, moderately or completely from Beijing Chinese (for details see below). We predict that such a task will be easy for native Beijing listeners but difficult for listeners who speak a language which is completely different from Beijing Chinese (for instance, a non-tone language) and for Beijing aphasic listeners with unilateral damage to the LH.

So far, there are few studies on speech perception by tone-language speakers with a lexical tone deficit (but see Naeser et al., 1980; Gandour et al., 1983 above), and the nature of the lexical tone deficit remains unclear. Several models have been put forward concerning the neurobiological mechanism underlying speech perception in general. For instance, hypotheses focusing on linguistic mechanisms, often called task-dependent or domain-dependent, claim that unique, neural mechanisms located in the LH are recruited for the speech domain (e.g., Van Lancker, 1980; Ross and Mesulam, 1979; also see above). Other models engaging both linguistic and acoustic processing stress dynamic interaction between the two hemispheres, i.e., processes in the left and the right hemispheres are assumed to interact dynamically during spoken-language comprehension. Two models are especially worth mentioning in this context, (i) a comprehensive model of speech perception that is mediated primarily by RH regions for complex-sound analysis but which is lateralized to task-dependent regions in the LH when language processing is required (Gandour, Tong, Wong, Talavage, Dzemidzic, Xu, Li and Lowe, 2004), and (ii) a dynamic dual-pathway model of speech comprehension that claims that segmental, lexical (i.e. tone) and syntactic information is processed in the LH but suprasegmental sentence-level information (i.e. intonation) in the RH; this model additionally claims that the involvement of the LH increases as either the stimulus or the task assumes a more linguistic nature (Friederici and Alter, 2004).

We shall address the perceptual issue of lexical tone deficit by asking what mechanism enables Standard Chinese (i.e. Beijing dialect) aphasic patients with LH damage to perceive lexical tone. Do they perceive lexical tones as if they are non-native learners of Beijing with (i) a non-tone L1 or (ii) some tone-language L1?

1.1. Assumptions and hypotheses

Our assumption is that early exposure to a language has a lasting impact on speech processing routines in adults, i.e., listeners use a processing apparatus specially tuned to their mother tongue. Consequently, they have difficulty in dealing with sound structures that are alien to the language they learnt as infants. That is to say, non-native (henceforth L2) speakers have a structural deficit when dealing with phonological contrasts that do not occur in their native language.

L2 speakers with non-tone L1 will not have any clearly defined and well-established tonal categories. Therefore, they will show a more severe structural deficit in lexical tone perception than L2 speakers with a tone L1. Non-tone language listeners, then, will be at a disadvantage when having to identify tones. L2 speakers with a tone L1 have a clearly defined, life-long established representation of their lexical tones, which, however, will not be in a one-to-one mapping with the L1 Beijing categories. We assume that the degree of the phonological 'deafness' to L2 contrasts increases with the phonological distance between the learner's L1 and the target L2.

Our hypothesis is that, relative to Beijing L1 speakers, L2 speakers have a structural deficit in lexical tones of Beijing dialect but do not have obvious processing limitations. L2 speakers of non-tone L1 will show a more severe deafness to tonal contrasts in comparison with L2 speakers of tone L1, and similarly, L2 speakers of a closely related tone L1 will have less 'deafness' than those of a distantly related tone L1.

We set up a laboratory-controlled lexical tone identification experiment to investigate the effects of different strategies on the categorical perception of speech sound continua. We recruited tone-language patients with LH damage only (Beijing dialect speakers), and four groups of healthy controls. Controls differed in terms of native-language type (tone vs. non-tone language) and in terms of phonological similarity (closely-related tone dialect vs. distantly-related tone dialect). For L2 speakers of a non-tone L1, we chose Uygur speakers, who live in China but speak a

Turkic language belonging to the Altaic family. As for L2 speakers of a tone L1, we operationalised distance in terms of mutual phonological intelligibility between Chinese dialects. Research indicates that Changsha and Beijing dialect speakers are mutually less intelligible than Nantong and Beijing dialect speakers (61% vs. 91%, respectively, cf. Cheng, 1984). Accordingly, we recruited four groups of control speakers, (i) Beijing L1 speakers, (ii) Beijing L2 speakers with a closely-related tone-language L1 (Nantong), (iii) Beijing L2 speakers with a distantly related tone-language L1 (Changsha), and (iv) Beijing L2 speakers with a non-tone language L1 (Uygur).

We predict that all the L2 speakers will show a structural deficit in lexical tone perception in Beijing dialect but will not have any obvious processing limitation. The Uygur speakers will have the severest structural deficit in lexical tone perception as their L1 is a non-tone language that is not related to Beijing dialect. Within the two tone-language groups, although both belong to Sinitic family, we predict that the identification of Beijing tones will be more difficult for Changsha speakers than for Nantong speakers as the latter variety is more closely related to Beijing dialect than the former.

More specifically, we expect that the Beijing speakers will display better-defined cross-overs and less randomness in their responses in their perception patterns than the L2 speakers, and that the L2 speakers with a tone-language L1 (Nantong, Changsha) will do better than those with non-tone L1 (Uygur). Two predictions are possible for the aphasic patients: (i) if their processing ability is limited, their responses will resemble those of native Beijing listeners except that their decision times will be longer or they will show a high degree of randomness as they are unable to access the underlying tone system; but (ii) if their perceptual categories are compromised due to a structural deficit, their responses will have less clearly defined cross-overs and more random variation, which is characteristic of L2 speakers.

Relative to healthy controls (L1 vs. L2 speakers; L2 with tone L1 vs. L2 with tone L1 speakers), we examined the perception patterns of the aphasic listeners to test the hypothesis that Beijing aphasic patients with LH damage may be less sensitive to categorical features, as predicted by the task-dependent view (i.e. linguistics being the driving force underlying hemispheric lateralization, cf. Van Lancker, 1980; Ross et al., 1979).

1.2. Short introduction to the tone inventories of Beijing, Nantong and Changsha dialect

Before turning to our experiment, it may be useful to first present a brief description of the tone inventories of the Beijing, Nantong and Changsha dialects. Beijing has four lexical tones, described as 'high level', 'rising', 'low dipping', and 'falling', which are traditionally also referred to as Tone 1, Tone 2, Tone 3, and Tone 4, respectively. In Chinese linguistics, tones are generally described in a five-level system (Chao, 1930), in which the numbers 1 to 5 stand for the relative pitch and each single digit represents a relatively short tone, e.g. [55] is high tone and roughly twice as long as [5]. Table 1 presents the tone inventories of Beijing, Nantong and Changsha dialects.

Table 1: Tone inventories of Beijing, Changsha and Nantong dialects transcribed in terms of Chao's (1930) five-level system.

Register	**Beijing Dialect**		**Nantong Dialect**	**Changsha dialect**
High	Tone 1	55	55	55
			5	
			4	
	Tone 2	35	35	24
	Tone 4	51	42	41
Mid				33
Low	Tone 3	21(4)	21(3)	13
				11

As we can see from Table 1, the four tones in Beijing dialect are described as [55] 'high level', [35] 'rising (unidirectional), [214] 'falling-rising' (bidirectional), and [51] 'falling' (unidirectional). The Nantong and Changsha dialects have more tones than Beijing dialect, i.e. seven and six lexical tones, respectively. Besides, Changsha dialect seems different from Beijing and Nantong dialect in that Changsha dialect has more level tones in its inventory such as [55] (high level), [33] (mid level) and [11] (low level) and lacks the bidirectional dipping tone.

Prediction of interference is difficult. Nantong and Beijing are closely related, on the strength of the argument that both share the feature of a bidirectional tone as well as the absence of the mid-level tone [33]. Changsha dialect is more distantly related to Beijing, since it has level and unidirectional contour tones only, the former type including a mid-level tone [33]. However, since the tone inventory of Changsha is richer than that of Nantong dialect, it should be easier for a Changsha learner of Beijing dialect to find tones in his L1 that may serve as adequate substitutes in the L2.

1.3. Experimental design

Choice of pitch continua. Lexical tones are distinguished by fundamental frequency (F0) in terms of height, contour and direction (Gandour, 1981). The four tones of Beijing dialect are distinguished primarily by the F0 contour during the vowel (Howie, 1976). There is a long-standing debate in the linguistic literature whether contour tones should be considered single units (e.g. Abramson, 1978; Clark, 1978; Pike, 1948; Wang, 1967) or sequences of two level units (e.g., Duanmu, 1994; Gandour, 1974; Leben, 1973; Woo, 1969; Yip, 1991). For instance, Tone 2 is analyzed as a single rise in the former ('unitary') approach while it is analyzed as a low followed by a high in the latter ('compositional') approach. The compositional approach views any Chinese tone as a sequence of two sub-syllabic timing units, or 'morae', each of which functions as a tone-bearing unit. We will endorse the more recent, compositional approach, as it affords a more economical and insightful account of tone inventories cross-linguistically.

In previous literature, among the four lexical tones in Beijing dialect, it is uncontroversial to regard Tone 1 (high level) as a static tone, and Tone 2 (a rising tone) and Tone 4 (a falling tone) as contour tones. However, there has been some debate as to whether Tone 3 should be regarded as a static tone (low tone), a unidirectional contour tone (low followed by high), or even a bidirectional contour tone (high-low-high). Experiments show that perceptual tone confusion may arise from differences in fundamental frequency (F0) between tone onset and offset (Lee, Chiu and van Hasselt, 2002), indicating that tone perception is not only influenced by average pitch but also by pitch change over the course of a syllable.

Accordingly, we constructed three types of continua, varying *level* and *contour;* the latter was further divided into a *rise*, a *fall*, and a *fall-rise* series. With the *level* continuum, we intend to test how crucial the mean height is in perceptually defining Tone 1 (high) versus Tone 3 (low, at least in some accounts). With the *rise, fall* and *fall-rise* continua, we aim to determine the magnitude of the *rise, fall* or *fall-rise* needed to differentiate the contour tones, such as Tone 2 (a rising tone), Tone 4 (a falling tone), and Tone 3 (fall-fall, at least in some accounts) from the high level Tone 1.

A number of studies in child tone acquisition have revealed that the high level tone is produced earlier and is easier to identify (Cheung, 1995; Ching, 1988; Ching, 1990; Fok, 1984; Lee, Cheung, Chan and van Hasselt, 1997; Tse, 1978; Varley and So, 1995; Lee et al., 2002). Moreover, a production study with a Beijing dialect speaker with LH damage has shown that Tone 1 is far better identified than the other three tones (88%, 9%, 20% and 16% for Tone 1, 2, 3 and 4, respectively), which indicates Tone 1 is the least vulnerable tone (Liang and van Heuven, 2004b). In studies with hearing-impaired children both Fok (1984) and Ching (1988) found that (mean) pitch height was perceptually more salient than pitch change (contour) for listeners with hearing impairment. However, with normal-hearing adults Ching, Williams and van Hasselt (1994) found that pitch contour was the more salient cue. Based on these considerations, one extreme of all our continua was invariably Tone 1. Continua involving contour tones were implemented as straight lines, by varying the pitch between three pivot points, two of which were kept constant in any one continuum.

Varying time pressure. Our listeners were asked to perform lexical tone identification. This should be an easy task for healthy native listeners but will be increasingly difficult as the listener's conception of the Beijing tone system is poorer, or if the listener suffers from a processing limitation that prevent rapid access to the tonal representation in long-term memory, or both. In order to get a grip on the cause of the tonal deficit that we hope to find in some of our listener groups, we varied the identification task such that it made heavier or lighter demands on the listener's processing resources. When relatively light demands are made, we expect listeners, especially aphasic listeners, to access the stored tone representations more successfully than when the task makes heavy processing demands. However, when the tone representation in long-term memory is itself defective, varying the task demands will not affect the listener's performance. We believe that forcing the listener to access the mental representations of the tones within a shorter time interval ('time pressure') prevents the listener from fully extracting the relevant information from the auditory stimulus so that the abstract representations in long-term memory are compared with incomplete auditory traces of the input. Assuming that auditory bottom-up processing is adequate for native and non-native listeners alike, applying time pressure should not greatly affect the identification scores of healthy listeners. However, if the listener is an aphasic patient, time pressure will lead to incomplete feature extraction. Therefore, if we find a negative effect of time pressure on the identification of our aphasic listeners, we will take this as evidence that the nature of the tone deficit is a processing limitation. If, on the other hand, the aphasic listeners' task performance is poor regardless of time pressure, then we will conclude that their abstract representation of the tones in long-term memory is compromised (a structural deficit).

2. Methods

We set up a four-factor experiment with listener group, continuum, step number, and time pressure as the principal independent variables.

2.1. Listener group

Ninety-five Chinese listeners from five different groups volunteered to participate in the experiments. All of them were right-handed. The distribution of listeners over the groups is illustrated in Table 2.

Table 2: Information of the listener types

Listeners	Native language			
	Non-tone language	Tone-language		
	Uygur	Beijing dialect (Four tones)	Changsha dialect (Six tones)	Nantong dialect (Seven tones)
Aphasic		N = 14		
Healthy	N = 30	N = 30	N = 13	N = 8

The patients' non-verbal communication was still effective, and apart from their aphasia, they were able to carry out the activities of everyday life without difficulty. All of them suffered from unilateral damage in the left frontal and/or parietal lobe (detailed information presented in Table 3) and showed normal hearing sensitivities at 0.5, 1, and 2 kHz following a pure-tone air-conduction screening. None of the participants had been diagnosed with neurological or psychiatric illness prior to the experiments, apart from a single-event cerebrovascular accident (CVA) with damage in the LH. They participated in this experiment from January, 2002 to January, 2003.

The thirty healthy Beijing listeners were native speakers of Beijing dialect, aged between 21 and 70 (average 40), 17 male and 13 female. All of them had normal hearing and at least 12 years of formal education. They took part in the experiments in September 2002.

The thirty Uygur listeners were students of Xinjiang Normal University, age 18–20, who had just finished a one-year intensive training course in Standard Chinese (Beijing dialect) and intended to pass the HSK[27], which is obligatory for them if they wish to continue their higher education in China. They took part in the experiments in July, 2002.

The eight Nantong dialect listeners were construction workers, age 22–48, who had just spent their first year in Beijing and had learned Beijing dialect through natural exposure to the language. They took part in the experiments in October, 2002.

[27]The HSK is for China's Hanyu Shuiping Kaoshi 'Chinese Proficiency Test', a standardized test at the state level designed and developed by the HSK Center to measure Chinese proficiency of foreigners, overseas Chinese and students from Chinese national minorities (http://www.hsk.org.cn).

Table 3: MRI or CT scan findings of the lesion site in the left hemisphere for individual patients

Patients	Initials (across)	PYF	ZZL	XWC	CRY	HSQ	QCF	YSY	CH	ZHR	YJJ	FZJ	ZLJ	TDH	LKM
	Age	39	50	68	43	47	63	69	80	31	54	50	69	56	52
	Sex	f	m	f	m	m	f	f	f	m	m	m	m	m	m
	Time after onset (months)	23	25	4.5	5.5	5.5	3.5	2	12	2	6	12	4	35	39
Frontal lobe	Precentral gyrus lower part	●			●	●				●				●	●
	Superior f. gyrus posterior								●						
	Inferior gyrus pars triangularis	●	●			●					●		●	●	
	Broca's area pars opercularis						●	●					●		●
Parietal lobe	Postcentral gyrus lower part	●	●	●	●							●	●		
	Supramarginal anterior							●	●						
	Gyrus posterior						●								
	Posterior						●								
	Superior lobule										●				
Cingulated	Gyrus posterior			●					●						

The thirteen Changsha dialect listeners were first-year students at Hunan University, age 19–21, who also intended to pass the HSK, which is required for a teacher certification in China. They had learned Standard Chinese either through formal training or self-study. They participated in the experiments in June 2003.[28]

2.2. Stimuli

Four words, /ma1 ma2 ma3 ma4/, 'mother, hemp, horse, scold', respectively, were recorded by a male native speaker of Beijing dialect, experienced in sound recording, in a sound-insulated booth in the Phonetics Laboratory at Leiden University, on digital audio tape (DAT) using a Sennheiser MKH-416 unidirectional microphone. The recording was then transferred to computer memory, and downsampled to 16 kHz (16 bit amplitude resolution) using the Praat speech processing package (Boersma and Weenink, 1996). Using the Praat software the pitch contour of the word /ma1/ was stylized with three points (onset, temporal mid point and offset) with straight lines in a log-frequency (semitone, ST) by linear time representation. For the level continuum, the pitch level was decremented by nine 2-ST decrements (yielding a 10-step continuum), which adequately covered the range

[28] Here the authors thank Prof. Ning Chunyan (in Changsha), Prof. Li Bing (in Changsha), Prof. Guo Weidong (in Ulumqi), Miss Li Xianghui (in Tianjin), and Mr. Li Gang (in Beijing) for their kind help in recruiting subjects. We also thank all our subjects for participating in the experiments.

we established in the production of Tone 1 (high level) and the low pivot point of Tone 3 (low) tokens by our speaker (see Level panel in Figure 1).

For contour tones, i.e. rise or fall continua, two ten-step continua were generated by (i) incrementing the terminal pitch of the Tone-1 token in nine steps of 2 ST and (ii) decrementing the terminal pitch by nine 1-ST steps, while keeping the pitches at the onset and mid points unchanged (see *Late Rise* and *Late Fall* panels in Figure 1).

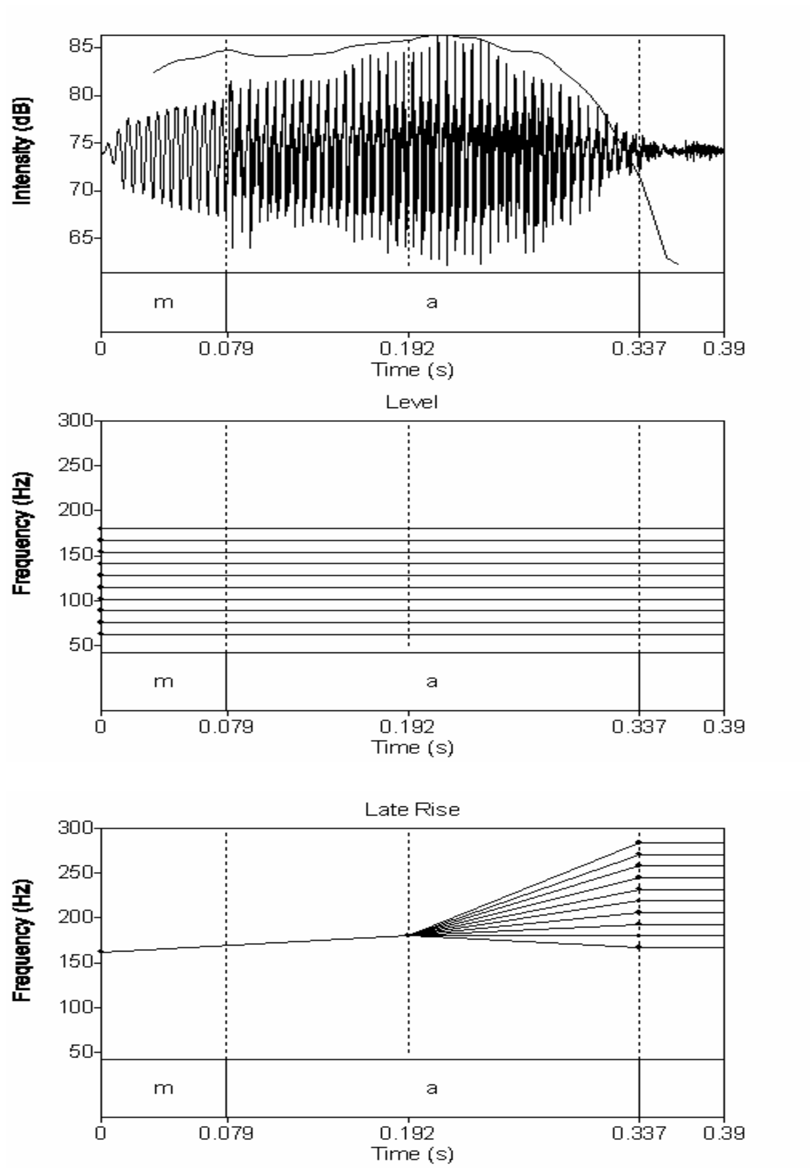
The *Early-rise* continuum was generated by incrementing the onset pitch of the Tone-1 token in nine steps of 2 ST while keeping the pitches at the mid and end points constant (see *Early Rise* in Figure 1). The *Mid-Low* continuum was generated by decrementing the pitch at the mid point of the Tone-1 token in nine steps of 2 ST while keeping the pitches at the onset and end points constant (see the Mid-Low panel in Figure 1).

As our contour tones started from the second half of the syllable, i.e. from 192 to 337 ms, the first half of the syllable (0 – 192 ms) still carried Tone-1 information. Therefore, two types of tonal feature might compete in tone identification, i.e. *high* (first part) vs. *rise* or *fall* (second part). Since the portion of the vowel that is located in the second half of the syllable is both longer and louder than that in the first half of the syllable, we expect the pitch information in the second half of the syllable to carry more perceptual weight.

In all there were five different stimulus continua, each comprising 10 steps (but sharing one type between all continua) yielding 50 stimuli (46 types). These were generated from one word /ma^1/ 'mother' (Tone 1) through PSOLA resynthesis (e.g., Moulines and Verhelst, 1995). By manipulating only the frequency values but leaving the time coordinates unaltered, the 46 different pitch patterns were generated according to the following schema (see Table 4). The five continua were named *Level*, *Late rise*, *Late fall*, *Early rise* and *Mid low* (see Figure 1).

Table 4: Five continua generated by manipulating pivot points in one stylized natural token of Tone 1. Each continuum spans 10 steps; stepsize in semitones (ST) indicated in parentheses.

Stimuli		Pitch values at three manipulation points		
		Initial	Middle	Final
Original	/ma^1/	162 Hz	180 Hz	170 Hz
Continuum	Level	Down (2 ST)		
	Late rise			Up (2 ST)
	Late fall			Down (1 ST)
	Early rise	Down (1 ST)		
	Mid low		Down (2 ST)	


Intensity (dB)
85
80
75
70
65
m
a
0
0.079
0.192
0.337
0.39
Time (s)
Level
Frequency (Hz)
300
250
200
150
100
50
m
a
0
0.079
0.192
0.337
0.39
Time (s)
Late Rise
Frequency (Hz)
300
250
200
150
100
50
m
a
0
0.079
0.192
0.337
0.39
Time (s)

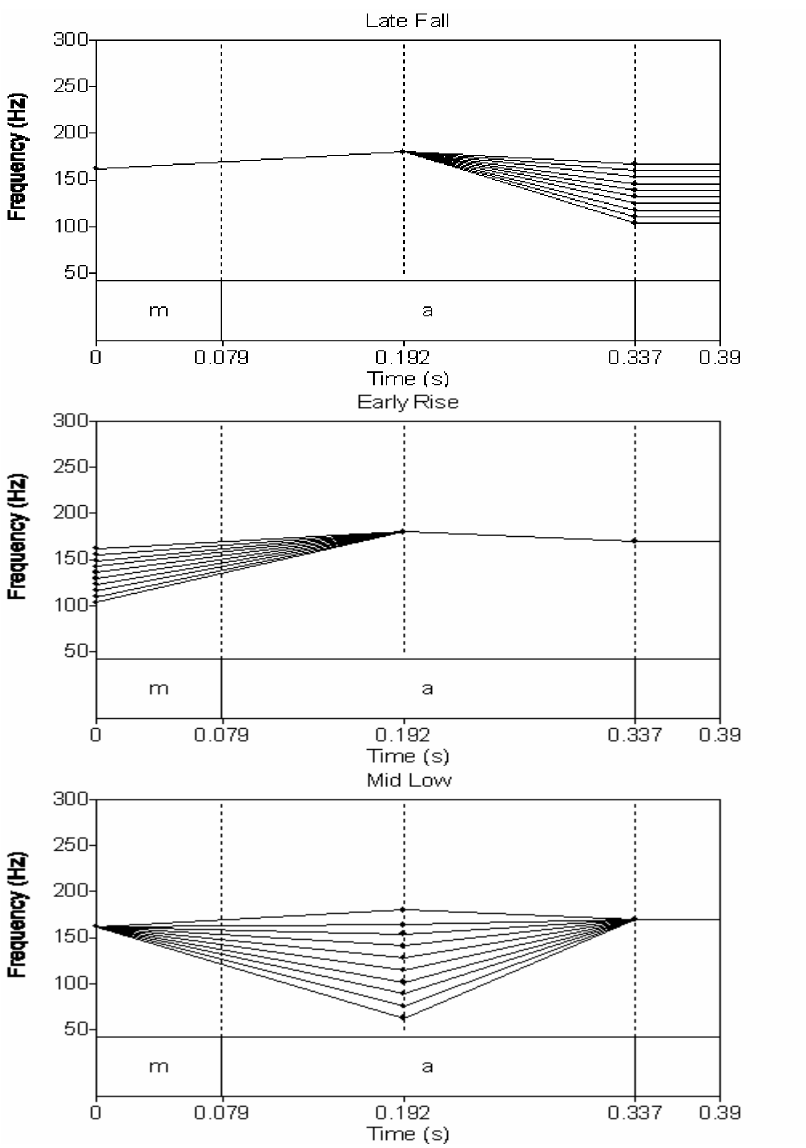


Figure 1: Schema of the experimental design: original sound and steps along the level, late rise, late-fall, early-rise and mid-low continua (from top to bottom). Note that step1 in the level continuum does not feature the slight rise-fall pattern that it has in the contour continua. In the level continuum the natural token of tone 1 was stylized by a straight horizontal line though the high pivot at the temporal midpoint of the syllable.

2.3. Procedure and task

The stimuli were randomized and presented by computer binaurally over headphones (Sony MDR-V3) at a comfortable listening level. Listeners were tested one at a time in a quiet room, in the case of patients often in the subject's home. A specially designed keyboard was developed with four buttons marked with the corresponding Chinese characters, /ma^1/, /ma^2/, /m^3/ and /ma^4/, 'mother', 'hemp', 'horse' and 'scold' for tone identification.[29] The four buttons, were evenly spaced across the top row of the response box; a black key for start/ continue was located in the center of the bottom row. Before the experiment specific instructions were presented on the computer screen and explained orally as well.

The 50 lexical-tone stimuli were presented to the listeners twice in two blocks (once without time pressure and a second time with time pressure) in order to locate the boundary for category judgments. Stimulus presentation and response collection were controlled by E-prime software.[30]

Listeners decided which of four words they had heard by pressing one of the four buttons on the response box each time they heard a stimulus. They were asked just to avoid errors in the first stage of the experiment (no time pressure). In the second stage the subjects were instructed not only to avoid errors but also to perform the task as quickly as they could manage (time pressure).

The experimental task was preceded by a short practice session with four practice trials and lasted about 20 minutes. Decisions made and reaction times (with a precision of 1 ms measured from the offset of the stimulus) were stored in computer memory. When there was no time pressure, there was a fixed 3000-ms inter-stimulus interval (ISI) after the offset of the stimulus, irrespective of the reaction time. If a subject did not respond within the ISI, s/he timed out, and the next stimulus was presented. In the sections with time pressure imposed on the listener, the next stimulus started 1000 ms after the response. The (seemingly) shorter ISI in the time-pressure condition prompted the subjects to speed up their reaction time. In the debriefing after the experiment, subjects confirmed that they had felt pressured; the effects of pressure are also apparent from the results: not only were reaction times much faster, but also the number of time-out responses was higher (even though the ISI had not been reduced).

[29] The keyboard was designed and built by J.J.A. Pacilly at the Universiteit Leiden Phonetics Laboratory.

[30] The E-prime script for stimulus presentation and response collection was written by J.J.A.Pacilly.

3. Results

In this section, we report the results on lexical tone identification obtained with the five groups of listeners. As all the stimuli were derived, through parametric manipulation, from a Tone-1 token, there is no pre-given norm here as to which stimulus type should be identified as Tone 1, Tone 2, Tone 3 or Tone 4 except that tone identification in each continuum type should start with Tone-1 identification well above chance level. Therefore, in our data analysis we will just analyze the percentage of 'Tone1', 'Tone 2', 'Tone 3' or 'Tone 4' responses and the decision latencies for each; we expect longer latencies as the choice between the response alternatives is more evenly balanced (which would be a sign of ambiguity in the stimulus).

3.1. Overall analysis

The 50 stimuli from five types of continuum with ten steps in each (yielding five identical high level stimuli) were presented in two sections, i.e. without vs. with time pressure, which resulted in 100 stimuli for each of 95 listeners. The total number of valid responses was 9,355, i.e. 98.5% of the total 9,500 with 145 missing cases.

The distribution of lexical tone responses as a function of listener type is presented in Figures 2 (no time pressure) and 3 (time pressure). In each of the two graphs, distribution of overall responses and those from each of the five continua (*Level, Late rise, Late fall, Early rise* and *Mid low*) is presented in separate panels.

Panel A in Figure 2 shows that all the five groups of listeners are very close to each other in Tone 4 identification (between 11 and 14%), but differ greatly in Tone-1 and Tone-3 responses. The comparatively high percentage for Tone 1 can be explained by the fact that all the five continuum types started from Tone 1, thereby overrepresenting Tone 1 in the stimulus set. The Uygurs' response distribution was closest to chance level, i.e. 25% for each tone, with $\chi^2 = 14.4$ ($p = 0.002$); the aphasic's responses was the second-most even distribution ($\chi^2 = 23.3$, $p < 0.001$). The tone-language groups had much more uneven response distributions with χ^2-values between 44.4 and 56.2 ($p < 0.001$).

In Figure 2, from Panel B to Panel F, each of the four tones was predominantly identified in one of the continua, Tone 1 in *Level,* Tone 2 in *Late rise*, Tone 4 in *Late fall*, Tone 1 in *Early rise*, and in *Mid low* Tone 3 for the Beijing, aphasic and Uygur groups but Tone 2 for the Nantong and Changsha groups. Again, in each of the five continua, the Uygurs' distribution of tone responses was most evenly spread, followed by that of the aphasic group.

Figure 3 shows that the response distributions obtained under time pressure were similar to those without time pressure. In Panel A, all the five groups of listeners are very close to each other in Tone-4 identification (between 13 and 18%), but differ greatly in Tone-1 and Tone-3 identification. The Uygurs' identification of the four tones was closest to chance level, with $\chi^2 = 15.4$ ($p < 0.001$) and the aphasic ($\chi^2 = 13$, $p < 0.001$) was next to the Uygur group. The tone-language groups had much more uneven response distributions with χ^2-values between 24.8 and 50.8 ($p < 0.001$). The same predominant tone response is observed in Panels B to F, and again in each continuum the Uygur listeners' tone responses is closest to chance level with the aphasics second only to the Ugyur group.

In order to determine whether listener group, continuum, and step number had any effect on tone identification, we carried out four-way RM-ANOVAs with listener group (five groups) as a between-subject factor, with continuum (Level, Late rise, Late fall, Early rise and Mid low), step number (ten steps) and time pressure (no, yes) as within-subject factors. The RM-ANOVAs and Scheffé post-hoc tests were run separately for percent 'Tone 1', 'Tone 2', 'Tone 3' and 'Tone 4' responses as the dependent variable. The results of the analyses and the Scheffé post-hoc tests are presented in Table 5 and Table 6.

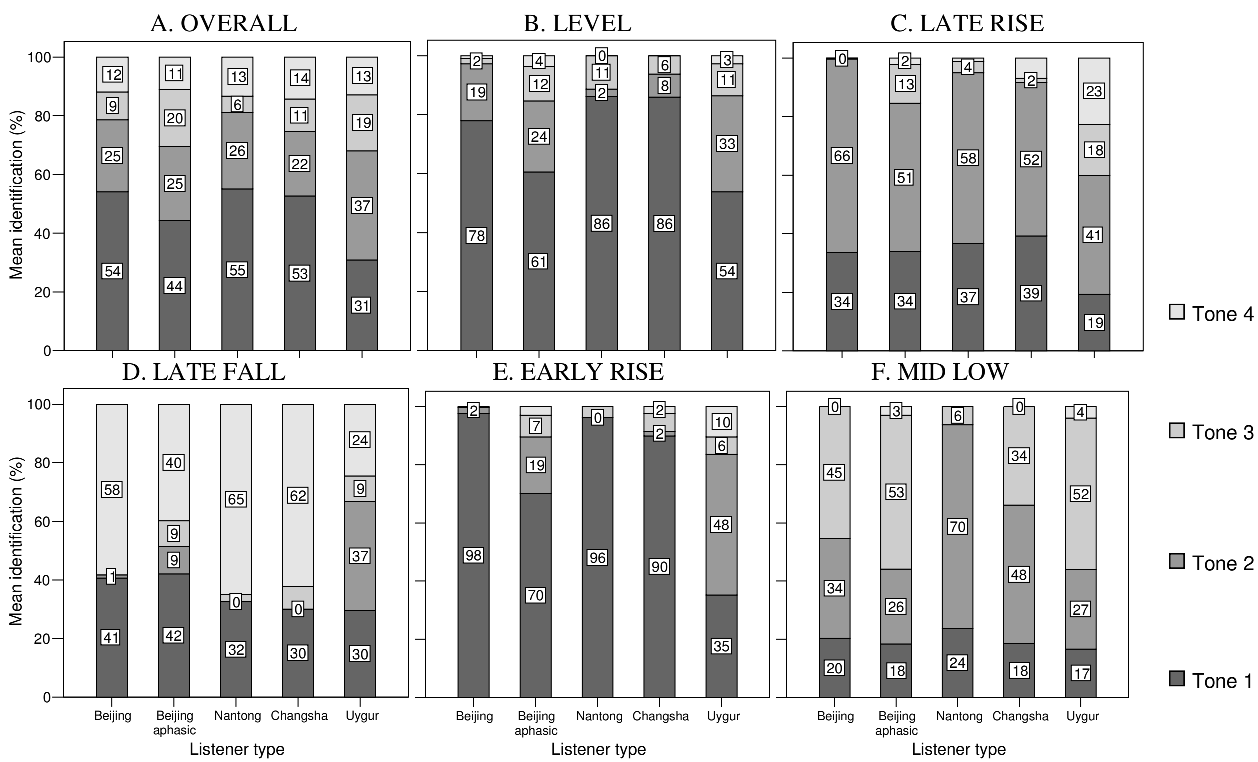


Figure 2: Lexical tone identification as a function of listener type separated by Over-all, Level, Late rise, Late fall, Early rise and Mid low without time pressure.

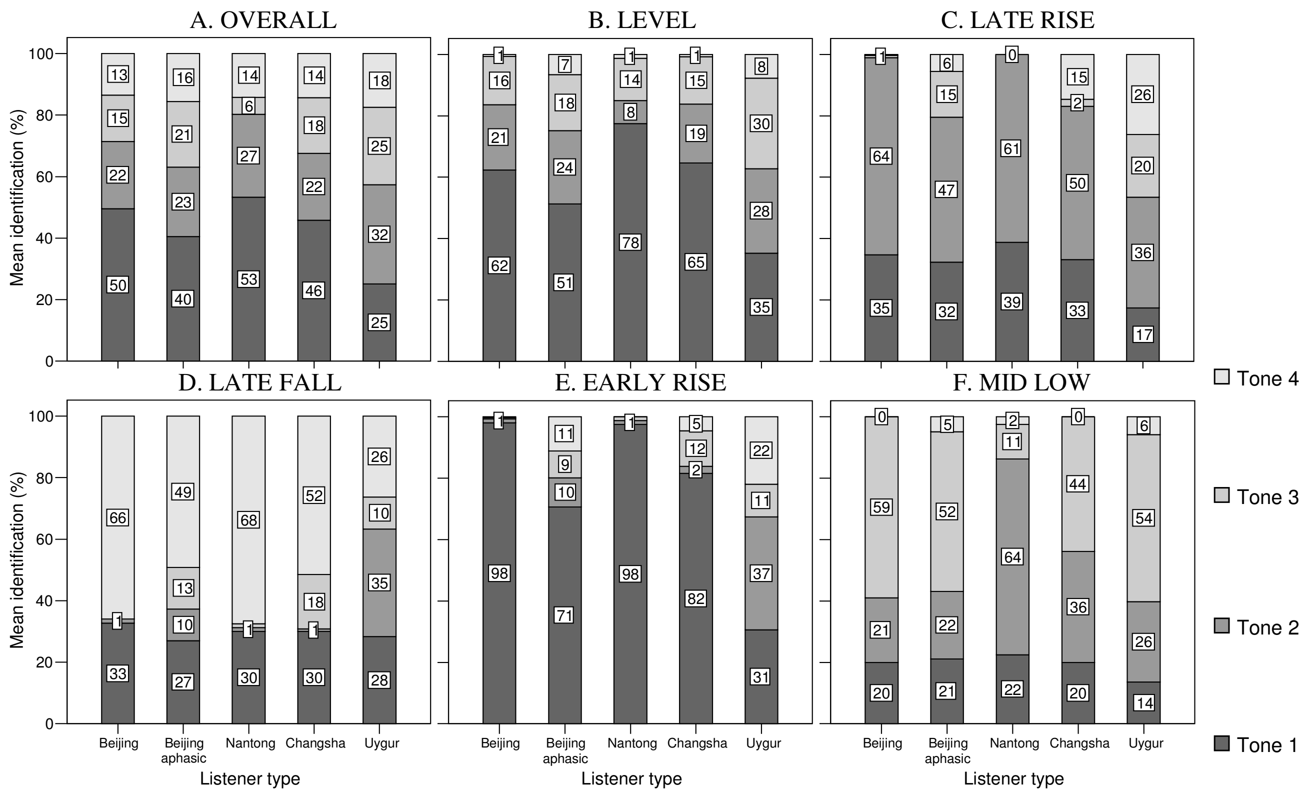


Figure 3: Lexical tone identification as a function of listener type separated by Over-all, Level, Late rise, Late fall, Early rise and Mid low with time pressure.

Table 5: Summary of RM-ANOVA with listener type as a between-subject factor and continuum, step number and time pressure as within-subject factors on the percent Tone-1, Tone-2, Tone-3 and Tone-4 responses.

Source	**Tone 1**				**Tone 2**				**Tone 3**				**Tone 4**			
	df1	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**
Listener type (L)	4.0	90	32.9	0.000	4.0	90	22.2	0.000	4.0	90	14.8	0.000	4.0	90	1.1	0.354
Continuum (C)	3.6	320	266.0	0.000	3.7	328	138.3	0.000	2.4	214	106.2	0.000	2.0	180	247.6	0.000
Step (S)	6.9	622	269.5	0.000	9.0	810	54.4	0.000	7.3	654	45.0	0.000	8.7	778	36.9	0.000
Time pressure (T)	1.0	90	16.7	0.000	1.0	90	8.7	0.004	1.0	90	27.6	0.000	1.0	90	6.4	0.013
C * L	14.2	320	17.7	0.000	14.6	328	25.8	0.000	9.5	214	6.0	0.000	8.0	180	19.6	0.000
S * L	27.7	622	13.6	0.000	36.0	810	8.4	0.000	29.1	654	3.3	0.000	34.6	778	4.4	0.000
T * L	4.0	90	0.5	0.728	4.0	90	2.3	0.069	4.0	90	3.4	0.012	4.0	90	1.2	0.301
C * S	27.5	2471	32.1	0.000	26.2	2354	28.7	0.000	22.9	2057	26.8	0.000	22.9	2057	36.8	0.000
T * C	3.7	335	10.0	0.000	3.9	353	3.4	0.010	3.3	293	6.0	0.000	2.8	255	0.6	0.585
T * S	8.3	746	1.6	0.122	8.9	800	1.7	0.097	8.6	776	1.8	0.071	8.6	774	2.5	0.009
C * S * L	109.8	2471	3.0	0.000	15.7	353	1.4	0.146	13.0	293	2.5	0.003	11.3	255	1.9	0.038
T * C * L	14.9	335	1.1	0.396	35.6	800	1.4	0.088	34.5	776	1.3	0.111	34.4	774	0.7	0.915
T * S * L	33.1	746	1.5	0.047	104.6	2354	6.0	0.000	91.4	2057	3.1	0.000	91.4	2057	4.3	0.000
T * C * S	29.7	2669	2.9	0.000	30.0	2702	1.6	0.022	24.9	2243	2.0	0.002	22.1	1990	3.6	0.000
T * C * S * L	118.6	2669	1.2	0.045	120.1	2702	1.6	0.000	99.7	2243	1.8	0.000	88.4	1990	1.1	0.177

Table 6: Summary of Scheffé post-hoc tests on listener type

Listener type	N	Tone 1			Tone 2			Tone 3				Tone 4
		Subset			Subset			Subset				Subset
		1	2	3	1	2	3	1	2	3	4	1
Uygur	30	27.6					34.2				21.8	14.9
B.aphasic	14		39.9		22.6					18.6	18.6	12.0
Changsha	13		49.2	49.2	21.8				14.6	14.6		14.3
Nantong	8			54.0		26.5		5.5				13.8
Beijing	30			51.8	23.2	23.2			12.3			12.7

Table 5 shows that continuum type is the strongest effect by far, for each the four tones. Smaller but still sizable effects are observed for listener group and step number. Smallest effects are found for time pressure. Also, all second-order interactions reach significance except the interactions between time pressure and listener type, and between time pressure and step. As for the third-order interactions, there are significant effects for the interactions, viz. continuum by step by listener type on Tone-1 and Tone-3 identification, time pressure by step by listener type on Tone-2, Tone-3 and Tone-4 identification, and time pressure by continuum by listener type on Tone-1, Tone-3 and Tone-4. The four-order interaction, i.e. time pressure by continuum by step by listener type, is significant on Tone-2 and Tone-3 identification.

Table 6 shows that aphasics are grouped next to or together with the non-tone language group (Uygur) on Tone-1 and Tone-3 identification, while they are grouped together with the tone-language groups on Tone-2 identification except Nantong group, and did not differ significantly from any of the other groups in Tone-4 identification.

Three questions arise: (1) How effective was the continuum, in each of the five types, for tone identification? (2) How did the aphasic listeners perceive lexical tones in terms of pitch continua? (3) In what way did they behave like the non-tone language (Uygur) listeners? We shall address these questions by two types of analysis (within group and between group) of the tone responses from each continuum.

Table 7: Summary of RM-ANOVA with listener type as between-subject factor and continuum, step number and time pressure as within-subject factors on the associated reaction times for percent Tone-1, Tone-2, Tone-3 and Tone-4 responses.

Source	Tone 1				Tone 2				Tone 3				Tone 4			
	df1	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**	**df1**	**df2**	**F**	**Sig.**
Listener type (L)	4.0	90	6.3	0.000	4.0	90	27.8	0.000	4.0	90	17.5	0.000	4.0	90	2.3	0.063
Continuüm (C)	3.9	355	137.4	0.000	8.9	800	29.2	0.000	2.6	231	72.8	0.000	2.2	199	244.7	0.000
Step (S)	6.6	592	84.9	0.000	3.8	344	85.7	0.000	8.3	744	23.6	0.000	8.3	743	25.3	0.000
Time pressure (T)	1.0	90	134.9	0.000	1.0	90	126.1	0.000	1.0	90	2.6	0.114	1.0	90	9.0	0.003
C * L	15.8	355	5.8	0.000	15.3	344	15.7	0.000	10.3	231	5.0	0.000	8.8	199	14.2	0.000
S * L	26.3	592	3.5	0.000	35.6	800	5.2	0.000	33.1	744	2.2	0.000	33.0	743	2.5	0.000
T * L	4.0	90	0.2	0.951	4.0	90	11.6	0.000	4.0	90	2.4	0.057	4.0	90	0.6	0.694
C * S	21.7	1956	17.1	0.000	3.6	326	5.8	0.000	22.9	2058	12.5	0.000	18.0	1615	21.6	0.000
T * C	3.8	355	15.4	0.000	26.2	2356	14.7	0.000	2.9	259	6.1	0.001	2.3	204	17.9	0.000
T * S	8.3	747	5.5	0.000	8.7	784	2.8	0.003	8.9	804	1.0	0.448	8.7	779	3.5	0.000
C * S * L	15.3	355	1.0	0.454	104.7	2356	2.9	0.000	11.5	259	1.8	0.058	9.1	204	1.3	0.262
T * C * L	33.2	747	1.5	0.030	14.5	326	1.8	0.036	35.7	804	1.1	0.334	34.6	779	0.8	0.756
T * S * L	87.0	1956	1.9	0.000	34.8	784	1.4	0.083	91.5	2058	1.8	0.000	71.8	1615	2.9	0.000
T * C * S	25.8	2325	5.2	0.000	26.4	2378	2.9	0.000	23.7	2129	1.4	0.113	18.6	1674	4.2	0.000
T * C * S * L	103.3	2325	1.2	0.083	105.7	2378	1.4	0.011	94.6	2129	1.8	0.000	74.4	1674	1.5	0.003

We will present a rather detailed statistical analysis and discussion of the results obtained without time pressure (§§ 3.3.). We will not discuss the parallel data collected under time pressure. In order to keep this article within reasonable length, we will then propose summary measures that capture the degree of categoriality in the perception of the five continua and the degree of randomness in the responses (regardless of categoriality), and analyze the effect of time pressure only in terms of these summary measures (§§ 3.4). However, we will begin by analyzing the effect of time pressure on the overall reaction time with which our various types of listeners performed the identification task (§§ 3.2). This serves as a sanity check: if time pressure does not correspond with faster task completion, then the method is obviously wrong

3.2. Reaction time

We ran a similar RM-ANOVA on the associated reaction time, i.e. with listener type as a between-subject factor and with continuum, step and time pressure as within-subject factors. The results are presented in Table 7.

From Table 7, we've found significant effects for all the four factors (listener type, continuum, step and time pressure) on all the tone identification (Tone 1, Tone 2, Tone 3 and Tone 4) except in two cases where effect is neither significant for listener type on Tone-4 identification nor for time pressure on Tone-3 identification. Out of the significant effects, again continuum type is the strongest effect by far, for Tone-1, Tone-3 and Tone-4 identification. Immediately followed by is time pressure runs. Smaller but still sizable effects are observed for step number and listener type.

Also, all second-order interactions reach significance except the interactions between time pressure and listener type on Tone-1, Tone-3 and Tone-4 identification, and between time pressure and step number on Tone-3 identification. As for the third-order interactions, there are significant effects for the interactions, viz. continuum by step number by listener type on Tone-2, time pressure by continuum by listener type on Tone-1, time pressure by step number by listener type on Tone-1, Tone-3 and Tone-4, and time pressure by continuum by step number on Tone-1, Tone-2 and Tone-4 identification. The four-order interaction, i.e. time pressure by continuum by step number by listener type, is significant on Tone-2, Tone-3 and Tone-4 identification.

The results can be summarized as two main effects without any interaction, i.e. time pressure and listener type. We ran a two-way ANOVA on the subject means aggregated over steps and continua, with listener group and time pressure as fixed factors. There is a significant effect of listener group, [$F(4, 180) = 22$ ($p < 0.001$)]. Post hoc analyses show that the three tone-language groups do not differ from each other, the aphasics are slower than all other groups and the Uygur group differs from all groups except for the Nantong group (Scheffé procedure with $\alpha = 0.05$). The second large effect is due to time pressure [$F(1, 180) = 86.1$ ($p < 0.001$)]. When the listeners were instructed to give priority to speed they performed their task some 500 to 850 ms faster than when accuracy was emphasized. Although there are many small local effects figures 11 – 12 in appendix), often showing slower reaction times when the choice between two tones was ambiguous, there seems little point in analyzing these effects in detail.

Figure 4 presents mean reaction time broken down by the five listener groups and broken down further for task performance with and without time pressure.

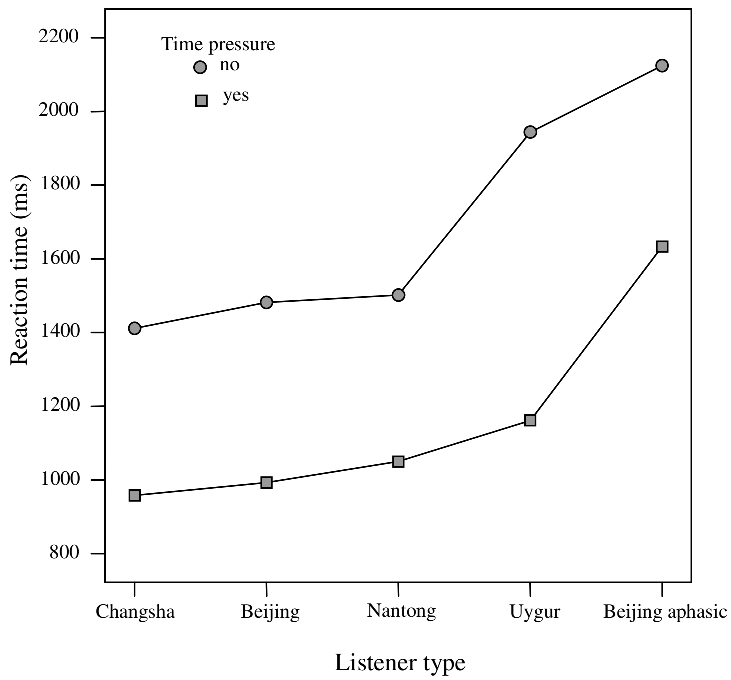


Figure 4: Reaction time (ms) for lexical tone identification for five groups of listeners with and without time pressure. Each mean is based on 10 (steps) times 5 (continua).

The results can be summarized as two main effects without any interaction. We ran a two-way ANOVA on the subject means aggregated over steps and continua, with listener group and time pressure as fixed factors. There is a significant effect of listener group, [F (4, 180) = 22 ($p < 0.001$)]. Post hoc analyses show that the three tone-language groups do not differ from each other, the aphasics are slower than all other groups and the Uygur group differs from all groups except for the Nantong group (Scheffé procedure with $\alpha = 0.05$). The second large effect is due to time pressure [F (1, 180) = 86.1 ($p < 0.001$)]. When the listeners were instructed to give priority to speed they performed their task some 500 to 850 ms faster than when accuracy was emphasized. Although there are many small local effects (figures 11 –

12 in appendix), often showing slower reaction times when the choice between two tones was ambiguous, there seems little point in analyzing these effects in detail.

In sum the reaction time results show, first of all, that our method of imposing time pressure on the subjects worked well: their response time decreased substantially when time pressure was on. Second, the results show that the task was relatively easy for listeners with a tone-language background. Non-tone language listeners (Uygur) cannot perform the task as quickly as tone language listeners can, and aphasic listeners are clearly slowest of all.

3.3. Within-group analysis of tone identification

In this section we will present some first impressions of the results based on informal inspection of the data. Within each section, we will break the presentation down by continuum, in the order *Level*, *Late rise*, *Late fall*, *Early rise* and *Mid low*. Within each continuum we will take the listener groups in a fixed order, i.e. Beijing, aphasics, Nantong, Changsha, Uygur. The results will be presented in two large tableaus, each displaying 25 graphs in a five-by-five matrix fashion. The figures are arranged in rows for each continuum, and in columns for listener type. In each panel four lines are drawn, showing the effect of the manipulation along the 10-step continuum on the dependent variable for each of the four tones separately. The first tableau (Figure 5) presents percent tone identification as the dependent variable obtained without time pressure, whilst the second tableau (Figure 6) presents those results obtained with time pressure. Their corresponding reaction times are presented in two tableaus (figures 11 – 12) in the appendix to this chapter.

Figure 5 shows that for the *Level* continuum the dominant response category is Tone 1. Yet, for the Beijing native listeners Tone 2 is an alternative when the level tone assumes a lower pitch but no cross-over is obtained, and the slope of the psychometric functions is shallow; at the lower pitch levels the percept is ambiguous between Tone 1 and Tone 2. The aphasic listeners resemble the Beijing controls, although the ambiguity between Tone 1 and Tone 2 is seen at earlier steps along the continuum. The Nantong and Changsha listeners behave very much alike. Tone 1 is the dominant response category for the whole continuum, with only marginal competition from Tone 2 and Tone 3 for the lower pitch levels. The Uygur listeners have a clearly more random distribution of responses. Only for the highest pitch values do they identify Tone 1 but the distribution is rather evenly spread over Tones 1, 2 and 3 for the lower values. Interestingly, Tone 4 is never an alternative for any listener group.

The *Late-rise* continuum yields a very clearly defined cross-over from Tone 1 to Tone 2, between step 3 and 4, for the Beijing listeners. The Nantong listeners show the same pattern but their cross-over is slightly more gradual and is shifted by one step to the right. The Changsha listeners seem identical to the Beijing, but after the cross-over between steps 3 and 4 the number of Tone-2 responses rises only reluctantly, and does not exceed the 75% criterion until step 8, yielding a large boundary width. The aphasic listeners also have a cross-over from Tone 1 to Tone 2, roughly at step 4, but their boundary is poorly defined. The non-tone Uygur listeners respond to the continuum in a more or less random fashion. No single tone ever reaches a convincing majority.

The *Late-fall* continuum basically crosses from Tone 1 to Tone 4 (from the high-level tone to the falling tone). The continuum effects a well-defined cross-over between steps 3 and 4 for the Beijing listeners. The boundary is at the same spot for the Nantong and Changsha listeners, but even better defined for the former group. The aphasics have a full albeit reluctant and poorly defined cross-over, with 75% Tone-4 responses delayed to step 9. The Uygur listeners' responses are randomly distributed among Tones 1, 2 and 4. There is not even a shimmer of a cross-over at any place along the continuum.

The *Early rise* continuum is completely ineffective for the three healthy groups of tone-language listeners, who identify Tone 1 for steps 1 through 10 (always over 75%). For the aphasics, too, Tone 1 is the preferred response but it has some competition from Tone 2 (around step 4, above step 7). The non-tone Uygur listeners come up with a random distribution of Tone-1 and Tone-2 responses.

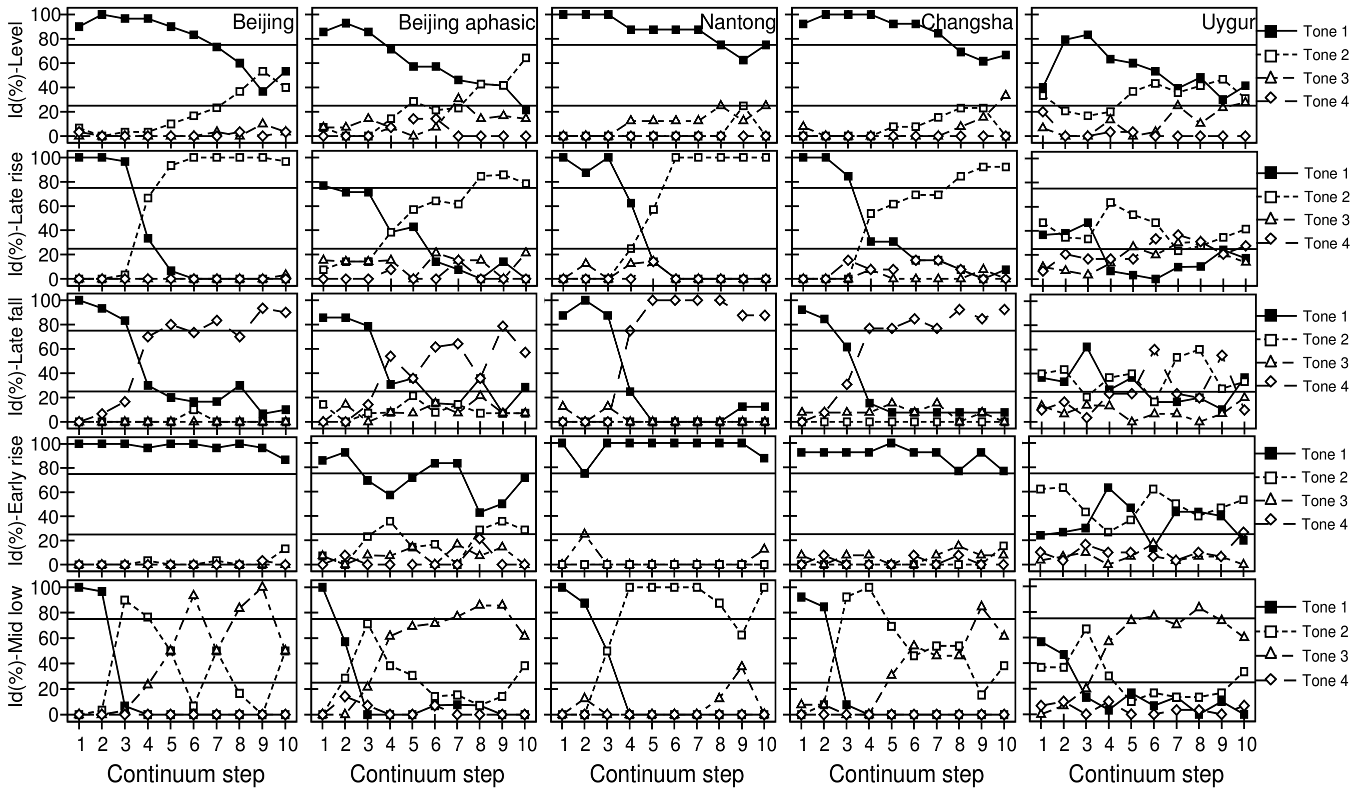


Figure 5: Percent tone identification as a function of step number for each combination of five listener groups and five continua presented without time pressure.

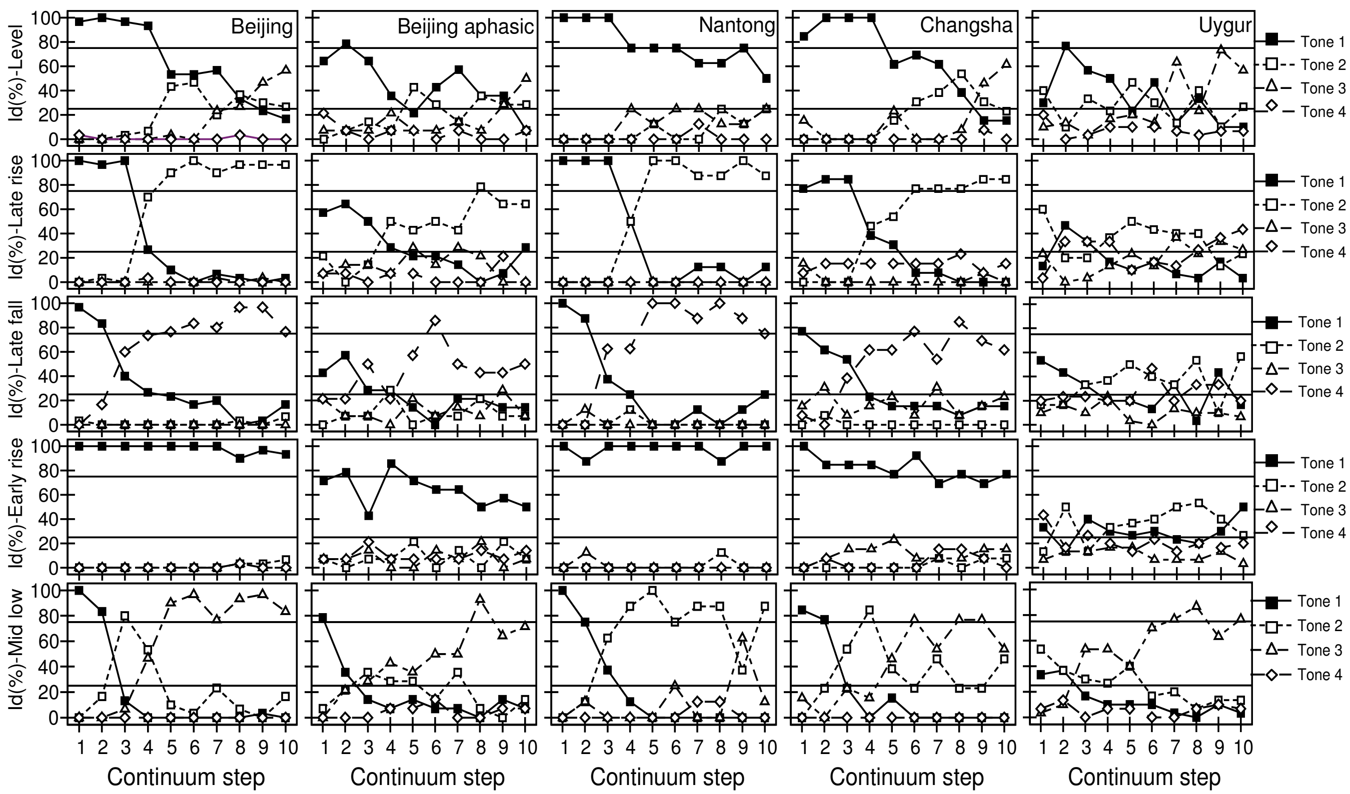


Figure 6: Percent tone identification as a function of step number for each combination of five listener groups and five continua under time pressure.

The *Mid-low* continuum seems to define a three-member contrast for the Beijing listeners. Of course, it begins with Tone-1 responses, and then crosses over to Tone 2 with a peak at step 3, which percept then yields to Tone 3 at step 6 and beyond.[31] The Nantong listeners have a very clear cross-over from Tone 1 to Tone 2 precisely at step 3, but for them no later cross-over is obtained from Tone 2 to Tone 3. The Changsha listeners have an even clearer cross-over from Tone 1 to Tone 2; however the second cross-over, from Tone 2 to Tone 3, although it is obtained very briefly, is poorly defined. The aphasics identify Tone 1 only for the first step. At step 3 there is a preponderance of Tone-2 responses (but less than 75%), which immediately shifts to Tone-3 responses. However, the cross-over to Tone-3 is very gradual and never convincing. The Uygur listeners behave like the aphasics except that they do not have a clear Tone-1 category at the beginning of the continuum.

Comparing Figure 6 (time pressure) with Figure 5 (no pressure), we found consistent patterns by all five groups of listeners in identifying lexical tones across five continua. For instance, the three tone-language groups behave very much similarly in *Level* with time pressure, i.e. Tone 1 is the dominant response category, to those without time pressure, and the Uygur and aphasic listeners show similar patterns too, i.e. more random distribution of responses without any clearly defined tones.

3.4. Cross-group analysis of tone identification

We will now make an attempt at formalizing the differences in the identification responses between the five listener groups, and at the same time evaluate the effect of time pressure on the quality of the identification. Is seems to us that the non-tone Uygur listeners are characterized overall by a more random distribution of responses, i.e. their responses are more evenly spread over the four tones.

Ambiguity between competing tones is also considerable for the aphasic listeners, while the healthy tone-language listeners generally appear to have clearly defined tonal categories along all five continua, which makes their response distributions less random. A first quantification of these variations in randomness in the response distributions would be in terms of entropy in the distribution of responses over the four alternatives (Tone 1 through Tone 4) for each step along the continuum, and then averaged over all steps. With four alternatives the theoretically maximum entropy would be 2 bits, i.e. in the case of a perfectly even distribution (25% in each tone category).

This would reflect completely random behaviour on the part of the listeners. When all the listeners in a group are in complete agreement (i.e., if they all choose the same response alternative) for each step (although the choice may vary from step to step), the entropy equals zero. So, the smaller the entropy, the greater the

[31] There is a local ambiguity between Tone 2 and Tone 3 at step 7, which we cannot explain.

determination or stability in the responses for the group of listeners. In terms of entropy we would expect the Uygurs to be located near the noisy extreme and the healthy Beijing listeners (who should have the clearest perceptual norms for the four tones of their language) near the deterministic end.

However, we need a second parameter in order to describe the responses to our continua, such that this parameter captures the sensitivity of the listeners to a change in the continuum. This parameter would differentiate continua with cross-overs from those that lack a changing percept. Transmitted information (in bits) would provide a good measure of this sensitivity. The clearer the division of the continuum into discrete perceptual categories (and the more categories are distinguished along the continuum), the higher the amount of transmitted information, again with a theoretical maximum of 2 bits representing the situation where all four tones are perfectly distinguished ('transmitted from stimulus to response category'). Since in our data the continua typically afford a binary choice (only the mid-low continuum suggests three tone categories, at least for native listeners), the amount of transmitted information will in practice not be in excess of 1 bit.

We have computed mean response entropy and transmitted information for each of the five continua for each of the five listener groups separately (formulae in appendix, see also Shannon and Weaver, 1949; Attneave, 1959). In Figure 7, the upper panels show the mean entropy in the responses per listener group while the lower panels present the results for transmitted information, with separate panels for absence (left) versus presence (right) of time pressure.

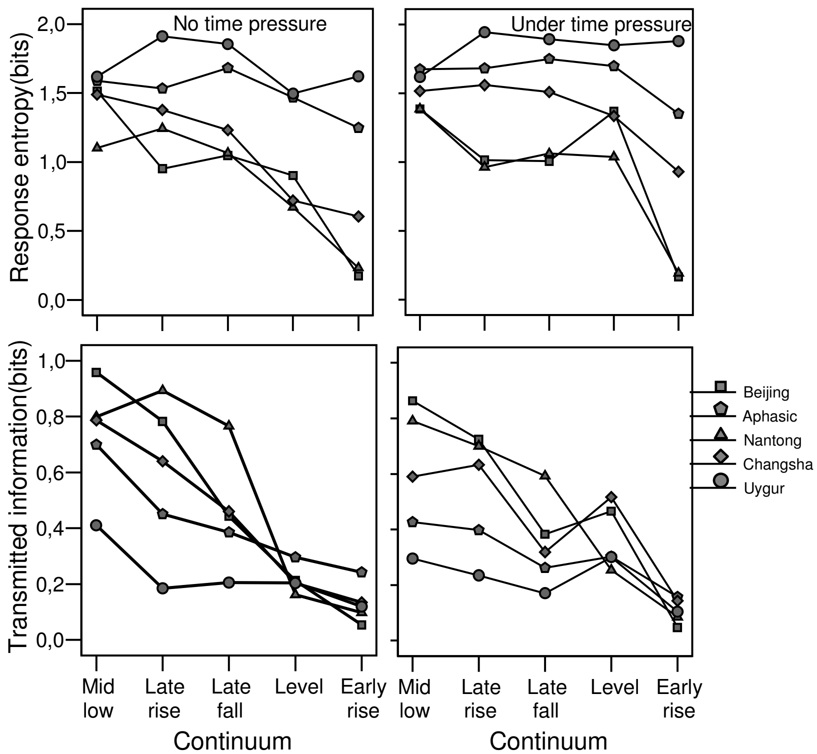


Figure 7: Entropy (upper panels) and transmitted information (lower panels) as a function of continuum broken down by listener type and separated presence versus absence of time pressure.

The upper two panels in Figure 7 show that the chaos in the responses differs both by continuum and by listener group. Without time pressure, the *Mid-low* continuum elicits the most chaotic response behaviour (across listener type), almost equal to the *Late-rise* and the *Late-fall* continua. More determination is visible in the responses to the *Level* continuum and the most invariant response behaviour is obtained for the *Early-rise* manipulations. Under time pressure, there is no effect of continuum, except for the *Early rise,* which shows more determination on the part of the listeners. It indicates that listeners traded accuracy for speed when time pressure was applied. Regardless of time pressure, the Uygur group shows the most chaotic response behaviour throughout the five continua; the three tone-language groups have the least random response behaviour and the aphasic listeners find themselves in between tone and non-tone listeners.

The lower two panels show that the transmitted information carried by the responses differs both by continuum and by listener group. Without time pressure, the order, from most to least amount of transmitted information, is *Mid low*, *Late rise*, *Late fall*, *Level* and *Early rise*. Under time pressure, we observe a reduced amount of transmitted information for all continua except the *level* continuum, which now clearly shows more transmitted information, especially for Beijing and Changsha listeners. The Uygur group transmits the least amount of information, the aphasic are next to the Uygur, and the three tone-language groups transmit the largest amount of information.

The aphasic group is quite close to the Uygur group in both dimensions, except in the *Mid-low* continuum, where the amount of transmitted information is close to that of Changsha listeners. That is to say, in four out of five continua, the aphasic listeners' responses are characterized by a low transmitted-information rate and massive ambiguity, almost as much as those of the non-tone language group. Overall, the Beijing native group is most informative and unambiguous in its response behaviour, the Uygur group is the least; the Nantong group comes close to Beijing, immediately followed by the Changsha group. The aphasic group is in between the Changsha and the Uygur groups but clearly closer to the latter.

In order to have a more comprehensive view of the differences in behaviour exhibited by the five groups of listeners, we have plotted, in Figure 8, the location of the five continua, separately for each of the five listener groups without and with time pressure in separate scatterplots with transmitted information along the Y-axis and response entropy along the X-axis.

A first inspection of the configuration of data points in Figure 8 reveals a strong negative correlation within each listener group between transmitted information and response entropy. This makes sense, of course. If a continuum effects a clear cross-over (or multiple cross-overs), the randomness in the response distribution diminishes. No (negative) correlation, however, is observed for the Uygur listeners. For this listener group all five continua are characterized at the same time by little transmitted information and by random response distributions.

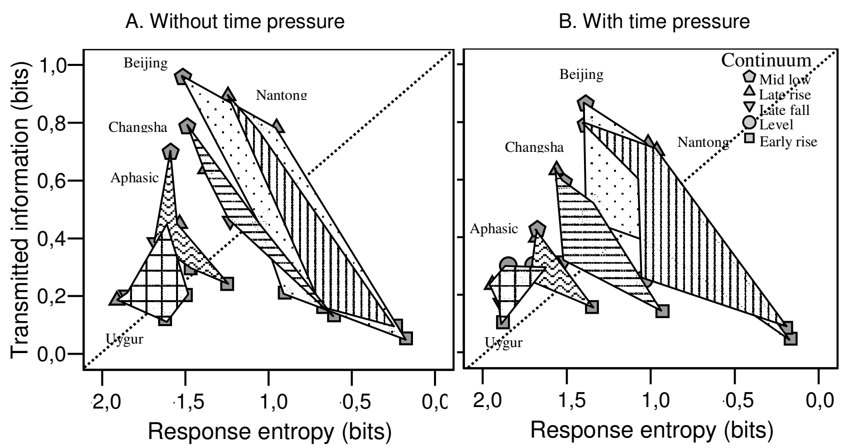


Figure 8: Transmitted information against response entropy broken down by five continua and five groups of listener type. Polygons were drawn joining the five continua per listener type without time pressure (A. left panel) and with time pressure (B. right panel).

Prima facie, it appears that the negative correlations are stronger when time pressure is absent. That is to say, the continuum either displays clear cross-overs or extreme determination to stick to just one category. This suggests that the listeners use more clearly defined categories when a longer time interval elapses between hearing the stimulus and giving the response. The (partial) correlations are generally smaller when time pressure is applied. This indicates that the responses are noisier, and therefore display less clearly defined cross-overs. Table 8 summarizes the correlation coefficients found between transmitted information and entropy within each listener group without and with time pressure.

Table 8 shows high negative correlation coefficients between transmitted information and response entropy for the three tone-language groups when no time pressure was applied, which indicate that their identification patterns contain clearly defined cross-overs (or even multiple cross-overs), i.e. a large amount of transmitted information and determination in their responses. With time pressure, the negative correlation coefficients become slightly weaker. However, regardless of time pressure, the non-tone language Ugyur listeners find themselves at the other end of the scale since there is basically no correlation between transmitted information and response entropy for this group of listeners.

There is only a weak negative correlation for the aphasic group, which ranks them intermediate between the tone-language and non-tone language groups both without and with time pressure. That is to say, there are fewer clear cross-overs (witnessed by little transmitted information) and a lot of ambiguity (high entropy values) in their responses.

Table 8: Correlation coefficients between transmitted information and response entropy without and with time pressure (: $p \leq 0.05$; **: $p \leq 0.01$).*

Listener type	Without time pressure		With time pressure	
	Correlation	p =	Correlation	p =
Beijing	−0.84	0.076	−0.81	0.097
aphasic	−0.59	0.298	−0.67	0.219
Nantong	−0.95*	0.015	−0.81	0.099
Changsha	−0.98**	0.003	−0.79	0.110
Uygur	0.18	0.769	0.51	0.380

We have seen from Figure 4 and Figure 5 that tone identification was indeed influenced by the various continua (*Level, Late rise, Late fall, Early rise,* and *Mid low*). The question *to what extent* identification is correlated with continuum step, and how the correlations differ between the continua, is interesting since such correlations can be used for computing overall similarity between the listener types.

As a classification of the five groups of listeners as a function of the five continua, we employ a projection scheme, whereby points from each of the five continua in Figure 8 were projected onto the diagonal (dotted lines in Figure 8A–B) and the results are presented in Figure 9 broken down by listener type. The diagonals capture the directionality of the clouds; projecting data points onto the diagonal maintains near-optimal distances among the points along a single generalized dimension to which each of the original two dimensions, i.e. entropy and transmitted information, contribute equally. Figure 9 plots the projection scores obtained with time pressure (horizontally) against the same obtained without time pressure.

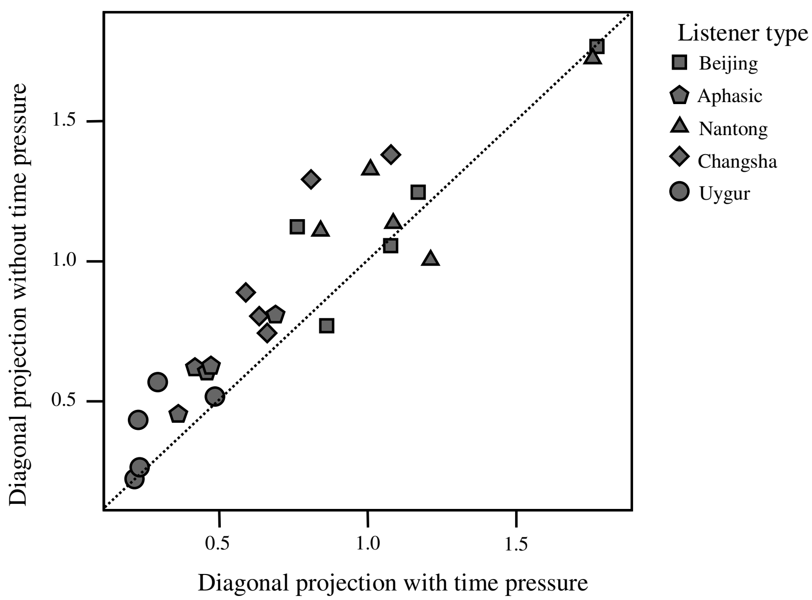


Figure 9: Diagonal projection scores for each of the five continua for each of five listener groups obtained without time pressure (vertically) plotted against the same scores with time pressure (horizontally).

We observe that regardless of time pressure, the three tone-language groups are higher on the projection scale than the non-tone language group (Uygur), the aphasic group is somewhere between them. The finding was supported by statistical analysis: a two-way ANOVA with listener type and continuum as fixed factors found a significant effect for listener type [F (4, 25) = 70 ($p < 0.001$)], continuum [F (4, 24) = 16.6 ($p < 0.001$)], and for the interaction between listener type and continuum [F (15, 25) = 3.1 ($p = 0.005$)]. No significant effect was found for time pressure. Scheffé post-hoc tests divided the listener groups into three subsets: (1) Uygur aphasic, (2) Changsha, and (3) Beijing and Nantong (Scheffé procedure with $\alpha = 0.01$). Figure 9 also reiterates that the responses obtained without time pressure transmit more information and reveal more determination than those obtained under time pressure, since most of the data points in Figure 9 are above the diagonal (which represents equal performance for both tasks).

In a last step in the data reduction, we plotted mean projection points (means over five points per listener group in figure 9) as a function of correlation between transmitted information and entropy (as shown in Table 8) for each of the five listener groups in Figure 10, separately for means obtained in the time-pressure task (filled symbols) and in the no-pressure task (open symbols).

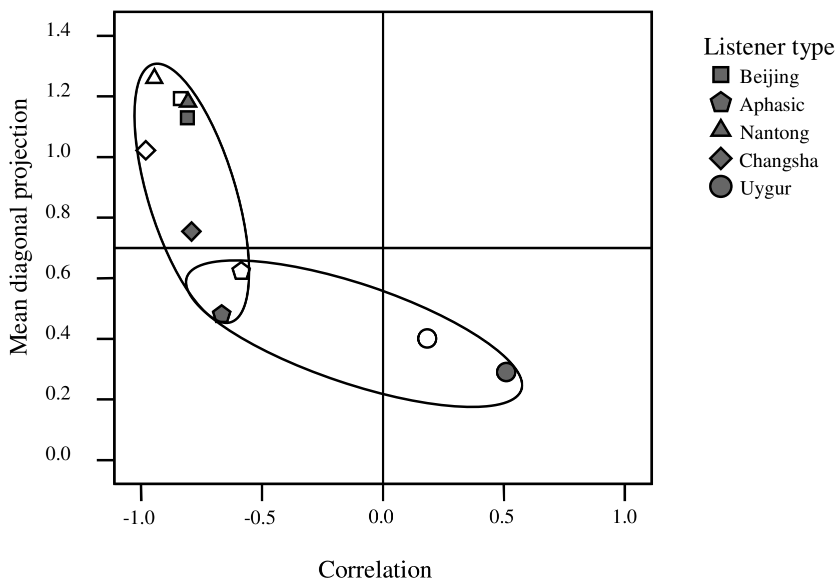


Figure 10: Mean projection points as a function of the correlations between transmitted information and entropy broken down by the five listener groups. Open symbols represents means obtained without time pressure; filled symbols were obtained with time pressure.

Figure 10 captures what is in common between the aphasic and the three tone-language groups is the negative correlation, which means that whenever there are more and better-defined cross-overs, there is less randomness in the responses. What is shared by the aphasic and Uygur groups is little transmitted information accompanied by a lot of randomness in the responses. This then reveals the nature of the aphasics' task performance relative to that of the other groups of listeners. In one respect the aphasics behave as if they are speakers of a non-tone language (random responses with few and poorly defined category boundaries) but in another respect they behave in much the same way as tone-language listeners (if there are no categories, then they respond with great determination).

4. Discussion

We asked Chinese nationals to identify stimuli, derived from a single spoken syllable through systematic manipulation of selected properties of the stylized pitch pattern, as instances of Beijing lexical tones 1 through 4. Our stimuli differed by some tone-related feature such as the height of the pitch (high ~ low) or direction of the pitch change (rise ~ fall). Aphasic Beijing listeners were found to make a larger number of ambiguous tone decisions than healthy native speakers of Beijing dialect and L2 speakers of Beijing whose native language was a tonal dialect of Chinese. The question now is, what might be the cause of the poor tone identification results for the Beijing aphasic listeners.

4.1. Lexical tone characterization: sharp vs. gradient cross-overs

Some researchers have argued that speakers of tone languages exhibit a decreased sensitivity to frequency differences within a lexical tone category (Stagrau and Downs, 1993; Tanner and Rivette, 1964). Our lexical tone identification tests show that tone-language listeners (Beijing, Nantong and Changsha) display rather gradient curves within a lexical tone category but rather sharp and well-defined cross-overs across lexical tone categories. Taking together, these findings support that idea that experience in making categorical, phonological decisions on the basis of perceived pitch may lead to the categorical perception of F0 continua, i.e., increased sensitivity to pitch differences across a tone boundary but decreased sensitivity to pitch differences within one phonological tone category.

The fact that there are no convincing cross-overs but rather gradient curves in the aphasics' response patterns can be interpreted as evidence of deficit in phoneme identification. Similar gradient curves for Tone 3 in *Mid Low* continua were present in both aphasics' and the non-tone listeners' responses but absent in the tone-language listeners. These findings suggest that the acoustic characteristics of the stimulus, rather than linguistic knowledge, may govern perceptual judgments by aphasic and non-tone language listeners. Our results, then, are in line with previous findings showing deficit in categorical perception of phonemes, i.e. deficit in both in phoneme identification and discrimination tasks (Caramazza, Berndt and Basili1983; Halleand Stevens1991; Stevens, 1991; Gow and Caplan, 1996).

To summarize, our findings show that the aphasic patients display some deviation in toneme identification in comparison with healthy Beijing listeners. Before we draw any conclusion, however, we ask ourselves whether this deviation in toneme identification fundamentally changes the mental tone representation of a native speaker.

4.2. Mental representation of lexical tones: native vs. non-native

We assumed, in our introduction, that early exposure to a language has a lasting impact on speech processing routines in adults. We predicted that L2 speakers of Beijing dialect suffer from a structural deficit in their conception of the Beijing lexical tone contrasts, and that the severity of the deficit would increase with the phonological difference between Beijing dialect and the learner's native language.

We have found in our study what we predicted, i.e., all L2 speakers show less transmitted information and more randomness in their identification of the Beijing tones than L1 speakers. Within L2 speakers, listeners belonging to the tone-language groups have better-defined tone categories than listeners with a non-tone language background. Within the tone-language groups, the closely-related tone-language (Nantong) group displayed a perception pattern which is closer to that of the Beijing group than that of the distantly-related group (Changsha). These results are consistent with the hypothesis that tonal category boundaries in the target language are at least partly determined by tone categories in the mother tongue (or by the absence of such tones).

Crucially, the response pattern of our Beijing aphasic listeners was very different from that of the healthy controls. The aphasic listeners behave very much like non-native speakers of Beijing, and in fact find themselves in between the group of L2 speakers with a tone-language background and L2 speakers with a non-tone language (Uygur). Specifically, all L2 listeners in our experiment suffer from a structural deficit in the mental representation of the lexical tones of Beijing dialect, as shown by the poor rate of transmitted information in the responses. The aphasic listeners, although being native speakers of Beijing, show the same response pattern as the non-native listeners.

4.3. Pitch processing: structural deficit vs. processing limitation

In an attempt to account for the nature of the deficit observed in Broca aphasic speech, two main approaches have been developed (see introduction). The structural deficit hypothesis postulates that the deficits are caused by damage to grammatical representations (Caplan, 1983; Hagiwara, 1995; Friedmann and Grodzinsky, 1997). In the alternative view, the deficit is caused by processing limitations (Friederici and Frazier, 1992; Kolk, 1995; Roo, Kolk, Hofstede, 2003). The claim would then be that the mental representations of the lexical tones are still intact and the observed impairment is due to processing limitation, which makes it difficult for patients to access their mental representations. We argue that our study rules out the processing account for two reasons.

First, most of the decisions made by our aphasic listeners took more than 2500 ms, which points to processing limitation. It is well-known, for instance, that a faithful and detailed mental representation of the auditory stimulus does not remain available for more than 250 ms (Crowder and Morton, 1969; Massaro, 1974). When more time elapses between hearing the stimulus and issuing the response, the listener will have to recode the auditory pattern to some more abstract representation, such as a linguistic feature or category (be it a phoneme, a word or a lexical tone). Therefore, it must have been the case that our aphasic patients had to recode the auditory pattern to some higher-order representation, thereby abstracting from phonetic details.

Second, once the auditory pattern has faded from auditory memory, the patients would have to gamble among the four tones. This would have resulted in more or less random choices among the four lexical tones. However, the results show that the aphasic patients share a confusion pattern with the tone-language groups to a great extent, which suggests that the aphasic peripheral auditory mechanisms were still functioning normally.

Taken together, these findings indicate that the lexical-tone impairment is caused by a structural deficit or the delay in reaction time is due to either a structural deficit or a processing limitation at some higher-order stage in the auditory pathway.

4.4. Possible neural mechanisms for lexical tone processing

Now let us return to the hypothesis concerning possible neural mechanisms. The reduced sensitivity to pitch differences across a tonal boundary in the Broca aphasics (lesions to the left hemisphere) seems to support the task-dependent hypothesis, i.e., the left hemisphere is recruited for speech. However, the fact that the patients are almost as sensitive as, or even more sensitive than, tone-language listeners are to the pitch differences within a tonal category suggests that the pitch processing is not restricted to the left hemisphere.

Our results, therefore, support the hypothesis engaging both linguistic and acoustic processing, which claims the existence of dynamic interactions between the two hemispheres. That is to say, during spoken-language comprehension processes in the left and the right hemispheres interact dynamically. To be more specific, our findings can be explained by a comprehensive model of speech perception that is mediated primarily by RH regions for complex-sound analysis but which is lateralized to task-dependent regions in the LH when language processing is required (Gandour et al., 2004).

Appendix

Reaction times (s) for responses 'Tone 1', 'Tone 2', 'Tone 3' and 'Tone' as a function of stimulus step for each of five continua broken down by five listener groups, obtained without time pressure (Figure 11) and with time pressure (Figure 12). These reaction times correspond with the percentages of tone identification presented in figures 5 and 6 in the main text.

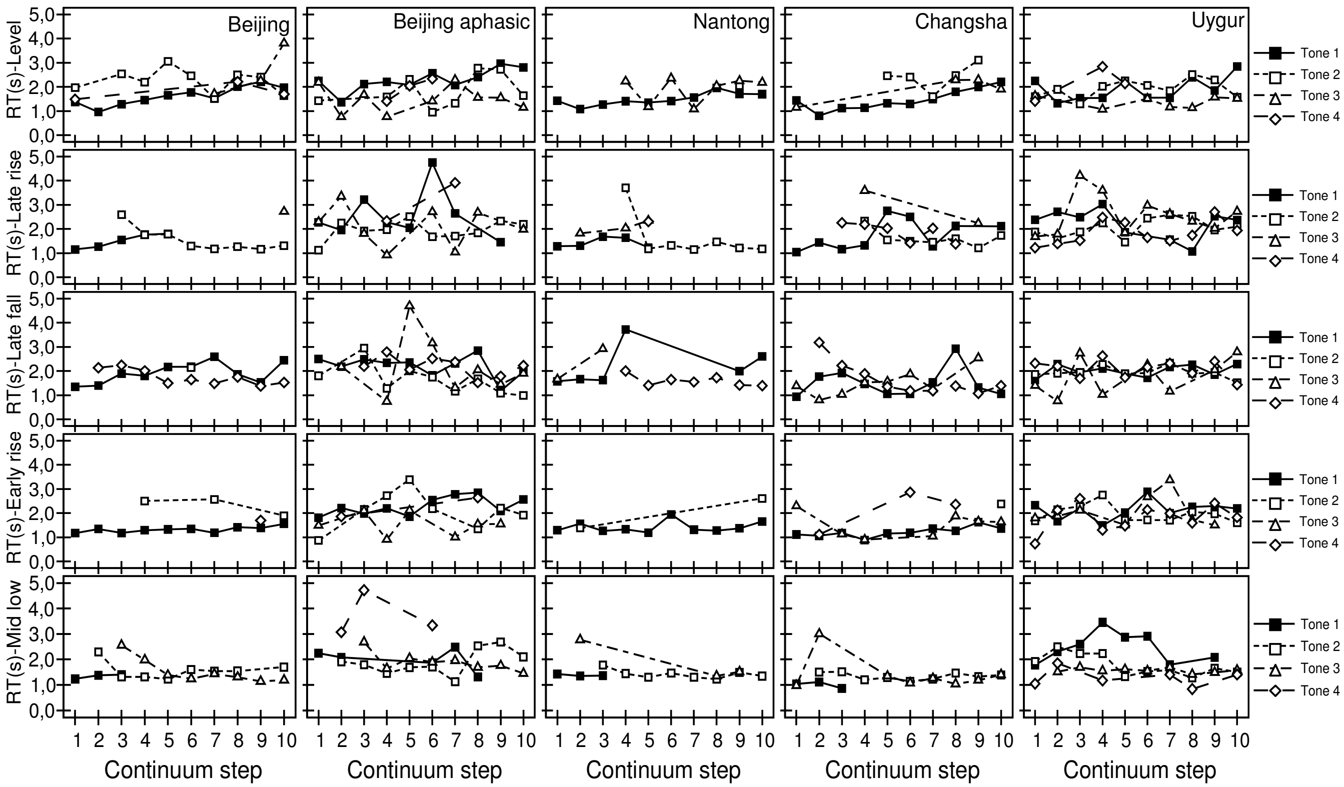


Figure 11: Reaction time (ms) as a function of step number for each combination of five listener groups and five continua presented without time pressure.

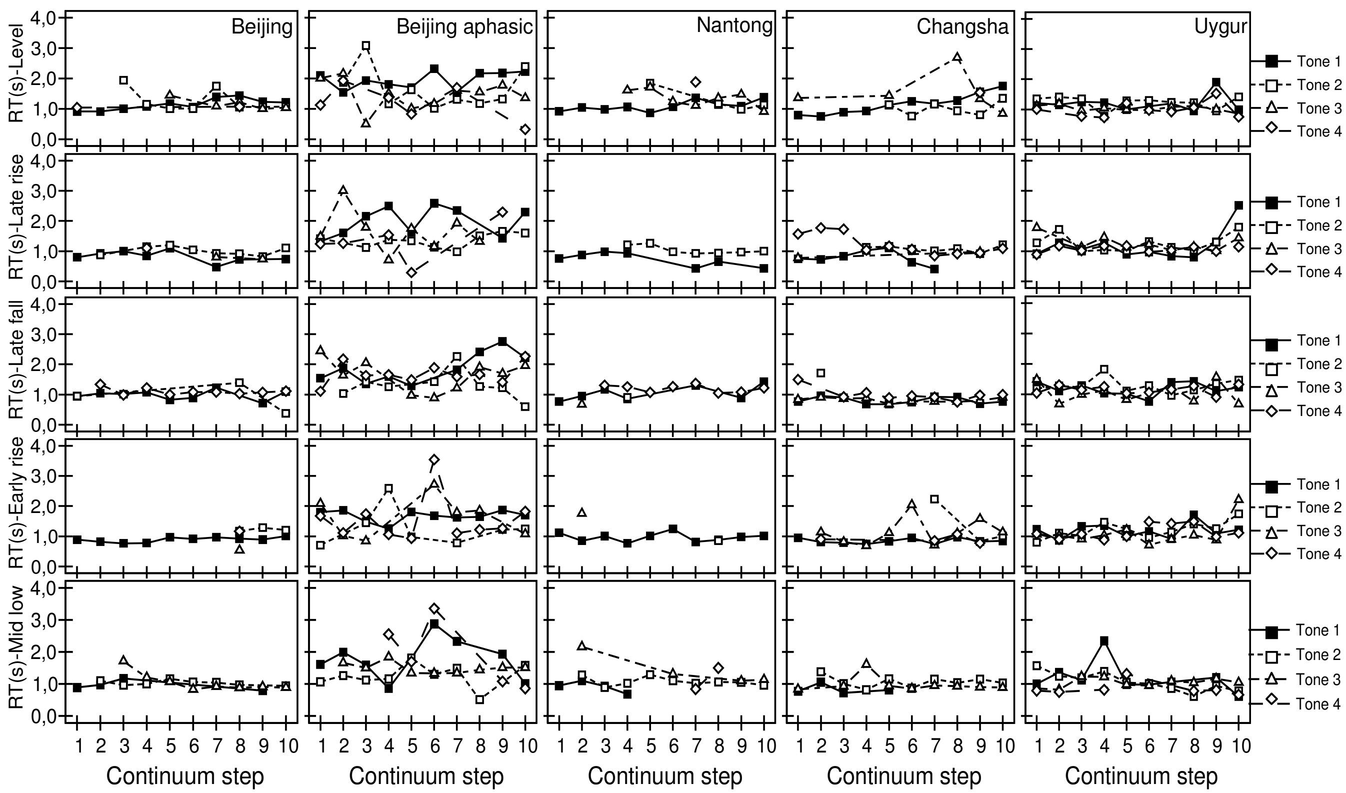


Figure 12: Reaction time (ms) as a function of step number for each combination of five listener groups and five continua under time pressure

Chapter Six

Conclusion

In the thesis, a series of laboratory-controlled experiments have been carried out to investigate the effects of brain lesion on prosodic processing in Chinese Broca's patients with respect to the suprasegmental aspects pertaining to (i) lexical pitch features, i.e. lexical tones in Chinese; (ii) post-lexical pitch features, i.e. sentence intonation. The basic overall question, i.e., what is the place of word-prosody (lexical tone) in the mental architecture, is it separate from segmental structure and is it separate from sentence-prosody, has been addressed throughout the thesis.

In Chapter One, we have raised five basic issues, which have been addressed one by one from Chapters Two to Chapter Five as well as in Appendix I, respectively. In the present chapter, we will first review the main findings which are relevant to those issues, draw some general conclusions based on them, and then provide some suggestions for future research.

1. Main findings

First, let us summarize the findings reported in previous chapters (chapters 2 to 5 and Appendix I) addressing the five main issues raised in Chapter One §§ 5.4.

(1) To what extent does brain lesion affect lexical tones in the speech production of Broca's patients? Are all lexical tones in Beijing dialect equally impaired in speech production? What happens to the vowels that are the primary tone-bearing segments for lexical tone?

In Chapter Two, an acoustic study of segmental and prosodic properties of words produced by one of our patients, the severest case, revealed that this aphasic speaker suffered severe loss of tones, but retained all of the vowel phoneme contrasts (about as well as the control speakers), even though both phenomena serve to maintain lexical contrasts in Chinese. Tone 1 was the least impaired of the four lexical tones in Beijing dialect (88 vs. 9, 20 and 16 % correct identification for Tone 1 to 4, respectively).

We conclude from these findings that the specification of segmental and tonal aspects of lexical entries in Chinese, and probably in tone languages in general, are located or processed separately in the brain.

(2) To what extent does lesion affect linguistic prosody in speech perception? Are lexical tone and sentence intonation equally impaired in the perception of Chinese aphasic patients, because pitch variation is used both at word and sentence level in Chinese?

In Chapter Three, a comparative group study on the perception of fourteen Broca's patients and 30 control speakers showed that relative to their healthy control listeners the aphasic patients were significantly poorer (78 vs. 99.3%) and slower (1870 vs. 1000 ms) in lexical tone identification but were not significantly different in intonation identification (statement vs. question) and were even faster in terms of reaction time.

This would indicate that the representation and/or processing of word-level prosody is separate from those of sentence-level prosody, i.e., production and perception of prosodic phenomena in speech, such as lexical tone and sentence intonation, engage multiple areas, which may comprise a large-scale spatially distributed network.

(3) To what extent does phonological 'deafness' affect linguistic prosody in speech perception? Are lexical tone and sentence intonation equally impaired in the perception of L2 speakers of Beijing dialect?

In Chapter Four, we repeated the experiment run in Chapter Three but this time the experimental listeners were L2 learners of Beijing Chinese – with or without a native tone-language, and performing their perception task with and without time pressure. The results of the experimental groups were compared with those obtained in Chapter Three for the healthy Beijing control group. The results indicated that Chinese learners with a non-tone native-language background (Uygur) proved perceptually less sensitive to lexical tones but were at the same time more sensitive to intonational boundary tones than all three groups of tone-language listeners (Beijing, Nantong and Changsha). The Uygur learners confused the lexical tones of Beijing Chinese more often than any of the other learner groups and needed more time to make a decision. In the perception of sentence intonation, however, the Uygurs were as sensitive to manipulations of the final boundary tone as the other groups (including the Beijing controls) but – surprisingly – the Uygurs' task performance was clearly faster.

(4) What are the characteristics of lexical tone categorization by Chinese aphasic patients relative to healthy controls on the one hand, and to non-native speakers on the other?

The experiment in Chapter Five was set up in order to study whether Broca's patients identify the lexical tones of Beijing Chinese differently than either healthy Beijing native-language controls or than non-native learners of Beijing Chinese –

either with or without a tone-language L1. We found that the patients not only exhibited less well-defined cross-overs and more randomness in their responses than L1 speakers did, but also they performed more poorly than L2 speakers with a tone-language L1.

The combined results presented under (3) and (4) above suggest that aphasic speakers with unilateral damage in the left hemisphere suffer from a structural deficit in the mental representation of the lexical tones, which can be compared with the structural deficit of L2 speakers with a non-tone language L1.

(5) In line with Question 4, what role does time pressure play in lexical tone categorization? Is it true that the use of prosody by patients with Broca's aphasia deteriorates progressively when they are allowed less time to perform language processing tasks? This would allow us to conclude whether lexical tone deficit is due to processing limitation or a structural deficit.

In Appendix I, we reported an analysis in order to establish the effect of time pressure on tone identification and found that (i) time pressure increased the amount of randomness in the identifications for both Beijing patients and healthy controls, while (ii) categoriality in the responses was greatly enhanced for the controls only but not for the patients. This was an exploratory study that targeted only one of the five stimulus continua used in the experiment reported in Chapter 5. Future analyses will reveal whether the method can be extended to the other four continua.

We conclude from the preliminary findings that the perceptual system of both patients and healthy controls tries to elaborate a phonological representation with detailed acoustic-phonetic information when more time is allowed, which blurs the clear cross-overs defined by mental representations.

2. Main conclusions

To recapitulate, the findings obtained from our lesion studies as well as the comparative studies with second-language learners lead us to draw the following conclusions:

(1) The specification of segmental and tonal aspects of lexical entries in Chinese, and in tone languages in general, are located or processed separately in the brain.
(2) The representation and/or processing of word-level prosody is separate from those of sentence-level prosody, i.e., production and perception of prosodic phenomena in speech, such as lexical tone and sentence intonation, engage multiple areas, which may comprise a large-scale spatially distributed network.
(3) Aphasic speakers with unilateral damage in the left hemisphere suffer from a structural deficit in the mental representation of the lexical tones, which can be compared with the structural deficit of L2 speakers regardless of their exposure

to the target language (long or short), and their native language (tone or non-tone language).

(4) The perceptual system tries to elaborate a phonological representation with detailed acoustic-phonetic information when more time is allowed, which blurs the clear cross-overs defined by mental representations.

3. Suggestions for future research

As was stated in the introductory chapter, it is now uncontroversial that the classical model of aphasia described by Broca is problematic in that it cannot account for the range of aphasic syndromes — sets of symptoms that co-occur often enough to suggest a single cause. However, the study of Broca's aphasia remains a fascinating research area for hypotheses that link the conceptual and technical apparatus of linguistics with the machinery of neuroscience, i.e. linking hypotheses that bridge concepts such as 'distinctive feature' with the mechanisms of neurobiology (e.g. concepts such as 'receptive field'). Experimental research on Broca's aphasia has contributed much to a new functional anatomy of language.

In the following sections, suggestions for future research on suprasegmental aspects specifically (i) lexical pitch features and (ii) post-lexical pitch features, in Broca's aphasic patients are put forward. Some of our suggestions aim to remedy certain infelicities in the experiments that were carried out in the project; others extend the earlier line of research. We present our suggestions in three subsections pertaining to subject selection (§ 3.1), perception (§ 3.2), and production (§ 3.3).

3.1. Subject selection

To investigate the contributions of the left and right cerebral hemispheres to the processing of linguistic prosody, it is important to select patients with unilateral damage either in the left or right hemisphere. To evaluate the deficit of patients' speech we recommend enlisting both native and non-native healthy control listeners. Accordingly, we propose the following suggestions.

(1) Speakers of a tone language with unilateral brain damage in the right hemisphere should be recruited to investigate the role of the right hemisphere in processing linguistic prosody.

(2) Speakers of a non-tone language (e.g. Dutch or English) with unilateral brain damage either in the right and left hemisphere should be included to examine the effect of lesion on prosodic processing at the word level vs. the sentence level, e.g. word stress vs. sentence intonation.

(3) As control subjects, speakers of other tone languages than Chinese – such as Thai, Vietnamese or Yoruba – should be included in order to have a better control of speakers' exposure to the Beijing dialect. In our comparison of Beijing Chinese learners with either a tone-language or a non-tone language L1, it is very likely that the Nantong and Changsha L2 speakers unduly benefited

from their exposure to Standard Chinese through the media (radio and television), which gave them an advantage over the Uygur speakers.

3.2. Perceptual studies

In our study, only one function of sentence intonation (i.e. the signalling of sentence type – statement versus question) was tested, while at the same time all four lexical tones of Chinese were tested both in citation forms and in context (yielding sandhi tones). This seems a poorly balanced comparison between word and sentence prosody. Therefore we should test other functions of sentence intonation in addition to sentence typing. Some examples are suggested below.

(1) Apart from sentence-type as being studied in the present research, other functions of intonation could be investigated such as the marking of focus, of topic, demarcation of phrases, etc., which are conveyed through F0 and which affect F0 as well.
(2) Apart from lexical tones, i.e., the prosodic variation at the word level in a tone language, its counterpart in a non-tone language, i.e. word stress, could also provide an excellent opportunity to test hypotheses about hemispheric lateralization of the processing of linguistic prosody.
(3) To test the true nature of linguistic prosody, it should be compared with the perception of paralinguistic prosody. So, future experiments should involve tasks carried out by patients and control subjects involved in the perception not only of linguistic prosody (lexical tone, stress, sentence intonation) but also of paralinguistic prosody signalling affect (at the sentence level only).

3.3. Production studies

The last, and probably the most complicated, phenomenon is the speech production by the patients. Based on a better understanding of the patients' perception of speech prosody, more research on production should be carried out. The findings presented in the previous chapters show that the production of pitch and of segmental cues can be separately affected by brain damage. Further research might determine how separate or connected the prosodic system is from other components of the grammar, and the extent to which dissociations may occur within the prosodic system itself. For instance, although dissociations were demonstrated in the perception of lexical tone and sentence intonation, no studies exist that show the same dissociation in the speech production of Chinese Broca's patients. Therefore, specifically, the following two points could be taken into consideration in future research.

(1) In the study reported in Chapter Two we analysed in great detail the recordings by one patient, whose speech production seemed most affected of all. The results of this patient indicated a severe loss of lexical tone production but

hardly any deficit in the production of the segments, especially of the vowels. Informal auditory screening by qualified listeners indicated that for all 14 patients lexical tones seemed more compromised than the segmental structure. As a first item on the research agenda, of course, we should run a detailed formal analysis of the remaining 13 patients in the same way we did for the single patient in Chapter Two. However, 14 patients is a relatively small sample so that the possibility cannot be ruled out that a larger sample might reveal the existence of Broca's patients whose segmental structure might be more compromised that the lexical tones. The existence of such patients would present clear evidence for a double dissociation of segmental structure and lexical tones – even though both are lexical properties.

(2) In none of the experiments reported in this dissertation did we study patients with brain damage in the right hemisphere. Yet, such unilaterally RH damaged patients would be needed if we wish to learn the role of right hemisphere in the patients' production of pitch variation at word level and sentence level.

Appendix I

Phonetic and phonological processing of pitch levels: A perception study of Chinese (aphasic) speakers[32]

1. Introduction

Chinese is a lexical tone language. Words in such languages do not only differ in the sequence of vowels and consonants (segments) but also by word melody ('tone'). Standard Chinese (Mandarin, Beijing dialect) has four lexical tones: Tone 1 (high level, 55), Tone 2 (rising, 35), Tone 3 (dipping, 214) and Tone 4 (falling, 51) Whenever there is little movement in F0 (or none at all), listeners tend to hear Tone 1 (Whalen and Xu, 1992), regardless of the mean pitch of the syllable, showing that absence of pitch change is a more powerful cue than mean pitch. Accordingly, we assume that Tone 1 – the only tone that is not a contour tone – is the unmarked or default tone.

What happens when a Chinese speaker suffers from a brain lesion (e.g. as a result of a stroke) in the left hemisphere, the dominant brain half for language processing? Packard (1986) demonstrated that left-hemisphere (LH) damaged non-fluent aphasic speakers of Chinese experience a tonal production deficit. However, impairment may affect tones to different degrees, e.g. Tone 1 is the least impaired, as shown by Liang and van Heuven (2004b), suggesting that Tone 1 is indeed the default

To account for the nature of the deficit in aphasic speech, two main approaches have been developed. The structural-deficit hypothesis postulates that deficits are caused by damage to abstract representations (Caplan, 1983; Hagiwara, 1995; Friedmann and Grodzinsky, 1997). Alternatively, the account is not in terms of a representational deficit, but the problem is rather caused by processing limitations (Friederici and Frazier, 1992; Kolk, 1995).

---

[32] This chapter has been published as J. Liang and V.J. van Heuven (2005) Phonetic and phonological processing of pitch levels: A perception study of Chinese (aphasic) speakers. In J. Doetjes and J. van de Weijer (eds.), *Linguistics in the Netherlands,* 125–137, Amsterdam: John Benjamins.

Our question for the present paper is whether the lexical tone impairment in Chinese is due to a structural deficit or to some acquired processing limitation that prevents patients from effectively accessing their linguistic knowledge. If there is a processing limitation, we generally observe better task completion when the patient is allowed more time. When the application of time pressure has no effect on the patient's performance, then the problem is typically a structural deficit. When a structural deficit is the underlying problem we need to know how words in Chinese are stored in long-term memory.

At least two representational formats have been proposed: words may be represented as abstract phonological representations (McClelland and Elman, 1986; Norris, 1994), or as detailed acoustic traces (Goldinger, 1996; Klatt, 1979; Pisoni, 1996). We call the former a phonological approach and the latter a phonetic approach.

We test this abstract vs. concrete format by varying the effect of time pressure (accuracy vs. speed) on responses to pitch-level manipulation in lexical tone identification. It is commonly accepted that phonetic processing is continuous in nature while phonological processing is discrete. Research shows that a faithful and detailed mental representation of the auditory stimulus does not remain available for more than 250 ms (Crowder and Morton, 1969; Massaro, 1974). When more time elapses between hearing the stimulus and issuing the response, the listener will have to recode the auditory pattern into some more abstract representation, such as a linguistic category (be it a phoneme, a word or a lexical tone).

We assume there will be a significant difference in response pattern when the listener has little time (time pressure) to recode the auditory stimulus into some higher-order abstract representation compared with a situation where such time pressure is absent. If the former pattern reflects a more gradient curve while the latter yields more clearly defined perceptual categories, we have reason to believe the representations are stored in terms of acoustic-phonetic properties. Otherwise, we would accept the view that categories are stored as abstract representations.

Accordingly, we predict that if Chinese Tone 1 is stored as a detailed acoustic-phonetic representation, Tone-1 identification would employ more acoustic-phonetic details (yielding a gradient curve) under time pressure than it would without time pressure. Alternatively, if Tone 1 is stored as an abstract phonological representation, we would expect that under time pressure Tone 1 will be identified with less acoustic-phonetic detail but with better defined cross-overs (characterized by discreteness) than when there is ample time. That is to say, under time pressure, access to Tone-1 representations is facilitated by phonological processing but inhibited by phonetic processing.

As for aphasic patients with lexical tone impairment, we would expect no effect of time pressure if there is damage to mental representations. Moreover, if the brain lesion prevents the extraction of detailed acoustic-phonetic information from the auditory stimulus, we expect that the aphasic listeners would not be able to follow acoustic traces in the stimuli. If the abstract representations are damaged, the aphasic patients will still display a gradient pattern in their responses rather than a pattern with well defined cross-overs.

2. Methods

Stimuli. Four words (covering the tone inventory in Beijing dialect) /ma^1 ma^2 ma^3 ma^4/, 'mother, hemp, horse, scold' were recorded by a male native speaker of Beijing dialect onto digital audio tape (DAT) using a Sennheiser MKH 416 unidirectional condenser microphone, transferred to computer disk (16 kHz, 16 bits) and digitally processed using the Praat speech processing software. The tone patterns of the four words were stylized with three points (onset, midpoint and offset) defining straight lines in a log-frequency (semitone, ST) by linear time representation. Using PSOLA analysis-resynthesis (Moulines and Verhelst, 1995), ten different tone patterns were generated from /ma^1/ 'mother' by decrementing the overall pitch level in 2-ST steps, which adequately covered the range we established in the production of the four lexical tones of the speaker, as shown in Figure 1. Here the low pivot point in the dipping Tone 3 defined the bottom of the speaker's pitch range in the continuum.

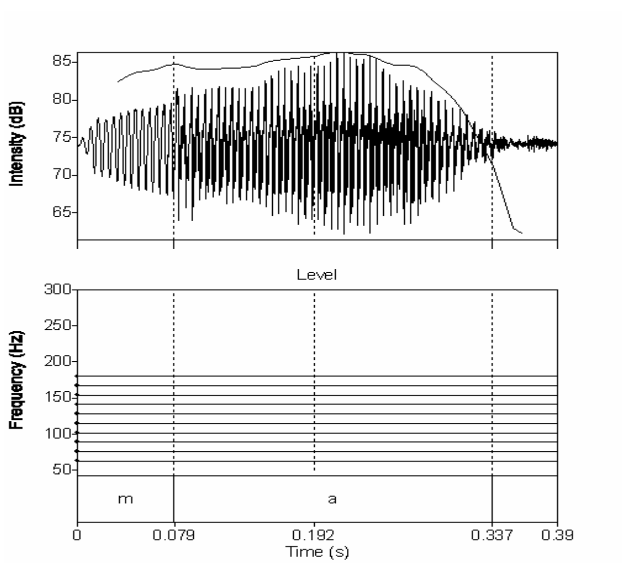


Figure 1: Steps 1 through 10 along a resynthesized continuum differing in overall pitch level by 2-ST increments.

Procedure. Stimuli were randomized and presented by computer binaurally over headphones (Sony MDR-V3) at a comfortable listening level. Listeners were tested individually in a quiet room, in the case of patients often in the subject's home. A keyboard was designed with four buttons marked with the corresponding Chinese characters for tone identification.[33] Listeners decided which of four words they had heard, by pressing one of the four buttons on the response box each time they heard a stimulus. They were asked just to avoid errors in the first stage of the experiment (no time pressure). In the second stage they were instructed not only to avoid errors but also to perform the task as quickly as they could manage (time pressure).

Before the experiment specific instructions were presented on the screen and explained orally as well. The ten lexical-tone stimuli were presented to the listeners twice in two blocks (once without and a second time with time pressure).

Before the experiment specific instructions were presented on the screen and explained orally as well. The ten lexical-tone stimuli were presented to the listeners twice in two blocks (once without and a second time with time pressure).

The experimental task was preceded by a short practice session with four trials. Decisions made and reaction times (with a precision of 1 ms measured from the offset of the stimulus) were stored in computer memory by E-prime software.[34]

When there was no time pressure, there was a fixed 3000-ms inter-stimulus interval (ISI) after the offset of the stimulus, irrespective of the reaction time. If a subject did not respond within the ISI, s/he timed out, and the next stimulus was presented. In the sections with time pressure imposed on the listener, the next stimulus started 1000 ms after the response. The seemingly shorter ISI in the time-pressure condition prompted the subjects to speed up their reaction time (see results).

Subjects. Thirty healthy Beijing listeners and fourteen Beijing aphasic patients participated in the experiment. The healthy controls were native speakers of Beijing dialect, aged between 21 and 70, average 40, 17 male and 13 female. All of them had normal hearing and at least twelve years of formal education. They took part in the experiments in September 2002.

The fourteen Beijing aphasic listeners, native Beijing speakers from Tianjin, P. R. China, aged 39–80, were diagnosed as non-fluent Broca aphasics characterized by word-finding difficulties, incomplete syntactic constructions, and errors in sound production.[35] Production studies on the tones and vowels of one of the patients, the severest case, showed that lexical tones were seriously damaged while the vowels (and consonants) were comparatively well preserved (Liang and van Heuven, 2003, 2004).

---

[33] The keyboard was designed and built by J.J.A. Pacilly at the Universiteit Leiden Phonetics Laboratory.

[34] The E-prime script for stimulus presentation and response collection was written by J.J.A.Pacilly.

[35] We thank Professor Zhang Banshu from Tianjin General Hospital for her invaluable help in this matter.

The patients' non-verbal communication was still effective, and apart from their aphasia, they were able to carry out the activities of everyday life without difficulty. All of them suffered from unilateral damage in the left frontal and/or parietal lobe (detailed information presented in Table 1) and showed normal hearing sensitivities at 0.5, 1, and 2 kHz following a pure-tone air-conduction screening.

None of the participants had been diagnosed with neurological or psychiatric illness prior to the experiments, apart from a single-event cerebrovascular accident (CVA) with damage in the LH of the brain. The patients participated in this experiment in 2002–2003.

Table 1: MRI or CT scan findings of the lesion site in left hemisphere for individual patients

Patients	Name	P	Z_1	X	C_1	H	Q	Y_1	C_2	Z_2	Y_2	F	Z_3	T	L
	Age	39	50	68	43	47	63	69	80	31	54	50	69	56	52
	sex	f	m	f	m	m	f	f	f	m	m	m	m	m	m
	time post onset (mo.)	23	25	4.5	5.5	5.5	3.5	2	12	2	6	12	4	35	39
Frontal lobe	precentral gyrus lower part	●			●	●				●				●	●
	superior f. gyrus posterior								●						
	inferior gyrus pars triangularis	●	●			●					●		●	●	
	Broca's area pars operculalis						●	●					●		●
Parietal lobe	Postcentral gyrus lower part	●	●	●	●							●	●		
	Supramarginal ant.							●	●						
	gyrus posterior						●								
	Posterior						●								
	Superior lobule										●				
Cingulated	gyrus posterior			●					●						

3. Results

Since the stimuli were presented for identification once with and once without emphasis on speed of response (time pressure), we predict that listeners traded accuracy for speed when the time pressure was on, i.e., they were prepared to gamble in the case of ambiguous stimuli in order to gain speed. Therefore, in our data analysis we will just analyze the percentage of 'Tone-1', 'Tone-2', 'Tone-3' or 'Tone-4' responses and the decision latencies for each; we expect longer latencies as the choice between the response alternatives is more evenly balanced (which would be a sign of ambiguity in the stimulus). In the results obtained for the present continuum the dominant response is Tone 1 (high level tone), while Tones 2 (rising tone) and 3 (low, dipping tone) are alternatives for lower pitch levels.[36] Significance testing was done on a dichotomized response variable: 'Tone 1' or 'not Tone 1'.

3.1. Tone-1 identification by Beijing healthy listeners

A two-way ANOVA with time pressure and stimulus step as fixed factors was carried out on Tone-1 identification. Significant effects were found for time pressure [F (1, 580) = 26.3 (p < 0.001)], step number [F (9, 580) = 29.4 (p < 0.001)], and for the interaction between these factors [F (9, 580) = 2.8 (p = 0.003)]. The same analysis on the associated reaction times yielded a significant effect for time pressure [F (1, 580) = 166.3 (p < 0.001)], for stimulus step [F (9, 580) = 11 (p < 0.001)], as well as for the interaction [F (9, 580) = 6.9 (p < 0.001)].

The significant interaction is crucial, since it reveals differences in boundary width, such that the psychometric function should be steeper as listeners depend more on abstract phonological representations than on phonetic detail. The tone identification scores of the Beijing healthy listeners and the associated reaction times broken down by presence vs. absence of time pressure are presented in separate panels in Figure 2.

The two panels in the left column of Figure 2 show that when there is no time pressure, percent Tone-1 identification drops gradually from 100 to 40, while percent Tone 2 goes up from 10 to 50, as the physical stimuli decrement continuously in 2-ST steps. This demonstrates that Beijing listeners are sensitive to the manipulation of pitch level when discriminating between Tone 1 (high level tone) and Tone 2 (rising tone). Tone 3 and Tone 4, however, were never reasonable response alternatives for this continuum.

We may also observe a trend in the reaction time data: responses are fastest towards the left-hand side of the scale (representing high pitch levels) but gently rise as the pitch level assumes lower values. The steady increase in reaction time reflects the rising ambiguity in the choice between Tone-1 and Tone-2 responses towards the right-hand side of the continuum.

The two panels in the right-hand column show that, with time pressure, Tone-1 identification clearly differs between the left and right part of the continuum. There is a sharp discontinuity in the responses after step 4 (suggesting a tonal category boundary). The highest pitch levels are unambiguously perceived as exemplars of Tone 1. From step 5 onwards the responses are scattered among Tones 1, 2 and 3 whilst Tone 4 is never even an option. From step 7 onwards, however, Tone 3 is an

[36] This experiment is part of a much larger study in which several other, more complex, continua were synthesized between Tone 1 and each of the three other tones of Beijing Chinese (for a full account see Chapter Five)

increasingly attractive alternative but never yields a clear majority response, suggesting nevertheless that low average pitch is a characteristic of Tone 3.

Clearly, then, the significant difference found by the ANOVA between the two perception patterns (without and with time pressure) is caused by different perceptual strategies, i.e. one is more dependent on phonetic variation while the other depends more on abstract representations.

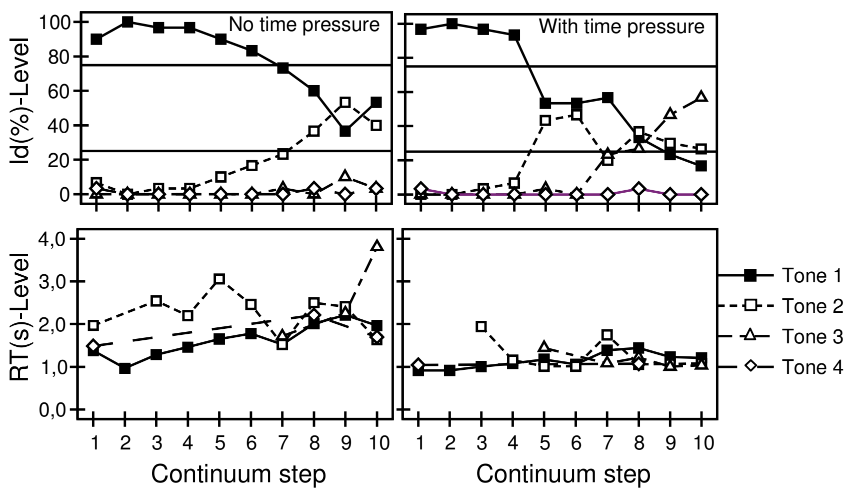


Figure 2: Beijing listeners' tone identifications (percent, upper panels) and reaction time (seconds, lower panels) as a function of excursion size of pitch levels (semitone) divided by time pressure (left vs. right). Gaps in the reaction-time curves occur when the response alternative was never chosen by any of the listeners.

In fact, it seems to us that the response strategy which is followed under time pressure, is an averaging operation on the part of the listener. The listener chooses the response category whose average pitch (roughly) matches that of the stimulus. However, if the listener is allowed more time, he separately evaluates the extent to which the stimulus fits such acoustic details as the presence of a fall, presence of a rise, in addition to average pitch.

3.2. Tone-1 identification by Beijing aphasic listeners

A similar two-way ANOVA was carried out on Tone-1 identification scores and reaction times collected from the aphasic listeners. However, for Tone 1 identification, only the effect of step number proved significant [F (9, 260) = 6.2 (p < 0.001)]; for the associated reaction time, there was a small effect of time pressure [F (1, 238) = 9.7 (p = 0.002)].

Since no significant effect was found for the interaction between time pressure and step number, the statistics suggests that the aphasic listeners behaved similarly under the two conditions (without and with time pressure). Figure 3 presents the results.

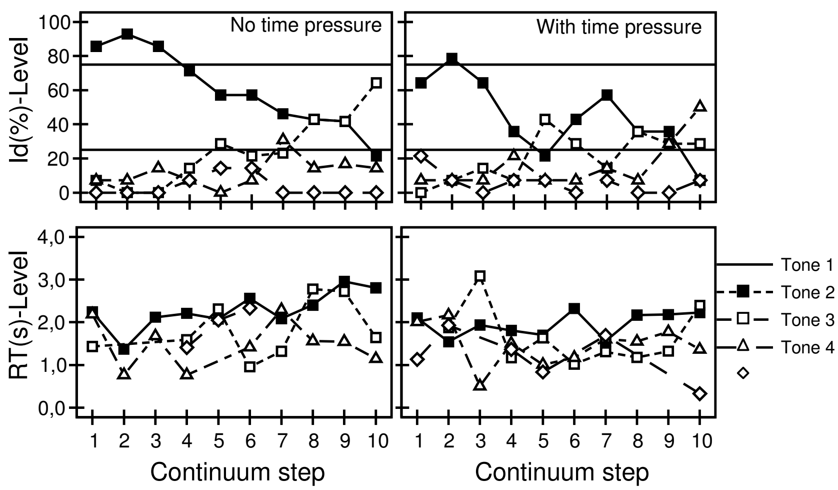


Figure 3: Beijing aphasic listeners' tone identifications (percent, upper panels) and associated reaction time (million second, low panels) as a function of excursion size of pitch levels (semitone) divided by time pressure (left vs. right).

In the absence of time pressure (left two panels in Figure 3), Tone-1 identification gradually decreases from 90 to 20% as the physical stimuli decrement by 2-ST steps; at the same time we observe a gradient upward trend in Tone-2 identification (from 10 to 60%). This demonstrates that Beijing aphasic listeners are sensitive to manipulation of pitch level when discriminating between Tone 1 (high level tone) and Tone 2 (rising tone) or Tone 3 (fall-rise; dipping tone); Tone 4 (falling tone), however, was never a reasonable response alternative for this continuum.

Reaction time of Tone-1 and Tone-2 identification increases somewhat (from 1500 ms to 2500 ms) for lower pitch levels. The trend is significant for Tone-1 responses only ($r = 0.689$, $p = 0.025$, two-tailed). The findings indicate that the listeners depend on the change in the stimuli for their perception of both Tone 1 (from a good exemplar to ambiguity) and Tone 2 (from non-Tone 2 to ambiguity).

When there is time pressure, as shown in the two rightmost panels, the highest level pitches were predominantly perceived as Tone 1, even though Tone 1 is never identified over 75%. There seems to be an early cross-over between 6 and 10 ST, where percent Tone-1 identification drops by almost 50% (from 75 to 29%).

However, if there is a tone boundary here, it is not reflected in locally increased reaction time. Also, Tone-1 identification rises to over 50% and reaches a second cross-over at the low end of the continuum. Therefore, we conclude that neither tone was categorically perceived.

3.3. Cross-group analysis

Our visual inspection can be summarized as a better defined cross-over for Tone-1 identification under time pressure, as opposed to a smooth, continuous identification pattern for Tone 1 and Tone 2 when the identification task was performed without time pressure for the Beijing healthy listeners. That is to say, the Beijing listeners displayed a continuous perception when time pressure was off and categorical perception with the time pressure on.

Our data indicate that high level pitch somewhere in the top 6-ST band of the speaker's pitch range is a primary feature for Tone 1. Mid and low pitch are only secondary cues for Tone 2 and Tone 3, respectively. The aphasic listeners showed a similar pattern regardless of time pressure. The gradual pattern suggests that the aphasic listeners were still capable of phonetic processing but the reduced identification scores and the absence of a clearly defined cross-over suggests that the aphasic listeners no longer have clearly defined tonal categories, i.e. they suffer from a structural deficit.

We will now formalize the differences in the responses between the two listener types. The aphasic listeners are characterized, overall, by a more random distribution of responses. Ambiguity between the competing tones is also considerable for the aphasic listeners, while the healthy Beijing listeners generally appear to have better defined tonal categories along the continuum, which makes their response distributions less random.

A first quantification of these variations in randomness in the response distributions would be in terms entropy in the distribution of responses over the four alternatives (Tone 1 through Tone 4) for each step along the continuum, and then averaged over all steps. With four alternatives the theoretically maximum entropy would be 2 bits, i.e. in the case of a perfectly even distribution (25% in each tone category).

This would reflect completely random behaviour on the part of the listeners. When all the listeners in a group are in complete agreement (i.e. all choose the same response alternative) for each step (although the choice may vary from step to step) the entropy equals zero. The smaller the entropy, therefore, the greater the determination (or: stability) in the responses for the group of listeners. In terms of entropy we would expect the aphasic listeners to be located near the noisy extreme and the healthy Beijing listeners (who should have the clearest perceptual norms for the four tones of their language) near the deterministic end.

However, we will need a second parameter to describe the listeners' responses to our continuum, such that this parameter captures the sensitivity of the listeners to a change in pitch level. This parameter would differentiate identification patterns with cross-overs from those that lack a changing percept.
Transmitted information (in bits) would provide a good measure of this sensitivity. The clearer the division of the continuum into discrete perceptual categories (and the more categories are distinguished along the continuum), the higher the amount of transmitted information, again with a theoretical maximum of 2 bits representing the situation where all four tones are perfectly distinguished ('transmitted').

We computed the mean response entropy and the transmitted information for the continuum for each listener group separately. [37] Figure 4 plots transmitted information (as a measure of categoriality in the responses) along the Y-axis against response entropy (as a measure of determination in the responses) along the X-axis broken down by listener type (Beijing vs. aphasic) and time pressure (accurate vs. speed).

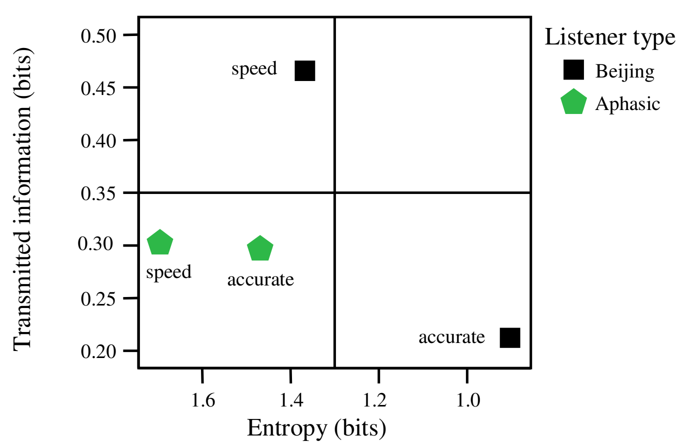


Figure 4: Transmitted information against response entropy broken down by listener type (Beijing vs. aphasic) and time pressure (accurate vs. speed).

In Figure 4, we find what we predicted, i.e., the Beijing listeners reveal stronger categorical identification of Tone 1 under time pressure than without time pressure. However, the difference was greatly reduced in case of the aphasic listeners, which indicates that the patients were as much confused under time pressure as without time pressure.

---

[37] Transmitted information (see also Shannon and Weaver, 1949; Attneave, 1959): T(x;y) = H(x) + H(y) – H(x,y), where H is the entropy. H(x) = −∑(pi 2log pi), where i is an index ranging over the categories 1 through 10 along our stimulus dimension (x). H(y) = −∑(pj 2log pj), where j ranges over the response categories 1 to 4. H(x,y) = −∑(pij 2log pij).

We also found that, although both the Beijing and aphasic listeners show more entropy (or ambiguity) under time pressure, the time-pressure effect is smaller for the aphasic than the Beijing listeners. In comparison with the patterns of Beijing listeners, the distance between the two patterns was reduced much more in terms of transmitted information than entropy. That is to say, time pressure caused more randomness but greatly improved categoriality for Beijing listeners while time pressure only caused some randomness but hardly improved categorical information for the aphasic listeners.

As for the time pressure effect, we found a much bigger effect for the healthy Beijing listeners than for the aphasic patients in terms of difference in reaction time between the two parts of the experiment, i.e. 334 vs. 101 ms; a one-way analysis of variance in the difference in individual reaction time between the two conditions (with ~ without time pressure) shows a significant effect for listener group (healthy vs. aphasic) [$F(1, 42) = 6.5$ ($p = 0.015$)].

4. Conclusions and discussion

Our results indicate that, under time pressure, Beijing listeners use variation in the frequency of a level pitch to distinguish among three lexical tones. High level pitch within the upper 6-ST part of the speaker's pitch range is categorized as Tone 1, the middle 6-ST band is considered representative for Tone 2, while the lowest 6-ST band is at least an option (but yielding a-typical exemplars) for Tone 3. This use of pitch height sheds light on the status of pitch level as a Tone-1 feature, i.e., whether it is a high or a mid-high tone. Although, presumably, speakers will typically realize Tone 1 with high level pitch [55], the perceptual tolerance is such that the listener will also accept a mid-high level pitch [44] as an adequate token of Tone 1.

Our data show that time pressure is an effective technique for tapping abstract representations in response patterns. Our results suggest that the perceptual system tries to elaborate a phonological representation with detailed acoustic-phonetic information when more time is allowed, which blurs the clear cross-overs defined by mental representations.

As time pressure was found to have no effect on the aphasics' identification pattern, suggesting that the aphasic listeners were equally confused with or without time pressure, our data strongly support the structural deficit account. Therefore, our aphasic patients may have had to recode the auditory pattern into a more abstract representation, thereby abstracting from phonetic details.

We have presented a laboratory-controlled perception experiment the results of which point to the existence of pre-lexical phonological processes in word recognition, suggesting that spoken words are accessed using a phonological code, and offer response patterns from aphasic patients which can be interpreted as evidence for a deficit in the abstract phonological representation of the lexical tones.

References

Abramson, A.S., (1978). The phonetic plausibility of the segmentation of tones in Thai phonology. In W.U. Dressler and W. Meid (eds.), *Proceedings of the 12th International Congress of Linguistics* (pp. 760–763). Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

Amunts, K., (2004). Broca's region. Retrieved from http://ibronew.alp.mcgill.ca/Pub_Main_Display.asp?Main_ID=15.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B.M., and Zilles, K., (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology,* 412, 319–341.

Attneave, F., (1959). *Applications of information theory to psychology*. Holt, New York.

Bastiaanse, R., Rispens, J., Ruigendijk, E., Rabadán, O.J., and Thomspon, C.K., (2002). Verbs: Some properties and their consequences for agrammatic Broca's aphasia. *Journal of Neurolinguistics*, 15, 239–264.

Bates, E., and MacWhinney, B., (1982). Functionalist Approaches to grammar. In E. Wanner and L. R. Gleitman (eds.), *Language Acquisition: The State of the Art* (pp. 173–218). New York: Cambridge University Press.

Bates, E., Friederici, A., and Wulfeck, B., (1987). Grammatical morphology in aphasia: evidence from three languages. *Cortex,* 2, 545–574.

Baum, S., and Pell, M., (1997). Production of affective and linguistic prosody by brain-damaged patients. *Aphasiology,* 11, 177–198.

Baum, S., and Pell, M., (1999). The neural basis of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13, 581–608.

Behrens, S.J., (1989). Characterizing sentence intonation in a right hemisphere-damaged population. *Brain and Language,* 37*,* 181–200.

Beijing University (1962/1989). *Chinese Dialect Character Pronunciation List* (漢语方音字匯). Beijing: Wenzi Gaige Chubanshe.

Bezooijen, R. van, and Gooskens, C., (2005). How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? In J. Doejes and J. van de Weijer (eds.), *Linguistics in the Netherlands*, 13–24.

Binder, J.A., Frost, J.R., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., and Possing, E.T., (2000). Human temporal lobe activation by speech and non-speech sounds, Cerebral. *Cortex,* 10, 512–528.

Bladon, R.A.W., and Lindblom, B., (1981). Modelling the judgment of vowel quality differences. *Journal of the Acoustical Society of America*, 69**,** 1414–1422.

Blumstein, S., (1998). Phonological aspects of aphasia. In M. Sarno (ed.), *Acquired aphasia* (pp. 157–185). New York: Academic Press.

Blumstein, S.E., Baker, E., and Goodglass, H., (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, 15, 19–30.

Boersma, P., and Weenink, D., (1996). Praat, doing phonetics by computer. *Report nr.* 132, Institute of Phonetic Sciences, University of Amsterdam.

Bolinger, D., (1978). Intonation across languages. In J. Greenberg (ed.), *Universals of Human Language. Vol II: Phonology* (pp. 471–524). Palo Alto, CA: Stanford University Press.

Broca, P.P., (1861). Perte de la parole; ramolissement chronique et destruction partielle du lobe antérieur gauche de cerveau. *Bulletins de la Société d'Anthropologie de Paris*, 62, 235–238. [English: Broca, Paul. Remarks on the seat of the faculty of articulate language, followed by an observation of aphemia. In *Some Papers on the Cerebral Cortex*. Translated by von Bonin, G. Springfield, Illinois: Thomas Publisher, 1960, 49–72.]

Broselow, E., Hurtig, R.R., and Ringen, C., (1987). The perception of second language prosody. In G. Ioup and S.H. Weinberger (eds.), *Inter-language phonology, the acquisition of second language sound system* (pp. 350–362). Cambridge: Newbury House.

Brown, C., Hagoort, P., and Swaab, T., (1996). Neurophysiological evidence for a temporal disorganization in aphasic patients with comprehension deficits. In W. Widdig, I.M. Ohlendorff and J.-P. Malin (eds.), *Aphasietherapie im Wandel* (pp. 89–122). Freiburg: Hochschul Verlag.

Bryan, K., (1989). Language prosody and the right hemisphere. *Aphasiology*, 3, 285–299.

Bryden, M.P., and Murray, J.E., (1985). Toward a model of dichotic listening performance. *Brain and Cognition*, 4, 241–257.

Caplan, D., and Hildebrandt, N., (1988). *Disorders of syntactic comprehension*. Cambridge, MA: MIT Press.

Caplan, D., (1983). Syntactic competence in agrammatism – a lexical hypothesis. In M. Studdert-Kennedy (ed.), *Psychobiology of language* (pp. 177–187). Cambridge: MIT Press.

Caramazza, A., and Zurif, E.B., (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, 3, 572–582.

Caramazza, A., Berndt, R.S., and Basili, A., (1983). The selective impairment of phonological processing: A case study. *Brain and Language*, 18, 128–174.

Chao, Y.R., (1930). A system of "tone letters". *Le Maitre Phonetique,* 45, 24–27.

Chen, A., (2005). *Paralinguistic meaning of intonation: universal or language-specific?* Doctoral dissertation, Nijmegen University.

Chen, S., and Bates, E., (1998). The dissociation between nouns and verbs in Broca's and Wernickes's aphasics: Evidence from Chinese. Special issue on Chinese aphasia. *Aphasiology*, 12, 5–36.

Cheng, C.-C., (1994). Calculation of Chinese dialect mutual intelligibility. (汉语方言沟通度的计算方法). *Zhongguo Yuwen,* 1**,** 35–43.

Cheung, P.S.P., (1995). A *speech therapy activity book on Hong Kong Cantonese phonology*. Hong Kong: Transaudio.

Ching T.Y., Williams, R., and Van Hasselt A., (1994). Communication of lexical tones in Cantonese alaryngeal speech. *Journal of Speech Hearing Research,* 37, 557–63.

Ching, Y.C., (1988). Lexical tone perception by Cantonese deaf children. In I.M. Liu, H.C. Chen and M.J. Chen (eds.), *Cognitive Aspects of the Chinese language* (pp.). Hong Kong: Asian Research Service.

Ching, Y.C., (1990). Tones for profoundly deaf tone-language speakers. *Papers in Linguistics,* 2, 1–22. Hong Kong: City University.

Clark, M., (1978). *A dynamic treatment of tone, with special attention to the tonal system of Igbo*. Doctoral dissertation. University of Massachusetts.

Cooper, W.E., Soares, C., Nicol, J., Michelow, D., and Goloskie, S., (1984). Clausal intonation after unilateral brain damage. *Language and Speech*, 27, 17–24.

Coster, D., and Kratochvil, P., (1984). Tone and stress discrimination in normal Beijing speech. In B. Hong (ed.), *New Papers on Chinese Language Use* (pp. 119–132). Canberra: Australia National University.

Crowder, R.G., and Morton, J., (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics,* 5, 365–373.

Cutting, J.E., (1974). Two left-hemisphere mechanisms in speech perception. *Perception and Psychophysics*, 16, 601–612.

Danly, M., and Shapiro, B., (1982). Speech Prosody in Broca's aphasia. *Brain and Language*, 16, 171–190.

Danly, M., Cooper, W.E., and Shapiro, B., (1983). Fundamental frequency, language processing, and linguistic structure in Wernicke's aphasia. *Brain and Language*, 19, 1–24.

De Francis, J.F., (1963). *Beginning Chinese*. New heaven: Yale University Press.

Démonet, J.F., Fiez, J.A., Paulesu, E., Petersen, S.E., and Zatorre, R.J., (1996). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language*, 55, 352–379.

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., and Frackowiak, R., (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753–1768.

Démonet, J.-F., Price, C., Wise, R., and Frackowiak, R.J., (1994). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain*, 117, 671–682.

Duanmu, S., (1994). Against contour tone units. *Linguistic Inquiry,* 25, 555–608.

Dupoux, E. and Peperkamp, S., (2002). Fossil markers of language development: phonological ‘deafnesses’ in adult speech processing. In J. Durand and B. Laks (eds.), *Phonetics, Phonology, and Cognition* (pp. 168–190). Oxford: Oxford University Press.

Dwyer, J., Blumstein, S.E., and Ryalls, J., (1982). The role of duration and rapid temporal processing on the lateral perception of consonants and vowels. *Brain and Language*, 17, 272–286.

Eng, N., Obler, L.K., Harris, K.S., and Abramson, A.S., (1996). Tone perception deficits in Chinese-speaking Broca’s aphasics. *Aphasiology*, 10, 649–656.

Fiez, J., Raichle, M., Miezen, F., Petersen, S., Tallal, P., and Katz, W., (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, 7, 357–375.

Fok, A. C.Y., (1984). The teaching of tones to children with profound hearing impairment. *British Journal of Disorders of Communication*, 19, 225–236.

Friederici, A.D, and Alter, K., (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89, 267–276.

Friederici, A.D., and Kilborn, K., (1989). Temporal constraints on language processing: Syntactic priming in Broca's aphasia. *Journal of Cognitive Neuroscience*, 1, 262–272.

Friederici, A.D., and Frazier, L., (1992). Thematic analysis in agrammatic comprehension: Syntactic structures and task demands. *Brain and Language,* 42, 1–29.

Friederici, A.D., and Alter, K., (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and Language,* 89, 267–276.

Friedmann, N., and Grodzinsky, Y., (1997). Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language*, 56, 397–425.

Fry, D.B., (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 30, 765–769.

Fry, D.B., (1958). Experiments in the perception of stress. *Language and Speech,* 1, 126–152.

Gandour, J., (1974). On the representation of tone in Siamese. *UCLA Working Papers in Phonetics,* 27, 118–146.

Gandour, J., (1981). Perceptual Dimensions of Tone: Evidence from Cantonese. *Journal of Chinese Linguistics,* 9, 20–36.

Gandour, J., (1983). Perception of tone in Far Eastern languages. *Journal of Phonetics,* 11, 149–175.

Gandour, J., (1998). Phonetics and phonology. In B. Stemmer and H. Whitacker (eds.), *Handbook of Neurolinguistics* (pp. 207–219). New York: Academic Press.

Gandour, J., (2000). Frontiers of Brain Mapping of Speech Prosody. *Brain and Language,* 71, 75–77.

Gandour, J., and Dardarananda, R., (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language*, 18, 98–114.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., and Luritob, J., (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language,* 84*,* 318–336.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., and Lowe, M., (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage,* 23, 344–357.

Gandour, J., Holasuit Petty, S., and Dardarananda, R., (1988). Perception and production of tone in aphasia. *Brain and Language*, 35, 201–240.

Gandour, J., Larsen, J., Dechongkit , S., Ponglorpisit, S., and Khunadorn, F., (1995). Speech prosody in affective contexts in Thai patients with right hemisphere lesions. *Brain and Language*, 51*,* 422–43.

Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., Boonklam, R., and Potisuk, S., (1992). Lexical tones in Thai after unilateral brain damage. *Brain and Language*, 43, 275–307.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., and Li, X., (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping*, 18, 149–157.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., and Hutchins, G.D., (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience,* 12, 207–222.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., and Lurito, J., (2002). Neural circuitry underlying perception of duration depends on language experience. *Brain and Language*, 83, 268–290.

Gao, S.-R., (1993). *Aphasia.* Beijing: Beijing Medicine University and China Xiehe Medicine University Joint Press.

Gårding, E., and Abramson, A.S., (1965). A study of the perception of some American English intonation contours. *Studia Linguistica*, 19*,* 61–79.

Gårding, E., (1985). Constancy and variation in Standard Chinese tonal patterns. *Lund University Working Papers,* 28, 19–51.

Gårding, E., (1987). Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica,* 44, 13–29.

Gårding, El, Kratochvil, P., Svantesson, J.-O., and Zhang, J., (1986). Tone 4 and Tone 3 discrimination in modern Standard Chinese. *Language and Speech*, 29, 281–293.

Goldinger, S.D., (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22, 1166–1183.

Goldsmith, J., (1990). *Autosegmental and metrical phonology.* Oxford, UK: Basil Blackwell.

Goldstein, K., (1933). L'analyse de l'aphasie et l'etude de l'essence du langage. *Journal de Psychologie*, 30, 430.

Goodglass, H. (1976). Agrammatism. In H. Whitaker and H. A. Whitkaker (eds.), *Studies in Neurolinguistics* (Vol. 1, pp. 237–260). New York, NY: Academic Press.

Goodglass, H., and Kaplan, E., (1983). The *assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.

Gooskens, C., and Heeringa, W., (2004). Perceptive Evaluation of Levenshtein Dialect Distance Measurements using Norwegian Dialect Data. In D. Sankoff, W. Labov and A. Kroch (eds.), *Language Variation and Change* (pp. 189–207). Cambridge University Press, New York,

Gow, D.W., and Caplan, D., (1996). An examination of impaired acoustic-phonetic processing in aphasia. *Brian and Language*, 52, 386–407.

Grodzinsky, Y., (1989). Agrammatic comprehension of relative clauses. *Brain,* 37, 480–499.

Gussenhoven, C., (2004). *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

Gussenhoven, C., and Chen, A., (2000). Universal and language-specific effects in the perception of question intonation. In B. Yuan, T. Huang and X. Tang (eds.), *Proceedings of the 6th International Conference on Spoken Language Processing ICSLP 2000* (pp. 91–94). Beijing: China Military Friendship Publish.

Haarmann, H.J., and Kolk, H.H.J., (1994). On-line Sensitivity to Subject-Verb Agreement Violations in Broca′s Aphasics: The Role of Syntactic Complexity and Time. *Brain and Language,* 46**,** 493–516.

Hadding-Koch, K., (1961). *Acoustico-phonetic studies in the intonation of southern Swedish*. Lund: Gleerup.

Hagiwara, H., (1995). The breakdown of functional categories and the economy of derivation. *Brain and Language*, 50, 92–116.

Hagoort, P., (1997). Semantic Priming in Broca's Aphasics at a Short SOA: No Support for an Automatic Access Deficit. *Brain and Language,* 56, 287–300.

Hagoort, P., (1993). Impairments of Lexical-Semantic Processing in Aphasia: Evidence from the Processing of Lexical Ambiguities. *Brain and Language,* 45, 189–232.

Hagoort, P., Brown, C.M., and Swaab, T.Y., (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. *Brain,* 119, 627–649.

Halle, M., and Stevens, K.N., (1991). Knowledge of language and the sounds of speech. In J. Sundberg, L. Nord and R. Carlson (eds.), *Music, language, speech, and brain* (pp. 1–19). London: MacMillan.

Hart, J. ‘t, Collier, R., and Cohen, A., (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech perception.* Cambridge: Cambridge University Press.

Hartman, L.M., (1944). The segmental phonemes of the Peiping dialect. *Language,* 20, 28–42.

Hayward, K., (2000). *Experimental phonetics*. Harlow, UK: Pearson Education.

Head, H., (1926). *Aphasia and Kindred Disorders of Speech.* London: Cambridge University Press.

Hermes, D.J., and van Gestel, J., (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America,* 90, 97–102.

Hickok, G., and Poeppel, D., (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131–138.

Ho, A.T., (1976). Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, 4, 1–13.

Hocket, C.F., (1947). Peiping phonology. *Journal of the American Oriental Society,* 67, 253–267.

Hofstede, B., (1992). *Agrammatic speech in Broc’s aphasia, strategic choice for the elliptical register*. Ph.D. Dissertation, University of Nijmegen, The Netherlands.

Howie, J.M., (1974). On the domain of tone in Mandarin. *Phonetica*, 30*,* 129–148.

Howie, J.M., (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Hsieh, L., Gandour, J., Wong, D., and Hutchins, G., (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, 76, 227–252.

Hughes, C.P., Chan, J.L., and Su, M.S., (1983). Aprosodia in Chinese patients with right cerebral hemisphere lesions. *Archives of Neurology,* 40, 732–736.

Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H., and Tonoike, M., (1998). Task-dependent laterality for cue decoding during spoken language processing. *Neuro Report,* 9, 899–903.

Institute of Linguistics, Chinese Academy of Social Science., (1999). *Word list for dialect investigation (revised).* Beijing: Commercial Press.

Ip, K., and Hoosain, R., (1993). Dichotic listening of Chinese and English words. *Psychologia*, 36, 140–143.

Ivry, R., and Lebby, P., (1993). Hemispheric differences in auditory perception are similar to those found in visual perception. *Psychological Sciences*, 4, 41–45.

Kimura, D., (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, 3, 163–168.

Klatt, D.H., (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics,* 7, 279–312.

Klecka, W.R., (1980). *Discriminant analysis*. Beverly Hills, London: Sage.

Klein, D., Zatorre, R., Milner, B., and Zhao, V., (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, 13*,* 646–653.

Kolk, H., and Weijts, M., (1996). Judgments of Semantic Anomaly in Agrammatic Patients: Argument Movement, Syntactic Complexity, and the Use of Heuristics. *Brain,* 54, 86–135.

Kolk, H.H.J., (1995). A time-based approach to agrammatic production. *Brain and Language*, 50, 282–303.

Kolk, H., and Heeschen, C., (1992). Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes*, 7, 89–129.

Koski, L., Wohlschlager, A., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., and Iacoboni, M., (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex,* 12, 847–855.

Kratochvil, P., (1998). Intonation in Beijing Chinese. In D. Hirst and A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 417–431). Cambridge University Press.

Ladd, D.R., (1981). On intonational universals. In T. Myers, J. Laver and J. Anderson (eds.), *The Cognitive Representation of Speech* (pp. 389–397). Amsterdam: North Holland Publishing.

Ladd, D.R., (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

Ladd, D.R., and Terken, J., (1995). Modelling intra- and inter-speaker pitch range variation. In K. Elenius and P. Branderud (eds.), *Proceedings of the thirteenth International Congress of Phonetic Sciences, Stockholm* (pp. 386–389).

Laver, J., (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Leather, J., (1987). F0 pattern interference in the perceptual acquisition of second language tone. In A. James and J. Leather (eds.), *Sound patterns in second language acquisition* (pp. 59–81). Dordrecht: Foris.

Leben, W.R., (1973). *Suprasegmental phonology*. Doctoral dissertation. University of Massachusetts.

Lee, K.Y.S., Cheung, D.M.C., Chan, B.Y.T., and van Hasselt, C.A., (1997). Cochlear implantation: Implication for tone language. In I. Honjo and H. Takahashi (eds.), *Cochlear implant and related sciences update* (pp. 254–257). Basel: Karger.

Lee, K.Y.S., Chiu, S.N., and van Hasselt., C.A., (2002). Tone Perception Ability of Cantonese-Speaking Children. *Language and Speech*, 45, 387–406.

Lee, Y., Vakoch, D., and Wurm, L., (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25, 527–544.

Li, P., Bates, E., and Macwhinney, B., (1993). Processing a language without inflections: a reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language,* 32, 169–192.

Liang, J. (2004). Review on Chinese aphasic studies, *Contemporary Linguistics*, 2, 154–158.

Liang, J., and van Heuven, V.J. (2003). Evidence for separate tonal and segmental tiers in the lexical specification of words: A production and perception study of a brain-damaged Chinese speaker. *Brain and Language*, 87, 116–117.

Liang, J., and van Heuven, V.J. (2004a). Evidence for separate lexical tone and sentence intonation: A perception study of Chinese aphasic patients. *Brain and Language*, 91, 60–61.

Liang, J., and van Heuven, V.J. (2004b). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language*, 91, 282–293.

Liang, J., and van Heuven, V.J. (2005). Phonetic and phonological processing of pitch levels: A perception study of Chinese (aphasic) speakers. In J. Doetjes and J. van de Weijer (eds.), *Linguistics in the Netherlands* (pp. 125–137). Amsterdam: John Benjamins.

Liang, J. and van Heuven, V.J. (accepted). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven and T. Riad (eds.), *Tone and Intonation.* Berlin: Mouton de Gruyter.

Liang, J. and van Heuven, V.J. (under revision). Separate representations for lexical tone and sentence intonation? A perception study of Chinese aphasic patients. *Aphasiology*.

Lieberman, P., (1960). Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America,* 22, 451–454.

Lu, C., Bates, E., Li, P., Tzeng, O., Hung, D., Tsai, C., Lee, S., and Chung, Y., (2001). Judgment of grammaticality in aphasia: The special case of Chinese. *Aphasiology*, 14, 1021–1054.

Luria, A., (1964). Factors and Forms of Aphasia. In A. de Reuck and M. O'Conner (eds.), *Disorders of* Language (pp. 143–167). London: J. and A. Churchill Publishers.

Luria, A., (1947/1970). *Traumatic aphasia*. The Hague: Mouton. Reprinted in translation

Massaro, D.W., (1974). Perceptual Units in Speech Recognition. *Journal of Experimental Psychology*, 102, 349–353.

McClelland, J., and Elman, J., (1986). The TRACE model of speech perception. *Cognitive Psychology,* 18, 1–86.

Meulen, I. van der, (2004). *Syntactic movement and comprehension deficits in Broca's aphasia.* LOT PhD Dissertation 95, Utrecht.

Milberg, W., Blumstein, S.E., and Dworetzky, B., (1987). Processing of lexical ambiguities in aphasia. *Brain and Language,* 31, 38–150.

Milberg, W., Blumstein, S. Katz, D. Gershberg, F., and Brown T., (1995). Semantic facilitation in aphasia: Effects of time and expectancy. *Journal of Cognitive Neuroscience,* 7, 33–50.

Moen, I., (1993). Functional lateralization of the perception of Norwegian word tones — Evidence from a dichotic listening experiment. *Brain and Language,* 44, 400–413.

Moen, I., and Sundet, K., (1996). Production and perception of word tones (pitch accents) in patients with left and aright hemisphere damage. *Brain and Language,* 53, 267–281.

Moulines, E., and Verhelst, W., (1995). Time-domain and frequency-domain techniques for prosodic modification of Speech. In W.B. Kleijn and K.K. Paliwal (eds.), *Speech Coding and Synthesis* (pp. 519–555). Amsterdam: Elsevier Science.

Naeser, M.A., and Chan, W.-C.S., (1980). Case study of a Chinese aphasic with the Boston Diagnostic Aphasia Exam. *Neuropsychologia,* 18*,* 389–410*.*

Nooteboom, S.G., (1997). The prosody of speech: melody and rhythm. In W. J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences* (pp. 640–673). Basil Blackwell Limited, Oxford.

Packard, J., (1986). Tone production deficits in nonfluent aphasic Chinese speech. *Brain and Language*, 29, 212–223.

Packard, J., (1993). *A linguistic investigation of aphasic Chinese speech*. Dordrecht: Kluwer Academic Publishers.

Pallier, C., Colomé, A., and Sebastián-Gallés, N., (2001). The influence of native-language phonology on lexical access: Concrete vs. abstract lexical entries. *Psychological Science*, 12, 445–449.

Paulesu, E., Frith, C.D., and Frackowiak, R.J., (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R.S.J., and Frith, C.D., (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain*, 119, 143–157.

Petrides, M., and Pandya, D.N., (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Boller and J. Grafman (eds.), *Handbook of Neuropsychology* (pp. 17–58). New York: Elsevier.

Pike , K.L ., (1948). *Tone languages.* Ann Arbor: University of Michigan Press.

Pisoni, D.B., (1996). Word identification in noise. *Language and Cognitive Processes,* 11, 681–687.

Poeppel, D., and Hickok, G., (2004). Towards a new functional anatomy of language. *Cognition,* 92, 1–12.

Poeppel, D., (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55, 317–351.

Poeppel, D., (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25, 679–693.

Pöppel, E., (1997). A hierachical model of temporal perception. *Trends Cognitive Sciences,* 1, 56–61.

Pugh, K.R., Shaywitz, B.A., Shaywitz, S.E., Constable, R.T., Skudlarski, P., Fulbright, R. K., Bronen, R.A., Shankweiler, D.P., Katz, L., Fletcher, J.M., and Gore, J. C., (1996). Cerebral organization of component processes in reading. *Brain*, 119, 1221–1238.

Roo, E. de, (1999). *Agrammatic grammar, functional categories in agrammatic speech.* LOT PhD Dissertation 24, Holland Academic Graphics, The Hague.

Roo, E.M. de, Kolk, H., and Hofstede, B., (2003). Structural properties of syntactically reduced speech: a comparison of normal speakers and Broca's aphasics. *Brain and Language*, 86, 99–115.

Ross, E.D., and Mesulam, M.M., (1979). Dominant language functions of the right hemisphere? Prosody and emotional gesturing. *Archives of Neurology,* 36*,* 144–148.

Ross, E.D., Edmondson, J., and Seibert, G., (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics*, 14*,* 283–302.

Ryalls, J., (1982). Intonation in Broca's Aphasia. *Neuropsychologia*, 20, 355–360.

Ryalls, J., and Reinvang, I., (1986). Functional lateralization of linguistic tones: Acoustic evidence from Norwegian. *Language and Speech*, 29, 389–398.

Saffran, E.M., Marin, O.S.M., and Yeni-Komshian, G.H., (1976). An analysis of speech perception in word deafness. *Brain and Language*, 3, 209–228.

Sagart, L., (1999). The origin of Chinese tones. In S. Kaji (ed.), *Proceedings of the Symposium: Cross-Linguistic studies of Tonal Phenomena Tonogenesis, Typology, and Related Topics* (pp. 91–103). Tokyo: Tokyo University of Foreign Studies.

Schuell, H., Jenkins, J.J., and Jimenez-Pabon, E., (1964). *Aphasia in adults: Diagnosis, prognosis and treatment*. New York: Harper and Row.

Schwartz, M., and Tallal, P., (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 207**,** 1380–1382.

Seddoh, S.A., (2002). How discrete or independent are "affective prosody" and "linguistic prosody"? *Aphasiology,* 16, 683–692.

Sergent, J., Zuck, E., Le´vesque, M., and MacDonald, B., (1992). Positron emission tomography study of letter and object processing: Empirical findings and methodological considerations. *Cerebral Cortex*, 2, 68–80.

Shannon, C.E., and Weaver, W., (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Shapiro, B., and Danly, M., (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and Language*, 25, 19–36.

Shen, X.S., and Lin, M.C., (1991). A perceptual study of Mandarin Tone 2 and Tone 3. *Language and Speech*, 34**,** 145–156.

Shih, C.-L., (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83–109.

Shipley-Brown, F., Dingwall, W.O., Berlin, C.I., Yeni-Komshian, G., and Gordon-Salant, S., (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brian and Language,* 33, 16–26.

Stagray, J.R., and Downs, D., (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguisti*cs, 21, 143–163.

Stevens, K.N. (1992). Lexical access from features. *Speech Communication Group Working Papers,* 8, 119–144. Research Laboratory of Electronics. Massachusetts: Massachusetts Institute of Technology.

Swaab, T.Y., (1997). Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. *Journal of Cognitive Neuroscience,* 9. 39–66.

't Hart, J., Collier, R., and Cohen, A., (1990). *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Tallal, P., Miller, S., and Fitch, R.H., (1993). Neurobiology of speech perception. *Annals of the New York Academy of Science,* 682, 27–47.

Tallal, P., and Newcombe, F., (1978). Impairment in auditory perception and language comprehension in dysphasia. *Brain and Language*, 6, 13–24.

Tanner, W.P., and Rivette, G.L., (1964). Experimental study of 'tone deafness'. *Journal of the Acoustical Society of America,* 36, 1465–1467.

Thorsen, N.G., (1980). A study of the perception of sentence intonation − Evidence from Danish. *Journal of the Acoustical Society of America,* 67, 1014−1030.

Tsao, W., (1967). Question in Chinese. *Journal of Chinese Language Teachers' Association*, 2*,* 15–26.

Tse, J.K.P., (1978). Tone acquisition in Cantonese: A longitudinal case study. *Journal of Child Language,* 5, 191–204.

Tseng, C., (1981). *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Ph.D. dissertation, Brown University.

Tyler, L.K., Ostrin, R.K., Cooke, M., and Moss, H.E., (1995). Automatic Access of Lexical Information in Broca's Aphasics: Against the Automaticity Hypothesis. *Brain,* 48**,** 131–162.

Van Heuven, V.J., and Haan, J., (2000). Phonetic correlates of statement versus question intonation in Dutch. In A. Botinis (ed.), *Intonation: Analysis, Modelling and Technology* (pp. 119–144). Dordrecht: Kluwer.

Van Heuven, V.J., and Haan, J., (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven and N. Warner (eds.), *Papers in Laboratory Phonology* VII (pp. 61−86). Berlin: Mouton de Gruyter.

Van Lancker, D., (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication,* 13, 201–277.

Van Lancker, D., and Fromkin, V.A., (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics,* 1, 101–109.

Varley, R., and So, L.K.H., (1995). Age effects in tonal comprehension in Cantonese. *Journal of Chinese Linguistics*, 23, 76–97.

Wang, W. S.-Y., (1967). Phonological features of tone. *International Journal of American. Linguistics,* 33**,** 93–105.

Wang, W. S.-Y., (1972). The many uses of F_0. In A. Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 487–503). The Hague: Mouton.

Wang, W. S.-Y., and Li, K.-P., (1967). Tone 3 in Pekinese. *Journal of Speech Hearing Research,* 10*,* 629–636.

Wang, Y., Behne, D., Jongman, A., and Sereno, J., (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, 26, 449–466.

Wang, Y., Sereno, J.A., and Jongman, A., (in press). L2 acquisition and processing of Mandarin tone. In P. Li, E. Bates, L. Tan and O.J.L. Tzeng (eds.), *The Handbook of East Asian Psycholinguistics*. Cambridge: Cambridge University Press.

Wang, Y., Sereno, J., Jongman, A., and Hirsch, J., (2003). fMRI evidence for cortical modification during language learning. *Journal of Cognitive Neuroscience,* 15, 1019–1027.

Warren, R.M., (1999). *Auditory perception.* Cambridge: Cambridge University Press.

Weintraub, S., Mesulam, M.-M., and Kramer, L., (1981). Disturbances in prosody: A right-hemisphere contribution to language. *Archives of Neurology*, 38, 742–744.

Whalen, D.H., and Xu, Y., (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica,* 49, 25–47.

Wiese, R., (1997). Underspecification and the description of Chinese vowels. In J.L. Wang and N. Smith (eds.), *Studies in Chinese Phonology* (pp. 219–249). The Hague: Mouton.

Woo, N., (1969). *Prosody and phonology*. Doctoral dissertation. University of Massachusetts.

Xu, Y., (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, 95, 2240–2253.

Xu, Y., (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics,* 25, 61–83.

Xu, Y., and Xu, C.X., (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159–197.

Yip, M., (1991). *Tone tonal phonology of Chinese.* New York: Garland Press.

Yiu, E., and Fok, A., (1995). Lexical tone disruption in Cantonese aphasic speakers. *Clinical Linguistics and Phonetics,* 9, 79–92.

Yuan, J., Shih, C., and Kochanski, G.P., (2002). Comparison of declarative and interrogative intonation in Chinese. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 711–714). Aix-en-Provence, Laboratoire Parole et Langage.

Zatorre, R., Evans, A., Meyer, E., and Gjedde, A., (1992). Lateralization of phonetic and pitch processing in speech perception. *Science,* 256, 846–849.

Zatorre, R.J., Meyer, E., Gjedde, A., and Evans, A.C., (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex,* 6, 21–30.

Zhou, X., and Marslen-Wilson, W., (1997). The abstractness of phonological representation in the Chinese mental lexicon. In H.-C. Chen (ed.), *Cognitive Processing of Chinese and Related Asian Languages* (pp. 3–26). Hong Kong: The Chinese University Press.

Zurif, E., Swinney, D., Prather, P., Solomon, J., and Bushell, C., (1993). An On-Line Analysis of Syntactic Processing in Broca's and Wernicke's Aphasia. *Brain,* 45**,** 448–464.

Summary

This book investigates effects of brain lesions on prosodic processing in Chinese patients with Broca's aphasia with respect to the supra-segmental aspects pertaining to (i) lexical pitch features, i.e. lexical tones in Chinese and (ii) post-lexical pitch features, i.e. sentence intonation. The basic overall question is: Where and how is word prosody (lexical tone) represented in the mental architecture? Is it separate from segmental structure and is it separate from sentence prosody?

The goal of this study is to find out whether linguistic prosody at the word-level and sentence-level are processed equally well by Chinese patients with Broca's aphasia relative to normal subjects and second-language learners of Chinese. Laboratory-controlled experiments were carried out with Beijing aphasic patients who had a unilateral damage in the left hemisphere of the brain. We aimed to test how and to what extent they process Chinese lexical tones and intonation relative to healthy Beijing speakers and learners of Beijing dialect with and without a tone language as their native (L1) language.

We rest our investigation on the following three hypotheses. We started out from the assumption that language users may employ different mechanisms when processing speech in a tone language, depending on their native-language background. We investigated the degree of difficulty in language processing, i.e. the impairment of lexical tones and intonation, by patients with Broca's aphasia and compared this with the degree of processing difficulty experienced in the acquisition of Beijing Chinese as a tone language by healthy non-native learners from varying linguistic backgrounds. The idea is that native-language speakers and second-language (L2) learners have different mental lexicons and that this difference will be reflected in their speech behaviour. Among L2 learners, difference in L1 – e.g. whether tone language or not – will also lead to a different mental lexicon and the difference will in turn be observed in their speech behaviour.

Hypothesis One: Research has demonstrated that language acquisition begins at a very early age and proceeds at an amazingly fast pace. Early exposure to a language has a lasting impact on speech processing routines in adults such that the adults will be functionally deaf to new sound contrasts in an L2. Our hypothesis is that the degree of phonological 'deafness' increases as the phonological similarity between L1 and L2 decreases. For instance, L2 speakers of Beijing Chinese with a tone-less L1 will display a greater deafness to the tone contrasts in comparison with L2 speakers with a tone-language L1. Moreover, L2 speakers with a closely related

tone- language L1 will be more sensitive to the L2 tone contrasts than L2 speakers with a less closely related tone-language L1.

Hypothesis Two: If the locations in the left and right hemisphere are of paramount importance in determining hemispheric specialization, then unilateral lesions should result in uniform behavioural deficits regardless of the linguistic status of the prosodic elements. If, on the other hand, different linguistic functions result in hemispheric specialization, then we may expect behavioural deficits to vary accordingly.

Hypothesis Three: If, as suggested, language abilities are functions of the left hemisphere, lexical tone, which is phonemic and can distinguish between lexemes in tone languages, should play the same linguistic role as segments (e.g. vowels) do. Therefore, like vowels, lexical tone will also be subject to deficit following damage to the left hemisphere.

More generally, the hypotheses to be tested are whether lexical tone and linguistic intonation are separate functions with separate locations in the brain and to what extent brain lesion may affect linguistic prosody at the word and sentence levels. Although the importance of speech melody is generally well understood, relatively little attention has been paid to prosodic deficits following focal (i.e. restricted to a compact area) brain damage.

Following the above hypotheses, we carried out a comparative study on lexical tone and sentence intonation of Beijing dialect by patients with Broca's aphasia and two types of healthy speakers, i.e. L1 speakers (Beijing) as well as L2 Beijing speakers with either a tone-language L1 (Nantong, Changsha) or a tone-less L1 (Uygur).

Five main research questions were raised in Chapter One as follows:

(1) To what extent does brain lesion affect lexical tones in the speech production of Broca's patients? Are all lexical tones in Beijing dialect equally impaired in speech production? What happens to the vowels, which are the primary tone-bearing segments?
(2) To what extent does the lesion affect linguistic prosody in speech perception? Given that pitch variation is used at both the word and sentence level in Chinese, are lexical tone and sentence intonation equally impaired in the speech perception of Chinese aphasic patients?
(3) To what extent does phonological 'deafness' due to interference from the L1 affect linguistic prosody in speech perception? Are lexical tone and sentence intonation equally impaired in the perception of L2 speakers of Beijing Chinese?
(4) What are the characteristics of lexical tone identification by Chinese aphasic patients relative to healthy controls as well as to non-native speakers?
(5) In line with Question 4, what role does time pressure play in lexical tone identification? Is it true that the use of prosody by patients with Broca's aphasia

deteriorates progressively when they are allowed less time to perform language-processing tasks? If so, this would allow us to conclude that lexical tone deficit is due to processing limitation rather than to structural deficit.

Findings addressing these five main issues were reported in Chapters Two to Five and in Appendix I. The findings obtained from our lesion studies as well as the comparative studies with L2 learners lead us to the following conclusions:

(1) The specification of segmental and tonal aspects of lexical entries in Chinese, and in tone languages in general, are located/processed separately in the brain.
(2) The representation and/or processing of word-level prosody is separate from sentence-level prosody, i.e., production and perception of prosodic phenomena in speech, such as lexical tone and sentence intonation, engage multiple areas, which may comprise a large-scale spatially distributed network.
(3) Aphasic speakers with unilateral damage in the left hemisphere suffer from a structural deficit in the mental representation of the lexical tones, which can be compared with the structural deficit of L2 speakers regardless of their exposure to the target language (long or short), and their native language (tone or non-tone language).
(4) The perceptual system tries to elaborate a phonological representation with more detailed acoustic-phonetic information as more time is allowed, which blurs the clear cross-overs defined by mental representations.

Samenvatting

Dit boek onderzoekt effecten van hersenbeschadiging op de verwerking van prosodie (in dit onderzoek vooral toonhoogteverloop in spraak) door Chineestalige patiënten met Broca's afasie, op gebied van (i) lexicaal bepaalde eigenschappen van de toonhoogte, d.w.z. woordtonen in het Chinees en (ii) postlexicale eigenschappen, in het bijzonder zinsmelodie. De meer algemene onderzoeksvraag luidt: Waar en hoe is de woordprosodie (lexicale toon) gerepresenteerd in de architectuur van de menselijke geest? Is die representatie wel of niet gescheiden van de segmentele structuur enerzijds en van de zinsprosodie anderzijds?

Het doel van dit onderzoek is om na te gaan of linguïstische prosodie op woordniveau en op zinsniveau door Chineestalige patiënten met Broca's afasie even goed wordt verwerkt als door normale proefpersonen en door tweede-taalverwervers van het Chinees. Daartoe zijn laboratoriumexperimenten uitgevoerd met Chinese afasiepatiënten die het dialect van Beijing spraken en die een eenzijdige hersenbeschadiging hadden in de linker hersenhelft. We hebben onderzocht hoe (goed) zij de woordtonen en zinsmelodie van het Chinees konden verwerken in vergelijking met gezonde sprekers van het Beijing dialect en in vergelijking met leerders van het Beijing Chinees die zelf een moedertaal hadden met of zonder woordtonen.

Ons onderzoek is gebaseerd op de volgende drie hypotheses. We gaan uit van de aanname dat taalgebruikers verschillende mechanismen kunnen toepassen als zij spraak moeten verwerken in een toontaal, afhankelijk van hun moedertaal (L1). Het idee is dat moedertaalsprekers een ander mentaal lexicon hebben dan tweedetaalsprekers (L2) en dat dit verschil weerspiegeld wordt in het taalgedrag. Bij L2-leerders zal een verschillende L1-achtergrond – met name de vraag of hun moedertaal woordtonen heeft of niet – ook leiden tot een ander mentaal lexicon, en dat verschil zal op zijn beurt waarneembaar zijn in het taalgedrag van deze leerders.

Hypothese Een: Onderzoek heeft aangetoond dat taalverwerving op zeer vroege leeftijd begint en zich daarna verbazend snel ontwikkelt. Vroeg taalaanbod heeft blijvende gevolgen zodat volwassenen gebruik maken van ingeslepen spraakverwerkingsprocedures die hen functioneel doof maken voor nieuwe klankcontrasten in een L2. Onze hypothese is dat de ongevoeligheid voor nieuwe klankcontrasten toeneemt naar mate L1 en L2 fonologisch meer op elkaar lijken. Zo zullen L2-spekers van het Beijing Chinees met een L1 zonder woordtoon ongevoeliger zijn voor tooncontrasten dan L2-leerders die van huis uit een L1 hebben met woordtonen. En L2-leerders die zelf een L1 beheersen met een woordtoonsysteem dat sterk lijkt op dat van de doeltaal zullen gevoeliger zijn voor de toonverschillen in de L2 dan

L2-leerders die een toontaal spreken. En L2-leerders met een toontaal L1 zullen gevoeliger zijn voor de tonen in de doeltaal dan leerders met een L1-toontaal die verder afstaat van de doeltaal.

Hypothese Twee: Als de locatie in de linker of rechter hersenhelft doorslaggevend is voor hemisferische specialisatie, dan zullen eenzijdige beschadigingen leiden tot invariante uitvalverschijnselen ongeacht de linguïstische versus paralinguïstische status van de prosodische verschijnselen. Als daarentegen verschillende taalfuncties leiden tot hemisferische specialisatie, dan kunnen we verwachten dat de uitvalverschijnselen variabel zijn.

Hypothese Drie: Als, zoals gesuggereerd, taalstructurele vaardigheden functies zijn die zetelen in de linker hersenhelft, dan moet woordtoon (waarmee lexemen van elkaar onderscheiden worden in toontalen) dezelfde taalkundige rol spelen als de segmenten (b.v. klinkers). We voorspellen dat lexicale toon evenals klinkers onderhevig zijn aan uitval als gevolg van een beschadiging in de linker hersenhelft.

Meer algemeen, de vragen die we onderzoeken zijn of lexicale toon en linguïstisch gebruik van zinsmelodie gescheiden functies zijn met aparte locaties in de hersenen, en in welke mate een hersenbeschadiging het taalkundig gebruik van prosodie kan aantasten op woord- en zinsniveau. Hoewel het belang van de melodische eigenschappen van spraak redelijk goed bekend is, hebben prosodische uitvalverschijnselen als gevolg van een focale (compacte) hersenbeschadiging nog maar weinig aandacht gekregen.

Naar aanleiding van bovenstaande hypotheses hebben we een vergelijkend onderzoek uitgevoerd naar het gebruik van lexicale tonen en zinsmelodie door patiënten met Broca's afasie die Beijing Chinees als moedertaal hebben en twee typen gezonde controlesprekers, d.w.z. L1-sprekers van het Beijing dialect alsmede L2-leerders van het Beijing dialect die zelf ofwel een Sinitische toontaal als L1 hadden (Nantong, Changsha) ofwel een in China gesproken L1 zonder woordtonen (Uygur).

(1) In Hoofdstuk Een zijn de volgende vijf hoofdvragen aan de orde gesteld:
In welke mate tast hersenbeschadiging de woordtonen aan in de spraakproductie van patiënten met Broca's afasie? Worden alle woordtonen in hun Beijing dialect in gelijke mate aangetast? En wat gebeurt er met de klinkers, die de primaire toondragende segmenten zijn?
(2) In hoeverre tast de hersenbeschadiging het talig gebruik van prosodie aan in de spraakwaarneming? Gegeven dat toonhoogteveranderingen in het Chinees zowel op woord- als op zinsniveau worden gebruikt, worden lexicale tonen en zinsintonatie dan in gelijke mate aangetast in de spraakwaarneming door Chinese afasiepatiënten?
(3) In hoeverre heeft de waarneming van talig gebruik van prosodie in een L2 te lijden van fonologische 'doofheid' als gevolg van interferentie vanuit de

moedertaal? Zijn lexicale toon en zinsintonatie in dezelfde mate aangetast in de waarneming van L2-leerders van het Beijing Chinees?

(4) Hoe verhoudt zich de identificatie van lexicale tonen door Chinese afasie-patiënten tot die van gezonde controlepersonen en van L2-leerders?

(5) Als vervolg op vraag 4, welke invloed heeft tijdsdruk op de identificatie van lexicale tonen? Verslechtert de verwerking van prosodie bij patiënten met Broca's afasie disproportioneel als die minder tijd krijgen om hun taak uit te voeren? Een antwoord op deze vraag staat ons toe te concluderen of uitval van lexicale tonen het gevolg is van een beperking op de verwerkingscapaciteit of eerder teruggaat op een gebrek in de structurele representatie van de tonen.

De bevindingen met betrekking tot de hoofdvragen worden gerapporteerd in de Hoofdstukken Twee t/m Vijf en in Appendix I. De resultaten van ons onderzoek met patiënten en met L2-leerders leiden tot de volgende conclusies:

(1) De specificatie (en verwerking) van segmentele en tonale aspecten van lexicale eenheden (woorden, morfemen) in het Chinees, en in toontalen in het algemeen, vindt plaats op gescheiden locaties in het brein.

(2) De representatie en/of verwerking van woordprosodie is gescheiden van die van zinsprosodie, met andere woorden: productie en perceptie van prosodische verschijnselen in spraak, zoals woordtonen en zinsmelodie, verloopt via een veelheid van hersengebieden die tezamen een grootschalig ruimtelijk gedistribueerd netwerk vormen.

(3) Afatische sprekers met een eenzijdige beschadiging in de linker hersenhelft hebben te kampen met een gebrek in de structurele representatie van de woordtonen, dat vergelijkbaar is met de gebrekkige structurele representatie bij L2 sprekers, ongeacht de ervaringstijd met de doeltaal en ongeacht de aard van hun moedertaal (al dan niet een toontaal).

(4) Het perceptief systeem poogt een meer gedetailleerde fonologische representatie uit te werken naar mate meer verwerkingstijd beschikbaar is, waardoor de scherpe perceptieve grenzen tussen tooncategorieën in de mentale representatie vervagen.

Index of Subjects

Q

R

S

T

V

W

X

Y

Curriculum Vitae

Liang Jie was born on 6th May, 1960 in Urumqi, Xinjiang, P.R. China. In 1982, she obtained a Bachelor's Degree at Xinjiang University, majoring in English. In 1990 she graduated from Beijing University of Science and Technology specializing in General and Applied Linguistics, and in 2002 she obtained a Doctor's degree in Arts from the Chinese Department of Nankai University. In de academic year 2000/01 she worked as a research associate in the Leiden University Phonetics Laboratory with a grant from the China Scholarship Council. From 2002 until 2005 (with interruptions) she was 'assistent-in-opleiding' (paid PhD researcher) at the Leiden University Centre for Linguistics (LUCL) working on the research project that is the topic of the present dissertation. Liang Jie is now professor of Linguistics and dean of the College of Foreign Languages at Xinjiang Normal University.