Phonetic and phonological processing of pitch levels

A perception study of Chinese (aphasic) speakers

Jie Liang and Vincent J. van Heuven
Universiteit Leiden Centre for Linguistics

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 & 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 & Grodzinsky, 1997). Alternatively, the account is not in terms of a representational deficit, but the problem is rather caused by processing limitations (Friederici & Frazier, 1992; Kolk, 1995). Our question

Linguistics in the Netherlands 2005, 125–137.
ISSN 0929–7332 / E-ISSN 1569–9919

for the present paper is whether the lexical tone impairment in Chinese is due to 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 & 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 & 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 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 (Boersma & Weenink, 1996). 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 & 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

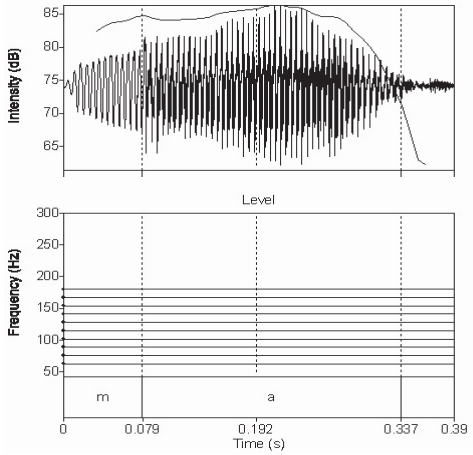

Figure 1. Steps 1 through 10 along a resynthesized continuum differing in overall pitch level by 2-ST increments.

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.

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.[1] 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).

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.[2] When there was no time pressure, there was a fixed 3000-ms interstimulus 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.[3] 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 & van Heuven,

Table 1. MRI or CT scan findings of the lesion site in left hemisphere for individual 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
Patients	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			●					●						

2004a, 2004b). 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.

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.[4] 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 [$F(9, 580) = 11.0$ ($p < 0.001$)], 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 slow down 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.

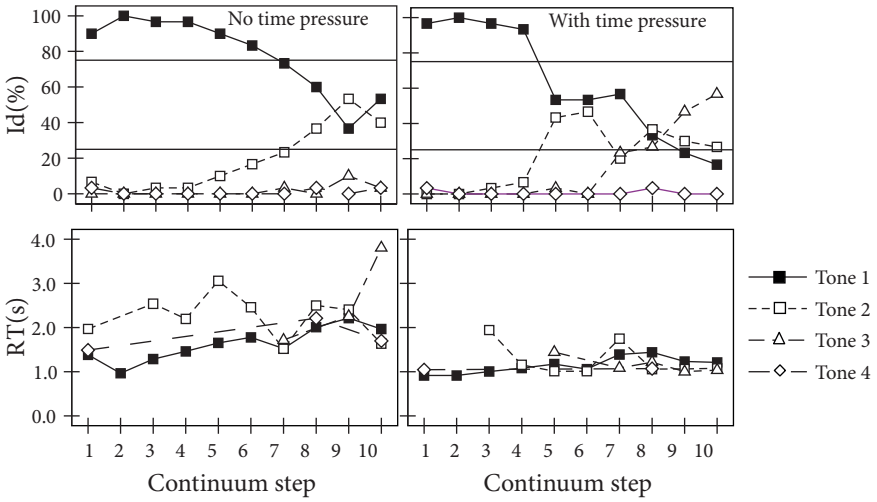

Figure 2. Beijing listeners' tone identifications (percent, upper panels) and reaction time (seconds, lower panels) as a function of pitch level (steps) broken down by time pressure (left vs. right). Missing data points in the reaction-time curves occur where a response alternative was never chosen by any of the listeners.

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 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.

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.

In the absence of time pressure (left 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, 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

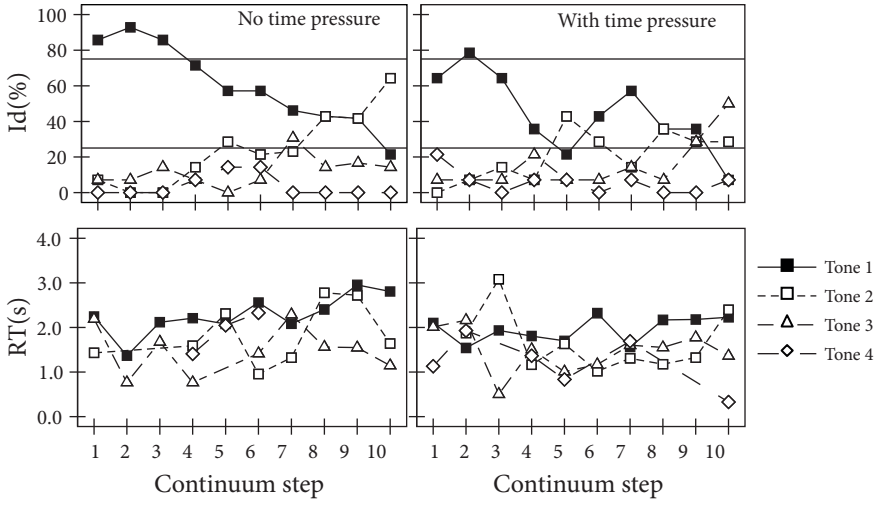

Figure 3. Beijing aphasic listeners' tone identifications (percent, upper panels) and associated reaction time (seconds, lower panels) as a function of pitch level (steps), broken down by time pressure (left vs. right).

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. Their 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.[5] 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).

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. We also find 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

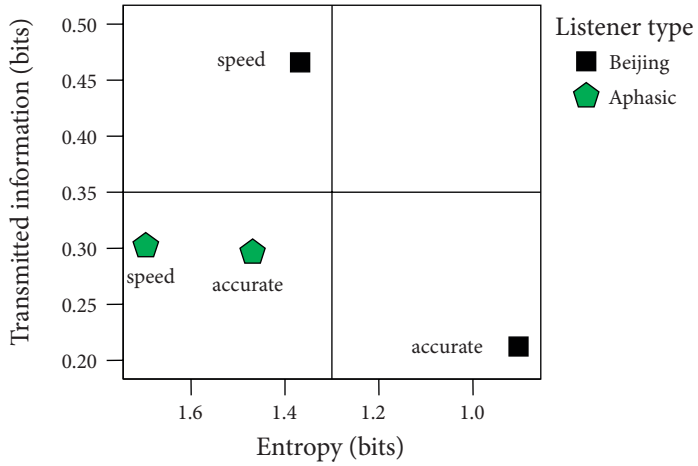

Figure 4. Transmitted information plotted against response entropy broken down by listener type (Beijing vs. aphasic) and time pressure (accurate vs. speed).

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 on 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 realise 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 prelexical 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.

Notes

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

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

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

4. 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 Liang, 2006).

5. Transmitted information (see also Shannon & Weaver, 1949; Attneave, 1959): $T_{(x;y)} = H_{(x)} + H_{(y)} - H_{(x,y)}$, where H is the entropy. $H_{(x)} = -\Sigma(p_i\ ^2\log p_i)$, where i is an index ranging over the categories 1 through 10 along our stimulus dimension (x). $H_{(y)} = -\Sigma(p_j\ ^2\log p_j)$, where j ranges over the response categories 1 to 4. $H_{(x,y)} = -\Sigma(p_{ij}\ ^2\log p_{ij})$.

References

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

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

Caplan, D. (1983) 'Syntactic and semantic structures in agrammatism.' In M.L. Kean (ed.) *Agrammatism*. Academic Press, New York.

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

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

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

Goldinger, S.D. (1996) 'Words and voices: Episodic traces in spoken word identification and recognition memory.' *Journal of Experimental Psychology: Learning, Memory, & Cognition* 22, 1166–1183.

Hagiwara, H. (1995) 'The breakdown of functional categories and the economy of derivation.' *Brain and Language* 50, 92–116.

Klatt, D.H. (1979) 'Speech perception: A model of acoustic-phonetic analysis and lexical access.' *Journal of Phonetics* 7, 279–312.

Kolk, H.H.J. (1995) 'A time-based approach to agrammatic production.' *Brain and Language* 50, 282–303.

Liang, J. (2006). *Experiments on the modular nature of word and sentence phonology: comparative study of Chinese and Dutch aphasics*. LOT dissertation series, Utrecht (in preparation).

Liang, J. & V.J. van Heuven (2004a) Evidence for separate lexical tone and sentence intonation: A perception study of Chinese aphasic patients. *Brain and Language*, 91, 60–61.

Liang, J. & V.J van Heuven (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.

Massaro, D.W. (1974) 'Perceptual Units in Speech Recognition.' *Journal of Experimental Psychology* 102, 349–353.

McClelland, J. & J. Elman (1986) 'The TRACE model of speech perception.' *Cognitive Psychology* 18, 1–86.

Moulines, E. & W. Verhelst (1995) 'Time-domain and frequency-domain techniques for prosodic modification of Speech.' In W.B. Kleijn & K.K. Paliwal, eds., *Speech Coding and Synthesis*, 519–555. Elsevier Science, Amsterdam.

Packard, J.L. (1986) 'Tone production deficits in nonfluent aphasic Chinese speech.' *Brain and Language* 29, 212–223.

Pallier, C., A. Colomé & N. Sebastián-Gallés (2001) 'The influence of native-language phonology on lexical access: Concrete vs. abstract lexical entries.' *Psychological Science* 12, 445–449.

Pisoni, D.B. (1996) 'Word identification in noise.' *Language and Cognitive Processes* 11, 681–687.

Shannon, C.E. & W. Weaver (1949) *The mathematical theory of communication*, University of Illinois Press, Urbana.

Whalen, D. H., & Xu, Y. (1992). 'Information for Mandarin tones in the amplitude contour and in brief segments'. *Phonetica* 49, 25–47.