Lund University, Dept. of Linguistics
Working Papers 51 (2005), 257–274

Syllable boundary investigation of some word pairs in Standard Chinese

Cuiling Zhang and Gösta Bruce

Abstract
In this paper we compare the acoustic difference between the two words in three categories of minimal pairs in Standard Chinese and make statistical analyses. The words in a pair are easily confused when spoken in isolation. The pairs in category 1 and 2 consist of two disyllabic words which have exactly the same phonemes and tonal combinations, but different syllable boundary. The pairs in category 3 consist of one monosyllabic and one disyllabic word that have the same phonemes and phonetically similar tones. All words were recorded and wideband spectrum analysis was made in *Praat*. Detailed investigations were conducted on the duration, formant pattern, pitch and intensity for each pair, and finally their effects on the determination of the syllable boundary and discrimination of the two words in a pair were investigated briefly, which can be useful for automatic speech recognition and foreign language learning.

1. Introduction

In Standard Chinese there are some word pairs with exactly the same phonemes and tonal combinations, but different syllable boundaries and meanings. These words can easily be distinguished auditorily by native speakers but may be a little difficult for people who are not native speakers of Chinese. Furthermore, both for native and non-native Chinese speakers, it is difficult to distinguish them from spectra only, without supporting auditory analysis, especially to determine the syllable boundaries. Automatic syllable segmentation is difficult as well. In this sense it is worthwhile to study these pairs to find acoustical cues for distinguishing them and for detecting syllable boundaries which can be used for machine recognition of speech as well as for foreign language learning.

For our investigation we collected three types of minimal pairs. Pairs of the first category consist of two disyllabic words such as /chai'iou/ ~ /cha'iou/. The first syllable of the first word ends with a vowel (/i/ in this example) or a nasal and the second syllable starts with the same segment, so that there are double vowels or nasals around the syllable boundary. The second word has the same phoneme combination as the first one except that there is a single vowel or nasal in the middle. A pair of the second category

also consists of two disyllabic words, such as /da'ni/ ~ /dan'i/, which have exactly the same phonemes but different positions of the syllable boundary. The third category is different from the former two. A pair in this category consists of a monosyllabic word and a disyllabic word such as /bian/ ~ /bi'an/. But they have also the same phoneme combinations and aurally similar tones. It is quite easy to confuse them when they are spoken in isolation at high speed.

At present few similar studies on this have been reported for Standard Chinese. Our study aims to find certain acoustical cues for distinguishing the pairs and detecting the syllable boundary as exactly as possible by aural and visual methods.

2. Experimental method

2.1 Material

Our speech material consists of 12 pairs of type 1, 7 pairs of type 2 and 12 pairs of type 3. Table 1 lists all pairs in the three categories. All words are given in a phonemic transcription based on the Chinese *pinyin* transcription. There are four tones in Standard Chinese: 1. high tone, 2. rising tone, 3. falling–rising tone, and 4. falling tone. The numbers following syllables in the tables and text below represent the tonal combinations.

2.2 Speaker

The speaker who is 36 years old comes from northeastern China and speaks Standard Chinese without evident dialect or special speech habits.

2.3 Procedure

All words in the speech material were ordered randomly and read by the speaker as isolated words. His speech was recorded five times on Fuji digital tape with a SV-3700 Panasonic recorder in the studio at the Department of Linguistics and Phonetics, Lund University. All recorded syllables were then analyzed in *Praat*. The sample rate was 16 kHz and wideband spectra were made and analyzed. Detailed investigations were conducted on the duration, formant pattern, pitch and intensity for each pair, especially for their effects on the determination of syllable boundary and the discrimination of the two words in a pair. The durations of all phonemes in the pair and the first four formant frequencies at three points (the start, mid and end points) as well as the pitch value at the syllable boundary were extracted and analyzed statistically.

Table 1. Minimal pairs in the three categories

Pairs in category 1		Pairs in category 2		Pairs in category 3	
chai'iou^{22}	'diesel oil'	da'ni^{33}	'beat you'	bian^{3}	'flat'
cha'iou^{22}	'tea and oil'	dan'i^{33}	'whisk the chair'	bi'an^{34}	'the other shore'
chai'i^{14}	'servant'	da'nü^{43}	'elder daughter'	dian^{4}	'but'
cha'i^{14}	'difference'	dan'ü^{43}	'pelter'	di'an^{14}	'dike'
mai'i^{34}	'personal name'	na'ni^{23}	'take you'	huan^{4}	'change'
ma'i^{34}	'personal name'	nan'i^{23}	'difficult to'	hu'an^{44}	'revetment'
zhai'iao^{14}	'abstract'	ji'niao^{34}	'emiction'	jian^{4}	'cheap'
zha'iao^{14}	'injection'	jin'iao^{34}	'important'	ji'an^{14}	'long-unsolved cases'
jian'nan^{12}	'hardship'	qi'nü^{23}	'rare woman'	jiou^{3}	'alcohol'
jia'nan^{12}	'difficult family'	qin'ü^{23}	'language of Qin dynasty'	ji'ou^{13}	'odd and even'
sun'nan^{12}	'personal name'	xi'ni^{12}	'watery mud'	lian^{4}	'practise'
su'nan^{12}	'person's name'	xin'i^{12}	'doubt'	li'an^{44}	'register'
jun'nan^{42}	'handsome man'	xi'ni^{44}	'delicate'	liou^{2}	'keep'
ju'nan^{42}	'great man'	xin'i^{44}	'faith'	li'iou^{32}	'reason'
nan'nai^{24}	'intolerabile'			piao^{3}	'glance'
nan'ai^{24}	'difficult to love'			pi'ao^{23}	'leather clothes'
pan'ni^{43}	'expecting your coming'			tuan^{2}	'corps'
pa'ni^{43}	'be afraid of you'			tu'an^{24}	'pattern'
lan'nü^{23}	'blue woman'			shuan^{4}	'rinse'
lan'ü^{23}	'blue rain'			shu'an^{14}	'desk'
tian'nü^{13}	'heavenly woman'			tian^{2}	'field'
tian'ü^{13}	'rain from sky'			ti'an^{24}	'resolution'
xia'an^{11}	'blind installation'			xiou^{1}	'shy'
xi'an^{11}	'a city name'			xi'ou^{11}	'Western Europe'

3. Results and discussion

3.1 Minimal pairs in category 1

3.1.1 Duration. We compared the absolute duration of words for each minimal pair in category 1 (see Table 2). According to our expectation, the total duration of the first word of a pair, which contains double vowels or nasals around the syllable boundary, should be longer than the duration of the second word because there is one more phoneme in the first word. Only 2 of 12 pairs (shown in bold), however, are coincident with our expectation, 9 pairs are not, and the two words of one pair (/jun'nan^{42}/ ~ /ju'nan^{42}/) have almost equal duration. It seems less effective to compare the absolute duration of syllables for each minimal pair since no evident and consistent

Table 2. Duration of minimal pairs in category 1

	total dur. (ms)	first cons.	syll. 1	syll. 2	S1/S2	mid phoneme(s)
chai'iou^{22}	667	22%	58%	42%	1.4	28%
cha'iou^{22}	736	26%	54%	46%	1.2	15%
chai'i^{14}	**635**	**24%**	62%	38%	1.6	52%
cha'i^{14}	**534**	**19%**	56%	44%	1.3	44%
mai'i^{34}	520	**13%**	52%	48%	1.1	67%
ma'i^{34}	560	**5%**	50%	50%	1.0	50%
zhai'iao^{14}	510	5%	54%	46%	1.2	28%
zha'iao^{14}	540	5%	50%	50%	1.0	21%
jian'nan^{12}	605	10%	44%	56%	0.8	17%
jia'nan^{12}	617	9%	40%	60%	0.7	17%
sun'nan^{12}	739	**25%**	52%	48%	1.1	16%
su'nan^{12}	794	**33%**	52%	48%	1.1	9%
jun'nan^{42}	575	**7%**	33%	67%	0.5	24%
ju'nan^{42}	573	**12%**	41%	58%	0.7	12%
nan'nai^{24}	**517**	10%	53%	47%	1.1	19%
nan'ai^{24}	**489**	11%	62%	38%	1.6	13%
pan'ni^{43}	715	17%	44%	56%	0.8	19%
pa'ni^{43}	728	17%	47%	53%	0.9	18%
lan'nü^{23}	684	8%	42%	58%	0.7	14%
lan'ü^{23}	781	9%	44%	56%	0.8	6%
tian'nü^{13}	778	16%	51%	49%	1.0	14%
tian'ü^{13}	820	17%	53%	47%	1.1	10%
xia'an^{11}	706	23%	54%	46%	1.2	62%
xi'an^{11}	736	24%	57%	43%	1.3	39%

tendency was found between the two disyllabic words. Besides, the duration of syllables can change a lot under different speech conditions, and even under the same condition.

We also normalized the duration for each word, syllable and phoneme in the pair and obtained the duration percentage of each phoneme and syllable in the whole word. Table 2 lists the absolute duration of the whole word in each pair, the relative durations of the two syllables, and their duration ratios. The duration of the first consonant, of the doubled vowel or nasal in the middle part of the first word, and their counterparts in the second word is also shown.

The result shows that there is little duration difference between the initial consonants in most pairs but greater difference (more than 5%) in a few cases (shown in bold). There are also certain differences for other phonemes, but

no regularity was found. The duration ratios show that the two words in each pair have similar duration distributions for the two syllables. It is possible that this is due to the tonal combinations in the words. Thus it is difficult to distinguish the two words in a pair by the duration except for the middle phoneme. There is a consistent tendency that the duration of the double phonemes in the first word in a pair is longer than its single counterpart (except for the pair /jian'nan^{12}/ ~ /jia'nan^{12}/ with equal durations). This seems to be the only cue for distinguishing the words of a pair by their duration. On the average the duration of the double phoneme is 9% longer than its single counterpart, ranging from 0% (/jian'nan^{12}/ ~ /jia'nan^{12}/) to 17% (/mai'i^{34}/ ~ /ma'i^{34}/). Figure 1 shows spectrograms of one pair in category 1 to illustrate the difference between the two words.

3.1.2 Formant pattern. Examination of the formant patterns for all pairs in category 1 indicates that there is no great difference between the two words in a pair as regards the voiced consonants in the first syllable and the vowels in the second syllable. The most evident difference of formant structure exists in the vowels of the first syllable and in the interval between the two syllables in each pair, especially on the first three formants. Generally the transition between the phonemes in one syllable is more natural and gradual than that between phonemes belonging to two syllables or close to the syllable boundary. The first four formant frequencies of all phonemes were extracted automatically at three points (start, mid and end) to make statistical analyses. Generally speaking, these three points can represent the formant trajectory of each phoneme. Table 3 lists the frequencies of the first two formants, and some data were corrected manually.

Observation of the formant distribution shows that the formant frequencies and the distances between them differ more or less between the two words for the main segments, especially the distance between F1 and F2 and between F2 and F3. The result of paired sample two-tailed *t*-tests for all four formant frequencies at these three points in the two words of each pair is that only the *p*-value of F2 (0.002) shows a significant difference at the 0.01 level, which indicates that F2 is more important than the other three formants for distinguishing the words of a pair.

For further exploring the difference in formant patterns between the words in a pair we also calculated the absolute distances between the formant frequencies (F2–F1, F3–F2 and F4–F3) at three points for the main phonemes. We also calculated the relative distances between them. Table 4

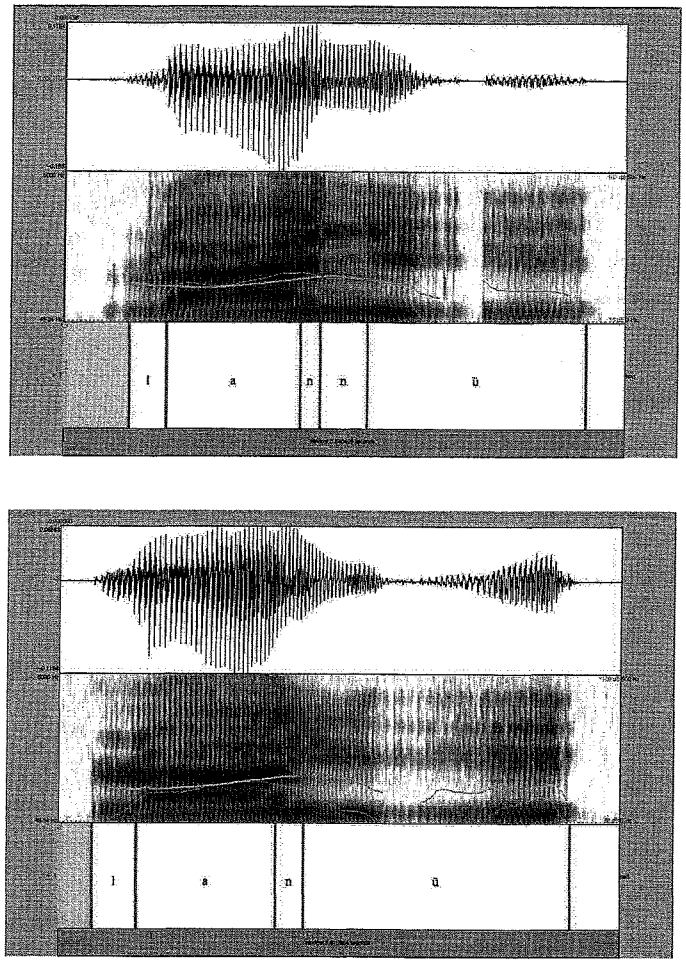

Figure 1. Wideband spectrograms for a pair of category 1: /lan'nü^{23}/ (top) vs. /lan'ü^{23}/ (bottom)

lists the ratios (F3–F2)/(F2–F1) = *ratio1*, and (F4–F3)/(F2–F1) = *ratio2*, which show the differences of the formant distribution for each pair more clearly.

Table 4 shows these ratios for the phonemes of each pair. The results of paired sample *t*-tests for all ratio values from the two words at three points in each pair are that the *p*-values are 0.002 for *ratio1* and 0.006 for *ratio2*,

Table 3. The first two formant frequencies for minimal pairs in category 1 (Hz)

	syllables	F1			F2				syllables	F1			F2		
		start	mid	end	start	mid	end			start	mid	end	start	mid	end
a	chai'i^{14}	617	666	595	1319	1651	1863	a	mai'i^{34}	695	821	714	1341	1571	1697
	cha'i^{14}	990	802	488	1808	1168	1790		ma'i^{34}	669	816	444	867	1130	1910
I	chai'i^{14}	577	241	228	1894	2167	2231	i	mai'i^{34}	714	260	227	1697	2154	2118
	cha'i^{14}	412	214	222	1878	2009	2109		ma'i^{34}	372	210	224	2164	2144	2193
a	chai'iou^{22}	762	696	595	1575	1659	1792	a	nan'nai^{24}	591	933	563	1537	1629	1611
	cha'iou^{22}	744	800	545	1056	1140	1658		nan'ai^{24}	423	926	605	1550	1607	1455
I	chai'iou^{22}	563	274	478	1824	2133	1343	n	nan'nai^{24}	398	332	318	1615	1659	1515
	cha'iou^{22}	456	278	369	1824	2060	1812		nan'ai^{24}	466	541	675	1500	1424	1436
i	jian'nan^{12}	347	420	480	1859	1844	1817	i	tian'nü^{13}	299	362	445	2166	2164	1966
	jia'nan^{12}	479	705	900	1643	1490	1387		tian'ü^{13}	341	396	426	1877	1931	1912
a	jian'nan^{12}	497	727	398	1795	1657	1573	a	tian'nü^{13}	455	624	361	1941	1854	1989
	jia'nan^{12}	915	810	624	1324	1171	1331		tian'ü^{13}	445	635	338	1873	1711	1936
n	jian'nan^{12}	284	316	369	1613	1795	1512	n	tian'nü^{13}	325	335	286	1952	1957	1665
	jia'nan^{12}	307	307	325	1339	1755	1487		tian'ü^{13}	349	320	298	1934	1998	2191
u	sun'nan^{12}	403	343	330	1427	1828	2162	i	xia'an^{11}	444	603	770	1871	1745	1510
	su'nan^{12}	357	357	357	1088	775	1244		xi'an^{11}	212	266	512	2171	2080	1899
n	sun'nan^{12}	310	310	257	1766	1724	1628	a	xia'an^{11}	798	890	629	1424	1221	1552
	su'nan^{12}	278	304	290	1164	2000	1573		xi'an^{11}	577	808	496	1828	1435	1876
u	jun'nan^{42}	305	344	340	2116	1763	2397	a	pan'ni^{43}	609	936	813	1597	1665	1728
	ju'nan^{42}	304	301	331	2232	2030	1704		pa'ni^{43}	952	943	436	1126	1240	1972
n	jun'nan^{42}	332	275	305	2428	1685	1384	n	pan'ni^{43}	761	287	279	1792	1717	2006
	ju'nan^{42}	109	282	305	1692	2123	1443		pa'ni^{43}	314	297	283	1958	1746	1121
a	zhai'iao^{14}	528	674	608	1538	1621	1809	a	lan'nü^{23}	602	867	643	1411	1510	1780
	zha'iao^{14}	675	804	563	1375	1165	1651		lan'ü^{23}	583	902	625	1555	1514	1620
i	zhai'iao^{14}	583	280	359	1842	2219	2061	n	lan'nü^{23}	545	318	308	1855	1663	1924
	zha'iao^{14}	497	402	725	1783	2157	1181		lan'ü^{23}	578	410	327	1660	1762	2004

values which are far less than the 0.01 level of significance. This means that the difference in formant distribution between two words in a pair is significant and this should be an important cue for distinguishing the pair.

3.1.3 Pitch, intensity and syllable boundary. The pitch contours for the two words of a pair are very similar since they have the same tonal combination. No obvious syllable boundary was found in the pitch. Table 5 lists the pitch values at the syllable boundaries in words of category 1 and 2; some data are absent because the pitch contour is discontinuous and cut off at the syllable

Table 4. Formant distribution ratios for pairs of category 1

		ratio1			*ratio2*					*ratio1*			*ratio2*		
		start	mid	end	start	mid	end			start	mid	end	start	mid	end
a	chai'iou^{22}	0.9	0.7	0.6	1.2	1.1	0.8	a	chai'i^{14}	0.7	0.7	0.6	2.1	1.0	0.7
	cha'iou^{22}	3.7	3.1	0.9	4.0	3.5	0.8		cha'i^{14}	1.7	2.9	0.8	1.8	3.5	0.6
I	chai'iou^{22}	0.6	0.5	1.3	0.8	0.4	1.4	I	chai'i^{14}	0.6	0.5	0.5	0.6	0.3	0.2
	cha'iou^{22}	0.6	0.5	0.5	0.7	0.3	0.7		cha'i^{14}	0.8	0.8	0.7	0.4	0.1	0.2
i	jian'nan^{12}	0.6	0.6	0.6	0.6	0.7	0.8	a	mai'i^{34}	1.3	1.0	0.6	1.8	1.4	1.1
	jia'nan^{12}	0.9	1.1	1.7	0.7	1.4	2.6		ma'i^{34}	8.2	4.3	0.5	4.7	3.1	0.7
a	jian'nan^{12}	0.6	1.0	0.8	0.8	1.1	1.0	i	mai'i^{34}	0.6	0.5	0.5	1.1	0.4	0.3
	jia'nan^{12}	2.2	3.1	1.7	3.0	3.4	1.7		ma'i^{34}	0.5	0.6	0.5	0.2	0.3	0.4
n	jian'nan^{12}	0.7	0.5	0.9	1.0	0.5	1.4	i	tian'nü^{13}	0.3	0.4	0.4	0.4	0.4	0.6
	jia'nan^{12}	1.1	0.6	0.9	1.2	0.4	1.3		tian'ü^{13}	0.5	0.4	0.5	0.7	0.7	0.6
a	nan'nai^{24}	1.2	1.3	1.0	1.1	1.5	1.0	a	tian'nü^{13}	0.4	0.6	0.6	0.6	0.7	0.4
	nan'ai^{24}	0.8	1.2	1.0	1.2	1.5	1.3		tian'ü^{13}	0.5	0.8	0.5	0.7	0.9	0.5
n	nan'nai^{24}	0.8	0.6	0.8	0.8	0.6	1.1	n	tian'nü^{13}	0.6	0.7	1.0	0.4	0.8	0.2
	Nan'ai^{24}	0.8	1.0	1.2	1.1	1.4	1.5		tian'ü^{13}	0.6	0.6	0.5	0.5	0.3	0.3
u	sun'nan^{12}	1.1	0.4	0.3	0.8	0.5	0.5	i	xia'an^{11}	0.7	0.8	1.4	0.6	0.8	1.3
	su'nan^{12}	0.7	0.8	1.6	0.0	0.3	1.0		xi'an^{11}	0.6	0.6	0.6	0.4	0.3	0.7
n	sun'nan^{12}	0.5	0.6	0.6	0.8	0.6	1.2	a	xia'an^{11}	1.7	3.4	0.9	1.5	3.3	1.3
	su'nan^{12}	1.6	0.4	0.8	1.1	0.6	0.9		xi'an^{11}	0.7	1.7	0.7	0.9	1.7	0.7
u	jun'nan^{42}	0.2	0.4	0.2	0.6	0.7	0.1	a	pan'ni^{43}	0.8	1.2	1.0	0.8	1.3	0.9
	ju'nan^{42}	0.4	0.2	0.7	0.3	0.5	0.6		pa'ni^{43}	8.1	4.1	0.6	7.4	4.1	0.6
n	jun'nan^{42}	0.2	0.6	0.9	0.1	0.6	1.5	n	pan'ni^{43}	0.8	0.8	0.6	0.9	0.5	0.5
	ju'nan^{42}	0.6	0.2	0.9	0.5	0.6	1.2		pa'ni^{43}	0.6	0.2	2.2	0.4	0.6	0.0
a	zhai'iao^{14}	0.7	0.8	0.6	0.7	1.0	0.9	a	lan'nü^{23}	1.3	1.3	0.7	0.7	1.9	0.8
	zha'iao^{14}	1.0	2.8	0.9	1.2	3.2	0.9		lan'ü^{23}	0.4	1.5	0.9	0.9	1.8	1.0
i	zhai'iao^{14}	0.5	0.5	0.3	0.8	0.3	0.6	n	lan'nü^{23}	0.6	1.0	0.7	0.6	0.3	0.3
	zha'iao^{14}	0.7	0.4	2.9	0.7	0.5	2.4		lan'ü^{23}	0.8	0.6	0.4	0.9	0.6	0.3

boundary. The data in Table 5 shows that the pitch values at the syllable boundary are close to each other for the two words in a pair, and there is no statistically significant difference between them. There are a few exceptions where the pitch differences are more than 18 Hz (shown in bold). The reason for this is not yet clear. In pairs of category 1, the average F0 difference at the syllable boundary is 10 Hz, ranging from 3 Hz for the pair /chai'i^{14}/ ~ /cha'i^{14}/ to 21 Hz for /jian'nan^{12}/ ~ /jia'nan^{12}/. In pairs of category 2, the average F0 difference at the syllable boundary is 9 Hz, ranging from 1 Hz for /na'ni^{23}/ ~ /nan'i^{23}/ to 23 Hz for /xi'ni^{12}/ ~ /xin'i^{12}/.

It seems difficult to discriminate the words in a pair by their pitch contours. Also the peaks and valleys of the pitch contour are not as effective cues for segmenting the two syllables as we expected in this case. No more information which could help in detecting the syllable boundary was found in

Table 5. F0 at the syllable boundary in words of categories 1 and 2 (Hz)

Pairs in category 1				Pairs in category 2	
chai'iou^{22}	**132**	jun'nan^{42}	127	da'ni^{33}	160
cha'iou^{22}	**151**	ju'nan^{42}	112	dan'i^{33}	167
chai'i^{14}	159	nan'nai^{24}	173	da'nü^{43}	111
cha'i^{14}	162	nan'ai^{24}	165	dan'ü^{43}	118
mai'i^{34}	104	pan'ni^{43}	**126**	na'ni^{23}	159
ma'i^{34}	113	pa'ni^{43}	**108**	nan'i^{23}	158
zhai'iao^{14}	159	lan'nü^{23}	149	ji'niao^{34}	
zha'iao^{14}	155	lan'ü^{23}	158	jin'iao^{34}	
jian'nan^{12}	**147**	tian'nü^{13}	157	qi'nü^{23}	156
jia'nan^{12}	**168**	tian'ü^{13}	162	qin' ü^{23}	166
sun'nan^{12}	154	xia'an^{11}	132	xi'ni^{12}	**123**
su'nan^{12}	159	xi'an^{11}		xin'i^{12}	**146**
				xi'ni^{44}	124
				xin'i^{44}	116

the pitch contour though sometimes the peaks or valleys are quite close to the syllable boundary.

The intensity contours of the two words in a pair behave in a way similar to the pitch contours. No special regularity was found, and no evident cue for detecting the syllable boundary effectively for the words in a pair was found. This can be done only by aural and visual methods including listening and examination of the waveform. Judging from listening, the syllable boundary is not a single point but an interval, within which any point can be aurally accepted as the syllable boundary. It is, however, rather difficult to locate the start and end points of the syllable boundary interval exactly and to quantify them by listening. In some cases the examination of waveforms is also less effective for detecting the syllable boundary or the boundary between phonemes because of strong coarticulation.

3.2 Minimal pairs in category 2

3.2.1 Duration. In order to investigate the duration difference between the two words in a pair of category 2 we measured the duration of each phoneme and calculated their relative durations. Table 6 lists the absolute duration of the whole word, the relative duration of each phoneme and of the two syllables in a word as well as the duration ratio of the first and second syllable in the same word.

It can be seen from the data in Table 6 that the absolute durations of the two words in a pair are different even though they have the same phonemes, but no evident regularity was found. The relative duration for each phoneme

Table 6. Duration of minimal pairs in category 2

	duration (ms)	relative duration						
	word	d	a	n	i	S1	S2	S1/S2
da'ni^{33}	659	2%	38%	23%	37%	40%	60%	0.7
dan'i^{33}	755	2%	34%	10%	54%	46%	54%	0.9
		d	a	n	ü			
da'nü^{43}	611	2%	41%	15%	42%	43%	57%	0.8
dan'ü^{43}	667	2%	28%	11%	59%	41%	59%	0.7
		n	a	n	i			
na'ni^{23}	649	8%	35%	19%	38%	43%	57%	0.7
nan'i^{23}	688	11%	31%	7%	51%	49%	51%	0.9
		j	i	n	i	a	o	
ji'niao^{34}	636	14%	34%	16%	17%	8%	11%	48% 52% 0.9
jin'iao^{34}	533	12%	13%	17%	24%	24%	10%	42% 58% 0.7
		q	i	n	ü			
qi'nü^{23}	769	23%	23%	10%	44%	46%	54%	0.8
qin'ü^{23}	806	19%	12%	18%	51%	49%	51%	1
		x	i	n	i			
xi'ni^{12}	878	21%	38%	13%	28%	59%	41%	1.4
xin'i^{12}	783	20%	15%	17%	48%	52%	48%	1.1
		x	i	n	i			
xi'ni^{44}	714	32%	27%	14%	27%	59%	41%	1.5
xin'i^{44}	622	27%	16%	18%	39%	61%	39%	1.6

shows that the initial consonants of the first syllables in each word of a pair are not very different, but a greater difference exists in the following parts of the word (vowels and voiced consonants). For each pair the duration of the second phoneme in the first syllable is always longer in the first word than in the second word. This is due to the relatively stable duration ratio for the two syllables in one word. In one disyllabic word the durations of the two syllables are roughly equal and each phoneme in a syllable takes a certain duration percentage. So the duration for the vowel phoneme in a syllable with two phonemes is longer than its counterpart in a syllable with three phonemes. For instance, in the first pair /da'ni^{33}/ ~ /dan'i^{33}/, the relative duration of /a/ in /da'ni^{33}/ (38%) is longer than that in /dan'i^{33}/ (34%). This tendency is consistent for each pair and can be used to a certain degree for distinguishing the words of a pair.

Furthermore, the duration of a syllable final nasal is generally shorter than a syllable-initial nasal, which is also a cue for discriminating the pair. However, our data also shows counterexamples such as the pairs /qi'nü^{23}/ ~ /qin'ü^{23}/ and /xi'ni^{12}/ ~ /xin'i^{12}/. It seems that the duration of a nasal depends

on the neighbouring vowels. For example in the words /qin'ü^{23}/ and /xin'i^{12}/, strong coarticulation between /i/ and the final nasal causes the /in/ segment to have nearly the same waveform and formant pattern throughout. This results in relatively longer duration of this nasal than that in other pairs. For other phonemes in the same syllable the obvious difference lies in the duration of the last phoneme of the second word in each pair (except for the pair /ji'niao^{34}/ ~ /jin'iao^{34}/). Its duration is always longer than its counterpart in the first word because the only phoneme constitutes the whole second syllable. Hence, for this kind of pairs the duration of each phoneme, especially of vowels and voiced consonants, can be used for distinguishing the pair. Figure 2 shows the spectrograms of one pair in category 2 to illustrate the difference between the two words in a pair.

3.2.2 Formant pattern. Examination of the formant patterns for all pairs in category 2 indicates that the formant distributions of all phonemes (nasals and vowels) in the words of a pair are substantially different though they consist of exactly the same phonemes. The greatest differences can be seen on the nasals. For words whose first syllable ends with a nasal the transition between the vowel and nasal is smooth and without any sudden discontinuity, rise or fall. In some cases it is difficult to segment the vowel and the final nasal accurately. On the other hand, words whose second syllable starts with a nasal show a very distinct nasal segment and a discontinuous transition between the two syllables. Therefore, the syllable boundary can be located at the starting point of the initial nasal of the second syllable. Furthermore, for each pair the distance between formants differs more on the vowel and final nasal in the first syllable than in the second syllable and it also varies with the preceding vowel. For instance, in the pair /da'ni^{33}/ ~ /dan'i^{33}/, the distance between the first three formants in /a/ of /dan'i^{33}/ is larger than that in /da'ni^{33}/, especially the distance between F1 and F2. In the pairs /xi'ni^{12}/ ~ /xin'i^{12}/ and /xi'ni^{44}/ ~ /xin'i^{44}/ the difference is, however, not as obvious because of the strong coarticulation between /i/ and /n/.

Paired sample two-tailed *t*-tests of formant frequencies for all phonemes in the two words across 7 pairs gave the *p*-values 0.007, 0.049, 0.033 and 0.000 on the first four formants, which means that the difference in formant patterns between two words in each pair is significant at least at the level 0.05 (two of them at the 0.01 level). The task of distinguishing the two words in a pair of category 2 is thus easier than for pairs of category 1.

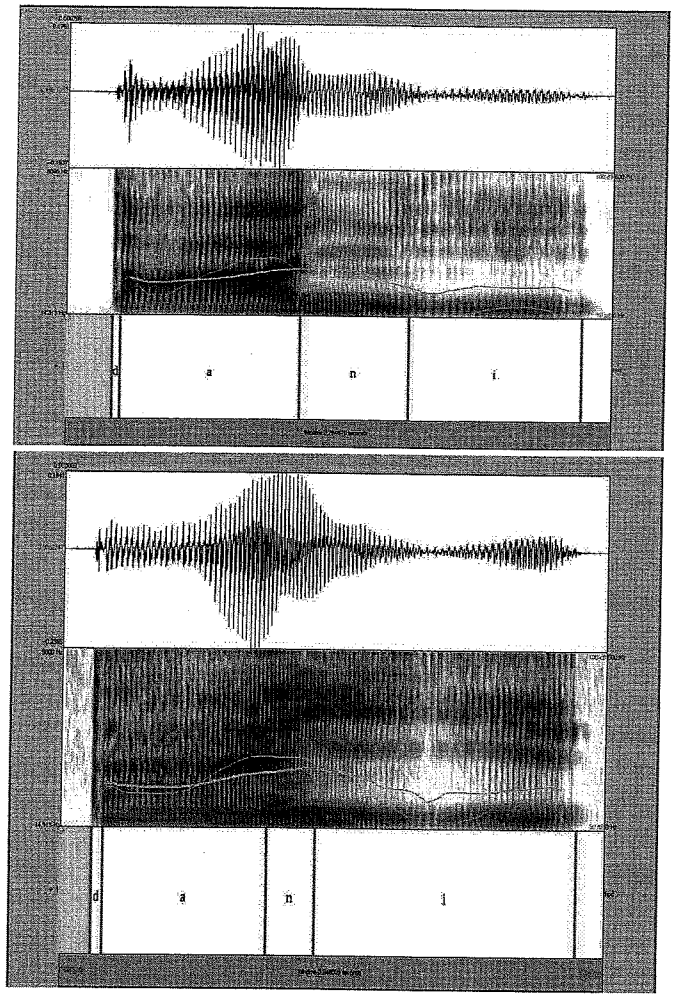

Figure 2. Wideband spectrograms for a pair of category 2: /da'ni^{33}/ (top) vs. /dan'i^{33}/ (bottom)

3.2.3 Pitch, intensity and syllable boundary. The pitch contours of the two words in a pair of category 2 are quite similar and no obvious pitch-based syllable boundary was found (see also Table 5). In some cases the peak or the valley of the pitch contour is in an interval around the syllable boundary, as is the case with pairs in category 1. The intensity contours of the pair are similar too, and no special regularity was found. It is easier to locate the

syllable boundary for words of category 2 than category 1. However, for a single pair such as /ji'niao^{34}/ ~ /jin'iao^{34}/, the syllable boundary of /jin'iao^{34}/ is very difficult to locate because /n/ is between two /i/s and the formant pattern of /n/ is nearly the same as for the two /i/s due to strong coarticulation. In sum, it is more difficult to locate the syllable boundary in words whose first syllable ends with a nasal than in words whose second syllable starts with a nasal because there are rather evident sharp transitions in the formant pattern of the nasal for the latter.

3.3 Minimal pairs in category 3

3.3.1 Duration. In order to find the duration differences between the two words in pairs of category 3 we measured the duration of each phoneme in the word and calculated the total duration of the whole word as well as the duration ratio of the monosyllable and disyllable in the pair (see Table 7).

The data in Table 7 indicates that the durations of the two words are different in each pair as well as in each phoneme because they are monosyllabic and disyllabic words, although they have exactly the same phonemes. The average total duration of a monosyllabic word is 36% shorter than the corresponding disyllabic word but the difference varies with different phonemes and tonal combinations. The first pair /bian^{3}/ ~ /bi'an^{34}/ is special because the durations of these words are nearly equal (15 ms difference), which should be due to their tonal combination. The third tone is generally the longest and the fourth tone the shortest in Standard Chinese. The relationship between word duration and tonal combination is not clear yet, however, since not enough words were investigated in our study. Furthermore, a few disyllables, such as /ji'ou^{13}/, /li'an^{44}/, and /shu'an^{14}/, have relatively great pauses or breaks between the first and second syllable. The syllable boundary of disyllables in pairs of category 3 seems obvious enough to be detected and labeled more easily than those in the other two categories. We are not sure if this is a common feature for this kind of pair or if it is a idiosyncratic speech phenomenon of the speaker. But generally speaking, discrimination of these pairs can be made through comparison of their duration and temporal overlapping of phoneme combinations. Figure 3 shows the spectrograms of one pair in category 3 to illustrate the difference between the two words.

3.3.2 Formant pattern. Observation of the formant structures of pairs in category 3 shows that the formant distributions of phonemes are different

Table 7. Duration in minimal pairs of category 3

		absolute duration (ms)					total (ms)	monosyllable/ disyllable ratio
		b	i	a	n			
1	bian^{3}	28	173	189	57		**447**	**0.97**
	bi'an^{34}	21	189	210	42		**462**	
		d	i	a	n			
2	dian^{4}	16	116	40	48		220	0.42
	di'an^{14}	18	279	198	32		527	
		h	u	a	n			
3	huan^{4}	50	63	68	53		234	0.46
	hu'an^{44}	77	189	190	52		508	
		j	i	a	n			
4	jian^{4}	62	77	90	71		300	0.61
	ji'an^{14}	76	204	187	28		495	
		j	i		o	u		
5	jiou^{3}	69	96		149	133	447	0.55
	ji'ou^{13}	101	246	45	150	277	819	
		l	i		a	n		
6	lian^{4}	57	138		85	32	312	0.59
	li'an^{44}	46	171	99	170	42	528	
		l	i	o	u			
7	liou^{2}	44	152	76	135		407	0.61
	li'iou^{32}	55	346	123	148		672	
		p	i	a	o			
8	piao^{3}	136	68	150	157		511	0.70
	pi'ao^{23}	186	198	216	131		731	
		t	u	a	n			
9	tuan^{2}	101	78	243	21		443	0.85
	tu'an^{24}	129	213	118	64		524	
		sh	u		a	n		
10	shuan^{4}	158	56		79	49	342	0.47
	shu'an^{14}	226	228	62	173	40	729	
		t	i	a	n			
11	tian^{2}	156	106	167	25		454	0.79
	ti'an^{24}	159	220	155	43		577	
		x	i	o	u			
12	xiou^{1}	191	75	49	128		443	0.70
	xi'ou^{11}	165	205	131	132		633	

between the two words in a pair, especially for F2, F3 and F4. The transition between phonemes in monosyllabic words is stable and smooth, while for disyllabic words the transition between the two syllables is rather evident. In some cases there is a break between the two syllables and in other cases the formant trajectory seems to converge towards the syllable boundary. This provides an easy way to detect the syllable boundary and also facilitates distinguishing the words in a pair.

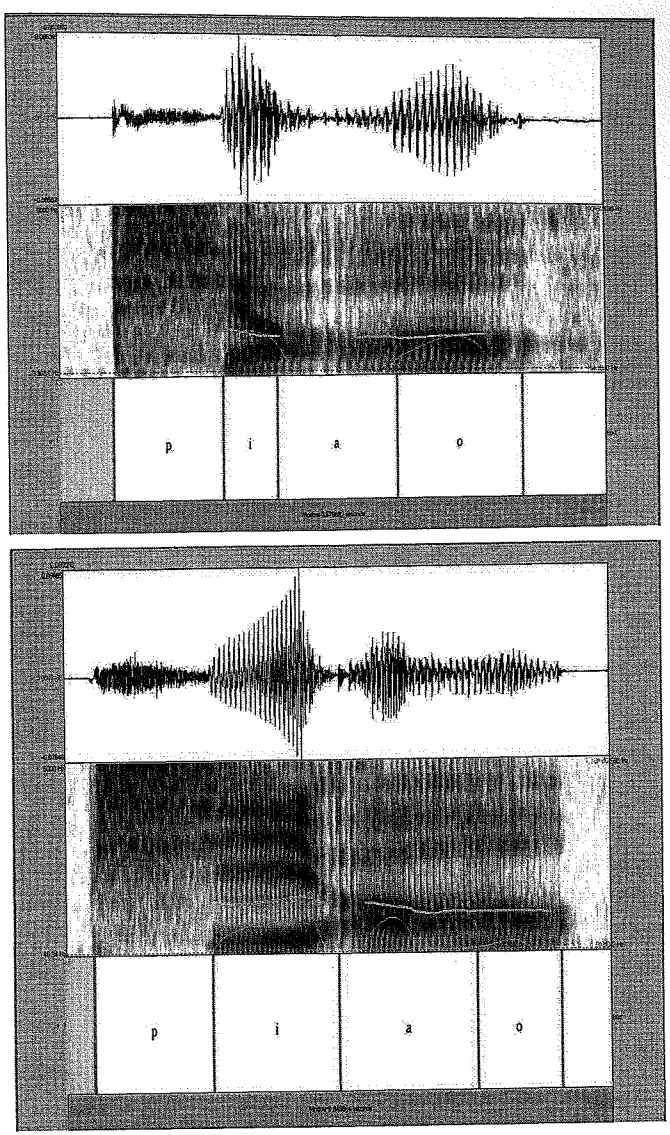

Figure 3. Wideband spectrograms for a pair of category 3: /piao^{3}/ (top) vs. /pi'ao^{23}/ (bottom)

Furthermore, paired sample *t*-test of formant frequencies for all phonemes in the two words across 12 pairs gave the *p*-values 0.928, 0.000, 0.004 and 0.044 for the first four formants, which indicates that the difference of formant pattern between the two words in a pair is significant for F2, F3 and F4 at the level 0.05 (two of them are significant at the level 0.01). We also calculated the absolute distances F2–F1, F3–F2, F4–F3 at three points for all phonemes and the relative distance between them in the same way as for words in category 1 (cf. Table 4). The ratios (F3–F2)/(F2–F1) = *ratio1* and (F4–F3)/(F2–F1) = *ratio2* denote the magnitude of the difference of formant distribution for each pair rather clearly. Paired sample two-tailed *t*-tests for all ratio values from two words in each pair indicates that the *p*-values are 0.004 for *ratio1* and 0.000 for *ratio2*, which shows that the difference in relative formant distances between words in a pair is statistically significant at the 0.01 level. This is useful for distinguishing the pairs.

3.3.3 Pitch, intensity and syllable boundary. For the monosyllabic words in pairs of category 3 the pitch contours are continuous in most cases but they are discontinuous for some of the disyllabic words. There is a pause in the pitch contour between the two syllables in many of these. There is also more fluctuation in the pitch contours of disyllabic words than monosyllabic words. Furthermore, the intensity contours also differ between the two words in a pair. Obvious fluctuation of the intensity contour exists in disyllabic words while in monosyllabic words they are stable and smooth. For the location of the syllable boundary, we can rely on the break of the pitch contour and formant trajectory between the two syllables in addition to listening and examination of waveforms. In some cases the converging of formant trajectories towards the syllable boundary is another cue for detecting the boundary.

3.4 Effect of pitch contour on syllable perception

In order to investigate the effect of the pitch contour on syllable perception we did experiments to try to transform one word to its counterpart in a pair through changing the positions of the points on their pitch contours in semitones. Aural perception indicates that the transformation is impossible for pairs of category 1 and 2, since they have the same phoneme and tonal combinations. Pitch change can only cause the change of tones and syllable boundaries, but has no effect on the formant patterns, which decide the nature and inherent characteristics of syllables and words to a greater extent. For

pairs in category 3, both monosyllabic and disyllabic words can be turned into their counterparts by changing the pitch contour. However, the manipulated words sound unnatural and more like machine speech. A disyllabic word obtained by manipulating its monosyllabic counterpart sounds a little hurried because the duration is not as long as a normal disyllabic words. On the other hand, a monosyllabic word obtained by manipulating its disyllabic counterpart sounds drawly. Further study on the aural perception of such pairs is necessary.

4. Conclusion

Observation of the pairs in the three categories shows that there is no great duration difference between the initial consonant segments in the first syllable, but certain differences between other phonemes in each pair, mainly in the region of the syllable boundary. It seems that it is not relevant to compare the absolute duration of words in each pair, and relative duration should be a better choice, especially for pairs in category 1 and 2. The absolute duration is more significant for pairs in category 3 because a greater difference between the words in a pair can be found for them. The relative durations of the first and second syllables are similar for the two words in all pairs of the first and second categories, although the first word has double vowels or nasals in pairs of category 1. In each pair the two words have consistent duration ratio of the first and second syllable as well. Generally speaking, the duration of each phoneme is more or less different for words in a pair, which can be used to distinguish the words of a pair to a certain degree.

The formant patterns for all phonemes (nasals and vowels) in the two words of a pair are substantially different though the words consist of exactly the same phonemes. The largest differences are found in the vowels of the first syllable of disyllabic words, and in the region of the syllable boundary, and mainly on the first three formants. Comparison of the distance between the formants seems to be a better way to discriminate the two words in a pair than direct comparison of the formant frequencies. On the whole the difference in formant distribution is statistically significant, which indicates that the formant pattern is an important cue for distinguishing the words of a pair. Furthermore, for all syllables containing a nasal, the duration of the final nasal is shorter than that of the initial nasal, which also is a cue for discriminating the pair.

As for pitch contour and intensity, there is generally no great differences between the two words in pair, but the situation is a little different among the three categories. The peaks or valleys on pitch and intensity contours do not seem to be good cues for detecting the syllable boundary. The location of the syllable boundary can be detected in waveforms and by aural perception in this case. Also, sometimes the intervals of pitch contour and formant trajectories between two syllables are also helpful. In some cases the converging of formant trajectories towards the syllable boundary is another cue for detecting the boundary.

Finally, our study indicates that the synthetic transformation from one word to its counterpart in the pair seems impossible in category 1 and 2 though they have the same phonemes and tonal combinations. Pitch change can only cause the change of tones and syllable boundaries but has no effect on the formant patterns, which determines the nature of words in greater extent. This result shows that the effect of changing the pitch contours on the syllable boundary and word perception is limited. For pairs in category 3, both monosyllabic and disyllabic words can be turned into their counterparts by manipulating the pitch contour. However, the changed words sound unnatural. Further study on the aural perception of these pairs is necessary.

Cuiling Zhang 张翠玲, Institute of Chinese Linguistics, Nankai University and Department of Forensic Science, China Criminal Police College <zhangcuiling1972@sohu.com>
Gösta Bruce <Gosta.Bruce@ling.lu.se>