INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Acoustic characteristics of lexical tone disruption in Mandarin speakers after brain damage

Wenjun Chen[1], *Jeroen van de Weijer*[2], *Shuangshuang Zhu*[3], *Qian Qian*[3], *Manna Wang*[3]

[1]Shanghai International Studies University, China
[2]Shenzhen University, China
[3]Shanghai Sunshine Rehabilitation Center, China

0184101205@shisu.edu.cn, jeroen@szu.edu.cn,
doubleshuang91@163.com, flowersinmyheart@126.com, wang18321292690@163.com

Abstract

This study identifies the acoustic characteristics of tones produced by Mandarin brain-damaged patients. We investigate the F0 characteristics of the patients' tone productions and compare them with a control group of healthy speakers. The results show tone disruption in patients with brain damage in either the left or the right hemisphere. Even patients' tone productions that were correctly identified by Mandarin native speakers were acoustically different from the ones produced by healthy speakers. The patterns of tone disruption in Mandarin brain-damaged patients might be caused by damage to the motor function in the brain.

Index Terms: Chinese, tone, brain damage, acoustic measures

1. Introduction

Studies that investigate tone production tend to involve left hemisphere damage (LHD) more than right hemisphere damage (RHD). Lesions to the LH are associated with tone disruption, whereas abnormal tonal production hardly occurs in patients with RH damage [1-3], because tones, involving rapidly changing temporal cues, have been linked with processing in the LH [4]. However, fMRI research shows that tone production is less lateralized than other elements of speech sounds [5], and there are studies demonstrating that both LHD and RHD groups were equally impaired in tone production tasks in comparison to controls [6-7], which raises the question that where tone is processed in the brain.

It is generally believed that lexical tone disruption is a generalized sign of aphasia, observed in many specific types [8]. Tone production problems were also identified in a number of studies in which patients with different types of brain damage were recruited to produce tones, awaiting further research on how, in general, tones produced by speakers with brain lesions are different from those of healthy people, which could lead to further debate on the possible patterns of tone impairment. For example, Naser and Chan [1] found that of the four Mandarin lexical tones, the falling-rising tone is most likely to be impaired; Gandour et al [3] concluded that certain tones are more impaired than others, with the dynamic tones (tones where there is a change in pitch direction or movement) more easily impaired than static tones (tones that only change in F0 height). By contrast, Yiu and Fok [8] proposed, based on observations on lexical tone disruption in Cantonese aphasic speakers, that there is no general pattern for the disruption of tones in patients at all. As previous research has shown various results in regard to characterizing patterns of tone impairment, more evidence is needed as to whether there are any patterns of tone disruption in brain-damaged patients.

One way to address these issues is to perform acoustic analysis on the tones produced by healthy speakers and brain-damaged patients, and compare the acoustic parameters between the two. Detailed acoustic data on patients with brain damage will provide us with a sensitive measure for finding any fine-grained differences between patients and their healthy peers.

Mandarin is a tone language in which each syllable is a morpheme and carries one of the four full tones or a neural tone. In monosyllabic words, Tone 1 (T1), Tone 2 (T2), Tone 3 (T3) and Tone 4 (T4) have a high-level, low rising, falling-rising (dipping) and high falling pitch (F0), respectively [10].

By exploring the acoustic properties of tones produced by patients with brain damage and comparing them with healthy speakers, this study attempts to answer the following questions:

(1) Does tone disruption only occur in LHD patients?

(2) How do patients with brain damage differ from healthy speakers in terms of various F0 properties?

(3) Are there any regular patterns, in terms of F0, for Mandarin tone disruption in brain-damaged patients?

2. Method

2.1. Design

Native Mandarin speakers with brain damage and healthy speakers were recorded and their tone production was evaluated by ten native speakers of Mandarin Chinese. Tones that were identified correctly by 8-10 judges were scored as "correct", whereas those that were identified correctly by 0-4 judges were scored as "incorrect". In this way all the productions can be classified into three groups: "Healthy speakers correct productions" (HC), "Patients' Correct productions" (PC) and "patients' incorrect productions" (PI) (productions that were correctly identified by 5-7 judges were excluded from the analysis). We then measured and compared a number of acoustic properties of F0 of these productions. The aim was to see if the 'correct' tone productions by aphasics are acoustically different from those of healthy speakers, and whether the patients' 'incorrect' tone productions exhibit any acoustic regularity that distinguish them from the 'correct' ones.

http://dx.doi.org/10.21437/Interspeech.2019-2432

2.2. Subjects

Eight Mandarin speakers recruited from the Shanghai Sunshine Rehabilitation Center participated in this study. Based on a CT scan, four of the aphasic subjects had LHD and the others had RHD (Table 1). Patients with LHD were classified according to the Standard Aphasia Battery of Chinese [11], which is adapted from the Western Aphasia Battery [12]. The mean age of the patients was 61 years. In addition, four of the patients' spouses also participated in the study as a control group (mean age was 58 years). Both groups were balanced for gender. All subjects were native speakers of Mandarin, right-handed and without significant hearing loss (PTA<30dB HTL). All subjects had an education background of at least elementary school, and all patients preserved a mild degree of literacy after stroke.

Table 1: *Background Information on brain damaged patients, TPO = 'time post onset' (years; months).*

Patient	Age	TPO	Lesion site	Aphasia
LHD1	57	0;6	Left basal ganglia	Anomic
LHD2	67	1;0	Left temporal lobe infarct	fluent
LHD3	53	0;8	Left middle cerebral artery	Nonfluent
LHD4	66	1;8	Left fronto-parietal infarct	Broca's
RHD1	68	0;4	Right basal ganglia	No Aphasia
RHD2	57	0;5	Right basal ganglia	No Aphasia
RHD3	62	2;2	Right temporo parietal	No Aphasia
RHD4	54	1;4	Right temporo parietal	No Aphasia

2.3. Stimuli

The stimuli consisted of 24 (monosyllabic) words: six for each tone. The 24 words were selected based on a word familiarity test that was used with other aphasic patients in the rehabilitation center. These words are related to patients' daily routines and thus expected to be most recognizable to the patients, either by the orthography and/or their connotations. The words are provided in Table 2 in Chinese characters, Chinese alphabetic orthography (Pinyin), and English glosses.

Table 2: *Stimulus words.*

	Word	Pinyin	Gloss		Word	Pinyin	Gloss
T1	猫	mao	cat	**T3**	水	shui	water
	花	hua	flower		跑	pao	run
	七	qi	seven		笔	bi	pen
	妈	ma	mother		伞	san	umbrella
	灯	deng	light		写	xie	write
	鸭	ya	duck		脚	jiao	leg
T2	门	men	door	**T4**	画	hua	draw
	鱼	yu	fish		抱	bao	hug
	桃	tao	peach		药	yao	medicine
	球	qiu	ball		四	si	four
	毛	mao	fur		大	da	big
	盐	yan	salt		裤	ku	trousers

2.4. Procedure

Each participant was presented with 24 colored pictures representing 24 monosyllabic words in a random order. The Chinese character of each word was printed at the bottom right of the picture so that the patients could probe the connotations of the word either graphically or by words. The stimuli were targeted as monosyllabic words in isolation, to avoid coarticulation effects. The participants labeled each picture twice, with the examiner present. All responses were recorded by a PG42 Side Address condenser microphone with 16 bit rate and 22.1 kHz sampling rate. If the first production could not be used due to background noise or non-isolated productions, the second production was selected.

Ten native speaker judges were recruited in the rating procedure; six of these were also exposed to Shanghainese at work. The productions were low-pass filtered at 500Hz to eliminate to a certain extent the segmental information and retain the F0 information, and then normalized to the same intensity. The judges listened to all the filtered productions and categorized the tones in a forced four-choice task in a quiet room using a high-quality headphone.

2.5. Labeling and measurements

For the acoustic analysis, the rime (R) of each syllable, where the tone occurs, was excised from the original unfiltered stimuli. The onset of the rime was marked as the end of the four vocal cycles of the vowels and the offset of the rime was marked at the beginning of the fourth observable vocal cycle from the end of the rime in the waveform [13]. The F0 was extracted by a Praat script [14] and converted to semitones (St).

Based on the findings in Wong [13] that typical acoustic characters were key to the identification of different tones, seven acoustic parameters were then derived from the F0 extraction of each production for investigation. First, since T1 is a leveling tone, 'pitch shift' was used to measure and compare the flatness of the F0 contour in T1; also, the 'height of mean F0' and the 'height of min F0' indexed the mean and minimum level of pitch in the produced tone (token mean) relative to the mean F0 of the speaker (speaker mean), to measure how the speaker reached the high tonal target for T1 and low tonal target for T3 (dipping tone); second, since T2 and T4 are rising and falling tones, respectively, we used 'directional excursion' to measure and compare the degree of positive and negative F0 span and 'slope' to measure the range and steepness of the positive and negative slopes. Finally, 'timing of min F0' and 'timing of max F0' were used to measure and compare how early the minimum F0 and how late the maximum F0 occurred in T2 syllables, and how late the minimum F0 and how early the maximum F0 occurred in T4 syllables.

Pitch shift = maxF0 - min F0 (1)

Height of mean F0 = mean F0 (token) - mean F0 (speaker) (2)

Height of min F0 = min F0 (token) -mean F0 (speaker) (3)

Directional excursion = +/- (max F0 -min F0) (4)

Slope = Directional excursion/duration (max-min F0) (5)

Timing of max F0 = Duration (onset-max F0)/Duration (R) (6)

Timing of min F0 = Duration (onset-min F0)/Duration (R) (7)

3. Analyses and results

3.1. Judgment of tones

The raters identified 90 healthy speakers correct (HC) productions (24, 23, 20 and 23 productions for T1, T2, T3 and T4, respectively), 84 patients correct (PC) productions (31, 26, 7 and 20 productions for T1, T2, T3 and T4, respectively) and 55 patients incorrect (PI) productions (3, 9, 30 and 13 productions for T1, T2, T3 and T4, respectively). In the PC group, 41 and 43 correct productions were made by RHD and LHD patients, respectively; in the PI group, 23 and 32 incorrect productions were made by RHD and LHD patients, respectively. While the RHD patients seemed to make fewer incorrect productions than LHD patients, the Chi-square test reveals there was no significant difference in the number of productions made between LHD and RHD patients in either PC or PI group. (PC: $X^2(1) = 0.02$, $p>0.05$; PI: X2(1) = 0.602, $p>0.05$). The number of PC productions by none of the patients was above chance level, as indicated in Fig. 1.

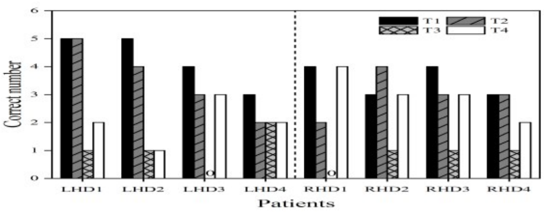

Figure 1: *Number of correct productions for each patient.*

3.2. Tone contours of HC, PC and PI productions

Prior to any statistical analyses, mean F0 curves were drawn to identify the general pattern of various effects. Fig.2 displays time-normalized F0 contours of the tone productions, with two examples selected from each group. Each line in each panel represents one production and each panel represents the F0 contours of the same word produced by the same speaker. The figures are derived from the data of healthy group Participants 1 and 4, Patients 1 and 7 (Correct) and Patients 2 and 5 (Incorrect). Genders are balanced in the selected figures of each group. Fig. 3 shows the average F0 plots of the four tones produced by the three groups.

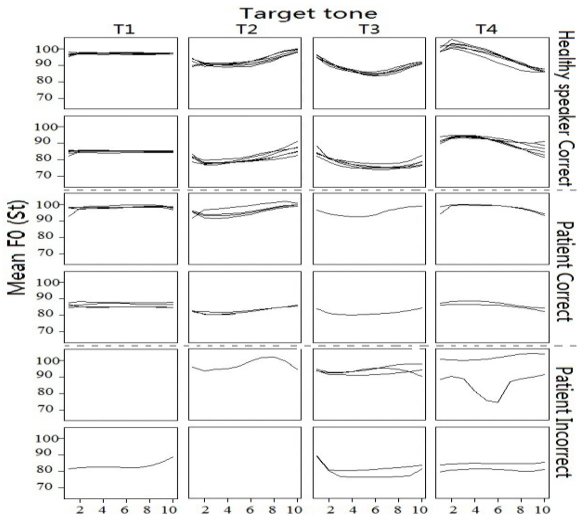

Figure 2: *F0 contours of tones T1-T4 by participant from different groups (two cases from each group).*

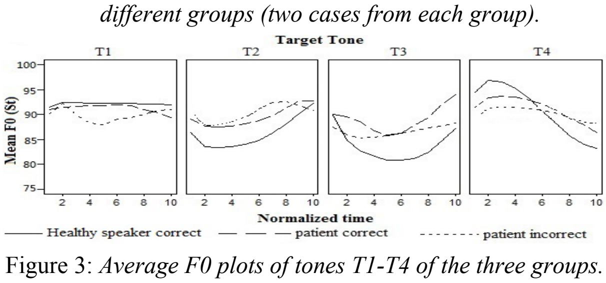

Figure 3: *Average F0 plots of tones T1-T4 of the three groups.*

Several facts should be noted: (1) T1 and T2 were produced relatively well, being identified as 'Incorrect' only in a small number of cases, the panels without any F0 contours indicate that none of the patients' productions fell into this category. (2) For healthy speakers, we observed very clear F0 contours that are high and level for T1, rising for T2, dipping for T3 and falling for T4. (3). The F0 contours of the patients' correct production were flatter overall than those of healthy speakers, with T1 more varied, T2 and T4 being less steep in PC group. (4) Large variation in F0 contours was observed in the patients' incorrect production; for example, no obvious dipping contour in T3 panels, a rising contour for T1, a rising or falling contour for T2 and a very flat contour for T4.

3.3. F0 analysis

Due to the violation of normality and homogeneity of variance for the parametric statistics, a Mann-Whitney-U test was conducted to compare the seven F0 parameters among different accuracy groups. Fig.4 exhibits the distribution of the statistics which is significant different among the three groups.

Patients' correct production of T1 were less level than that of healthy speakers, as indicated by significantly greater pitch shift (N=55, z=2.121, $p<0.05$); patients' correct production of T2 was not significantly different from the normal speakers in slope and the timing of minimum F0, but they did not span this tone as much as healthy speakers, as indicated by the value of directional excursion (N=49, z=-5.89, $p<0.001$), and their maximum F0 was also reached earlier than in the healthy speakers (N=49, z=-4.986, $p<0.001$), suggesting that the patients were not very good at keeping the F0 low and steep for an extended period of time; both of the two T3 parameters, 'height of mean F0' and 'height of min F0', were significantly smaller in the patients than in the healthy speakers (N=27, z=-3.873, $p<0.001$, for height of mean F0; z=2.379, $p<0.05$ for height of min F0), indicating that the patients had difficulties in maintaining the F0 either as low or as dipped as the healthy speakers. Patients correct T4 productions were less ranged and steep compared to healthy speakers, indicated by the significantly smaller value in directional excursion (N=43, z=-4.48, $p<0.001$) and slope (z=-3.29, $p<0.001$). The patients also showed late timing in reaching the maximum F0 compared to the healthy speakers (z=-5.6, $p<0.001$), suggesting that they might need a longer time to calibrate the high pitch outset and respond slowly in preparing the start of T4.

Patients' incorrect tone productions were also different from patients' correct tone productions. Patients' incorrect T1 productions have a significantly larger pitch fluctuation (N=34, z=2.823, $p<0.05$ for pitch shift) and a lower minimum F0 height (z=-.2.155, $p<0.05$) compared to PC productions (Figs. 2, 3). This makes PI production of T1 different from a leveling tone (Fig. 2). Patients' incorrect T2 productions ranged from significantly smaller excursion than in PC (N=35, z=-2.453, $p<0.05$) to having negative excursions, and the minimum F0 was also reached earlier (z=-2.076, $p<0.05$). The relative mean and minimum height of patients' incorrect T3 was also higher that of PC (N=37, z=-2.559, $p<0.05$ for the height of Mean F0; z=-3.955, $p<0.001$ for the height of Min F0). Patients' production T4 ranged from having positive or negative excursions (N=33, z=3.021, $p<0.05$) to reduced steepness (z=4.569, $p<0.001$). In short, for T1, T2, T3 and T4, PI productions were either flatter or more varied in the directions of the F0 curves than PC productions.

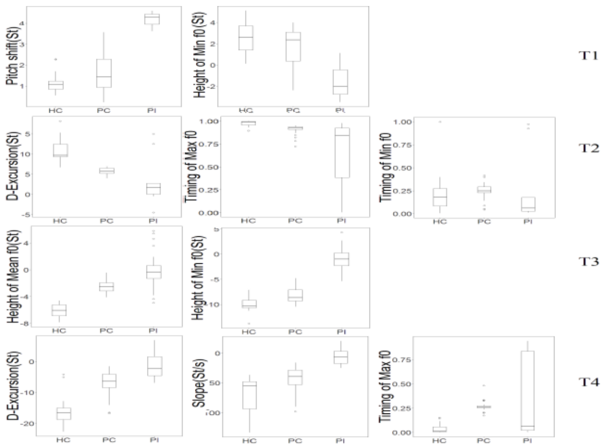

Figure 4: *Box plots of the distributions of the F0 derived parameters in the three groups.*

4. Discussion

This study investigated the production of tones by Mandarin aphasic speakers and compared their F0 information with that of healthy speakers. With regard to the nature of tone productions in Mandarin aphasic speakers, the following points deserve attention.

First, we observed a common disruption of tone production both in LHD and RHD Mandarin-speaking patients. Patients with RHD didn't show significantly better tone production than patients with LHD. Admittedly, it may be possible that RHD patients may still perceive Mandarin tones better than LHD patients because they might use more acoustic cues that they are sensitive to in the perception of the four tones. Nevertheless, the current study showed that both LHD and RHD patients were equally impaired in the tone production task in comparison to controls. This result accords with findings for lexical tone disruption in Shona [7], which showed that there is no difference in the tones produced by LHD vs. RHD Shona speakers.

Second, based on the acoustic parameters that were measured, even patients' tone productions that were identified correctly by Mandarin native speakers were different from the ones produced by healthy speakers. Patients' correct productions of T1 are less level than in healthy speakers, PC's productions of T2 and T4 did not span as much as healthy speakers and the maximum F0 of the patients' correct productions were reached earlier for T2 and later for T4 than in healthy speakers. In addition, PC's productions of T3 were neither as low nor as dipped as normal productions. Although identifiable to native Mandarin speakers, the tones produced by brain damaged speakers were never the same as their healthy peers. On the other hand, patients tone productions that were categorized as incorrect were further disrupted, as they were also significantly different from patients' correct tone productions for many F0 properties. Tokens of PI's T1 fluctuated more than in the PC group; conversely, PI's T2 and T4 productions were less spanned than that in PC, and there were both positive and negative excursions for PI's T2 and T4, suggesting that rising and falling of the curve were more variable in these patients' incorrect productions. T3 was even less lowered and less dipped in PI and PC, with many levelling contours.

Third, not all four Chinese tones are equally susceptible to disruption due to brain lesion, the high level tone (T1) appears to be most resistant to disturbance among patients. The accuracy ratings show that T1 was accurately identified in up to 74% (31) of the total stimuli. The F0 information also demonstrated that the patients produced T1 relatively well, with the only difference being the 'pitch shift' between HC and PC; on the other hand, the dipping tone (T3) appears to be especially vulnerable to disruption in comparison to the level (T1), rising (T2) and falling tone (T4); as few as 18% (7) of the T3 tokens were judged correctly and both of the patients' T3 F0 properties exhibited significant differences from the healthy speakers' correct productions. In the meantime, we found that the rising tone (T2) was also comparatively resistant to impairment, since 65% (26) of the patients' T2 stimuli were identified correctly and were different from HC by its excursion range and the timing when the maximum F0 was reached; by contrast, the patients' correct productions of T4 reached only 48% (20) of the stimuli, and these were different from the T4 production of healthy speakers in excursion range, steepness, and timing of the maximum F0 being reached. The general pattern of high level (T1)> low rising (T2) > high falling (T4) > dipping (T3) in terms of tone preservation was supported by comparing the proportions on the correctly identified productions of each type of tones (Chi-square: X^2= 10.502 (3), $p<0.05$), and was also observed in most patients (Fig.1). This pattern correlates with the claims made by Gandour that dynamic tones are relatively susceptible to disruption and that the dynamic complexity can be measured by F0 change, both in kind and degree. As the dipping tone, T3 involves the greatest directional changes in F0 contour, it was least preserved; T2 is also a dynamic tone, and it was better preserved by patients than the other dynamic tone T4, possibly because T2 involves a lesser degree of pitch change (slope, excursion) than T4 [15].

Various reasons can be proposed for the typical pattern of tone disruption observed in the current research. However, compared with the production of the static tone (T1), the observation that both the LHD and RHD patients underperformed in the dynamic tones coincides with the fact that impairment in motor control, which can be associated with either left or right brain, was prevalent in patients with brain damage [16], commonly seen as the impaired sensation and movement after stroke. Production of the dynamic tones might involve more complex coordination of muscles activated by the brain, in the performance of motor skill, and thus is more difficult for the patients to handle. We leave this aspect for further research.

5. Conclusions

The goal of the present study was to identify the characteristics of tones produced by Mandarin aphasic patients. We found that tone disruption in patients with either left or right brain damage was common, and that even the patients' tone productions that were correctly identified by Mandarin native listeners were acoustically different from the ones produced by healthy speakers. The patterns of tone disruption in Mandarin aphasic patients might be caused by the damage of motor functions in the brain. Since the sample in the current research is a comparatively small portion of the population on brain damage, further research may focus on how tones are different in terms of disruption degree based on a larger sample size.

6. Acknowledgements

This research was supported by the Shanghai Sunshine Rehabilitation Center of China in Shanghai. We thank all the therapists who helped with data collection.

7. References

[1] M. A. Naesar and S.W. Chan, "Case study of a Chinese aphasic with the Boston Diagnostic Aphasia exam," *Neuropsychologia*, vol.18, pp.389-410, 1980.

[2] J. L. Packard, Tone production deficits in nonfluent aphasic Chinese speech. *Brain and Language*, vol.29, pp.212-223, 1986.

[3] J. Gandour, S. Ponglorpisit and R. Dardarrananda, "Tonal disturbances in Thai after brain damage," *Journal of Neurolinguistics*, vol. 7, pp.133-145, 1992.

[4] J. Schwartz and P. Tallal, "Rate of acoustic change may underlie hemispheric specialization for speech perception," *Science*, vol. 207, pp.1380-1381, 1980.

[5] L. Li, et al., "Dissociation in the neural basis underlying Chinese tone and vowel production," *NeuroImage*, vol.29, pp.513-523, 2006.

[6] I. Moen and K. Sundet, "Production and perception of word tones (pitch accents) in patients with left and right hemisphere damage," *Brain and Language*, vol. 53, pp.267-283, 1996.

[7] M. R. Kadyamusuma, R. D. Bleser and J. Mayer, "Lexical tone disruption in Shona after brain damage," *Aphasiology*, vol.25, pp.1239-1260, 2011.

[8] E. M. Yiu and A. Y. Fok, "Lexical tone disruption in Cantonese aphasic speakers," *Clinical Linguistic & Phonetics*, vol. 9, pp.79-92, 1995.

[9] J. Gandour, S. Holasuit-Petty and R. Dardarananda, "Perception and production of tone in aphasia," *Brain and Language*, vol.35, pp. 201-240,1988.

[10] S. Duanmu, "Chinese (Mandarin): phonology", *Encyclopedia of Language and Linguistics*. Oxford, UK: Elsevier, 2006.

[11] S. Gao, "Standard Aphasia Battery of Chinese," *Chinese Mental Health Journal*, vol.6, pp.125-128, 1992.

[12] A. Kertesz, *Western Aphasia Battery*. New York: Grune & Stratton, 1982.

[13] P. Wong, "Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions," *Journal of Phonetics,* vol.40, pp.141-151, 2012.

[14] Y. Xu, ProsodyPro. Praat. Retrieved from <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>.

[15] S. Feng and P, Wang, "Statistical analysis of tone of monosyllabic word of Beijing dialect," *Contemporary Linguistics,* vol.8, pp.234-333, 2006.

[16] J. Cauraugh, K. Light, S.Kim, M. Thigpen and A. Behrman, "Chronic Motor Dysfunction after Stroke," *Stroke*, vol.31, pp.1360-1364, 2000.