PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Cochlear-Implanted Children's Perception of Mandarin Tones in Normal Speech and Whispered Speech

Xin DING[1]; Wentao GU[1,2]

[1] School of Chinese Language and Literature, Nanjing Normal University, China

[2] Nanjing Normal University and George Washington University Joint Laboratory of Speech, Hearing and Rehabilitation Sciences, China

ABSTRACT

Since cochlear implants lack an efficient coding of fine structures of speech including F0, it is generally difficult for cochlear-implanted (CI) children to perceive linguistic tones, of which F0 is the primary acoustic cue. In whispered speech, however, F0 is missing and no longer serves as an acoustic cue of tones. Using a picture selection task, we compared the identification of Mandarin four tones in isolated syllables between two groups of 4-to-5-year-old children, i.e., CI and normal-hearing (NH) groups, in both normal speech and whispered speech. For normal speech, NH showed significantly higher identification than CI, regardless of tone which showed no significant effect. For whispered speech, both groups identified T3 better than other three tones, possibly due to their use of other cues for T3; also, no significant difference was found between the two groups on all tones but T3, on which NH showed significantly higher identification than CI, suggesting that NH retrieved the voice quality cue for T3 better than CI. In both groups, identification rate was significantly lower for whispered speech than for normal speech, but the decrease in the rate was greater in NH than in CI, confirming that NH's tone perception relied more on F0.

Keywords: Cochlear-implanted children, Mandarin, Tone perception, Whispered speech

1. INTRODUCTION

Cochlear implants technology is a revolutionary advancement in biomedical engineering in the recent decades. It successfully employs electronic engineering techniques to simulate human auditory system, and has been widely used in hearing impaired children. Patients with severe sensorineural loss can regain hearing and communication skills after cochlear implantation. However, the cochlear implants technology has been less efficient in speakers of tone languages like Mandarin. To be specific, it is difficult for cochlear implants recipients to use fundamental frequency (F0) information to perceive tones and intonation. Because the speech coding strategy of cochlear implants was originally developed based on the Western non-tone languages, the consideration of the acoustic characteristics of the fundamental frequency is not thoughtful. Although duration, amplitude, and spectral envelope play a compensating role in the recognition of tones in the absence of the fundamental frequency, the performance of speech recognition for Mandarin users is still not satisfactory.

F0 is the most important acoustic cue for lexical tones, though there is also some redundant information carried by F0. For example, some previous studies found that even if Mandarin speech goes through a high-pass filter at 2.4 kHz, the rate of tone recognition was still high. In addition, Liu et al. (2004) found that even if the upside part of F0 for low tone (T3) was deleted, the recognition of T3 by normal people is still unaffected. Moreover, the role of other acoustic cues in tones, like duration, was still debated. Xu et al. (2002) reported that the recognition rate of four Mandarin tones reached 56.5% (above the chance level 25%) with duration information only for normal participants. In addition, the comparative experiment found that after removing the duration information, the recognition rate of tones was reduced by about 15%. In contrast, Kou et al. (2008) argued that duration was only a marginal cue, because their study showed that the recognition rate of tone was only 35% (the chance level is 25%) in duration-only condition.

[1] dingxin0417@126.com
[2] wtgu@njnu.edu.cn

In whispered speech, F0 information is missing, and listeners can only use the spectral envelope information to recognize tones. In this case, the rate of tone recognition by normal hearing people is about 60%-70% (Kong et al., 2006), which indicates that it is impossible to rely solely on duration and amplitude information. They believe that the formant frequency in the spectral envelope information is also used for tone recognition in whispers. Since there is no way to accurately control duration and amplitude contour information in whispered speech, it is still unknown how the spectral envelope information can be used to identify tones in whispered speech.

Many previous studies have focused on tone perception in children with cochlear implants (CIs). It has been shown that Cantonese-speaking children with CIs had a significantly lower perceptual accuracy of six Cantonese tones than the normal hearing (NH) children of the same age. For example, Ciocca et al. (2002) conducted a Cantonese tone recognition test on 17 children aged 4 to 9 years old with a Cantonese-speaking cochlear implant. The test sound was untreated natural speech. The experimental results showed that only two participants had recognition scores above the chance level. Lee et al. (2002) selected three tone pairs with distinct tonal differences in Cantonese, and the perceived accuracies were 92% and 64% for NH and CI children, respectively. Peng et al. (2004) investigated the perceptual ability for Mandarin tones of 30 Mandarin-speaking CI children, and the result was 72.8%, which was very similar to Xu et al. who reported 71% accuracy. Wei (2004) studied the tone recognition ability of 14 children with cochlear implants. They adopted the "forced-choice paradigm" in tone discrimination, and the results showed that their average recognition rate was 62.8%, and there were also individual differences.

Whisper is a special phonation type, in which the vocal cords do not vibrate, and thus F0 is missing. Also, whisper is associated with low intensity and low SNR. Liang (1963) found that the recognition rates of four tones in whispered speech were 53.5% (T1), 48.3% (T2), 79.1% (T3), and 73.3% (T4), respectively. Qian et al. (2000) found that the rate of tone recognition was 58.5% when only amplitude envelope was used, and reached 59.1% when both amplitude envelope and duration were employed, so it was concluded that amplitude envelope and duration played an important role in recognizing T3 and T4. Sha (2003) showed that tones in isolated whispered syllables could still be recognized, so amplitude envelope and duration are both important factors in the recognition of tones. In this study, we compared the ability of tone identification in CI children in normal speech and whispered speech.

2. Method

2.1 Participants

Two groups of participants, i.e., the cochlear implanted (CI) group and the normal hearing (NH) group, were recruited in the perceptual experiment. Each group consisted of 12 children including 9 boys and 3 girls. All participants were from native Chinese-speaking families, without any report of neurological or psychiatric diseases. The children in the CI group were recruited from the Speech and Hearing Rehabilitation Center of Jiangsu Province, aged between 4;7 and 5;7, with profound hearing losses (≥ 91 dB HL) in both ears prelingually. They were all implanted at the age of 1;5-3;0, and the duration of their using CIs was from 2;3 to 3;9. The children in the NH group were from Baoli Amethyst Mountain Kindergarten Affiliated to Nanjing Normal University, aged from 4;4 to 5;6. T-test did not show any significant difference in the age between the two groups.

2.2 Materials

For the four tones, there are six tone pairs. For each tone pair, we selected 3 pairs of monosyllabic words. Thus, there were altogether 18 pairs of monosyllabic words, as listed in Table 1. The two words in each pair shared the same segments, differing only in tone. For each word, a corresponding picture was also prepared.

2.3 Procedure

The tone identification experiment consisted of two tasks, one for normal speech and the other for whispered speech. Before the experiment, each participant received a training session to get familiar with the pictures and the experimental procedure. During the experiment, after each stimulus was played back through loudspeakers, the participants were asked to identify the picture corresponding to the stimulus.

In each task, the stimuli were presented in a random order using E-Prime 2.0. A laptop Lenovo new Air Pro with a screen resolution of 1366×768 was used, and a pair of loudspeakers Edifier R26T with

a signal-to-noise ratio ≥85dBA and distortion less than 0.5% were placed on both sides in front of the participants, with a 0.5m distance in between. The recording volume was set at about 70dB SPL.

Table 1 – Word list for tone identification

T1 vs. T2	chuāng	chuáng	qiāng	qiáng	tāng	táng
	窗	床	枪	墙	汤	糖
T1 vs. T3	sān	sǎn	shū	shǔ	wū	wǔ
	三	伞	书	鼠	屋	五
T1 vs. T4	bāo	bào	bēi	bèi	dēng	dèng
	包	报	杯	被	灯	凳
T2 vs. T3	hú	hǔ	bí	bǐ	yú	yǔ
	狐	虎	鼻	笔	鱼	雨
T2 vs. T4	wá	wà	yé	yè	tú	tù
	娃	袜	爷	叶	图	兔
T3 vs. T4	liǔ	liù	jiǎn	jiàn	mǐ	mì
	柳	六	剪	剑	米	蜜

3. Results

3.1 Normal speech experiment

Figure 1 shows the overall rates of tone identification for each participant in the condition of normal speech. All results are above the chance level 50%. The result of statistical analysis shows that the mean accuracy rate is significantly higher in the NH group (99.0%) than in the CI group (76.1%). Moreover, the average rates of identification into each tone type are shown in Table 2.

Repeated-measures ANOVA was conducted on the accuracy rates, with group as inter-subject factor, and tone as intra-subject factor. The results showed that group had a significant main effect ($F(1,23)=81.478$, $P<0.0001$, $\eta^2=0.78$), while tone did not have any significant effect ($F(3,69)=0.916$, $P=0.438$, $\eta^2=0.03$). Also, the interaction between tone and group was not significant ($F(3,69)=0.833$, $P=0.48$).

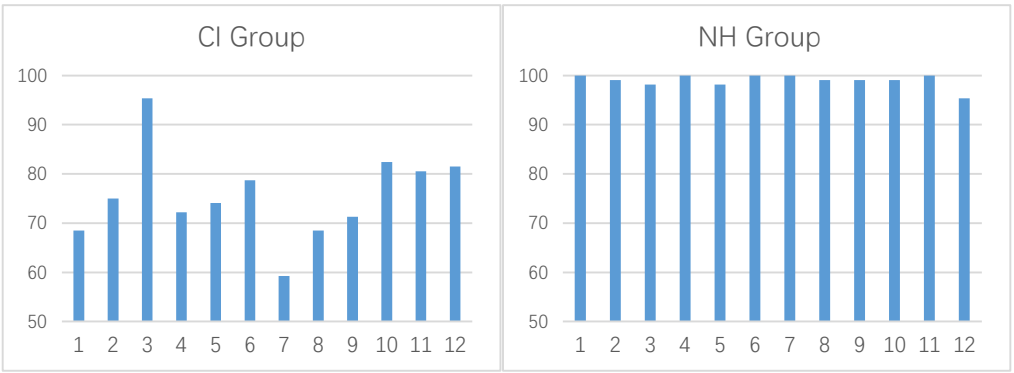

Figure 1 – Accuracy of tone identification for each participant in the condition of normal speech

Table 2 – Rates of tone identification for normal speech in the CI and NH groups

		T1	T2	T3	T4
CI	T1	**0.746**	0.105	0.090	0.059
	T2	0.077	**0.742**	0.146	0.034
	T3	0.031	0.143	**0.752**	0.074

	T4	0.109	0.041	0.050	**0.802**
NH	T1	**0.994**	0.000	0.000	0.006
	T2	0.003	**0.989**	0.009	0.000
	T3	0.009	0.003	**0.986**	0.003
	T4	0.003	0.003	0.003	**0.991**

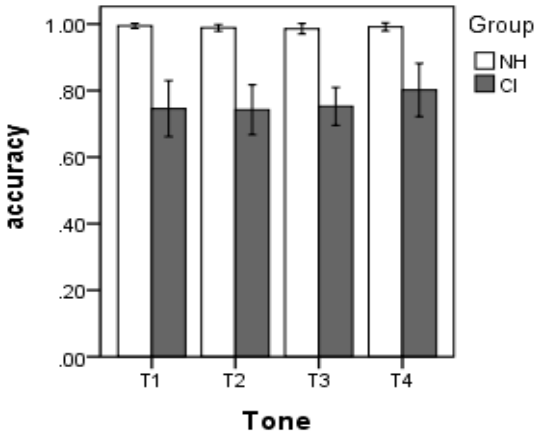

Figure 2 – Accuracy rates of tone identification for normal speech

3.2 Whispered speech

In the condition of whispered speech, the two groups showed a significant weaker tone recognition ability than in the condition of normal speech. The results of tone recognition for whispered speech is shown in Table 3. The scores of the NH group have dropped significantly compared with the previous (99.0%), with an average correct rate of 65.2%. The recognition scores of children in the cochlear implant group were mainly concentrated in the range of 45%-62%, the average correct rate was 55.5%, and the average correct rate under normal speech conditions was 76.1%. This indicates that this task is difficult for both groups of participants, but the rate of decline in the recognition rate of children in the hearing group (33.8%) is greater than that of the children in the cochlear group (20.6%).

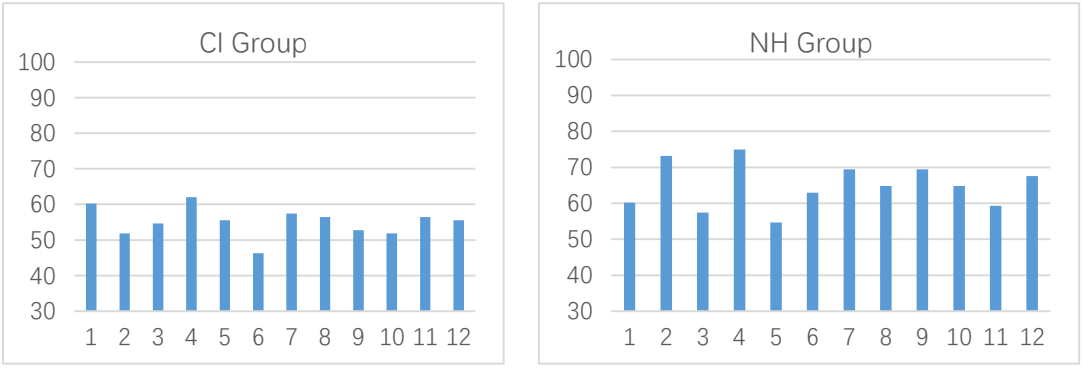

Figure 3 – Rates of tone recognition for whispered speech in CI and NH groups

Table 3 – Rate of tone identification for whispered speech in the CI and NH groups

		T1	T2	T3	T4
CI	T1	**0.466**	0.203	0.160	0.170
	T2	0.123	**0.566**	0.170	0.142

	T3	0.099	0.147	**0.617**	0.136
	T4	0.194	0.132	0.102	**0.571**
NH	T1	**0.521**	0.208	0.151	0.120
	T2	0.105	**0.618**	0.228	0.048
	T3	0.046	0.063	**0.835**	0.057
	T4	0.214	0.057	0.097	**0.632**

From the confusion matrix, although the recognition rate of the four tones of the two groups is high or low, there is still a certain rule, that is, the order of the four sounds from high to low is T3>T4>T2> T1. At the same time, there is a problem in the discrimination between the two groups of whispers. It is easy to judge T1 as T2, T2 is easy to judge as T3, T3 is easy to judge as T2 and T4, and T4 is easy to judge as T1. However, it is worth noting that the discrimination rate of T3 in normal children (83.48%) is significantly higher than that in CI children (61.73%), probably because T3 has the longest duration, and CI children cannot perceive the fine structure in tone.

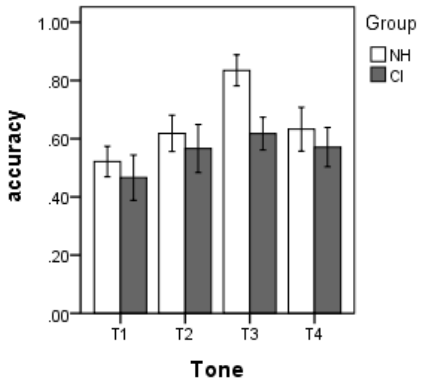

Figure 4 – Accuracy rate of tone recognition for whispered speech in CI and NH groups

Mixed ANOVA shows a significant main effect of group ($F(1,23)=21.14$, $P<0.0001$, $\eta^2=0.479$) – the NH group has a significantly higher accuracy rate (65.2%) than the CI group (55.5%). Tone also has a significant main effect ($F(3,69)=15.34$, $P<0.0001$, $\eta^2=0.40$) – T3 has a significantly higher accuracy rate (72.6%) than three other tones, and T4 has a significantly higher accuracy rate than T1 and T2.

There is also a significant interaction effect between group and tone ($F(3,69)=2.75$, $P=0.049$, $\eta^2=0.107$). Simple effect analysis shows that only in T3, the identification of the NH group and the cochlear group There was a significant difference between the rates (83.5% vs 61.7%), and there was no significant difference in the recognition rates between the two groups in the whispers of T1, T2 and Deaf. At the same time, we also found that among the four tones of the whispering group, the recognition rate of the upper voice (83.5%) was significantly higher than that of T1 (52.1%), T2 (61.8%), and de-sounding (63.2%). There was a significant difference between the upper acoustic (61.7%) and the T1 (46.6%) in the CI group, and there was no significant difference between T2 (56.6%) andT4 (57.1%).

3.3 Comparison: Normal speech vs. Whispered speech

The identification rates in the two experiments were analyzed by repeated-measures ANOVA, using phonation type (normal, whisper) and tone (T1, T2, T3, T4) as within-subject factors.

For the NH group, phonation type has a significant main effect ($F(1,12)=423.974$, $P<0.0001$, $\eta^2=0.972$) – the identification rate in normal speech (99%) is significantly higher than in whispered speech (65.2%). Tone also has a significant main effect ($F(3, 36) = 16.51$, $P < 0.001$, $\eta^2 = 0.579$). There is also a significant interaction effect between tone and phonation type ($F(3, 36) = 22.108$, $P < 0.0001$, $\eta^2 = 0.648$). Simple effect analysis further shows that there is no significant difference between four tones in normal speech, while in whispered speech T3 has a significantly higher rate than other tones, and T2 has a significantly higher rate than T1.

For the CI group, the main effect of phonation type was significant ($F(1,11)=45.442$, $P<0.0001$, $\eta^2=0.805$) – the identification rate in normal speech (76.1%) is significantly higher than in whispered speech (55.5%), while the main effect of tone was not significant ($P=0.171$). Also, the interaction between tone and phonation type is not significant ($P = 0.099$).

When the whispered tone is heard, their recognition rate will decrease, but the decrease in the NH group (33.8%) is more severe than in the CI group (20.6%). This indicates that the NH children's judgment on tones is more affected by F0 than the CI group.

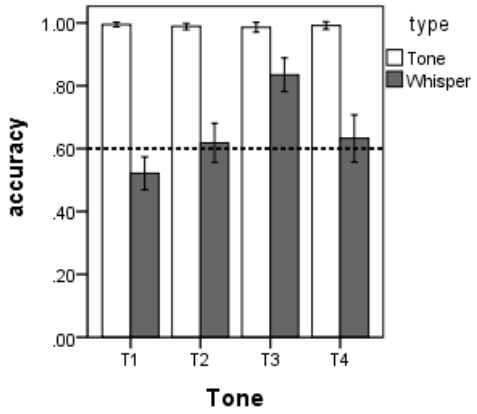

Figure 5 – Accuracy of tone identification for normal speech and whispered speech in the NH group

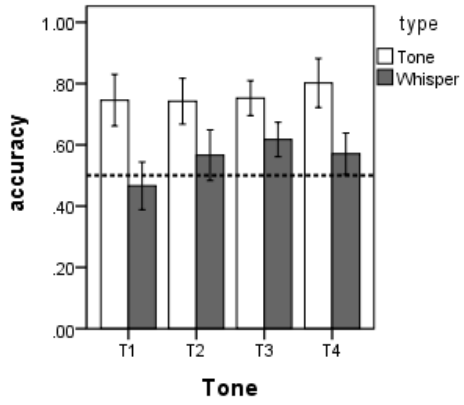

Figure 6 – The tone recognition rates for normal speech and whispered speech in the CI group

4. Discussion and conclusion

Because cochlear implants are not directed to tone languages, the presence of a deaf-speaking cochlear implant in a native language has been a concern for academics. Peng et al. (2004) showed that the rate of tone discrimination in CI children was 72.88%, in comparison to at least 95% for NH children. The present study found that children in the cochlear implant group were significantly weaker than the children in the hearing group, regardless of their ability to recognize natural speech in a quiet environment or the ability to identify whispers that lacked the fundamental frequency.

In the tone identification experiment for normal speech, the rate of recognition in the NH group was as high as 98.99%, and the correct rate of the cochlear group was 75.62%. This is basically consistent with the results of previous studies. At the same time, repeated-measures ANOVA showed that there was no significant difference between the two groups in recognize the four tones, and the ability to distinguish the four tones was relatively uniform. Probably because the current multi-cochlear cochlear implant uses the acoustic information extracted by the limited filtering channel on its corresponding electrode, which causes the cochlear implant to not distinguish the fundamental frequency well, and the primary information of the tone perception is the base. Frequency. Xu et al. used vocoder technology to simulate the cochlear implant and found that the number of filter channels is proportional to the tone discrimination ability of the cochlear implanter. This indicates that the low-frequency resolution may

be that the cochlear implanter perceives the tone as inferior. An important factor in hearing people.

However, by examining the vocabulary tonal recognition ability, we found that in the face of missing whispers of the fundamental frequency, the recognition rate of children in the NH group was only 64.89%, while the recognition rate of the CI group (55.09%) was also close to the chance level, indicating that this task was difficult for both groups. The experimental results of the two groups of children also have certain regularity. Their identification of T2 and T3 is easily confused with each other. The recognition rate of four sounds is ranked from high to low: T3>T4>T2>T1. This is also consistent with the results of previous studies on tone learning. Chen and other scholars have found that children with cochlear implants have a slower pace of tone learning, but the order of their tone acquisition is consistent with that of hearing children. T2 and T3 have similar acoustic characteristics, which is why the two are susceptible to confusion in perception. Through repeated measures analysis of variance, we found that there was a significant difference in the interaction between the group and the tone, especially in the recognition rate of the upper voice, the score of the hearing group (83.5%) was significantly higher than that of the cochlear group (61.7%). The highest recognition rate of the upper voice may be because the average duration is the longest, which indicates that the children in the hearing group can basically grasp the clues other than the fundamental frequency in the whispering perception, such as the duration and fine structure to perceive the tone, while the cochlear implant is implanted in the child because of the equipment. Defects do not perceive fine results well at baseband encoding, and they cannot be used to aid tone identification.

By comparison, the tone identification rate in normal speech is significantly higher than in whispered speech. The recognition rate of the NH group (91%) is significantly higher than that of T1 (75.8%), T2 (80.3%) and T4 (81.2%). At the same time, the recognition rate of T2 is also significantly higher than that of T1. For the CI group, T4 (68.6%) > T3 (68.5%) > T2 (65.4) > T1 (60.6%), but there was no significant difference between the four tones. However, the CI group have better recognition of the upper and lower sounds, indicating that the cochlear implant children can use duration information to distinguish between the upper and the lower sounds. Moreover, the decline in the recognition rate in the NH group from the ordinary tone experiment to the whisper tone experiment (33.8%) was greater than in the CI group (20.6%). This indicates that the NH group's judgment of tones is affected more severely by F0 than the CI group.

Tone recognition is important in speech communication for CI children. Because of the limited sample size, this experiment did not strictly control the auditory speech level and parental guidance of children before cochlear implants. In order to continue to investigate the specific factors affecting the tone perception ability of children with CIs, and if these factors interact, it is necessary to expand the sample size and conduct a large number of studies.

ACKNOWLEDGEMENTS

We are grateful to all participants, their parents and the schools who took part in the research. This research was supported by the Major Program of the National Social Science Fund of China (13&ZD189) and the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences (2017STD006).

REFERENCES

Sha D Q, Li X L, Xu B L. 耳语音声调特征的研究. 电声技术, 2003(11): 4-7.

Liu S, Samuel A G. Perception of Mandarin lexical tones when F0 information is neutralized. Language and Speech, 2004, 47(2): 109-38.

Fu Q J, Zeng F G. Identification of temporal envelope cues in Chinese tone recognition. Asia Pacific Journal of Speech, Language and Hearing, 2000, 5(1): 45-57.

Kong Y Y, Zeng F G. Temporal and spectral cues in Mandarin tone recognition. Journal of the Acoustical Society of America, 2006, 120(1): 2830-40.

Ciocca V, Francis A L, Aisha R, et al. The perception of Cantonese lexical tones by early-deafened cochlear implanters. Journal of the Acoustical Society of America, 2002, 111(1): 2250-6.

Lee K Y, van Hasselt C A, Chiu S N, et al. Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. International Journal of Pediatric Otorhinolaryngology, 2002, 63(2): 137-47.

Peng S C, Tomblin J B, Cheung H, et al. Perception and production of mandarin tones in prelingually deaf children with cochlear implants. Ear & Hearing, 2004, 25(3): 251-264.

Xu L, Chen X, Lu H, et al. Tone perception and production in pediatric cochlear implants users. Acta Otolaryngology, 2011, 131(4): 395-8.
Wei C G, Cao K L, Chen X W, et al. 语前聋儿童人工耳蜗植入患者的电极极间辨差阈与声调识别的初步研究. 中华耳鼻咽喉头颈外科杂志, 2004, 39(2): 73-76.
Liang Z A. 汉语普通话中声调的听觉辨认依据. 生理学报, 1963(2): 3-10.