Effect of vocal cord polyp on Mandarin tones recognition by native Chinese speakers

Bin Li
Department of Chinese Language and Literature
Peking University
Beijing, China
1701110707@pku.edu.cn

Infat Lo
Department of Chinese Language and Literature
University of Macau
Macau, China
infatlo@um.edu.mo

Jiangping Kong†
Department of Chinese Language and Literature
Peking University
Beijing, China
jpkong@pku.edu.cn

ABSTRACT

Intelligent Diagnosis for pathological voice contains two parts. One is intelligent detection, and the other is intelligent comprehension. Before the application of intelligent comprehension, it is important for us to know how human perceive pathological voice. This paper first investigated acoustically a patient who had vocal cord polyp read Chinese characters with Mandarin tones before and after the surgeries. Second, identification test was used to find out the effect of vocal cord polyp on Mandarin tones recognition. The results show that first, the effect of vocal cord polyp on Tone 1 and Tone 3 in F0 contour are statistically significant, but not Tone 2 and Tone 4. Second, vocal cord polyp does not affect identification of Mandarin tone types. Third, vocal cord polyp affects Tone 1 and Tone 2 in identification rate significantly, but not Tone 3 and Tone 4. It is concluded that vocal cord polyp has little influence on the intelligibility of Mandarin tone types. And our research results are referential to intelligent diagnosis for pathological voice.

CCS CONCEPTS

• Artificial intelligence → Natural language processing → Speech recognition

KEYWORDS

vocal cord polyp, lexical tones, perception, identification test, Mandarin

ACM Reference format:

Bin Li, Infat Lo and Jiangping Kong. 2020. Effect of vocal cord polyp on Mandarin tones recognition by native Chinese speakers. In *2020 International Symposium on Artificial Intelligence in Medical Sciences (ISAIMS 2020). ACM, Beijing, BJ, China, 5 pages.*

ISAIMS 2020, September, 2020, Beijing, China

https://doi.org/10.1145/1234567890

1 Introduction

An important subfield of AI is natural language processing, concerned with systems that understand language.[1] Take pathological voice for instance, if a patient had pathological voice, he / she would go to see a doctor. In diagnosis, the doctor might understand or not clearly hear his / her words phonetically. Provided that AI is used in diagnosis for patient with pathological voice, and the proceeding of which AI comprehend his / her words at the phonetic level mainly contains two parts. One is intelligent detection [2], and the other is intelligent comprehension. In order to make AI understand speech by patient with pathological voice, in the first place we would like to know how human perceive it.

Perceptual phonetics has been trying to find out how human beings perceive speech sound. From the perspective of whether the voice is normal or not, the speech sound we perceive could be classified into speech sound with normal voice and with pathological voice. Chinese is a tone language. Perceptual researches of Mandarin tones were mostly relevant to normal voice. For example, Wang and Li [3] studied tone 3 in different environments from perception experiment. However, perception of Mandarin tones with pathological voice is rare.

Vocal cord polyps [4], one of the frequently occurred pathological voices, which usually occur in adult men who use their voices excessively, are fluid-filled lesions that appear on the free edge of the vocal folds.

For native Chinese speakers, it is easy to identify Mandarin tone types. And if the tones were produced by a patient with vocal cord polyp, would they still be easy to recognize?

In classical perception experiment of tones, identification test and discrimination test are used. The former requires the participants to compulsorily choose a stimulus after they hear two stimuli as one of two tones shown on the screen. The latter requires the participants to decide whether two stimuli they heard are similar or not.

This paper investigates how vocal cord polyp affects the perception of Mandarin tones from identification test. And the results would be referential to medical diagnosis for pathological voice by artificial intelligent comprehension.

2 Experiment

2.1 Devices

Recording equipment were computer (ThinkPad e570c), external sound card (SBX Sound 5.1 Pro), audio console (XENYX302USB), microphone (SHURE SM58S), and EGG (ElectroGlottograph 7050A). The recording software was Adobe Audition 3.0 with 16 bits quantization using a 20 kHz sampling rate. Dual channel recording was used with the left channel as EGG signal and the right channel as speech signal.

2.2 Procedure

Procedure in this experiment mainly contained two sections. First, a patient with vocal cord polyp was recruited to record his reading of Chinese characters in CV syllables before and after the surgeries. And acoustic analysis was used to find out the influence of vocal cord polyp on F0 contour of Mandarin tones. Second, 20 Chinese participants were recruited to listen to Chinese characters mentioned above randomly. And perceptual results were analyzed to find out the effect of vocal cord polyp on identifying Mandarin tones.

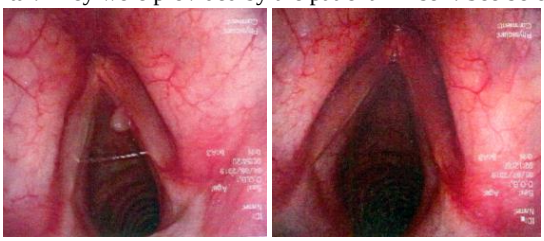
2.3 Participants

A 38-year-old male patient, who had vocal cord polyp, is from Macau, China, proficient in Mandarin and Cantonese. And the recording before and after the surgeries were done in a quiet room.

20 Chinese participants, whose ages are between 25 and 39, attended the identification test. And there were 10 male and 10 female participants. They all could speak Mandarin fluently and did not have hearing impairment. And they were paid for their work.

2.4 The images of vocal cord polyp

Before and after the surgeries, the images of vocal cord polyp which were from the patient's diagnosis report by Beijing Haidian Hospital. They were provided by the patient himself. See below.



(a) (b)

Figure 1: Images of vocal cord polyp before (a) and after (b) the surgeries

In Figure 1, the original images were reversed. And it was a unilateral polyp. It could be seen that before the surgery (see Figure 1(a)) the left vocal cord was smooth, while there was a small polyp in the front part of right vocal cord, causing a gap in vibration. After the surgery (see Figure 1(b)), the left and right vocal cords were smooth.

2.5 Reading list

The reading list for acoustic analysis and identification test were chosen from Contemporary Chinese [5]. And they are Chinese characters in CV syllables with plosives as the initial consonants and /a/, /o/, /e/, and /u/ as their following vowels. Here, /i/ is excluded because several Chinese characters with /i/ as their vowels were forgotten to list in reading list before recording. See below.

Table 1: Reading list for acoustic analysis and identification test

	Tone Types	Chinese Characters in CV Syllable
Tone 1	high level	/pa/, /ta/, /ka/, /po/, /p^ho/, /ke/, /k^he/, /pu/, /p^hu/, /tu/, /t^hu/
Tone 2	mid rising	/pa/, /ta/, /ka/, /po/, /p^ho/, /ke/, /k^he/, /pu/, /p^hu/, /tu/, /t^hu/
Tone 3	low falling rising	/pa/, /ta/, /ka/, /po/, /p^ho/, /ke/, /k^he/, /pu/, /p^hu/, /tu/, /t^hu/
Tone 4	high falling	/pa/, /ta/, /ka/, /po/, /p^ho/, /ke/, /k^he/, /pu/, /p^hu/, /tu/, /t^hu/

In table 1, the tone pitches from Tone 1 to Tone 4 are 55, 35, 214, and 51 respectively. And 1 is the lowest pitch, 5 the highest.

The Chinese characters in Table 1 were read twice. Totally, there were 176 valid samples before and after the surgeries. And the samples were segmented in Adobe Audition 3.0.

2.6 Extraction of acoustic parameters

The first recording of each character was used in acoustic analysis. Totally, there were 88 samples. We had recorded these samples in two channels with the left channel as EGG signal and the right channel as speech signal. And 88 samples in speech signal were used here. In acoustic analysis Praat was used to extract their F0 in normalization of which 20 points from initial measuring point to final measuring point were evenly extracted. The initial measuring point we selected depended on pulses shown in Praat. Normally, we chose the second pulse [6]. The final measuring point was before the end of intensity contour. And in measuring, elbow in head and drop in tail of a tone were out of consideration.

Then F0 would be converted to semi-tone [7] to transcribe in tone pitch. And the formula of semi-tone is below.

$$Semi - tone = 12 * \log 2\left(\frac{f1}{f2}\right) \quad (1)$$

where f1 is F0 of measuring points, and f2 is the lowest F0.

2.7 Perceptual test

The stimuli in perception experiment were from Table 1. And they were not changed acoustically. The first recording of each character was used in identification test. In totality, there were 88 samples. We had recorded these samples in two channels with the left channel as EGG signal and the right channel as speech signal. And 88 samples in speech signal were used in identification test.

The perception experiment in E-Prime 1.1 was identification test which was programmed by Linguistic Lab at Department of Chinese Language and Literature, Peking University.

20 participants were required to wear headphones to listen to 88 stimuli twice randomly. In E-Prime 1.1, stimuli were played

twice randomly. Each time the stimulus was played, then an option interface appeared to require the participant to choose the stimulus he / she heard as one of four Mandarin tones within 5 seconds. Participants needed to choose "←" (the left arrow), "→" (the right arrow), "↑" (the upper arrow), and "↓" (the lower arrow), which were shown on the screen, in the keyboard as Tone 1, Tone 2, Tone 3 and Tone 4 respectively. The test, which had exercise section before formal test, was conducted in a laptop in a quiet room. And each test lasted approximately 15 minutes.

2.8 Data analysis

Excel and SPSS 19 were used in data analysis. In Excel, data would be analyzed, meanwhile intercept function and slope function were used to analyse F0 contour in Mandarin tones. In SPSS 19, paired samples T test would be used to analyze acoustic data and perceptual results.

3 Acoustic analysis

The first recording of each character in table 1 was used in acoustic analysis. In totality, there were 88 samples. We had recorded these samples in two channels with the left channel as EGG signal and the right channel as speech signal. And 88 samples in speech signal were used in acoustic analysis.

There is creaky voice in Tone 3. Therefore, F0 could not be extracted, and there are five empty points. Creaky voice in Tone 3 was mentioned by Belotel-Grenié and Grenie [8], who using H2-H1 and F1-H1 analyzed four tones in Standard Chinese to find out that creaky voice is a redundant cue for Tone 3 and Tone 4.

In mean value, before the surgery, Tone 1 is 135.81Hz, Tone 2 is 118.3Hz, Tone 3 is 111.39Hz, and Tone 4 is 147.72Hz. After the surgery, Tone 1 is 143.64Hz, Tone 2 is 114.69Hz, Tone 3 is 92.4Hz, and Tone 4 is 153.45Hz. And T test (t(3)=0.37, p=0.736) shows that the mean value before and after the surgeries is not different significantly.

In the F0 range, before the surgery, Tone 1 is 15.27Hz, Tone 2 is 21.6Hz, Tone 3 is 48.17Hz, and Tone 4 is 35.33Hz. After the surgery, Tone 1 is 8.92Hz, Tone 2 is 23.23Hz, Tone 3 is 16.52Hz, and Tone 4 is 52.12Hz. And T test (t(3)=0.483, p=0.662) shows that the mean value before and after the surgeries is not different significantly.

Next, before lexical tones in Hz would be converted to semi-tone (see (1)), we would like to introduce Mandarin tone types briefly. See below.

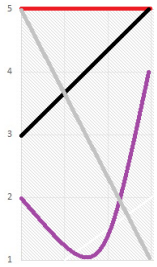


Figure 2. Mandarin tone types

In Figure 2, Tone 1, Tone 2, Tone 3 and Tone 4 are represented in red line, black line, purple line and grey line respectively. And in Y-axis numbers from 1 to 5 are pitch height. Tone 1 is a high-level tone, Tone 2 is a mid-rising tone, Tone 3 is a low falling- rising tone and Tone 4 is a high-falling tone. Tone 1 could be transcribed as 55, Tone 2 as 35, Tone 3 as 214, and Tone 4 as 51. In reality, Tone 1 might be 44.

In the following, lexical tones before and after the surgeries would be compared with tone types in Figure 2. In Figure 3 and Figure 4, the X-axis is point, and their correspondent semi-tone is in Y-axis.

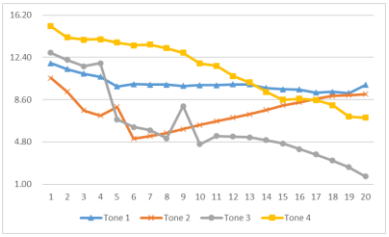


Figure 3. Semi-tones of lexical tones before the surgery of vocal cord polyp

From the above figure, we could see that after the F0 converted to semi-tones, the lexical tones are from 1 to 16.2 in semi-tone. If they are transcribed in tone pitch, Tone 1 is 33, Tone 2 is 323, Tone 3 is 421 and Tone 4 is 53. And generally, four tones before the surgery is relatively concentrated. Compared with tone types in Figure 2, before the surgery, Tone 1 is a mid-level tone with a falling trend a little bit. Contrary to our understanding, Tone 2 is a contour with a sharp turning point at point 6. Tone 3 falls to point 10 and then drops with a slope to point 20 and Tone 3 becomes a falling tone. Tone 3 in Mandarin should be a (low) falling-(high) rising tone, but Tone 3 before the surgery of vocal cord polyp does not rise in the final session. Two points of F0 in Tone 3 were deleted because they are outliers. Tone 4 is a falling tone, and from its head to its tail is relatively above middle part in Y axis. Before the surgery, Tone 2 is changed from a mid-rising tone to a contour, and Tone 3 is altered from a contour to a high falling tone. In all four tones, there were some unregular points extracted.

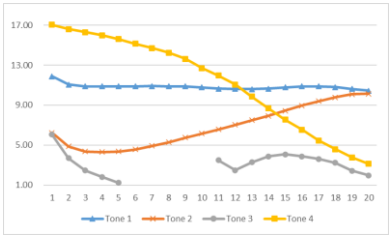


Figure 4. Semi-tones of lexical tones after the surgery of vocal cord polyp

From Figure 4, it could be seen that after the F0 converted to semi-tones, the lexical tones are from 1 to 17 in semi-tone. If they are transcribed in tone pitch, Tone 1 is 33, Tone 2 is 23, Tone 3 is 212 and Tone 4 is 52. Compared with tone types in Figure 2, after the surgery, Tone 1 is a mid-level tone. Tone 2 is lowered entirely and is a low-rising tone. Tone 3 is a low falling-low rising tone,

and it could be seen that the final parts should be elevated to a high point but still lowered. One point of F0 in Tone 3 was deleted because it is an outlier. Tone 4 is a high-falling tone but its tail is not lower than Tone 4 in Figure 3.

In Excel by slope function and intercept function, before the surgery, the slope of Tone 1, Tone 2, Tone 3 and Tone 4 are -0.75, 0.31, -3.38, and -3.7 respectively, and the intercept are 143.72, 115.09, 146.87 and 186.56 respectively. After the surgery the slope of Tone 1, Tone 2, Tone 3 and Tone 4 are -0.26, 2.18, -0.13 and -6.48, and the intercept are 146.35, 91.75, 93.86 and 219.84 respectively. And T test shows that the slope (t(3)=-0.548 p=0.622) and intercept (t(3)=0.55, p=0.621) do not differ significantly.

Mandarin tone types are that Tone 1 is high level (55), Tone 2 is mid rising (35), Tone 3 is falling-rising (214) and Tone 4 is high falling (51). Comparatively, before and after the surgeries of vocal cord polyp, as for tone shape (20 F0 points extracted by Praat) in Hz analysed in SPSS 19, Tone 1 (t (19) = -6.902, p=0.000) and Tone 3 (t (14) = 3.355, p=0.005) differ significantly, but not Tone 2 (t (19) = 1.853, p=0.079) and Tone 4 (t (19) = -0.526, p=0.605).

4 Analysis of identification experiment

A participant would listen to 88 samples twice randomly. Then, each tone has 22 samples which contain 11 samples before the surgery and 11 samples after the surgery. We would like to calculate the choices of 20 participants, from the perspective of each tone before or after the surgery. And there are 40 choices for each tone before or after the surgery. The rate of actual choices and total number of choices for a tone (which is 440) could be used to analyze perception results.

Table 2: Actual choices, total choices and rate of each tone before and after the surgeries

Tones	Before the surgery			After the surgery		
	AC	TC	R	AC	TC	R
Tone 1	394	440	89.55%	413	440	93.86%
Tone 2	401	440	91.14%	424	440	96.36%
Tone3	420	440	95.45%	426	440	96.82%
Tone 3	413	440	93.86%	417	440	94.77%

In Table 2, AC refers to actual choices, TC refers to total choices and R refers to rate. From the above table, we could see identification rates for all tones before or after the surgery are above 90% except that of Tone 1 before the surgery. The mean rate before the surgery is 92.5% and after the surgery is 95.45%.

And the rate of each tone in Table 2 could be shown in a bar chart as below.

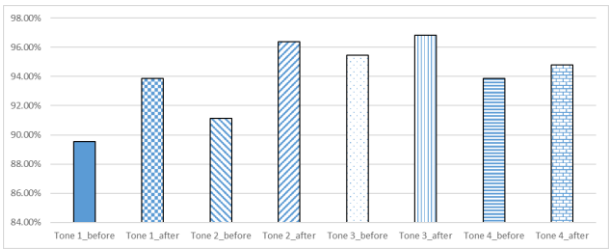


Figure 5: The rate of identifying each tone before and after the surgeries

It could be seen from the above figure that before and after the surgeries, identification rates of Tone 1 and Tone 2 ascend clearly, while Tone 3 and Tone 4 have little difference. And in rate of identifying lexical tones before the surgery, Tone 3 ranks first, followed by Tone 4, Tone 2 and Tone 1. And, after the surgery, identification rates of four tones all rise. Generally, Tone 3 and Tone 2 rank first, followed by Tone 4 and Tone 1. Comparatively, in identification test, Tone 3 and Tone 4 are higher than Tone 2 and Tone 1 in identification rate of lexical tones before the surgery of vocal cord polyp, and Tone 3 and Tone 2 are higher than Tone 4 and Tone 1 in identification rate of lexical tones after the surgery of vocal cord polyp.

5 Discussion

Before and after the surgeries, differences in fundamental frequency were mentioned in previous research, for example, Gan [9] analyzed 60 patients with vocal cord polyp before and after the operation and found that, after four weeks of the surgery, fundamental frequency is statistically significant in mean value, maximum value and range. The previous analysis was not based on Mandarin tones. But it could be seen from previous research that F0 differs before and after the surgeries for patient with vocal cord polyp.

In our analysis, the slope and the intercept of F0 before and after the surgeries are not statistically significant. In this research, we confirm that before and after the surgeries Tone 1 and Tone 3 differ in F0 contour significantly and Tone 2 and Tone 4 do not.

In identification test, as one of the perception experiments used here, the mean identification rates of four tones before and after the surgeries are all above 90% (except Tone 1 before the surgery). In mean identification rate, the rate after the surgery is only about 3% higher than the rate before the surgery. And this confirms that identifying Mandarin tone types are categorical, even though the tones were read by patient with vocal cord polyp. The ranking of identifying four lexical tones before the surgery of vocal cord polyp is that Tone 3 ranks first, followed by Tone 4, Tone 2 and Tone 1. Meanwhile, after the surgery, Tone 3 still ranks first, followed by Tone 2, Tone 4 and Tone 1. In SPSS 19, T test shows that before and after the surgeries, the identification rates are not statistically significant (t(3)=-2.767, p = 0.07). And it still could be seen that in the differences before and after the surgeries in the identification rate, comparatively Tone 2 and Tone 1 are higher than Tone 3 and Tone 4. Then, we calculated identification rate of each participant in Excel and analyzed in SPSS 19. T test shows that Tone 1 (t(19)=-2.373, p = 0.028) and Tone 2 (t(19)=-3.437, p = 0.003) differ significantly, and Tone 3 (t(19)=-0.972, p = 0.343) and Tone 4 (t(19)=-0.777, p = 0.447) do not.

It is summarized that acoustically vocal cord polyp influences Tone 1 and Tone 3 in F0 contour significantly, but not Tone 2 and Tone 4. Perceptually, from the perspective of tone types, Mandarin tone types recognition are not influenced by vocal cord polyp. From the perspective of each single tone before and after

the surgeries, vocal cord polyp affects Tone 1 and Tone 2 in identification rate significantly, but not Tone 3 and Tone 4.

In intelligent diagnosis for pathological voice, there are two parts. One is intelligent detection, and the other is intelligent comprehension. As for voice with vocal cord polyp, theoretically Tone 4 is easy to be detected and understood, Tone 3 is not easy to be detected but easy to be understood, Tone 2 is easy to be detected and is not easy to be understood relatively, and Tone 1 is not easy to be detected and comprehended relatively.

6 Conclusion

The main findings in this research are: first, the effect of vocal cord polyp on Tone 1 and Tone 3 in F0 contour are statistically significant, but not Tone 2 and Tone 4. Second, vocal cord polyp does not affect identification of Mandarin tone types. Third, vocal cord polyp has an effect on Tone 1 and Tone 2 in identification rate significantly, but not Tone 3 and Tone 4.

It is concluded that vocal cord polyp has little influence on the intelligibility of Mandarin tone types. And our research results are referential to intelligent diagnosis for pathological voice.

ACKNOWLEDGMENTS

This research was funded by Major Projects of Ministry of Education of China. Project Name: Language Ontology Research based on Multi-Modal. Project No. 17JJD740001.

REFERENCES

[1] K. Frankish and W.M. Ramsey. 2014. *The Cambridge Handbook of Artificial Intelligence*. UK: Cambridge University Press.

[2] A. Ali and S. Ganar (2018). Intelligent Pathological Voice Detection. *IJIRT*, 5(5), 92-95.

[3] W.S-Y. Wang and K-P. Li (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, 10(3), 629-636.

[4] Z. Ali, M.S. Hossain, G. Muhammad, and A.K. Sangaiah (2018). An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. *Future Generation Computer Systems*, 85, 19-28.

[5] B.R. Huang and X.D. Liao. 2011. *Contemporary Chinese*. Vol 1. Beijing: Higher Education Press.

[6] X.N. Zhu. 2010. *Phonetics*. Beijing: The Commercial Press.

[7] J.P. Kong. 2015. *Elements of Experimental Phonetics*. Beijing: Peking University Press.

[8] A. Belotel-Grenié and M. Grenié (1994), Michel. Phonation types analysis in standard Chinese. *International Conference on Spoken Language Processing*, Yokohama, 343-346.

[9] Q. Gan. 2017. *A Study on voice evaluation before and after operation of vocal cord polyp*. Southwest Medical University.

† Jiangping Kong (author for correspondence) [jpkong@pku.edu.cn], professor at Department of Chinese Language and Literature and Center for Chinese Linguistics, Peking University, No. 5 Yiheyuan Road, Haidian District, Beijing 100871, P. R. China.