

Analysis on Individual Characteristics of Stylistic Differences in Speech

Hongbing Zhang

Criminal Investigation Police University of China, Shenyang, China

zhanghongbing@cipuc.edu.cn

Keywords: Style; Tone; Voice; Pitch Patterns

Abstract: Individual characteristics of stylistic differences in speech play a particularly important role in the modern Chinese system, and the differences in pronunciation are especially obvious in different moods, places and environments. Based on such stylistic differences, this paper starts from three aspects: the long-term average power spectrum, the pitch pattern and the conventional acoustic characteristics, aiming to compare and analyze the differences between reading and speaking from a large number of data, and to study the individual characteristic changes of two different styles.

1. Introduction

With the development of society and the progress of science and technology, the voice forensic technology has been gradually reformed and improved. In the field of phonetic forensic expertise, the differences in the style of speech can be divided into reading and speaking. The natural oral communication and reading have obvious differences. The style of spoken language is fluent, including stress and rhythm, also accompanied by the phenomenon of rhotic accent. However, the rhythm of the reading style is not strong, and the intensity of each syllable is basically the same. Based on the basic method of voiceprint technology, this paper proposes an idea of analysis and comparison by comparing the different features of a large number of read and spoken corpora. It uses the long-time average power spectrum and pitch pattern spectrum to conduct the inspection and comparison, and analyzes the differences between the two types of language and the individual characteristics of the two sounds.

2. Causes of stylistic differences

2.1 External factors

The formation of stylistic differences is due to the changes in the environment, psychology, emotion and physical conditions, etc., but the indirect cause of the differences is affected by the changes in the objective carrier of voice acquisition. Usually collected at the scene of a crime, the voice is the natural communication between suspects in the spoken language, and in the process of handling a case, the voice is read in the form of a "scripted". The environment of field voice acquisition, the process of transmission and the selection of the collector are all different from the conditions in the later recording of sample voice, thus resulting in stylistic differences^[1].

2.2 Internal factors

In audio recordings, speakers often do not consciously cover up or disguise their speech patterns. In the presence of a police officer, the speaker's mental state is bound to change and his physical state to be different from his previous one. Some criminals deliberately use disguised voice reading, changing the tone, intensity, stress and other aspects of voice, resulting in the

^[1] Chen Liang, Zhang Xiongwei. A Novel Approach for Speech Camouflage Communication Based on Speech Quality Evaluation[J]. Signal Processing, 2003,19(5).

difference between reading and speaking^[2].

3. The expressive characteristics of stylistic differences

3.1 Tone and intonation variation

The main difference between reading and speaking is the tone. Intonation is closely related to tone. The main function of intonation to tone is to change the upper limit, lower limit and the width of tone sphere, which reflects the essential characteristics of intonation. To study the difference between reading and speaking is usually to compare the intonation from the perspective of sentences. In essence, it is necessary to grasp the close relationship between intonation and tone and compare the specific features of the range of tone. In the spoken and read sentences, the tone of the voice changes in a complex way. We can sense the characteristic changes in weight, intonation, rhythm of the two sounds^[3].

3.2 Features of "Rhotic accent"

The essence of "rhotic accent" is that the word "er" loses its independent status in the Chinese language and becomes an action of retro-flexion, which is combined with the preceding word for pronunciation, and the word "er" at the end of the word will be linked with the preceding word^[4]. The specific feature is that two syllables are fused into one syllable, which leads to phonetic variation. Due to the changes in the pronunciation environment and the different moods of the reader, the pronunciation of syllables in the corpus is relatively clear, while the phenomenon of "rhotic accent" in syllables is not obvious. The pronunciation of words and sentences is usually clear and stable, which is different from the pronunciation in the natural state of spoken English. In addition, when criminal suspects read the materials, some words that were originally "rhotic accent" were pronounced as rhotic accent, which is different from that in spoken language^[5].

3.3 Degree of pitch

Due to reading spoken language is reading in the rhythm of "scripted" one, the speaker pays less attention to the intensity of the words in the state. However, in the state of speaking, the degree of pitch will be more obvious. Therefore, the two styles show difference on degree of pitch, and the differences are specifically shown on the pitch and the duration, which is also a form of distinction.

3.4 Features of speech flow

The phenomena of dissimilation and weakening are most closely related to stylistic difference. Dissimilation is to avoid the phenomenon of tongue twister in pronunciation, which will obviously change the pronunciation of words. Therefore, it is not common in reading standard language. However, in spoken language, the pronunciation is produced according to people's habits, and some words will be dissimilated, which is in contrast with the reading style. In spoken English, the typical situation is that the tone is weakened, the reading of the language will be stronger, and the weakening phenomenon is not obvious^[6].

^[2] Zhang Guiqing, Jin Yizhu, Liu Hongwei, Cui Xiaoyi. Study on Changing Rules of Electronic Disguised Voice [J]. Evidence Science, 2010,10.

^[3] Li Sijing. A study on the record of Erhua sound in the history of Chinese phonology [J]. Chinese studies, 1981, (1).

^[4] Lin Tao. The difference of individual pronunciation in the rhyme of Beijing dialect [J]. Chinese studies, 1982, (2).

^[5] Sun Zhong. A new study on the phonetic nature of the phenomenon of "Er Hua" in Mandarin [J]. Journal of Chizhou Teachers College, 2005, (4).

^[6] Song Weiwei, Du Xinyu. A Simple Method about Third-order Curve Fitting[J]. Information Technology, 2008,6.

4. Experiment and analysis

4.1 Analysis of basic acoustic characteristics

Choose four students as informants and prepare four corpora. Each informant randomly selects a corpus for reading and speaking the pronunciation in the two states, keeping the same pitch in pronunciation. Meanwhile, choose and intercept four groups of typical words from the corpora of four informants in two styles ("hailang", "weixin", "chongfu" "yige"). Then, measure the average formant data of the pronunciation; extract the broad band visible spectrogram and the visible intensity contour and then carry on the analysis and comparison.

Comparison of the characteristics of the formant between the reading and speaking styles of the same informant:

Table 1. Comparison of the formant data of the two groups of corpora(unit: Hz)

| | “Hailang ” ”Reading | “Hailang ” Speaking | “Weixin ” ”Reading | “Weixin” Speaking | “Chongfu ” ”Reading | “Chongfu” Speaking | “Yige” Reading | “Yige” Speaking |
|----|------------------------|------------------------|-----------------------|----------------------|------------------------|-----------------------|-------------------|--------------------|
| F0 | 131 | 126 | 153 | 140 | 137 | 129 | 154 | 132 |
| F1 | 699 | 675 | 453 | 436 | 585 | 457 | 437 | 302 |
| F2 | 1489 | 1339 | 1916 | 1724 | 1325 | 1289 | 1600 | 1276 |

In reading and spoken styles, the influence between the spoken syllables is not obvious when reading, and the transition between syllables is stable and can be distinguished^[7]. In the spoken language style, the link between syllables is very obvious, and the two syllables are almost linked together on the spectrum. In addition, the frequency position of the end of the formant in the transition of syllables of the two styles is also different, and there are different degrees of difference between different people. Because the transition part between the vowel and the consonant is a characteristic representation, there are certain differences of the vowel-consonant characteristics between reading and natural speaking.

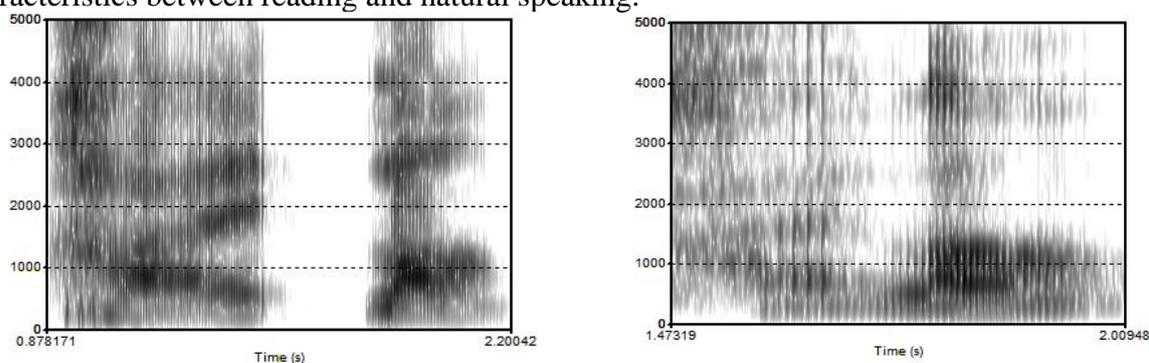


Figure 1. Corpora “hailing” reading and speaking

4.2 Long-term average power spectrum analysis

Praat speech analysis software is used to record the corpus of two styles, speaking and reading, of three speakers separately. Then the software was used to extract and measure the long-time average power spectrum, and the recorded speech time was divided into two cases, kept at about 30s and 7min, respectively, then the comparison and feature analysis between the spectra was performed^[8].

^[7] Feng Qingqing. Experimental study on the pitch pattern of Harbin dialect[J]. Liaoning Technical University Journal(social sciences version), 2008,5.

^[8] Zhang Cuiling. Research on court speech technology [M]. Beijing: China social press,

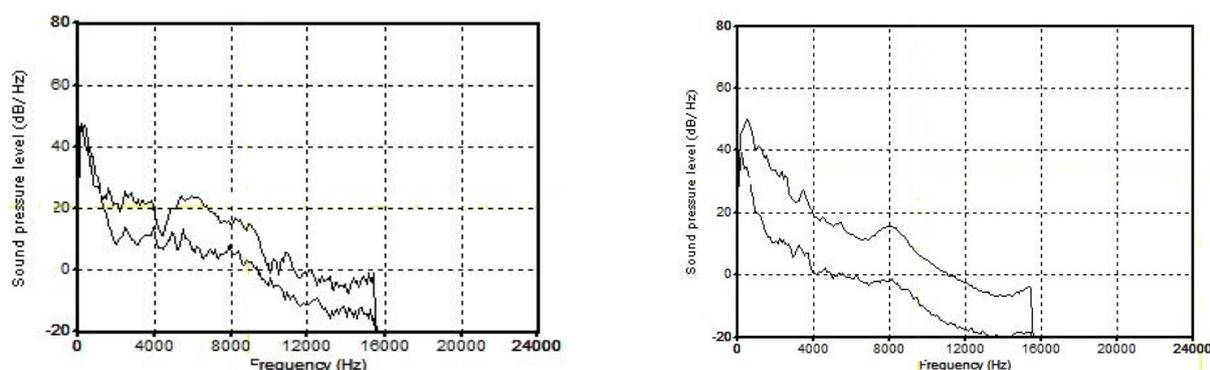


Figure 2. Comparison of long-term spectrum of two different styles

Table 2. measurement data of long-term average power spectrum of different styles (unit: dB)

| | style | maximum | maximum difference | median | standard deviation | slope |
|-----------------------------------|----------|---------|--------------------|--------|--------------------|--------|
| Informant1 | reading | 51.39 | 5.72 | 31.57 | 31.95 | -18.85 |
| | speaking | 45.67 | | 25.12 | 28.21 | -21.83 |
| Informant2 (different corpora) | reading | 50.08 | 10.65 | 33.75 | 32.48 | -12.33 |
| | speaking | 39.43 | | 20.03 | 25.84 | -21.02 |
| Informant3 | reading | 48.90 | 6.93 | 30.86 | 32.23 | -19.42 |
| | speaking | 41.97 | | 23.83 | 28.96 | -21.87 |

The comparison of the data and the analysis of the spectrum show that there is a certain difference in the long-term average power of the same speaker's reading and speaking styles, which generally exists between the speakers, but through the analysis of the trend of the data and the spectrum, the overall difference between the two styles is not too large^[9]. In addition, the change of reading and speaking corpus content will also affect the morphological characteristics of long-term average power. By analyzing the spectral form, standard deviation and slope of the two styles, we can find that there are differences between reading and speaking, but the differences are not very significant. There are different features of long-term average power among different speakers. The length of corpus time will have an impact on long-term average power^[10]. The longer the text time is, the more obvious the features of long-term average power spectrum will be.

Conclusion

This paper expounds the subjective and objective reasons for the formation of stylistic differences and the specific acoustic characteristics of stylistic differences, and explains the differences between reading and speaking. Then we use the analysis methods of long-term average power spectrum and formant spectrum to compare the differences between the two styles. By comparing the trend, shape and numerical value of long-term average power spectrum, we can grasp the differences between reading and speaking from the overall point of view, and

2009:34-45.

^[9] Zhuang Lin. speaker recognition using long-term average FFT power spectrum [J]. Journal of Shanxi Police Academy, 2011 (01).

^[10] Zhang Cuiling. Study on the effectiveness of long-term average power spectrum in forensic speaker identification [J]. Journal of China Academy of criminal police, 2004 (09).

provide help for judicial inspection.

Acknowledgements

This work was supported by Research on Dynamic Feature Extraction of Speaker Based on Deep Learning (Special funds for basic scientific research fees of Central Universities No.3242019012) and Research on the standardization of judicial application of audio-visual evidence in the context of Litigation Reform (No.2017LLYJXJXY040).

References

- [1] Chen Liang, Zhang Xiongwei. A Novel Approach for Speech Camouflage Communication Based on Speech Quality Evaluation[J]. Signal Processing, 2003,19(5).
- [2] Zhang Guiqing, Jin Yizhu, Liu Hongwei, Cui Xiaoyi. Study on Changing Rules of Electronic Disguised Voice [J]. Evidence Science, 2010,10.
- [3] Li Sijing. A study on the record of Erhua sound in the history of Chinese phonology [J]. Chinese studies, 1981, (1).
- [4] Lin Tao. The difference of individual pronunciation in the rhyme of Beijing dialect [J]. Chinese studies, 1982, (2).
- [5] Sun Zhong. A new study on the phonetic nature of the phenomenon of "Er Hua" in Mandarin [J]. Journal of Chizhou Teachers College, 2005, (4).
- [6] Song Weiwei, Du Xinyu. A Simple Method about Third-order Curve Fitting[J]. Information Technology, 2008,6.
- [7] Feng Qingqing. Experimental study on the pitch pattern of Harbin dialect[J]. Liaoning Technical University Journal(social sciences version), 2008,5.
- [8] Zhang Cuiling. Research on court speech technology [M]. Beijing: China social press, 2009:34-45.
- [9] Zhuang Lin. speaker recognition using long-term average FFT power spectrum [J]. Journal of Shanxi Police Academy, 2011 (01).
- [10] Zhang Cuiling. Study on the effectiveness of long-term average power spectrum in forensic speaker identification [J]. Journal of China Academy of criminal police, 2004 (09).