

## Research on the Pitch Pattern of Voice Variation

Hongbing Zhang

National Police University of China Shenyang, China

zhhbn@sina.com

**Abstract**—Differentiated voice is the common check material form in judicial voice testing field. There are many factors that lead to the voice variation; the more common approach is the voice channel, Stylistic variation, disguise and reverberation etc. The variation brings about changes in hearing, atlas and parameters for voice, which brings about many difficulties to speaker identification. According to the variety of speech variation mode we explore the variation of the internal principle, and get fundamental frequency variation rule before and after voice change of multiple corpuses by analyzing map and data. The results show that the fundamental frequency before and after voice variation exists a linearity relationship, we can realize speaker identification in voice variation field through comparing Chinese pitch pattern.

**Keywords**—voice variation; pitch pattern; fundamental frequency.

### I. INTRODUCTION

With the improvement of the scientific development and the construction of legal system, audiovisual materials is the law of criminal procedure as litigation evidence on the basis of many modern science and technology, speech evidence is an important part of audio-visual data. When the speech test technology becomes gradually mature, the inspectors start to pay more attention to the corpus of access mode. However, due to the hardware conditions and acquisition environment restriction it is difficult to guarantee the samples voices are collected by the same way in the case of speech identification process. When we record samples and samples, the differences in voice transmission mode urge the speech distortion in the hearing, for example, the samples is obtained through the internal recording method in mobile phone; the sample is obtained by direct recording mode. Secondly, the factors such as spatial reverberation have a great impact on the authenticity and reliability of voice, which will des troy the facticity of voice, thus which will affect the work of speech test. Therefore, we must consider the impact of these differences in the inspection process.

In addition, during the process of voice test, the collection of the samples and sample is affected by the subjective factors and objective environment, and there are different ways of expression in speech style. There is some speech phenomenon such as natural speech, speech in measured tones for signification and incomplete speech. Yet, the sample shows a complete but blunt dull tone speech, sometimes even the existence of non-coherent words speech.

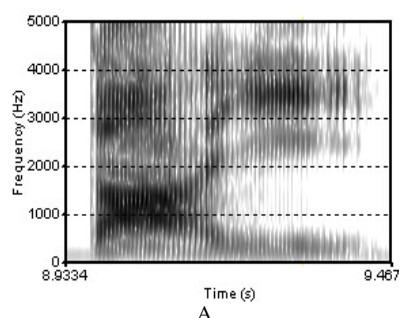
These differences in speech bring us a lot of trouble in hearing test, at the same time, which make the inspectors think of the influence on the mapping analysis and quantitative testing from the voice variation. In this paper, we will conduct a specific research to the above influence and draw the related conclusion through theoretical analysis, experimental comparison and statistical analysis of data etc. which will provide a powerful complement to the judicial speech identification.

### II. THE INFLUENCE ON THE WAY OF COLLECTION

#### A. The influence of reverberation on voice

Different from the outside, voice build the complex sound field in the propagation of sound in the indoor, the sound heard by us in the indoor consists of three basic components, such as the direct sound, early reflected sound and reverberation sound. Due to the presence of reverberation, the intensity of sound was strengthened, and the voice extended duration from the sense of hearing; When reverberation space is large, the sound will be strengthen or weaken in some frequency band due to the factors such as echo, focusing.

Analysis results have shown that the formant appear straight when the speech is affected by spatial transmission factor. The tail of formant produces the phenomenon of obvious delay, and certain syllables can produce new concentrated area of energy under the effect from the reverberation. These kinds of change are bound to affect or even mislead the identification from the identification of personnel to the formant morphology. According to measure and analyze the fundamental data from the reverberant speech, we found that the changes of the fundamental level from each syllable in different effect, especially the form of frequency curve hasn't changed markedly. (As shown in Fig.1)



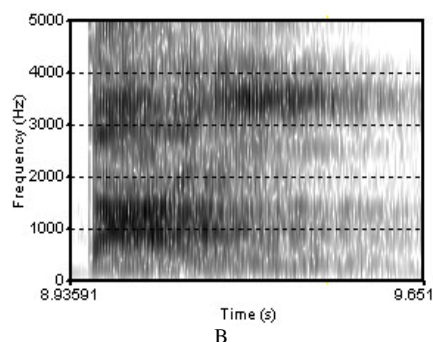


Figure 1. The broadband sonogram from different reverberation effect  
The A map is made from normal voice; the B map is made from moderate reverberation voice;

### B. The influence of channel on voice

Voice channel refers to the way of speech signal transmission, which can be divided into wired and wireless channel. Because of various practical factors, the band provided by the channel is always limited, and the channel's characteristics always existing imperfection. When the speech signal is transmitted in the channel, it always generates amplitude and phase distortion, and then generates the waveform distortion. When the speech signal is transmitted in the telephone channel, it usually can be affected by several factors such as the total attenuation, attenuation distortion, noise, frequency deviation and harmonic distortion.

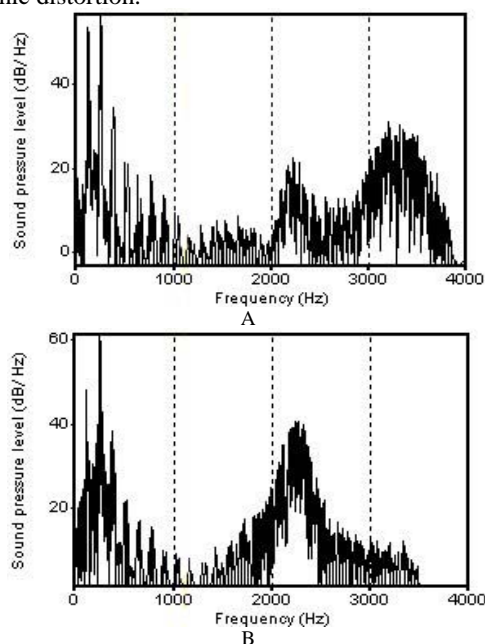


Figure 2. The frequency spectrogram of normal voice and channel voice  
The A map is made from the normal voice; the B map is made from the channel voice

The experimental results showed that, the voice generated distortion in the hearing after interfered by the channel, the frequency components of voice signal would be missing.

Through the comparison of the acoustic spectrum, we could find that the position and shape of formant haven't obvious change, the energy of the formant located in the high frequency region enhanced obviously, and the energy of the formant located in the low frequency region decreased obviously. The fundamental data from four tones wasn't obviously changed. The results of data analysis showed that, the formant's values of each syllable changed obviously because of the influence of channel. The reason for this phenomenon was that the Frequency distribution of voice changed, and the method for measuring mean also brought error. When we get such variations voice, we can't get an objective conclusion through the traditional examination method. So we can try to use the pitch pattern to avoid the interference brought by the variation of atlas.

### C. The influence of stylistic differences on voice

Style, as a basic concept in Phonetic Stylistics and Stylistics, has been widely used in many subjects. It is a language characteristic system that has been formed in applying the verbal behavior of the language used by all the people and fits in with the needs of different fields, goals and contents of communication. It is a historical outcome resulted from functional differentiation, which has been formed due to the long selective use of linguistic data (words, sentence patterns, phonetic devices, figures of speech and etc.) in the process of language using. In other words, different requirements in human communication are the social basis of the style, and functional differentiation of linguistic data is it's the material basis.

Affected by the speaker's subjective intention and environmental factors during the sample and specimen recording, there are always different ways of expression in speech style in the process of judicial speech identification. The pronunciation of the sample is natural acoustically, rhythmical for expressing ideas and has incomplete pronunciation. However, although it has complete pronunciation, the specimen shows the rigid and flat intonations and even non-consistency between words. Such differences of pronunciation has caused many difficulties for the judicial phonetic identification, and made it necessary for identifiers to take into consideration its effects on spectrum observation and quality analysis.

Tones of normal read speech, generally under the influence of speed of speech and emotions, stress cadence. Fast reading speed is used to express nervousness, agitation, happiness and excitement; low speed is used to show solemnness, sad, lament and depression; and normal speed is suitable for narration. Analyzing from linguistic standpoint, there are many elements that can cause the phonetic changes in spontaneous speech, including speed variations, personal speaking style, dialect, rhyme, context and vocabulary, etc. And this can be easily found in judicial phonetic identifying. Also, for the speaker with different expression purposes and less schooling, the phonetics of the sample often suffer the pronunciation distortions, including initial variation, finals variation, retroflex suffixation, and tone variation, etc.

The results of data comparison and analysis suggest that differences do exist and exist generally between the LTASS

of those two styles. The mean values and slopes of curve of the two styles are very close. The general mean value of changes shows that the differences between various styles are less than the differences between speakers. Long Term Average Spectrum (LTAS) analysis indicates that the differences between styles are more obvious than the differences caused by different time periods, which can be proved by power spectrum curve in later stage.

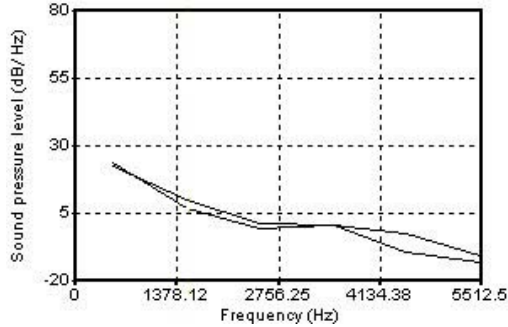


Figure 3. The Long Term Average Spectrum of two speakers' different styles

According to the change rules from the acoustic characteristics of above variation speech, and considering Chinese pitch mode theory, we analyzed deeply the regularity of the fundamental change in value, and compare the difference between the two pitch mode curves. We can provide scientific and effective methods for the inspection of such speech through the comparison of pitch modes.

#### D. The influence of electro camouflage on voice

In the process of experiment, two different voice changer applications were analyzed and used to create disguised voices. The acoustic features, such as pitch, formant, tone, energy, and zero-crossing rate, of both the disguised voices and the original voices were analyzed. Results showed that, compared to the original voices, the acoustic features of the disguised voices had their particular changing rules. The experiment also found that, according to the rules, the disguised voices could be easily recovered very well.

TABLE I. THE CHANGES OF THE FUNDAMENTAL DATA FROM FOUR SPEAKERS

	A	B	C	D	Average	STDEV
Raise	0.80	0.82	0.78	0.82	0.81	0.009
Reduction	1.32	1.30	1.33	1.31	1.32	0.01

We respectively changed the low tone and high tone of these four speaker using voice changer, measured the fundamental frequency value before and after voice changed and made statistical analysis, the statistical standard deviation was less than 0.05. That result showed that fundamental frequency value before and after voice changed and disguised had strong correlation, and had strong linear ratio relation that the fundamental frequency value from

voice changed can go back to the original voice by multiple relations.

### III. CHINESE PITCH PATTERN

Since Mandarin Chinese pitch changes in different, they form four tones in Mandarin Chinese. The Chinese pitch changes bear rich speech, linguistic information, these complex changes also reflects the speaker's pitch features. Generally we use tone shape and tone pitch to describe Chinese tone. Tone shape is a reflection of fundamental frequency changing morphology when the vocal cords vibrate. The fundamental frequency changing range is from the lowest tone to the highest tone which called tone domain. There are many internal difference existed in Chinese pitch. It is necessary to do a detailed decomposition of pitch change to analyze the various personality characteristics, so we can gain the specific differences from individual tones.

Fundamental frequency as one important acoustics character in speech check field has indispensable application value about speaker identification work. The pitch difference will not change tone, but the pitch change will change tone. Since every person has different compass, and check material is easy to camouflage, channel and voice extraction methods, etc. factors are different in judicial speech testing process, that speaker fundamental frequency is prone to some degree of variation, so the fundamental frequency data has not comparability. It's particularly difficult to make speaker identification because of variation speech which is the common check materiel form in judicial speech testing field, and electroacoustic camouflage is a typical individual pattern of the speech variation.

Because of the existence of speech variation, the comparability of the fundamental data was denied. In view of the above problems, some researchers have proposed a tone pattern theory for comparison of the speaker's pitch characteristics. We can firstly measure the fundamental data of different measuring points, and then the fundamental data will be normalized and relatively analyzed, some curves are used to characterize the changes in the pattern of speaker's four tones, which will avoid the interference of external factors on the fundamental data to some extent.

#### A. Fundamental frequency acquisition and processing

Now there are many methods of pitch extraction, someone simply adopt to choosing syllable region directly and get the average value or measure harmonic wave frequency value sometime and divided by harmonic wave number then make data analysis. Someone use complex algorithm to extract the fundamental frequency and consider the noise resistance performance and other factors. The former method's operation is simple but has large measurement error, and makes little sense for fundamental frequency comparison. The latter method has higher extraction precision but require complex algorithms and large number of data processing in later stage. However apply either algorithm to get fundamental tone periodic orbit that could not coincide with the true fundamental tone orbit perfectly. Based on this problem the author uses PRAAT software to make labeling and measuring analysis for

fundamental frequency, adjust the curve data at same time to ensure smooth of the curve.

Averagely select 10 sampling points from pitch curve which been extraction and smoothed, measure their fundamental frequency value respectively. About the same tone we use length average value ( $L$ ) to make normalization processing, and curve within  $L + 20\%$  length range as the measuring object, in order to avoid the influence brought by time length of fundamental frequency.

The dots mean fundamental frequency value of that time point which been analyzed by PRAAT software. If the fundamental frequency value is neat without sudden flip phenomenon, the analysis result is basically correct. If the fundamental frequency point free on the overall we need pay special attention and modification. Fundamental frequency value extraction errors typically occur near starting and ending position of fundamental frequency band. In addition if the fundamental frequency is relatively low or periodicity is not clear, that requires prior check about doubtful point of fundamental frequency value to ensure the error within 4Hz and do the accurate measurement with the help of narrow-band speech map.

Direct measurements of the fundamental frequency data have some floating ranges that do not meet the speaker pitch characteristic of overall pronunciation. Now the scientific method is combined with Zhao Yuanren's "five degree tone-marking method" as well as T value calculation formula of the tone pattern, to normalize the fundamental frequency data within 5 degree range. On fundamental frequency curve smoothing process, we find other types of curves appear the breakpoint region except the high-level tone curve. So consider the average value of high-level tone as the fundamental reference value to form pitch curve model is more accurate.

#### B. Pitch pattern curve fitting

After data measurement and five degrees normalization processing, the fundamental frequency data can been multiple fitted directly through Excel for minimum radius of curvature, etc. problems, thereby forming a smooth pitch pattern curve. Through the observation and comparison of the pitch pattern diagram, we can easily summarize speaker's pitch levels and characteristics. In addition to the comparison of their overall morphology change, also can make specific comparison analysis through parameters describe method. Such as, Upper limit value and lower limit value of pitch, average value of high-level tone, initial and ending value of each curve, the slope rate of curves rising section and descending section, etc. in the model. Otherwise special attention is need to the "elbow section" and "tail section" of the curve. Considering the starting and end point of pronunciation stability problem, when make comparing analysis we need to make the appropriate correction.

### IV. THE EXPERIMENT AND ANALYSIS

By selecting different speaker to make normal and variation speech from different corpus, we can analysis the fundamental frequency value change before and after camouflage, and use each corpus to fit their pitch model

curve. We can research & analysis the pitch pattern curve shape stability and Feasibility using pitch patterns for speaker identification through comparative analysis the curve shape and pattern parameter of pitch.

#### A. The experiment corpus and design

This experiment selects four speakers (three men and one woman) to do normal and electroacoustic disguised pronunciation about fifty syllables within four tones, that we form twenty corpus including two hundred syllables. Then we will select the professional recording studio as recording environment to ensure a high signal to noise ratio.

#### B. The experiment result analysis

The pitch patterns comparison of same corpus contents between different speakers. Comparison of four patterns can see that different speaker's curve pattern has certain difference. A and C curved in the end compared with B and D, that shows the former two speakers stop tone in rising tone end. A and C have obvious difference in the cross point of positive tone and falling-rising tone compared with B and D. The former speaker's parameter is above 4, especially C, and the latter speaker's parameter is under 4, especially D. This parameter shows their difference in their positive sound rising speed or falling-rising tone falling speed. In addition there is obvious difference in blank area which formed by each speaker's positive tone, rising tone and falling-rising tone curve crossed. This analysis results show that there is essential difference between each speaker's pitch pattern curves.

The pitch patterns comparison of the same speaker's corpus before and after variation. Fig.4 shows the pitch pattern curves respectively extract from normal and variation corpus.

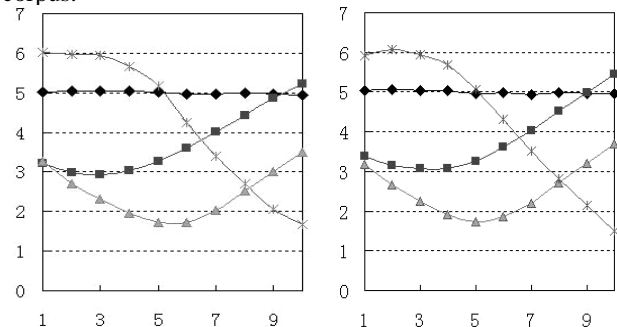


Figure 4. The pitch pattern curve from normal corpus and variant corpus

Although there are some differences between the two patterns, but from the overall shape and each pitch pattern parameters' comparative analysis, the pitch pattern did not change their quality before and after variation. The tape of curves from the figures is same basically, and the relative position relation between curves is very similar. Diapason of the pitch in the mode char distribute between 6.1 and 1.7 basically, and each curve is basically the same at the initial value. The end positions of the rising tone and falling tone have certain difference; it is due to the poor stability of end tone and measurement errors in the process of speech. The each features of two mode curve has consistent performance,

which the change rules of the each tone are the same before and after variation, the difference was caused by measurement error. As a result, the speaker identification method using pitch pattern is applicable to voice change camouflage speech.

## V. CONCLUSIONS

Chinese is a tone language, so comparative analysis for its pitch become important aspects of Implementation of speaker identification during judicial speech testing process, especially for speech camouflage, the comparative analysis of Chinese pitch pattern can more objectively and accurately reflect the speaker's pitch variation.

This article combines with five degree tone-marking method and tone pattern theory, using PRAAT and Excel and other conventional software to realization of speakers' pitch pattern measuring and fitting extraction. Based on statistical analysis of fundamental frequency values before and after voice change about every corpus, we know that speech fundamental frequencies before and after voice change exists strong linear ratio relation, and then respectively compared the pitch pattern of the same speaker in curve shape and model parameter before and after voice change. Comprehensive analysis of experimental results, I sum up that pitch patterns exist relatively stability of speaker-self; there are obvious difference in pitch pattern between different speakers. We can make speaker identification in variation speech field in help with

comparison of pitch patterns curve shape and parameter difference method.

## REFERENCES

- [1] Cui Jingxu. Audio-visual Materials Examination [M]. China Shenyang: The textbook of China Criminal Police University, 2004.
- [2] Nolan.F. The Phonetics Bases of Speaker Recognition [M].Cambridge University Press, 1998.
- [3] L. Lei, G. Zhichao, L. Youchun, M. Hao, T. Zhangwen, "A 975-to-1960MHz Fast-Locking Fractional-N Synthesizer with Adaptive bandwidth control and 4/4.5 prescaler for digital TV tuners", IEEE International in Solid-State Circuits Conference, pp. 396-397, 2009.
- [4] Zhou Wenzhen, "A Modified Electronic Ballast Circuit with Extended Conduction Time", IJACT: International Journal of Advancements in Computing Technology, vol. 4, no. 3, pp. 133-140, 2012.
- [5] Jerng and C. G. Sodini, "The Impact of Device Type and Sizing on Phase Noise Mechanisms", IEEE Journal of Solid-State Circuits, vol. 40, no. 2, pp. 360-369, 2005.
- [6] Ren Guocan, Wang Songliang, Shao Hua, "The Study on Embedded Memory Circuit Design in Low-power SoC Chips for Mobile Systems", IJACT: International Journal of Advancements in Computing Technology , vol. 4, no. 1, pp. 155-162, 2012.
- [7] Kurachi, S.; Yoshimasu, T.; Itoh, N.; Yonemura, K.; , "5-GHz Band Highly Linear VCO IC with a Novel Resonant Circuit", 2007 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, pp.285-288, 2007.
- [8] Chen Liang, Zhang Xiongwei. A Novel Approach for Speech Camouflage Communication Based on Speech Quality Evaluation[J]. Signal Processing, 2003,19(5).
- [9] Feng Qingqing. Experimental study on the pitch pattern of Harbin dialect[J]. LiaoNing Technical University Journal(social sciences version), 2008,5.