

The Experimental Analysis on the Change of Pitch

Hongbing Zhang

China Criminal Police University, Shenyang, 110854, China

zhhbn@sina.com

Keywords: Spectrum, Fundamental, Speech, Spectrogram

Abstract. The voice is a kind of composite wave, timbre is different depending on the different overtones. The timbre, namely audio overtones or harmonics components, is determined by the number of the overtone. The pitch is one of the most important parameters of speech. In the process of collecting phonetic evidence, we always find out the conditions of sampling rate are different, especially when the emotional state of speaker is unstable, we would use comparison of sound spectrograms and statistics to explore what influence has the change of the pitch had on sound spectrograms and the measurement of formants. The final conclusion of the study will have positive effect on the job of phonetic identification.

Introduction

As the most important one of the acoustic features in the field of judicial phonetic identification, the pitch has indispensable value of application in speaker identification work. When the speaker pronounces the tension of emotions and other factors are different because of the difference between its line degrees of the vocal, the differences between the pitch are very obvious. It determines the basic law of the speaker's individual characteristics. The pitch is changing during the different age of a same person because of the growth and development of vocal cord. The tone fluctuates frequently because of the influence of the psychological and emotional factors, which is due to the change of pitch. In the process of collecting samples, we found that the speaker's pitch changes obviously, which is a significant phenomenon. It is due to the suspects in crime are very nervous, for example: When the kidnappers call the victim's family to ask ransom, he often feels confused and stressed. In order to intimidate the victim's family, the suspect's "tone" will be improved deliberately. In addition, sometimes suspects will use some types of hardware and software deliberately to change the voice-frequency properties during the speaking process or the later processing procedure in order to cover up their pronunciation feature.

When a person answer the question from the police, whatever his attitude or mood are not same as which in the criminal process, the changes of pitch bring a lot of obstacles to the visual inspection. The suspect's pitch is passively changed, its main features is reflected in the changes of value (one high and one low). The pitch of speech samples is often higher in crime, while the pitch is lower when the suspect is interrogated. The significant differences between the speakers' pitches will have influences inevitably on voice map, and they would also bring about a lot of difficulties to judicial speech test. For exploring the influence of pitch changes on the voice map, and the method to restore the changed map, we have done a lot of researches, which has not been conducted comprehensive analysis of experiment in the judicial field. In this paper, we use the conventional voice analysis software to make a series of studies on voice sample which the pitch has been changed greatly, and investigate how variation of the pitch affect the maps of voice. On this basis, we have found a solution to improve the reliability of the measured values, thus providing the reference for related work in judicial speech test.

Effect of Pitch in the Voiceprint Identification

The adult male's vocal cord size is similar to a fingernail; it vibrates and generates source wave of speech because of the impact of airflow from the lungs, which is called the glottal wave. The glottal

wave is a sort of complex wave; we can break up the glottal wave into base band and a series of harmonics by using the Fourier transform. The glottal wave is only the source of voice, itself does not have any meaning of speech, but we can decompose the glottal base band into F_0 (sometimes also called the first harmonic) and a series of harmonics by using the Fourier analysis. Obviously, the pitch and its harmonics are the origin of constituted of speech.

The Timbre and the Pitch. As is known to all, the vibration brings waveform, the overall vibration send out the pitch, and the vibration of each part send out the overtone. The voice is a kind of composite wave, timbre is different depending on the different overtones, The timbre, namely audio overtones or harmonics components, is determined by the number of the overtone. Because there is a direct relationship between the level of the pitch and distribution of intensity of its harmonics, the spectral analysis is the distribution of the harmonic in effective energy range of speaker. Therefore, changes in the level of the pitch will have some influences on the distribution of spectrum. If we use a high frequency spectrum or envelope to measure the related numerical, which will bring us large errors.

Comparative Analysis of Different Fundamental Spectrum. The intensity of the harmonic distribution affects the concentration of resonant peak energy directly, in other words, it will affect the bandwidth of the formant. Therefore, the pitch changes in high and low, which will bring a lot of troubles in analyzing broadband and narrowband sonogram in later visual inspection. The broadband and formant spectrum with high pitch is not conducive to measure. When the speaker intentionally camouflages during the process of pronunciation, his pitch will often become high or higher. The suspects often change the tension of his vocal cords in order to change their pronunciation feature during the process of pronunciation. In addition, they also use voice-changer software and hardware (etc.) to achieve purposes of rising or reducing the pitch, which will bring many difficulties to the upper inspection and appraisal. Therefore, the viable treatment method needs to be come up with, which can improve the voice maps with abnormal pitch, and then we can use them to draw a correct conclusion.

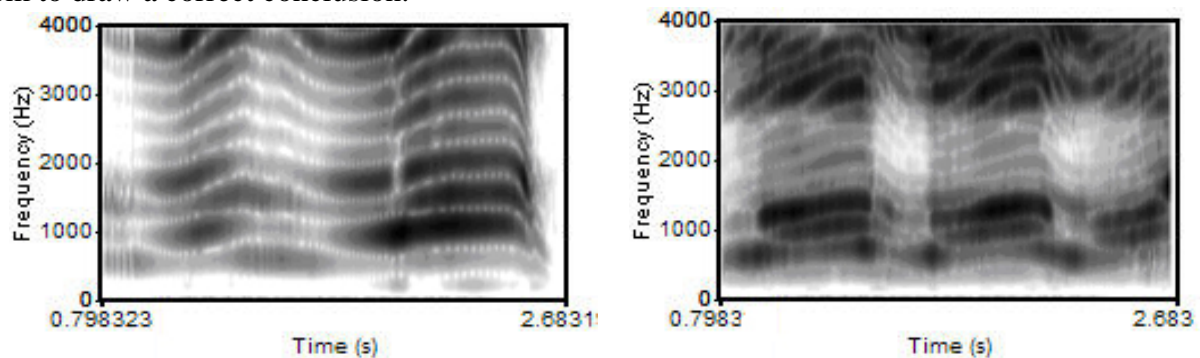


Fig.1 The spectrogram and the frequency spectrogram from the voice with high pitch
The left map and second map are made from a two-year-old girl, $F_0=460$ Hz; the right map and forth map are made from a twelve-year-old girl, $F_0=350$ Hz

The Experiment and Analysis

We chose different speakers who spoke in different pitch value to build speech database; and then we used several methods to measure the value of formants; at last, we put the value of formant from the normal pitch pronunciation as a reference, and tried to use software with the functions of parameter for tuning the high pitch speech become normal. During the process of this experiment, we chose four men, four women and two children as speakers. These people spoke out 50 syllables separately in both normal and camouflaged way, and then we built ten speech databases. We called these speech databases m1\m2\m3\m4\w1\w2\w3\w4\t1\t2. In the recording period, we asked speakers to control vocal state for making the pitch of each pronunciation rise successively. The audio files were mono, audio sample rate was 11025Hz, and digitalizing bit was 16 bit. We recorded these audio files in recording studio and to guarantee high signal to noise.

In this experiment we used the normalized method. Normalization was a simplified calculation method; the parameters with dimensions would be converted into the dimensionless parameters and became scalars by this method. This method were used in a variety of calculation, it was a dimensionless processing method, which turned the absolute value from physical experiment into relative values, and then simplified the calculation and reduced the amount of value. In order to analyze the change scope of pitch directly, we chose the lowest pitch in the speakers' speech databases as reference. The pitch value of other pronunciations would be normalized, and then we could obtain the changing depth of pitch represented by $\Delta F_{02} = F_{02}/F_{01}$ as shown in Table 1. syllables in the database m1, and calculated the changing depth of each pitch, then made the data into a table1.

Tab 1, Different pitch and changing depth

Syllable \ Pitch (Hz)	F ₀₁	F ₀₂	F ₀₃	F ₀₄	$\Delta F_{02} = F_{02}/F_{01}$	$\Delta F_{03} = F_{03}/F_{01}$	$\Delta F_{04} = F_{04}/F_{01}$
1	125	164	213	267	1.35	1.81	2.23
2	110	142	180	238	1.37	1.73	2.22
3	114	154	205	257	1.32	1.72	2.16
4	118	140	181	239	1.28	1.64	2.07
5	121	155	208	252	1.33	1.73	2.11

Tab1 illustrate that F02 distributes around 1.30, F05 distributes between 1.5 and 1.8, and F04 locates between 2.09 and 2.23. F02 is stable. According to such values, we know that the row of pronunciation state is close to the people speak normally. This row of syllable parameter which corresponds to formant is correct, and this can be set as standard. F04 is unstable and undulation. This shows that pronounce is not stable, and we can set this as upper limit. Majority of statistics in F04 exceed 2.0, and it is far away from the normal speaking condition, such camouflage is obvious.

The Pitch Change Influence on the Higher-order Formant. When the value of pitch is very low (F0=118Hz), only the first formant and the second formant appearing in the map from vowel (/a/), which cannot help us provide more information. All formants of /a/ are completely displayed by wideband spectrogram of F02 (F02=155 Hz), the wideband spectrogram of F04 (F02=290Hz) shows “feminine characteristics”, the higher-order formant becomes gradually clearer with the pitch increasing. The main reason is when pitch becomes higher, original parameter setting of normal phonetic analysis software cannot work anymore. The criminal's organs for pronunciation become muscular tension when camouflage by reducing pitch. The speaker must increase the intensity of pronunciation, then the overall energy of speech will be enhanced, this will enhance the harmonic attenuated under normal circumstances, especially in the high frequency region, originally weak energy becomes obvious, which make the formant spectrogram changed obviously. When the pitch is reduced, the organs of pronunciation become muscular relaxation. The speaker must reduce the intensity of pronunciation, and then the overall energy of speech will be weakened. These will weak the harmonic attenuated under normal circumstances, especially in the high frequency region. Therefore, decreased pitch often leads to reduction of energy, and finally leads to a lack of higher-order formant.

The Pitch Change Influence on Lower-order Formant. With the increase of the pitch and gradually rise, F1 rise gradually, especially when we use more visual examine and measure of frequency spectrum. In accordance with general knowledge of phonetics, when male pronounce vowel /i/, the first formant is located around 250Hz. Male's pitch normally distribute between 100Hz and 200Hz. Because pitch and first formant are close, there is a low energy band at timer shaft.

Visual method means that the staff estimates formant using visual method according to the height of formant in the voice maps of PRAAT software. The formant value from four pronunciations basically increases as the frequency increases. However, the fourth formant from F03 and F04 (about 3500Hz) are not in agreement with the fourth formant of F02. The reason lies in the pitch from the two groups of pronunciation are too large, with the advent of the harmonic in the voice map; the error will easily

come out by visual method. When we use ocular estimate to measure the central target value of formant roughly, with increasing of pitch, these formants of four pronunciations will increase. When pitch is on the high side, we find out that statistics of formants is not correct obviously by visual inspection. When we use software to measure statistics of formant, measuring error of low pitch is obviously. With increase of pitch, all formant trajectories are cleared firstly, and then begin to appear burr phenomena on the edge of higher-order formant, especially the fourth formant.

In a word, the changes of pitch affect the measure of formant. Generally, with the increase of pitch, formants are also increased. When the value of pitch is very low, it will cause the resonance peak deletion, which is well reflected in the three methods; high pitch values will produce feminine map, which will affect the vision measurement. In such three test methods, the numerical results of low-order formant measured by using spectral method are not reliable. In order to solve such problem, this paper tries to adjust bandwidth of high feminine map.

When M1 speaks /san/, the speaker's second pronunciation is close to the normal pronunciation. Results of visual method tell us, as the impact of some high frequency harmonic waves, the forth formant is lower than it is in what we get normally, so we need to use high frequency sub harmonic. We can place window length in the "0.002" by adjusting the option "Spectrogram settings" in the PRAAT software. We could get the value of formant (804Hz/1354Hz/2638Hz/3669Hz) from the adjusted approximation of the wideband spectrogram. Before the adjustment, there are big differences between the fourth formant of F04 and F02, but the measurement is the same as the value of F02 after being adjusted. We can work out the pitch's average of pronunciation from m3 according by adjusting the parameter "Formant shift ratio" and "New pitch median" in "Change gender", the first is 109Hz, and the fourth is 228Hz. By adjusting the option "New pitch median" in "change gender" to 109Hz, we can make the pitch of the 2nd pronunciation and the normal pronunciation equally. Secondly, the average value of third and fourth formant is only the equivalent of the 94% of fourth pronunciation. By adjusting "Formant shift ratio" in the option "Change gender" to 0.94, we can get the measurement results before and after adjustment.

Conclusion

We establish the speech corpus for different pitch, and use all kinds of methods to analyze what the impact on measuring formant and voice map when we change the pitch. The conclusion is below: when the pitch is low, the higher-order formant is weakened phenomenon obviously, even disappeared. When the pitch is high, it sounds like women. Broadband spectra also conform to character of women's voice. In the aspect of measuring formant, when pitch become very low or high, the measuring will be incorrect, especially when we use spectrogram to measure formant in low frequency range, it will be incorrect obviously.

In addition, as pitch rise, lower-order formant (especially F1 and F2, which are lower than 1000Hz) appears rising trend. By adjusting the parameter (Window length) of the option "Spectrogram settings" in PRAAT software, we can improve the quality of broadband spectra with higher pitch which is close to normal sonogram. Through adjusting the parameter "Change gender" in PRAAT software, the value of all formants with higher pitch is closed to the voice with lower pitch, but this method changed the sound samples, and its operability needs to be further studied.

References

- [1] Philip Rose, in: Forensic Speaker Identification. Taylor Francis, New York (2002).
- [2] Nolan. F. in: The Phonetics Bases of Speaker Recognition. Cambridge University Press (1998).
- [3] Xiong Ziyu. in: Manual of PRAAT(voice soft).Beijing: China Language Institute of the Academy of Social Sciences(2004).

- [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)