VOICE QUALITY ANALYSIS UNDER THE PITCH EFFECT*

[(1)]*Dan-Ning Jiang* [(2)]*Jian-Hua Tao* [(3)]*Lian-Hong Cai*

Department of Computer Science and Technology,
Tsinghua University, Beijing
[(1)]jdn00@mails.tsinghua.edu.cn {[(2)]jhtao, [(3)]clh-dcs}@tsinghua.edu.cn

ABSTRACT

Voice quality is the perceived timbre of speech. Considering the interaction between voice quality and prosody could improve the quality of acoustic processing in speech synthesis. This paper explores voice quality in accordance with various pitch conditions, and opens out the relationship between them. To ensure the accuracy of analysis, a rich experiment data set is prepared, which contains isolated vowels produced in modal quality, and statistic methods are applied. Acoustic parameter H1-A3 is used as the measurement of voice quality, since it has been proven to be a good indicator of perceived breathiness. The study found that, generally, the amount of breathiness increases with pitch in low pitch stage, and it decreases or changes little in median and high pitch stage. The interaction between voice quality and pitch is also relevant to vowel identities and individual characteristics. Quantitative changing function is estimated through regression analysis. Experiments prove the efficiency of the estimated changing function.

1. INTRODUCTION

Voice quality, which is perceived as the timbre of speech, brings various colors in speech. It changes across speakers, which causes parts of the individual characteristics [1]. However, voice quality is not constant within one speaker. The speaker often changes voice quality during utterances to show his emotion and attitude, and voice quality also interacts with other acoustic features, such as pitch. The present study explores the voice quality in accordance with various pitch conditions, and opens out the relationship between them. Here voice quality mainly refers to the amount of breathiness.

The primary motivation for the study is that a better understanding of the interaction between voice quality and pitch could increase the quality of acoustic processing in speech synthesis. Today's speech synthesis method of unit selection and concatenation, accompanying with large corpus, is proven to be able to get high intelligibility and natural prosody characteristics. However, the corpus is impossible to contain every condition in natural speech, and an acoustic processing procedure is necessary. In most systems, only prosody features are adjusted, while the corresponding changes in voice quality are ignored. This limits the naturalness of synthetic speech a lot. Another motivation is that speaker recognition or verification, and even speech recognition, can be improved with better knowledge of voice quality.

* Supported by 863 program (2001AA114072)

Most previous researches used parameter OQ (open quotient), or the corresponding spectral parameter H1-H2 (the strength relation between the fundamental frequency component and the second harmonic component in spectrum), to measure breathiness. These two acoustic parameters were considered to correlate positively with the amount of breathiness. The relations between the acoustic parameters and pitch were studied, but the conclusions were considerably ambiguous. Some researchers concluded that the amount of breathiness increases with pitch, some concluded that it decreases with pitch, and others concluded that voice quality roughly keeps constant with pitch [1] [5]. Thus, the problem is far from being solved. The primary reason for the ambiguous conclusions may be on account of the influence of other factors out of pitch. Voice quality changes in a very complex and undetermined way, and it could be also influenced by connected-speech, emotion, register of phonation, and so on. Another disadvantage of the previous researches is that they were done only in a qualitative way. However, only quantitative results could be applied in acoustic processing.

The present study emphasizes the influence of pitch and limits that of other factors through the analysis data set and methods. The data set only contains isolated vowels to eliminate the influence of connected-speech and emotional factors. All vowels are produced in modal quality, thus the register of phonation is constant. The data set is considerably large. Statistic methods are used for the indeterminate characteristics of the problem. In addition, the quantitative changing function between voice quality and pitch is achieved by regression analysis.

The paper is organized as follows: section 2 discusses the acoustic measurement of voice quality; the analysis procedure is explained in section 3; section 4 lists and discusses the experiment results.

2. ACOUSTIC MEASUREMENT

Usual acoustic measurements of breathiness could be classified into two main categories of source parameters and spectral parameters according to the estimate procedure [3]. Source parameters are estimated in a two-step procedure. The speech signal is first inverse-filtered to get the glottal source signal, and then a source model is matched to estimate the parameters. One general source model is the four-parameter LF model, whose parameters are EE (the excitation strength), RA (the measure of the return phase), RK (the measure of the symmetry/asymmetry of the glottal pulse), and RG (the measure of the opening branch of the glottal pulse). The familiar parameter open quotient (OQ)

is defined as (1+RK)/2RG. It has been found that breathy voice has high RA, RK, and OQ values [3].

Spectral parameters are estimated directly from the speech spectrum. General spectral characteristics correlated with breathiness are high relative strength of the fundamental frequency component, large spectral tilt, broad F1 bandwidth, noises in high frequency, and so on. Most of the researchers considered that parameter H1-H2 (the strength relation between the fundamental frequency component and the second harmonic component) plays an important role in breathiness, while spectral tilt and noises in high frequency are also crucial [1][2][3]. However, Hanson [4] studied the spectral parameters through listening tests, and found that H1-H2 is not the best measurement of perceived breathiness. Instead, parameter H1-A1 (the strength relation between the fundamental frequency component and the strongest harmonic component in F1 region) and H1-A3 (the strength relation between the fundamental frequency component and the strongest harmonic component in F3 region), which reflect both the relative strength of fundamental frequency component and spectral tilt, correlate with perceived breathiness better.

Though source parameters can describe the glottal characteristics directly, their estimate procedure is labour intensive, for estimating the inverse filter and source model automatically often causes errors and thus an interactive manual procedure is necessary. This makes it impossible to estimate source parameters on large data set [3]. On the other hand, spectral parameters can be estimated conveniently and accurately, and also measure breathiness well. According to the above considerations, spectral parameter H1-A3 is used as the acoustic measurement of voice quality in the study.

3. ANALYSIS PROCEDURE

In analysis procedure, the influence of pitch on voice quality should be emphasized, while that of other factors should be limited to minimum. We get it through the analysis data set and methods.

3.1 Data Set

The data set for analysis contains only isolated vowels to eliminate the influence of connected-speech and emotional factors. All vowels are produced in modal quality, thus the register of phonation is constant. To ensure the accuracy of the analysis results, the data set is considerably large. Speech of three speakers (two males and one female) are involved in the data set with about 200 productions of vowels for each speaker, which cover continuously over large pitch range.

Though voice quality is mainly determined by glottal characteristics and thus should be independent of vowel identities, four different typical vowels are analyzed to make the experiment results more comprehensive. The vowels are front-low vowel [a], front-high vowel [i], back-high vowel [u], and central vowel [e].

3.2 Methods

In analysis, the changing trends of voice quality and pitch are first estimated to find how voice quality is influenced by pitch. The changing trends are composed of H1-A3 values with the maximum probabilities in different pitch conditions, as in expression (1):

$$T(F_0) = \arg\max_{H1-A3} (P(H1-A3 \mid F_0)) \quad (1)$$

To estimate the changing trends, the pitch axis is divided into 30 small intervals from 100hz to 450hz, and Gaussian distribution is used to approximate the probability distribution in each interval. The statistic centers of H1-A3 in the small pitch intervals then form the changing trends. After being normalized in pitch, the changing trends are used to induce changing patterns between voice quality and pitch.

Regression analysis is used to estimate the quantitative changing function. Obviously, the simplest regression analysis method is linear regression, but it could only be applied on data set that fits linear relation. Fortunately, it is roughly satisfied in our data set. So linear regression analysis is performed to estimate the quantitative changing function, and the formula is shown as in expression (2):

$$H1-A3 = \beta_1(F_0) + \beta_0 + \varepsilon \quad (2)$$

4. EXPERIMENT RESULTS

The experiment results are presented in the following three subsections. Section 4.1 presents the changing patterns between voice quality and pitch. Section 4.2 shows the quantitative changing function estimated through linear regression analysis. Some discussions on the experiment results are discussed in section 4.3.

4.1 Changing Patterns

There are four changing patterns between voice quality and pitch induced in the data set, as shown in Fig.1.

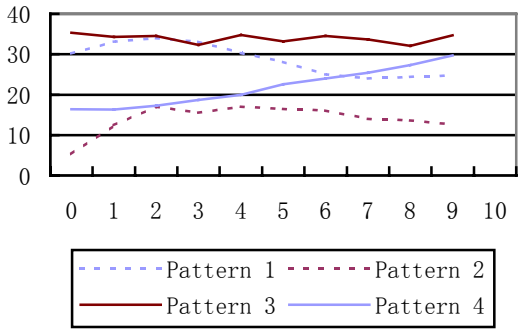

Fig.1. Four changing patterns between voice quality and pitch. The horizontal axis represents the normalized pitch, and the vertical axis represents the value of parameter H1-A3 in db.

Fig.1 illustrates the characteristics of each changing patterns. Pattern 1 has different changing fashions in low, median, and high pitch stage respectively. In low pitch stage, the amount of breathiness increases with pitch, and it then begins to decrease in the median pitch stage, and changes little in high pitch stage. Pattern 2 is some similar with pattern 1, but the amount of breathiness changes little in median pitch stage, and decreases in high pitch stage. In pattern 3, voice quality is roughly constant. The amount of breathiness increases monotonously in pattern 4.

Data satisfying pattern 1 and pattern 2 are basically composed of low vowel [a] and central vowel [e]. Pattern 3 corresponds to back-high vowel [u], and pattern 4 corresponds to front-high vowel [i]. The detailed correspondence between the changing patterns and data is listed in table 1. Representation as female-[a] means the data of vowel [a] produced by female speaker. Obviously, general cases are presented in pattern 1 and pattern 2, while pattern 3 and pattern 4 show some special characteristics caused by high vowels.

	Data
Pattern 1	Female-[a], female-[e], female-[i], Male2-[a], male2-[e]
Pattern 2	Male1-[a], male1-[e], male1-[i]
Pattern 3	Female-[u], male1-[u], male2-[u]
Pattern 4	Male2-[i]

Table 1. The detailed correspondence between the changing patterns and data.

4.2 Quantitative Changing Function

Since voice quality may change differently due to different pitch stage in some cases, linear regression analysis should be performed in each pitch stage respectively. Table 2 to Table 5 lists the linear regression coefficients β_1 and linear correlation coefficients r for data satisfying pattern 1 to pattern 4 respectively, where β_1 is the slope of the linear function as shown in expression (2), and r describes the strength of linear relation.

β_1 / r	Low pitch	Median pitch	High pitch
Female-[a]	0.05/0.42	-0.09/-0.35	-0.01/-0.09
Female-[e]	——	-0.11/-0.66	-0.04/-0.45
Female-[i]	——	-0.20/-0.85	0.02/0.12
Male2-[a]	0.16/0.13	-0.20/-0.34	0.01/0.01
Male2-[e]	0.10/0.06	-0.15/-0.12	-0.12/-0.11

Table 2. Linear regression coefficients β_1 and linear correlation coefficients r for data satisfying pattern 1.

β_1 / r	Low pitch	Median pitch	High pitch
Male1-[a]	0.50/0.53	0.01/0.07	-0.11/-0.37
Male1-[e]	0.53/0.43	0.04/0.30	-0.11/-0.45
Male1-[i]	0.38/0.34	0.08/0.56	-0.01/-0.09

Table 3. Linear regression coefficients β_1 and linear correlation coefficients r for data satisfying pattern 2.

β_1 / r	Low, median and high pitch
Female-[u]	-0.03/-0.59
Male1-[u]	0.04/0.60
Male2-[u]	0.01/0.49

Table 4. Linear regression coefficients β_1 and linear correlation coefficients r for data satisfying pattern 3.

β_1 / r	Low, median and high pitch
Male2-[i]	0.13/0.52

Table 5. Linear regression coefficients β_1 and linear correlation coefficients r for data satisfying pattern 4.

From table 2 to table 5, it is shown that most of the linear correlation coefficients are considerably large, which ensures the linear strength. The linear regression coefficients β_1 match the corresponding changing pattern, except for some particular cases.

MSE (db)	front-low [a]	central [e]	front-high [i]	back-high [u]
Female	1.74	1.30	1.64	0.73
Male1	0.71	0.54	1.38	0.85
Male2	0.47	1.11	0.71	1.11

Table 6. The mean square error (MSE) between the estimated parameters and the statistic centers on each data.

To evaluate the efficiency of the estimated quantitative changing function, we compare the parameters estimated by the changing function with the statistic centers of H1-A3 that form the changing trends. If the estimated changing function were valid, then they would match each other. Table 6 lists the mean square error (MSE) on each data. It is shown that the errors are relative small. The maximum error is 1.74db, which is about 17% of the parameter changing range (about 10db). Fig. 2 shows the matching status between the two parameters on data female-[a], which corresponds to the maximum error condition. It is shown that even when the error is the maximum, the two parameters

match each other well. In fact, the mean square error is mostly caused by the flutters on the changing trends.

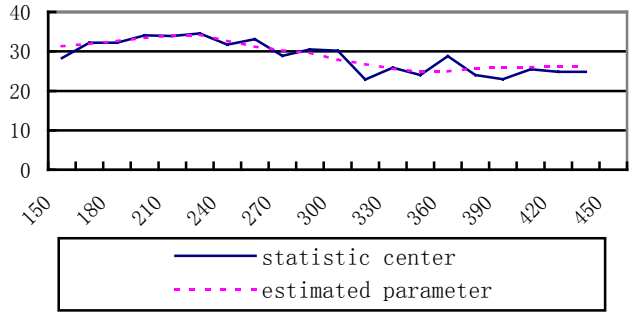

Fig.2. The matching status between estimated parameters and the statistic centers on data female-[a].

4.3 Discussions

Our experiment results show that voice quality could be influenced by pitch. Generally, the amount of breathiness increases with pitch in low pitch stage, and it decreases or changes little in median and high pitch stage. This is consistent with the laryngeal physiology [6]. Laryngeal muscles usually contract to increase pitch, while cause the amount of breathiness in voice decrease at the same time. Thus in median and high pitch stage, the amount of breathiness often decreases with pitch. Besides the laryngeal muscles contracting, pitch could also be increased by the augment of sub-glottal airflow strength. Thus voice quality may change little with pitch when the increase of pitch mostly caused by the augment of sub-glottal airflow strength. However, in low pitch stage, the mechanism is different. In this situation, laryngeal muscles do not contract distinctly when pitch is increased. On the contrary, they often contract to get a considerably low pitch, thus the degree of muscles contracting correlate negatively with pitch. So the amount of breathiness often increases with pitch in this stage.

In addition to the general cases, the interaction between voice quality and pitch is relevant to different vowel identities and individual characteristics. High vowel [i] and [u] often show different characteristics from general cases, which may be according to the influence of the vocal tract shape. A narrow gap is formed between the palate and the high-positioned tongue during the production of high vowels, and fricative noises are generated when sub-glottal airflow passes it. The additive fricative noises change the voice quality.

The quantitative changing function estimated through linear regression analysis has been demonstrated to be valid, for parameters estimated by it match the actual parameters well. However, voice quality could also be influenced by other prosody features, such as energy, duration, intonation, position in the utterance, and so on. Thus H1-A3 parameters show some deviations in each pitch interval, and the estimated quantitative changing function could not predict every H1-A3 parameters accurately, but the statistic centers instead.

5. CONCLUSION

This paper explored voice quality in accordance with various pitch conditions, and opened out the relationship between them. We have found that voice quality could be influenced by pitch. Generally, the amount of breathiness increases with pitch in low pitch stage, and decreases or changes little in median and high pitch stage. The interaction between voice quality and pitch is also relevant to vowel identities and individual characteristics. Quantitative changing function between voice quality and pitch is estimated through linear regression analysis. Parameters estimated by the changing function match the actual parameters well, which demonstrate the efficiency of the estimated changing function.

6. REFERENCES

[1] D. H. Klatt and L. C. Klatt. “Analysis, Synthesis, and Perception of Voice Quality Variations among Female and Male Talkers”, J. Acoust. Soc. Am. 87(2), pp. 820-857, February 1990.

[2] D. G. Childers. “Vocal Quality Factors: Analysis, Synthesis, and Perception”, J. Acoust. Soc. Am. 90(5), pp. 2394-2409.

[3] C. Gobl and A. N. Chasaide. “Acoustic Characteristics of Voice Quality”, Speech Communication, v 11, pp. 481-490, 1992.

[4] H. M. Hanson. “Individual Variations in Glottal Characteristics of Female Speakers”, ICASSP 1995, pp. 772-775.

[5] M. Swerts and R. Veldhuis. “the Effect of Speech Melody on Voice Quality”, Speech Communication, v 33, pp. 297-303, 2001.

[6] Philip Lieberman and Shelia E. Blumstein. Speech Physiology, Speech Perception, and Acoustic Phonetics, New York: Cambridge University Press, 1988.