Electroglottographic Analysis of the Voice in Young Male Role of Kunqu Opera

Li Dong 1,* and Jiangping Kong 2

1 School of Humanities, Shenzhen University, Shenzhen 518060, China
2 Department of Chinese Language and Literature, Peking University, Beijing 100871, China; jpkong@pku.edu.cn
* Correspondence: dongli@szu.edu.cn; Tel.: +86-1581-182-5016

Abstract: The phonation types used in the young male role in Kunqu Opera were investigated. Two national young male role singers volunteered as the subjects. Each singer performed three voice conditions: singing, stage speech, and reading lyrics. Three electroglottogram parameters, the fundamental frequency, contact quotient, and speed quotient, were analyzed. Electroglottogram parameters were different between voice conditions. Five phonation types were found by clustering analysis in singing and stage speech: (1) breathy voice, (2) high adduction modal voice, (3) modal voice, (4) untrained falsetto, and (5) high adduction falsetto. The proportion of each phonation type was not identical in singing and stage speech. The relationship between phonation type and pitch was multiple to one in the low pitch range, and one to one in the high pitch range. The sound pressure levels were related to the phonation types. Five phonation types, instead of only the two phonation types (modal voice and falsetto) that are identified in traditional Kunqu Opera singing theory, were concomitantly used in the young male role’s artistic voices. These phonation types were more similar to those of the young female roles than to those of the other male roles in the Kunqu Opera.

Keywords: young male role; singing; stage speech; electroglottography

1. Introduction
1.1. Young Male Role in Kunqu Opera

The Kunqu Opera is a traditional opera in China and inscribed in United Nations Educational, Scientific, and Cultural Organization’s list of “oral masterpiece and intangible heritage of humanity”. It has been handed down orally since the middle of the sixteenth century and is revered as the ancestor of all Chinese operas. There are at least 10 artists in a theatrical troupe, Jing (colorful face role), Guansheng (hat role), Jinsheng (kerchief role), Laosheng (aged male role), Fumo (second aged male role), Zhengdan (middle-aged woman role), Guimendan (young woman role), Liudan (young girl role), Fuchou (second clown role), and Xiaochou (clown role). Their voice timbres mirror the ages, genders, characters, and identities of the various personages.

Both Guansheng and Jinsheng belong to young male (YM) roles. A Guansheng performer, whose voice quality has been described as “broad and bright” having “a heavy oral resonance”, acts as a young king or a gifted scholar. Jinsheng performers often act in love stories, and his voice quality has been described as “brighter and lyrical” [1]. A YM singer can adjust his voice timbre to play the part of either Guansheng or Jinsheng. From the perspective of pitch range, a YM’s singing is similar to a Western baritone, while the stage speech is similar to a Western tenor [2]. In the traditional singing theory of Kunqu Opera, a YM role uses modal voice in the low pitch range and falsetto in the high pitch range to recite and sing. The passaggio of a YM role was not traditionally fixed. It was from B3 to #F4, which depended on the ages and identities of the personages. The younger the personage was, the lower the passaggio that was adopted [3]. However, the voice timbre in the low pitch range differs from that of the speech, and the falsetto deviates from both
the Western operatic tradition and untrained falsetto, which have been well described in previous research [4–8]. To reveal how the voice can be used in dramatic contexts to create the character of an ancient Chinese young man, it is necessary to investigate the details of the phonation type in scientific terms. The present study focuses on the phonation types from the perspective of electroglottogram (EGG) parameters and investigates (1) the distribution differences of parameters among three conditions, namely, singing, stage speech, and reading lyrics; (2) the phonation types used in singing and stage speech; and (3) the relationships between the pitch and phonation types.

1.2. EGG Analysis

Electroglottography is a non-invasive technique to measure variations in the contact area between the two vocal folds as a function of time. It is related complementarily to the glottal airflow [9]. The peaks in the derivative of the EGG signal correspond to the closing and opening events of the vocal folds [10–12]. A model [9] that pinpoints certain landmarks and their relation to the glottal airflow pulse during normal voicing is widely used to speculate the movement and position of the vocal folds during phonation [12]. Clinically meaningful alterations of vocal fold status and behavior also have been reported to result in different geometric characteristics in the EGG waveform [13]. Different phonation settings result in different phonation types and characteristic EGG shapes [4,14,15]. The production of modal voice is carried out with moderate adductive tension, medial compression, and longitudinal tension in the lower pitch range of a speaker. In a considerably higher pitch range, when the mass of the vocal fold is made stiff and less mobile, which often is accompanied by slightly abducted folds, a falsetto is produced. Unlike modal voice and falsetto, breathiness is a modificatory setting of vocal folds. The notion of breathy voice involves a type of phonation which can be produced over a very wide range of air flows [4].

Three EGG parameters, namely, fundamental frequency ($f_o$), contact quotient (CQ), and speed quotient (SQ), have been found to be associated with phonation types [2,5,16]. The CQ is defined as the ratio between the contact phase of the EGG signal and the fundamental period, as can be seen in Figure 1. The CQ and the closed quotient derived from the glottal flow are not necessarily equal, since transglottal airflow may occur during incomplete glottal closure [17]. For EGG waveforms with a single peak, a high CQ is typically related to a pressed voice, while a low one is commonly observed for breathy voice. However, this is not true for a double-peak EGG signal. The SQ was originally defined for glottal flow as the duration of the opening phase divided by the duration of the closing phase [18]. A high SQ corresponds to faster glottal contact and indicates that the voice has more high-frequency energy [19]. For EGG, the SQ is the ratio between the decontacting phase and the contacting phase (see Figure 1) in the EGG [5]. The SQ is related to the convergence (along with vertical phasing) in the glottis [20] and the degree of tension of vocal fold; the greater the tension (such as falsetto) is, the closer to 100% the SQ will be. In several Chinese dialects and minority languages, compared with modal voice, the special vocal fry showed a lower $f_o$, smaller CQ, and larger SQ; the breathy voice showed a lower $f_o$, smaller CQ, and smaller SQ; the pressed voice showed a lower $f_o$, larger CQ, and larger SQ; and the high-pitched voice showed a higher $f_o$, larger CQ, and smaller SQ [5]. In order to correctly calculate the EGG parameters, three methods have been previously applied in research to pinpoint the moments of glottal contact and of loss of glottal contact [21–25]. The criterion-level method [23] showed better applicability for double-peak EGG signals [14], weak signals with a large high-frequency noise component, and other special cases, and was more appropriate for the analysis of the complex EGG signals of YMs.
Two male national actors of Kunqu Opera (YM1 and YM2 for short) volunteered as subjects. The ages of YM1 and YM2 were 45 and 27 years, respectively, when the signals were collected. Their professional experiences were 27 and 9 years, respectively. Neither of them smoked or had a history of vocal cord disease. Both singers sang four songs (16 min), recited a section of stage speech (2 min) as they would on stage after a warm-up exercise for about 1 min, and read the lyrics of the songs and stage speech (3 min) in a modal voice. Two of the songs were for Jinsheng and were the South song, and the other two songs were for Guansheng and were the North song.

Both singers were recorded in a quiet living room, of a size of about 4 × 5 × 3 m³. Audio was picked up by a Sony Electret condenser microphone placed off axis at a measured distance of 15 cm from the mouth. Sound pressure level (SPL) calibration was carried out by recording a 1000-Hz tone, the SPL of which was measured at the recording microphone by means of a TES-52 sound level meter (TES Electrical Electronic, Corp., Taipei, China). The EGG signal was collected by an EGG system (Electroglottograph Model 6103; Kay, Montvale, NJ, USA). The signals were simultaneously recorded and digitized on 16 bits by recording a 1000-Hz tone, the SPL of which was measured at the recording microphone with the previous study.

The signals were divided into characters. The audio and EGG signals were analyzed character by character using the VoiceLab 1.0 [27] automatically. The SPL values were extracted from the audio signals and calibrated to 0.3 m in order to make an easier comparison with the previous study. The contacting and the decontacting events were approximated using the commonly used 35% of the EGG amplitude criterion [16,23,28], which is advantageous when vocal adduction is to be detected. Three phonatory parameters were calculated: (1) $f_o$, (2) EGG CQ (abbreviated as CQ), and (3) EGG SQ (abbreviated as SQ). The extracted data from each character were stratified by sampling into 30 data, since the duration of the character varied a lot. The song and the stage speech did not contain all speech sounds, but did contain most vowels of the language. The sample size in EGG cycles was 12,150 (YM1’s singing), 4350 (YM1’s stage speech), 11,400 (YM2’s singing), and 3870 (YM2’s stage speech). Statistical analyses were completed using SPSS 18. Since the data did not comply with the normal distribution, and the test of homogeneity of variances was significant ($p < 0.05$), and the distributions of the EGG parameters were compared by Mann–Whitney U tests.

Previous research indicated that the long-term-average spectrum of a YM role showed a large standard deviation, which implies a great variation of voice timbre [2]. To classify the phonation types, the EGG parameters were clustered by k-means method [16,29]. When clustering in three dimensions, at least two clusters in each dimension are needed to differentiate between high and low values. Hence, if the parameters vary independently of one another, then at least eight centroids may be needed. On the one hand, if the parameters of the two centroids in the clustering results are very close, they are considered to be of the same phonation type. On the other hand, the phonation types were determined consulting both parameter properties and characteristics of the waveform. For the purpose of discussing waveforms, the modal voice from reading lyrics was presumed as the neutral
phonation state. The phonation types were determined on the basis of the relationships between their EGG parameters and modal voice's EGG parameters.

3. Results

3.1. Distributions of EGG Parameters

The distributions of $f_o$ for different conditions of each singer are listed in Figure 2. The $f_o$ was transformed into semitones (re 55 Hz). The $f_o$ of reading lyrics was significantly lower than singing and stage speech ($p < 0.05$), but the difference between the $f_o$s of singing and stage speech was significant only for YM2. For both singers: reading lyrics presented the most concentrated distribution (which meant the smallest interquartile range), while stage speech showed the widest; the distribution of stage speech included that of singing and partly overlapped with reading lyrics; however, there was nearly no overlap between the distributions of singing and reading lyrics.

![Box plot of $f_o$.](image)

The CQ were also alike between singers (see Figure 3). For both singers, the median of the CQ was lowest in singing and highest in reading lyrics ($p < 0.05$). From the distance between the first and third quartile locations, which revealed the distribution width of typical data (the 50% data around the distribution center), in order of most concentrated, the CQ distribution was: reading lyrics, singing, and stage speech. With respect to the same condition, the distribution width of typical data in singing and stage speech was similar between singers; however, it was not in reading lyrics. Regarding the length of the whiskers as the reflection of the phonation diversity, for both singers, singing and stage speech showed greater CQ diversity than reading lyrics. For YM1, the CQ diversity of stage speech was slightly greater than that of singing, while the opposite was true for YM2, since he adopted more voices with large CQs.

Figure 4 shows the distributions of SQ for different singers and conditions. Most of the SQ values were above 100%, which implies that the decontacting phase was longer than the contacting phase. For both YM1 and YM2: the median SQ of reading lyrics was remarkably larger than that of singing and stage speech ($p < 0.05$); the median SQ of stage speech was a little smaller than singing ($p < 0.05$); the median was closer to the first quartile than to the third one, for both singing and stage speech. From the perspective of the distribution width of typical data, for both singers, singing showed the most concentrated distribution and stage speech showed the widest. In regard to the same condition, YM1 presented a similar typical data distribution width of reading lyrics with YM2; for YM1, the typical data distribution width of singing and stage speech were both wider than YM2. As for the phonation diversity, which is reflected by the length of the whiskers, singing showed a little greater SQ diversity than stage speech.
Figure 4 shows the distributions of SQ for different singers and conditions. Most of the typical data (the 50% data around the distribution center), in order of most concentrated, were modal voice, high adduction modal voice, pressed voice, untrained falsetto, high adduction falsetto, breathy voice.

The CQ were also alike between singers (see Figure 3). For both singers, the median CQ was a little greater than that of stage speech. The CQ diversity of reading lyrics was remarkably larger than that of singing and stage speech (the median was closer to the first quartile). Similar between singers; however, it was not in reading lyrics. Regarding the length of the typical data distribution width of modal voice, high adduction modal voice, pressed voice, untrained falsetto, high adduction falsetto, breathy voice.

As for the phonation diversity, which is reflected by the length of the whiskers, singing showed a little greater SQ diversity than stage speech. The SQ distribution was: reading lyrics, singing, and stage speech. With respect to the length of the SQ values were above 100%, which implies that the decontacting phase was longer than to the third one, for both singing and stage speech. From the perspective of the distribution and stage speech showed the widest; the distribution of stage speech included that of singing remarkably larger than that of singing and stage speech. From the EGG parameters, the SQ was transformed into semitones (re 55 Hz). The SQ values were above 100%, which implies that the decontacting phase was longer than the contacting phase. For both YM1 and YM2: the median SQ of reading lyrics was remarkably larger than that of singing and stage speech as the neutral phonation state. The phonation types were determined on the basis of the relationships between their EGG parameters and modal voice’s EGG parameters.

The EGG waveforms were more complicated than those in modal speech. Thus, the phonation types were determined by consulting both the parameters and characteristics of the waveform. The typical waveforms can be seen in Figure 5. The waveforms were all selected from YM’s singing and stage speech, except for Figure 5c, which was a typical waveform of the colorful face role [28]. The spectra of some audio clips which corresponded to the EGG waveforms in Figure 5 are shown in Figure 6.

Figure 4. Box plot of SQ.

3.2. Phonation Types

The EGG waveforms were more complicated than those in modal speech. Thus, the phonation types were determined by consulting both the parameters and characteristics of the waveform. The typical waveforms can be seen in Figure 5. The waveforms were all selected from YM’s singing and stage speech, except for Figure 5c, which was a typical waveform of the colorful face role [28]. The spectra of some audio clips which corresponded to the EGG waveforms in Figure 5 are shown in Figure 6.

Figure 5. Eight typical EGG waveforms: (a) modal voice, (b) high adduction modal voice, (c) pressed voice, (d) untrained falsetto, (e) high adduction falsetto, (f–h) breathy voice.
The waveform in Figure 5a was a modal voice, since it showed a similar $f_0$, CQ, SQ, and shape to the EGG parameters of YM1’s reading. The voice which had a similar EGG waveform as that in Figure 5b sounded much brighter and clearer than the modal voice, but not as tense as the pressed voice in other Kunqu male roles [28] (see Figure 5c). The CQ of this kind of voice was significantly larger than that of modal voice. In addition, as shown in Figure 6a, the high-frequency partials showed more energy than modal voice. Both of these indicated that the degree of posterior glottal adduction increased, which was based on the results of a previous research in which the experiment was documented by simultaneous laryngeal videostroboscopy, electroglottography, and by capturing the acoustic data. [30]. Meanwhile, the CQ was significantly smaller than that of the pressed voice in the same pitch range, which is around 70% for the other male roles in the Kunqu opera [28]. Thus, it was named modal voice with a high degree of posterior glottal adduction, or high adduction modal voice. The waveform in Figure 5d showed a significantly smaller CQ than modal voice and a SQ close to 100%. Thus, it was categorized as untrained falsetto. The waveform in Figure 5e showed similar geometric characteristics with that in Figure 5d in the contacting phase, but a larger CQ. It was named as falsetto with high degree of posterior glottal adduction, or high adduction falsetto. As shown in Figure 6b, the spectrum of high adduction falsetto had more high-frequency energy than that of untrained falsetto.
The waveforms in Figure 5f–h were all classified as breathy voice, although they had different shapes and parameter characteristics. In this case, low adductive tension and weak medial compression made the vocal folds never fully come together. There was a continuous glottal leakage with some audible frication noise compared with the modal voice [4]. The waveform in Figure 5f was a typical breathy voice, since it showed a smaller CQ than modal voice and a SQ larger than 100. The waveform in Figure 5g exhibited larger CQ and SQ than that in Figure 5d. A second peak, which occurred in the decontacting phase, suggested that the anterior and posterior parts of the vocal folds vibrated relatively independently [16,29,31]. Thus, it was another EGG signal pattern of breathy voice. A second peak in EGG may result in a sub-harmonic in the spectrum (see Figure 6c). The waveform in Figure 5h showed a similar CQ with the high adduction modal voice and a similar SQ with the untrained falsetto. However, the shape of the EGG signal could be considered as the combination of a main and a second peak. The spectrum is similar to that of Figure 5g (Figure 6c,d). The higher harmonics were replaced by aspiration noise in Figure 6c,d, which suggested the occurrence of the turbulence noise. Thus, it was also identified as breathy voice. More evidence for identifying these voices as breathy voice were found in the spectra if we ignored the vocal tract contribution to the spectra. The level difference between the fundamental component in the spectrum H1 and the second-formant amplitude A2 of a vowel /o/ was 6.4 dB lower for breathy voice than for modal voice. The first-formant bandwidth of the breathy /o/ was 170 Hz, and that for the modal /o/ was only 76 Hz.

3.3. Clustering Analysis

To determine the vibration modes of the vocal folds, all three EGG parameters needed to be taken into consideration. Table 1 illustrates the cluster centroids of EGG parameters and the percentage of data around the centroid of the two YM singers. A combination of the singer, the condition (S stood for singing, SS stood for stage speech), w and the number of the clustering centroid, such as “YM1_S_1” (the first clustering centroid of YM1’s singing parameters), was used to refer to each clustering centroid.

| Singer | Singing | Stage Speech |
|--------|---------|--------------|
|        | No.  | f_o/Semi | CQ% | SQ% | P% | PT | No. | f_o/Semi | CQ% | SQ% | P% | PT |
| YM1    | 1    | 25      | 43  | 318 | 10 | m  | 1   | 22      | 57  | 414 | 7  | M  |
|        | 2    | 26      | 33  | 216 | 15 | B  | 2   | 23      | 28  | 184 | 9  | B  |
|        | 3    | 27      | 46  | 166 | 7  | m  | 3   | 24      | 41  | 299 | 10 | m  |
|        | 4    | 29      | 50  | 269 | 6  | M  | 4   | 27      | 53  | 297 | 15 | M  |
|        | 5    | 29      | 25  | 115 | 18 | f  | 5   | 29      | 59  | 90  | 5  | B  |
|        | 6    | 32      | 36  | 139 | 11 | B  | 6   | 31      | 47  | 210 | 13 | m  |
|        | 7    | 34      | 31  | 105 | 21 | f  | 7   | 34      | 35  | 104 | 18 | f  |
|        | 8    | 38      | 39  | 126 | 11 | F  | 8   | 40      | 44  | 118 | 22 | F  |
| YM2    | 1    | 25      | 39  | 258 | 11 | m  | 1   | 21      | 37  | 289 | 9  | B  |
|        | 2    | 25      | 35  | 173 | 19 | B  | 2   | 22      | 25  | 146 | 10 | B  |
|        | 3    | 29      | 25  | 115 | 21 | f  | 3   | 24      | 49  | 335 | 7  | M  |
|        | 4    | 31      | 58  | 289 | 3  | M  | 4   | 26      | 36  | 200 | 15 | B  |
|        | 5    | 32      | 58  | 99  | 8  | B  | 5   | 27      | 53  | 132 | 8  | B  |
|        | 6    | 34      | 36  | 148 | 10 | B  | 6   | 28      | 47  | 240 | 16 | m  |
|        | 7    | 34      | 28  | 112 | 19 | f  | 7   | 37      | 28  | 109 | 18 | f  |
|        | 8    | 39      | 38  | 124 | 9  | f  | 8   | 41      | 35  | 111 | 17 | F  |

As shown in Table 1, there were eight effective clustering centroids for each singer in each condition. Up to 84% of the clustering centroids was distributed below the maximum
pitch of the YM roles’ passaggio. On the other hand, in the high pitch range (above B4), the vocal cords vibrated stably since there was only one clustering centroid. The vibration mode of the vocal fold did not show a one-to-one correspondence with pitch. The $f_0$ of YM1_S_4 and 5, YM2_S_1 and 2, and YM2_S_6 and 7 were same. However, the CQ and SQ of them showed great diversity. On the adjacent pitches, the CQ or SQ was also different. The proportions of the clustering centroids varied a lot. The smallest was 3% and the largest was 22%. In most cases, if the $f_0$ of two clustering centroids was close, the proportions of them displayed a relatively large difference.

Between singers, the clustering centroids appeared on different pitches in both singing and stage speech. Some clustering centroids presented the same or similar parameter characteristics, such as YM1_S_5 and YM2_S_3, and YM1_S_8 and YM2_S_8. The other centroids showed more obvious differences, such as YM1_SS_4 and YM2_SS_5, and YM1_S_6 and YM2_S_5. For both singers, multi differences were observed between conditions. The clustering centroids were located in different pitch ranges, especially for YM2. There were five clustering centroids in the pitch range of 29~34 in YM2’s singing, while there was none for his stage speech. Even if the clustering centroids showed the same $f_0$, the CQ and SQ of some clustering centers were different, such as YM1_S_5 and YM1_SS_5.

Eight clustering centroids did not equal eight phonation types. Regarding the influence of the $f_0$ on the clustering result, if two clustering centroids showed a similar CQ and SQ but a different $f_0$, they may be or may not be the same phonation type, which depended on the similarity degree between their EGG waveforms. Thus, combining the analysis for Figure 5, the phonation type of each clustering centroid can be determined; see the last column of each part in Table 1. For YM1 and YM2’s singing, the proportion of phonation type, from high to low, was untrained falsetto, breathy voice, modal voice, high adduction falsetto, and high adduction modal voice. The phonation types used were slightly different between singers in stage speech. YM1 used a higher percentage of high adduction modal voice, modal voice, and high adduction falsetto and less breathy voice and untrained falsetto than in singing. For YM2, compared with singing, a higher percentage of breathy voice, high adduction modal voice, modal voice, high adduction falsetto, and less untrained falsetto were employed in stage speech. A higher degree of posterior glottal adduction made the voice have stronger energy in stage speech than in singing, which was verified by previous research [26].

Taken together, high adduction modal voice, modal voice, breathy voice, and untrained falsetto were used in the low pitch range ($f_0$ from 25 to 34 for singing and from 21 to 37 for stage speech), while high adduction falsetto was used in the high pitch range ($f_0$ above 38 for singing and above 40 for stage speech), as shown in Figure 7. Abundant phonation types in the low pitch range diversified the voice, while high adduction falsetto made the voice timbre unify in the high pitch range. Modal voice, breathy voice, untrained falsetto, and high adduction falsetto were also adopted in different proportions by two young female roles in Kunqu Opera [16,29]. More modal voice and the use of high adduction modal voice made the voice of the YM role manlier. However, YM’s voice was much gentler than those of the colorful face role and the old man role, since they used pressed voice [28].

3.4. SPL Analysis

The medians of SPL were different between the phonation types and $f_0$s, as shown in Figure 8. The lowest SPL was found in breathy voice for each singer and condition, and the highest was observed in high adduction falsetto. In breathy voices, EGG waveforms with a single peak (Figure 5f,h) showed a low SPL, while EGG waveforms with double peaks (Figure 5g) showed a higher SPL, which was close to the SPL of the singer’s modal voice. The SPL of untrained falsetto was lower than that of high adduction falsetto. Except for the SPL of YM2’s singing, the SPL of untrained falsetto was higher than breathy voice in the proximate pitch range. For YM1, the SPL of high adduction modal voice was higher than that of modal voice. However, it was not true for YM2. From the above comparison, the
SPL was related to the phonation types, and extra contact or adduction increased the SPL in most cases. Between conditions, the median SPL of breathy voice was larger in singing. For other phonation types, in most cases, the median SPL was larger in stage speech.

![Figure 7](image_url)

**Figure 7.** Three-dimensional diagrams of the clustering centroids: (a) the parameters of YM1, and (b) the parameters of YM2. m: modal; M: high adduction modal; B: breathy; f: untrained falsetto; F: high adduction falsetto. The blue marks represent singing and the red ones represent the stage speech.

![Figure 8](image_url)

**Figure 8.** The relationship between $f_o$ and SPL medians for each cluster: (a) the parameters of YM1, and (b) the parameters of YM2. m: modal; M: high adduction modal; B: breathy; f: untrained falsetto; F: high adduction falsetto. The blue marks represent singing and the red ones represent the stage speech.

Taking all SPL and $f_o$ data into consideration, for most phonation types, the SPLs of singing showed more concentrated distributions (which corresponded to smaller interquartile range) than those of stage speech. Pearson tests were done between the $f_o$ and SPL values. $R^2$ was 0.41, 0.44, 0.30, and 0.66 for YM1’s and YM2’s singing and stage speech, respectively. The wide distribution range of $f_o$ in each pitch range resulted in the low $R^2$ values. It also led to the observation of multiple phonation types in the same pitch range, such as YM1_S_4 and 5, which showed the same $f_o$ but contrasting SPLs.

4. Discussion

Several phonation types were reported in this research. The determinations on the natures of the phonation types were based on three considerations. Firstly, the reference EGG signal was selected from each singer’s reading, since the parameters of modal voice
were different between singers. YM1 showed a higher CQ than YM2 in reading. The CQ of YM1’s stage speech was lower than that of his own reading, but similar to the CQ of YM2’s reading (see Figure 3). However, YM1’s stage speech sounded different from YM2’s reading. Thus, setting the singer’s own reading as the reference was important. Secondly, the CQ and SQ of certain phonation types varied with the \( f_o \); thus, the situation that the vocal folds were elongated as the \( f_o \) was rising should be considered while judging the phonation types. Previous research showed that the CQ had a positive correlation with the \( f_o \) in modal voice, and the SQ had a negative one [5]. The comparison of parameters need to be confined to the same pitch range. Thirdly, in the case of special phonation types, not only the EGG parameters, but also the shapes of the EGG signals and the spectrum characteristics should be taken into consideration. The EGG features can be explained with the geometric and kinematic parameters. The waveforms in Figure 5 can be explained by the combinations of four features, which are the pulse widening that occurred from adduction, the peak skewing caused by increased convergence in the glottis, the skirt bulging linked to medial surface bulging on the vocal folds, and the shirt ramping produced by vertical phasing [20].

Then, the waveform that showed a similar CQ can be explained to have different vocal fold vibration modes, such as the waveform of Figure 5b,h, Figure 5e,f. In addition, the SPL, instead of the amplitude of EGG signal, could be a supplement to the contact or adduction degree of the vocal folds. It could be used in the analysis of EGG signals with low signal-to-noise ratios [12].

Some double-peak EGG signals occurred in YM’s singing and stage speech, such as those in Figure 5g. Double-peak EGGs were also reported in the singing and stage speech of young woman and young girl roles [16,29]. The second peak in their EGG signals appeared in the contacting phase, was conspicuous, and visibly influenced the values of CQ and SQ. On the contrary, the second peak of YM’s EGG signal was easily ignored. Even so, the second peak corresponded to the vibration of some part of the vocal folds. It was necessary to point it out. However, neither the criterion-level method [23] nor the differential of EGG signal method [21,22] can solve this problem. New methods for double-peak EGG signal need to be introduced. One possible way is using the contact quotient by integration [32]. It could avoid the abrupt change of CQ when the second peak showed a similar level with the criterion, and the neglect of the second peak when its level was lower than the criterion. Another possible way is separating the second peak from the signal on the basis of the phase diversity. This was achieved by simultaneously analyzing multiple cycles of signals instead of analyzing one cycle at a time. Then, the higher CQ only corresponded to the more pressed voice, but not to the result of the second peak of the breathy voice. The phonation types used by YM singers were multitudinous and also observed in other types of singing voices. Above the passaggio, the voice had more high-frequency energy and a larger CQ than classic falsetto. It was more like the falsetto of a trained Western male singer [24,33]. Thus, the traditional term falsetto was loose. In and below the pitch range of passaggio, the phonation types were dramatically different from the term “modal voice” in the traditional singing theory of Kunqu Opera [1]. The singers varied phonation types not only with pitch, but also with conditions and personages. On the one hand, the extensive use of high adduction modal voice, breathy voice, and untrained falsetto made the voice timbre rich and various. On the other hand, the usage environment of each phonation type needs to be studied further. Some phonation types were found in similar pitch range, such as the untrained falsetto and the breathy voice in YM2’s singing, as can be seen in Figure 8. The position of a word in a sentence and the emotion of the sentence might be the reason for the adoption of a phonation type. Similar EGG signals of some phonation types can also be found in the singing voice of bel canto tenor. High adduction modal voice and high adduction falsetto showed a similar CQ and SQ with mezza voce and falsetto of tenor, respectively [33], though high adduction modal voice was used in lower pitch range than mezza voce.
5. Conclusions

The present study explored the phonation types used by YM roles in Kunqu Opera on the basis of the EGG parameters and the characteristics of EGG waveforms. The results showed that it is inaccurate to describe a YM role’s voice in the traditional terms modal voice and falsetto. The acoustic effects of YM role’s singing and stage speech were formed by complicated patterns of phonation types, which were breathy voice, modal voice, high adduction modal voice, untrained falsetto, and high adduction falsetto. The phonation types of YM were more similar to young female roles than to the other male roles in Kunqu Opera.

This study showed the complexity of analyzing the EGG signals of artistic voices. On the one hand, new method need to be introduced to solve the one-to-many problem between the parameters and the waveforms. On the other hand, the contributing factor of the second peak in the EGG waveform needs to be backed up by the other research that uses visible techniques, such as the laryngeal videostroboscopy, which can improve the credibility of the present conclusions and form a paradigm for double-peak EGG signal analyzing.

Author Contributions: Conceptualization, L.D.; methodology, L.D. and J.K.; software, J.K.; investigation, L.D.; resources, L.D. and J.K.; Writing—Original draft preparation, L.D.; Writing—Review and Editing, L.D.; funding acquisition, L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Office for Philosophy and Social Sciences, grant number 16CYY047.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to the voice experts for their gentle participation in this investigation. Thanks to Bradley Phillips for the language modification.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, X. Dictionary of Chinese Kunqu Opera; Nanjing University Press: Nanjing, China, 2002; pp. 565–568.
2. Dong, L.; Kong, J.; Sundberg, J. Long-term-average spectrum characteristics of Kunqu opera singers’ speaking, singing and stage speech. Logoped. Phoniatr. Vocal. 2014, 39, 72–80. [CrossRef]
3. Tian, S. Research on Kunqu Opera Singing Art; Zhejiang University Press: Hangzhou, China, 2013; pp. 6–18.
4. Laver, J. The Phonetic Description of Voice Quality; Cambridge University Press: London, UK, 1980.
5. Kong, J. On Language Phonation; Central Nationalities University Press: Beijing, China, 2001; pp. 159–189.
6. Fant, G. Speech Acoustics and Phonetics; Springer Netherlands: Berlin, Germany, 2005.
7. Sundberg, J. The Science of the Singing Voice; Northern Illinois University: DeKalb, IL, USA, 1987.
8. Welch, G.F.; Sergeant, D.C.; Maccurtian, F. Xeroradiographic-electrolaryngographic analysis of male vocal registers. J. Voice 1989, 3, 244–256. [CrossRef]
9. Rothenberg, M. Some relations between glottal air flow and vocal fold contact area. In Proceedings of the Conference on the Assessment of Vocal Pathology; American Speech and Hearing Association, Ed.; American Speech and Hearing Association: Rockville, MD, USA, 1981; pp. 88–96.
10. Baer, T.; Löfqvist, A.; McGarr, N.S. Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. J. Acoust. Soc. Am. 1983, 74, 1304–1308. [CrossRef] [PubMed]
11. Childers, D.G.; Naik, J.M.; Larar, J.N.; Krishnamurthy, A.K.; Moore, G.P. Electroglottography, speech, and ultra-high speed cinematography. In Vocal Fold Physiology and Biophysics of Voice; Titze, I.R., Scherer, R., Eds.; Denver Center of Performing Arts: Denver, CO, USA, 1983; pp. 202–220.
12. Herbst, C.T.; Schutte, H.K.; Bowling, D.L.; Svec, J.G. Comparing Chalk With Cheese—The EGG Contact Quotient Is Only a Limited Surrogate of the Closed Quotient. J. Voice 2017, 31, 401–409. [CrossRef]
13. Baken, R.J.; Orlikoff, R.F. Clinical Measurement of Speaking and Voice; Delmar Learning: New York, NY, USA, 2000.
14. Yoshinaga, I.; Kong, J. Laryngeal vibratory behavior in traditional Noh singing. Tsinghua Sci. Technol. 2012, 17, 94–103. [CrossRef]
15. Esling, J.H. A laryngographic investigation of phonation type and laryngeal configurations. *Work. Pap. Linguist. Circle Univ. Vic.* 1983, 14–36.
16. Dong, L.; Kong, J. Voice characteristics of young Girl Role in Kunqu Opera. *J. Voice* 2019, 33, e945.e19–e945.e25. [CrossRef] [PubMed]
17. Herbst, C.; Ternström, S. A comparison of different methods to measure the EGG contact quotient. *Logoped. Phoniatr. Vocol.* 2006, 31, 126–138. [CrossRef]
18. Timcke, R.; Von, L.H.; Moore, P. Laryngeal vibrations: Measurements of the glottic wave: Part II–Physiologic variations. *Ama. Arch. Otolaryngol.* 1959, 68, 438–444. [CrossRef]
19. Fant, G. Studies of the human voice source. *Acta Acust.* 1988, 13, 135–146.
20. Titze, I.R. Interpretation of the electroglottographic signal. *J. Voice* 1990, 4, 1–9. [CrossRef]
21. Childers, D.G.; Hicks, D.M.; Moore, G.P.; Alsaka, Y.A. A model for vocal fold vibratory motion, contact area, and the electroglottogram. *J. Acoust. Soc. Am.* 1986, 80, 1309–1320. [CrossRef]
22. Henrich, N.; D’Alessandro, C.; Doval, B.; Castelliengo, M. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J. Acoust. Soc. Am.* 2004, 115, 1321–1332. [CrossRef] [PubMed]
23. Rothenberg, M.; Mahshie, J.J. Monitoring vocal fold abduction through vocal fold contact area. *J. Speech Hear. Res.* 1988, 31, 338–351. [CrossRef] [PubMed]
24. Howard, D.M.; Lindsey, G.A.; Allen, B. Toward the quantification of vocal efficiency. *J. Voice* 1990, 4, 205–212. [CrossRef]
25. Howard, D.M. Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *J. Voice* 1995, 9, 163–172. [CrossRef]
26. Dong, L.; Sundberg, J.; Kong, J. Loudness and pitch of Kunqu Opera. *J. Voice* 2014, 28, 14–19. [CrossRef] [PubMed]
27. Kong, J. *VoiceLab 1.0*; National Copyright Administration of the People’s Republic of China: Beijing, China, 2012.
28. Dong, L.; Sundberg, J.; Kong, J. Formant and voice source properties in two male Kunqu Opera roles: A pilot study. *Folia. Phoniatr. Logo.* 2014, 65, 294–302. [CrossRef]
29. Dong, L. Voice characteristics of young woman role in Kunqu Opera. *Essays Linguist.* 2016, 54, 292–304.
30. Herbst, C.T.; Howard, D.; Schlömicher-Thier, J. Using electroglottographic real-time feedback to control posterior glottal adduction during phonation. *J. Voice* 2010, 24, 72–85. [CrossRef]
31. Ternström, S. *Laryngeal Dynamics and Physiological Models*; Peking University Press: Beijing, China, 2007.
32. Ternström, S. Normalized time-domain parameters for electroglottographic waveforms. *J. Acoust. Soc. Am.* 2019, 146, EL65–EL70. [CrossRef] [PubMed]
33. Mayr, A. Investigating the voce faringea: Physiological and acoustic characteristics of the bel canto tenor’s forgotten singing practice. *J. Voice* 2017, 31, 255.e13–255.e23. [CrossRef] [PubMed]