A measurement study on voice instabilities during modal-falsetto register transition

Yasufumi Uezu1,* and Tokihiko Kaburagi1,†

1Graduate School of Design, Kyushu University,
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan
2Faculty of Design, Kyushu University,
4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan

(Received 9 September 2015, Accepted for publication 22 April 2016)

Abstract: When one of the dominant harmonics (the fundamental frequency and its harmonic components) is close to the first formant frequency, the effect of the source-filter interaction can induce voice register transition, in which the vocal-fold vibration becomes unstable and the pitch jumps abruptly. We investigated the relationship between the dominant harmonics, the first formant frequency, and the pitch jump width in the modal-falsetto transition to examine the effect of source-filter interaction. We measured temporal patterns of the fundamental frequency and the first formant when subjects performed rising glissandi with /a/ and /i/ vowels. For the /a/ vowel, there were weak proximity relationships between the dominant harmonics and first formant during the transition, indicating that source-induced transition occurred. For the /i/ vowel, in contrast, the fundamental frequency was regularly close to the first formant in the transition, indicating that the acoustically induced transition was caused by the source-filter interaction. Additionally, it was found that the difference between these two mechanisms had little influence on the pitch jump width. Finally, we concluded that the source-filter interaction is a contributory factor of the modal-falsetto transition, in agreement with foregoing studies.

Keywords: Voice register transition, Source-filter interaction, External acoustic excitation

PACS number: 43.70.Gr [doi:10.1250/ast.37.267]

1. INTRODUCTION

When vocal pitch is raised gradually from a low pitch or lowered from a high pitch, the voice register can suddenly switch from the modal to falsetto register or from the falsetto to modal register, entailing a discontinuous pitch jump. The cause of the voice register transition has been studied thus far in terms of two different mechanisms, namely source-induced and acoustically induced mechanisms [1–8]. The source-induced mechanism explains the transition in terms of the change in the tension and the effective vibratory mass of the vocal folds. With the acoustically induced mechanism, in contrast, there is acoustic interaction between the voice-source system in the larynx and the acoustic filter of the vocal tract resulting in the induction of the transition.

In relation to the source-induced mechanism, van den Berg et al. [1] observed that the fundamental frequency increases moderately and then jumps abruptly upward to a higher frequency when the tension of the vocal folds in an excised larynx is gradually increased. Švec et al. [2] and Miller et al. [3] also studied the pitch jump in an excised human larynx and in three living subjects. Data obtained for the excised larynx revealed that a small and gradual change in the vocal-fold tension can cause an abrupt change in voice register and pitch; i.e., source-induced vocal-fold instabilities.

Acoustic interaction between the voice-source system in the larynx and the acoustic filter of the vocal tract can play an essential role as a mechanism of the voice register transition and associated change in the voice source characteristics. The source-filter interaction is interpreted as an extension and generalization of Fant’s source-filter theory [9]. The voice-source system and vocal tract filter in vivo are not in fact independent. They influence each other such that the acoustic load of the vocal tract affects the voice-source system in the larynx. This acoustic
interaction can then make the vocal-fold vibration unstable and cause sub-harmonics and pitch jump.

Ishizaka and Flanagan [10] showed the effect of the source-filter interaction using a two-mass model of the vocal folds and speech-generation simulation. Recently, Titze [4] conducted a computer simulation to show that the effect of the source-filter interaction is reinforced when the fundamental frequency is in the vicinity of the vocal-tract reactances, especially the first formant. It was found that the effect of the source-filter interaction is more accentuated when the cross-sectional area of the epilarynx tube is narrow. Moreover, it was suggested that accentuated source-filter interaction can induce a large pitch jump in the register transition. Since Titze’s study, vocal-fold vibration and voice production have been simulated when the fundamental frequency changes in time such that modal and falsetto registers are switched under the influence of the source-filter interaction. Tokuda et al. [5] used a four-mass model of the vocal folds to simulate such register transition and accompanying vocal instabilities. Kaburagi [6] carried out a computer simulation study using a voice production model that integrates a boundary layer analysis of glottal flow and the mechanism of the source-filter interaction.

From the results of these simulation studies, it is summarized that the source-filter interaction can take place when the fundamental frequency approaches the first formant and causes an acoustically induced register transition, resulting in an unstable phonation and pitch jump with subharmonics. In addition, it is suggested that the proximity relationship between the second or third harmonics and the first formant can also cause an acoustically induced transition. Therefore, it is important to confirm the relationship between the dominant harmonics (the fundamental frequency and its harmonic components, especially second and third harmonics) and the first formant during the register transition using an actual measurement. If one of the dominant harmonics is close to the first formant during the transition, it is expected that an acoustically induced pitch jump occurs, and if not, a source-induced register transition occurs as illustrated in Fig. 1.

Measurement studies of the voice register transition have been undertaken to investigate the effect of the source-filter interaction in vivo by observing the pitch contour. Titze et al. [7] measured the dominant harmonics when a number of subjects produced various combinations of pitch patterns and vowels. Zañartu et al. [8] measured the pitch, supra- and sub-glottal formants, and kymograms simultaneously when a subject performed upward and downward pitch glides. These studies confirmed that source-induced and acoustically induced instabilities are both responsible for the voice register transition. In addition, an acoustically induced instability tends to occur for vowels having a relatively low first formant frequency such as the /i/ vowel.

However, these experiments have problems with respect to the accuracy in measuring formant frequencies. Although Titze et al. [7] obtained the first formant frequency for the vocal fry, the vocal-tract configuration and first formant frequency may change during the production of a pitch glide. Zañartu et al. [8] procured the first formant frequency from speech signals of ascending pitch employing a covariance method of the linear predictive analysis technique. However, harmonic components of speech become sparse in pitch glides as the
fundamental frequency increases. Such sparseness of the frequency components can hinder an accurate formant frequency measurement using a signal processing technique.

To determine formant frequencies more accurately even for high-pitched speech signals, we use an acoustic excitation method of the vocal tract [11]. To acoustically stimulate the vocal tract, a noise signal is radiated from a loudspeaker and fed into the mouth through a horn and flexible tube. The noise signal has a wide frequency range that covers the frequencies of the lower formants. The acoustic characteristics of the vocal tract are then measured independently of the voice source of uttered speech using the response signal output from the mouth of the participant. This method has already been applied successfully in measurement studies of the vocal tract in the cases of speech, singing, and playing wind instruments [12–15].

The purpose of this study is to investigate the effect of the source-filter interaction in the register transition. We employ the acoustic excitation method of the vocal tract and precisely analyze the relationship between the dominant harmonics and the first formant frequency. First, the vocal-tract acoustic characteristics and the vocal-fold vibration are measured simultaneously when each participant produces ascending pitch glides from modal to falsetto registers. The first formant and fundamental frequencies are then analyzed to examine whether one of the dominant harmonics is close to the first formant frequency during the transition. Additionally, the pitch jump width is calculated from the fundamental frequencies before and after the transition to compare the widths of the source-induced and acoustically induced pitch jumps.

2. EXPERIMENT

2.1. Subjects and Task

Five Japanese males untrained in professional singing techniques participated in this study. Before the register transition measurement, the overlapping pitch range was examined for each participant. The range was determined by checking if each participant was able to produce both modal and falsetto voices with various pitches ranging from C3 (130.8 Hz) to G5 (784.0 Hz). Table 1 gives the ages and overlapping ranges of the subjects. As a result, it was confirmed that each participant was able to utter in a falsetto register and had an overlapping pitch range. We therefore decided that the subjects were suitable for the measurement of the modal-to-falsetto register transition.

The experiment was performed in a soundproof booth. The task was a rising glissando from modal to falsetto registers, because it is generally easier to produce a smooth change in voice pitch for a rising glissando than for a falling one. In this study, we instructed participants to produce the glissando in a natural way and did not instruct them explicitly to produce the glissando with or without the register transition, because we considered that such instruction might cause some artifact. In addition, each participant was instructed to produce the glissando in more than 20 trials for each of the /a/ and /i/ vowels by following a guide sound fed into the ear through an earphone. As shown in Fig. 2, the guide sound was composed of an ascending chirp signal and cue signals. The chirp signal was designed so that its instantaneous frequency varied logarithmically from 100 to 500 Hz over a period of 1,500 ms. Four cue signals consisted of the pure tone whose frequency was 500 Hz and the duration was 50 ms.

2.2. Measurement Method

Figure 3 is a block diagram of the measurement system used in this study. The vocal-tract acoustic characteristics were measured employing an external acoustic vocal-tract excitation method [11–15]. The vocal tract was stimulated by an external excitation signal; i.e., broadband white noise. The excitation signal was input from the mouth to

| Subject | Age | Overlap range |
|---------|-----|---------------|
| S1      | 27  | B3 (246.9 Hz)–C5 (523.3 Hz) |
| S2      | 26  | C4 (261.6 Hz)–F4 (349.2 Hz) |
| S3      | 24  | A3 (220.0 Hz)–F4 (370.0 Hz) |
| S4      | 23  | C4 (261.6 Hz)–E5 (659.3 Hz) |
| S5      | 23  | D4 (293.7 Hz)–E4 (329.6 Hz) |
the vocal tract while the subject uttered a sustained vowel. The acoustic response to the signal was then output from the vocal tract together with the subject’s own speech. These signals were recorded by a microphone placed in front of the subject’s mouth. Formant frequencies were derived by analyzing the frequency characteristics of the response signal.

The external acoustic excitation device was built from a speaker unit (FF165WK; Fostex) and an exponential horn having a length of 1,095 mm connected to a flexible tube having a length of 300 mm and inner radius of 7 mm. The lower cutoff frequency of this horn was set to 150 Hz, which sufficiently covered the range of the first formant for the /i/ vowel. The excitation signal was amplified by a power amplifier (TA-V55ES; Sony) and fed to the external excitation device to drive the vocal tract. The excitation signal then traveled through the vocal tract and radiated from the mouth as the response. A half-inch condenser microphone (Type 4191; Bruel & Kjaer), a preamplifier (Type 2669; Bruel & Kjaer), and a conditioning amplifier (Nexus 2690; Bruel & Kjaer) were used to record the output acoustic signal.

The excitation signal was generated by a computer as follows. First, an M-sequence signal with a bandwidth from 170 to 6,000 Hz was generated. The sampling frequency was 16,000 Hz. Next, the frequency characteristics of the external excitation device were calibrated. The M-sequence signal was input into the external excitation device and the output signal from the flexible tube was recorded by a microphone placed 5 mm away from the tube. The frequency characteristics of the external excitation device including the frequency characteristics of the speaker, exponential horn, and tube were obtained from this signal. A linear filter that had the inverse frequency characteristics of the output signal was then determined using the linear predictive coding method to cancel out undesired peaks and dips in the frequency characteristics. Finally, the excitation signal was generated by filtering the M-sequence signal with the inverse filter.

In the experiment, the microphone was set 10 cm away from the lips of the subject. Approximately 3 cm of the flexible tube was inserted into the subject’s mouth. While the subject performed the tasks, electroglottograph (EGG) and acoustic signals were recorded simultaneously and stored on the computer. The acoustic signal contained both the vocal-tract response to the excitation signal and the subject’s own speech.

Vocal fold vibration was measured by means of an EGG device (Model EG-2; Glottal Enterprises) with a pair of EGG electrodes fixed either side of the subject’s larynx. EGG and acoustic signals were gathered by a computer through an audio-interface device (Fast Track Ultra; M-AUDIO). This audio-interface device was also used to provide the broadband excitation signal to the external excitation device and the guide sound to the subject’s ear.

2.3. Analysis of the Fundamental Frequency

The fundamental frequency was obtained by applying the DECOM method to differentiated EGG (DEGG) signals as described by Henrich et al. [16]. First, a DEGG signal was generated by filtering the EGG signal with a differentiator filter that attenuated frequency components above the stopband frequency of 700 Hz. The glottal closure instant was detected from the positive peaks of DEGG signals. An interval of the adjacent glottal closure instants corresponded to the fundamental period. Next, the DEGG signal was separated into positive and negative parts and the fundamental period was then estimated by calculating the autocorrelation of the positive part. Finally, the fundamental frequency was calculated as the inverse of the estimated fundamental period. The length of the Hamming window was set adaptively to the quadruple of the fundamental period estimated from the previous analysis frame. The shift width of the analysis frame was set to twice the fundamental period. If the fundamental period could not be estimated in the previous frame, the window length and shift width were set to 40 and 5 ms, respectively.
2.4. Analysis of the Vocal-tract Acoustic Characteristics

Vocal-tract acoustic characteristics were obtained from the measured acoustic signal. However, the signal also contained the subject’s own speech, which was an undesired component to be eliminated. Cepstrum analysis and a liftering process were applied to the acoustic signal to remove the undesired component. First, a logarithm of the power spectrum was calculated from a windowed segment of the acoustic signal and then cepstral parameters were calculated. Next, the vocal-tract acoustic characteristics were calculated from cepstral components lower than the threshold value. Here, the length of the Hamming window was 30 ms, the shift width was 5 ms, and the liftering threshold was set to 2.5 ms. Finally, the temporal pattern of the first formant was estimated from the vocal-tract acoustic characteristics of successive frames employing a peak-picking method.

3. RESULTS

Figure 4 presents an example of the analysis results for the /a/ vowel. The figure shows a 300-ms portion of the entire data including the modal-falsetto transition. The top plot shows the time variation of the dominant harmonics and the first formant frequency. The middle plot shows the EGG signal waveform and the temporal change of the open quotient (OQ) value calculated using the DECOM method [16]. The bottom plot shows the EGG signal spectrogram focused on the transition. Focused on the time variation of the dominant harmonics and the first formant, the pitch jump in the transition occurred at about 1,100 ms, where the dominant harmonics and the first formant were not close. At this time, the amplitude of the EGG signal waveform declined greatly and the open quotient changed suddenly, which indicates the register transition arose. In the EGG signal spectrogram, the abrupt jump in the dominant harmonics was found. In addition, no clear sub-harmonics were found during the frequency jump. These features were observed in other subjects in the register transition for the /a/ vowel.

Figure 5 shows the results for the /i/ vowel including the transition. The time variations of the dominant harmonics and the first formant show that the pitch jump arose at about 970 ms. This pitch jump appeared to be more discontinuous than that for the /a/ vowel. When the pitch jump arose, the fundamental frequency was obviously close to the first formant. Then, the amplitude of the EGG signal decreased more sharply and the OQ value changed drastically, resulting in a break of the vocal-fold oscillation. After the break, stable vibration of the vocal folds was recovered. The stability of the vibration after the transition was firmer for the /i/ vowel than the /a/ vowel. The spectrogram shows the occurrence of frequency bifurcations and the subharmonics (indicated by the white arrows in the figure) associated with the irregular oscillation of the vocal folds during the transition. These features were found in other subjects in the transition for the /i/ vowel.
Figure 6 shows the relationships between the dominant harmonics and the first formant in the voice register transition with the /a/ and /i/ vowels for subjects S1 through S5. Each data sample plotted in Fig. 6 was selected from the entire data for which the response signal of the vocal tract driven by the external excitation was obtained sufficiently. The horizontal and vertical axes are frequency values of the dominant harmonics and the first formant, respectively. These frequency values were determined from the analysis data of the vocal-fold vibration and vocal-tract acoustic characteristics when the pitch jumped at the beginning of the modal-falsetto transition. The solid line indicates that the frequency of the dominant harmonics coincide the first formant frequency. Deep black markers near the line show that one of the dominant harmonics was very close to the first formant frequency. Here, these two frequencies were considered to be proximate when the difference between the two frequencies was less than 40 Hz.

For the /i/ vowel, it was found that the fundamental frequency was consistently near the first formant frequency for all subjects. In addition, most of the deep black markers are located on the right side of the solid line. This indicates that the fundamental frequency was slightly higher than the first formant frequency when the pitch jumped, which is in agreement with previous studies conducted by Titze et al. [7] and Lucelo et al. [17].

For the /a/ vowel, in contrast, it was found that there were no cases for any subjects that the fundamental frequency was close to the first formant frequency. However, there were a number of samples for which the second or third harmonics was close to the first formant. Figures 7 and 8 show such examples that the second or third harmonic approached the first formant in the transition for the /a/ vowel. It should be noted that data derived from the EGG signal showed specific features common to that for the /i/ vowel shown in Fig. 5. When the pitch jump occurred, the amplitude of the EGG waveform decreased sharply and the OQ value changed suddenly, and at the same time, the subharmonic components were apparent (indicated by the white arrows in each figure). Such features were especially observed in Fig. 8.

4. DISCUSSION

From the analysis results and the assumption of the relationship between the dominant harmonics and the first formant illustrated in Fig. 1, it is considered that the following factor causes the voice register transition: (1) In the most cases for the /a/ vowel, there were no proximity relationship between the dominant harmonics and the first formant in the transition. Therefore it is considered the source-induced mechanism (SI) is the main factor of the transition. (2) In almost all cases for the /i/ vowel and some cases for the /a/ vowel, it was found that the dominant harmonics and the first formant had the proximity relationship in the transition. Therefore, it is
considered that the acoustically induced mechanism (AI) is the main factor of the transition. The experimental results obtained for the /a/ vowel indicate that the source-filter interaction can take place even when one of the dominant harmonics approaches the first formant and the interaction has the potential for causing the voice register transition, in agreement with Titze's suggestion [4].

The analysis results shown in Fig. 6 were summarized in Table 2. It gives the mean and standard deviation of the first formant frequency at the beginning of the transition, those of the fundamental frequency before and after the transition, and the mean of the pitch jump width for each combination of the subject and vowel. Here, the pitch jump width was worked out in cent as $1200 \log_2 \left( \frac{f_{\text{post}}}{f_{\text{pre}}} \right)$.

The first formant frequency was approximately from 620 to 740 Hz for the /a/ vowel and from 250 to 270 Hz for the /i/ vowel. The fundamental frequency before the register transition was from 210 to 340 Hz for all vowels. It is noted that these frequencies depended on the subject.

The pitch before the transition tended to be higher than the first formant frequency for the /i/ vowel. Interestingly, the pitch jump width for the /a/ vowel with source-induced transition seemed to be lower than that for the /a/ and /i/ vowels with acoustically induced transition for each subject.

Next, we compare the pitch jump width during the voice register transition between the source-induced and the acoustically induced transitions in more detail. Titze [4] suggested that the source-filter interaction causes a large pitch jump when the epilarynx tube is narrow. However, the difference between the source-induced and acoustically induced transitions has not been clarified in terms of the pitch jump width. Here, we investigate whether the jump width caused by the acoustic mechanism for the /i/ vowel was different from that caused by the source mechanism for the /a/ vowel. An independent-samples t-test was conducted to compare the two pitch jump widths. As a result, no significant difference was found in the scores for the /i/ vowel.
vowel (mean of 412.56 Hz and S.D. of 117.16 Hz) and the /a/ vowel (mean of 380.83 Hz and S.D. of 97.40 Hz) where $r(100) = -1.41$ and the $p$-value was 0.163.

We can conclude that the voice register transition was caused not only by the source-induced instability mechanism but also by the acoustically induced instability...
mechanism. The acoustically induced transition was related to the source-filter interaction, which can be intensified when the fundamental frequency approaches the first formant in agreement with the results of foregoing studies based on computer simulations [4–6] and measurements [7,8]. In addition, we found that the pitch jump width in the voice register transition was not affected by the instability factors for the participating subjects.

5. CONCLUSION

In this study, we investigated the relationship between the dominant harmonics (the fundamental frequency and its harmonic components) and the first formant frequency in the voice register transition by making measurements of the vocal fold and vocal tract. The vocal-fold vibration pattern was measured by an EGG device and the fundamental frequency and the open quotient value were then obtained from the EGG signal employing the DECOM method. The vocal-tract resonance was measured employing an external acoustic excitation method that can determine the formant frequencies more accurately than ordinary signal processing methods when the fundamental frequency is high. For the formant extraction, cepstral analysis was also used to eliminate the subject’s own speech component included in the response signals of the excitation method. The relationship between the fundamental and first formant frequencies was then analyzed to determine the cause of the voice register transition and abrupt pitch jump.

Through the measurement and analysis, we confirmed two different types of voice register transition. For the /a/ vowel, there were weak proximity relationships between the fundamental and first formant frequencies in the transition. We thus concluded that the voice register transition and the pitch jump for the /a/ vowel were induced mainly by the instability of vocal-fold vibrations resulted from the laryngeal adjustment; i.e., by the source-induced mechanism. For the /i/ vowel, the fundamental frequency was regularly close to the first formant in the transition. This indicates that the voice register transition and the pitch jump for the /i/ vowel was caused by the effect of the source-filter interaction; i.e., by the acoustically induced mechanism. Finally, we conclude from the experimental results that the source-filter interaction is a contributory factor of the voice register transition, in agreement with the results of previous studies [4–8]. In addition, we found that these two transition mechanisms had little influence on the pitch jump width. Moreover, when the second or third harmonic was close to the first formant for the /a/ vowel, the pitch jump appeared in a similar manner for the /i/ vowel.

ACKNOWLEDGEMENT

This research was partly supported by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (Grant No. 23300071).

REFERENCES

[1] Jw. van den Berg, W. Vennard, D. Burger and C. C. Shervanian, Voice Production: The Vibrating Larynx (Instructional film, University of Groningen, The Netherlands, 1960).
[2] J. G. Švec, H. K. Schutte and D. G. Miller, “On pitch jumps
between modal and falsetto registers in voice: Data from living and excised human larynges,” *J. Acoust. Soc. Am.*, **106**, 1523–1531 (1999).

[3] D. G. Miller, J. G. Švec and H. K. Schutte, “Measurement of characteristic leap interval between modal and falsetto registers,” *J. Voice*, **16**, 8–19 (2002).

[4] I. R. Titze, “Nonlinear source-filter coupling in phonation: Theory,” *J. Acoust. Soc. Am.*, **123**, 2733–2749 (2008).

[5] I. T. Tokuda, M. Zemke, M. Kob and H. Herzel, “Biomechanical modeling of register transitions and the role of vocal tract resonators,” *J. Acoust. Soc. Am.*, **127**, 1528–1536 (2010).

[6] T. Kaburagi, “Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling,” *J. Acoust. Soc. Am.*, **129**, 1554–1567 (2011).

[7] I. R. Titze, T. Riede and P. Popolo, “Nonlinear source-filter coupling in phonation: Vocal exercises,” *J. Acoust. Soc. Am.*, **123**, 1902–1915 (2008).

[8] M. Zañartu, D. D. Mehta, J. C. Ho, G. R. Wodicka and R. E. Hillman, “Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study,” *J. Acoust. Soc. Am.*, **129**, 326–339 (2011).

[9] G. Fant, *Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).

[10] K. Ishizaka and J. L. Flanagan, “Synthesis of voiced sounds from a two-mass model of the vocal cords,” *Bell Syst. Tech. J.*, **51**, 1233–1268 (1972).

[11] J. Epps, J. R. Smith and J. Wolfe, “A novel instrument to measure acoustic resonances of the vocal tract during phonation,” *Meas. Sci. Technol.*, **8**, 1112–1121 (1997).

[12] E. Joliveau, J. R. Smith and J. Wolfe, “Vocal tract resonances in singing: The soprano voice,” *J. Acoust. Soc. Am.*, **116**, 2434–2439 (2004).

[13] Y. Swerdlin, J. R. Smith and J. Wolfe, “The effect of whisper and creak vocal mechanisms on vocal tract resonances,” *J. Acoust. Soc. Am.*, **127**, 2590–2598 (2010).

[14] J. M. Chen, J. R. Smith and J. Wolfe, “Do trumpet players tune resonances of the vocal tract?” *J. Acoust. Soc. Am.*, **131**, 722–727 (2012).

[15] N. Henrich, J. R. Smith and J. Wolfe, “Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in modal and falsetto registers,” *J. Acoust. Soc. Am.*, **135**, 491–501 (2014).

[16] N. Henrich, C. d’Alessandro, B. Doval and M. Castellengo, “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” *J. Acoust. Soc. Am.*, **115**, 1321–1332 (2004).

[17] J. C. Lucero, K. G. Lourenço, N. Hermant, A. Van Hirtum and X. Pelorson, “Effect of source-tract acoustical coupling on the oscillation onset of the vocal folds,” *J. Acoust. Soc. Am.*, **132**, 403–411 (2012).