

# ANALYSIS OF VOICE REGISTER TRANSITION FOCUSED ON THE RELATIONSHIP BETWEEN PITCH AND FORMANT FREQUENCY

Yasufumi Uezu and Tokihiko Kaburagi

Kyushu University, Fukuoka, Japan  
3DS14006W@s.kyushu-u.ac.jp, kabu@design.kyushu-u.ac.jp

## ABSTRACT

When the voice register transition (VRT) occurs, vocal-folds motion becomes unstable and the voice pitch jumps abruptly. In this article, we examine the relationship between the fundamental frequency  $f_0$  and the first-formant frequency  $F_1$  in VRT to reveal the influence of the source-filter interaction (SFI) on VRT. Five Japanese male speakers produced rising glissandos with vowels /a/ and /i/. The vibratory state of the vocal folds and the vocal tract resonances were measured simultaneously with an electroglottograph device and an external acoustic excitation method. We analyzed the temporal change in  $f_0$  from electroglottograph signals and in  $F_1$  using acoustic response signals. The relationship between  $f_0$  and  $F_1$  were then analyzed to determine the cause of VRT and abrupt  $f_0$  jump. As a result,  $f_0$  was very close to  $F_1$  when VRT arose in /i/, indicating the influence of SFI as a cause of VRT.

**Keywords:** voice register transition, source-filter interaction

## 1. INTRODUCTION

Voice register transition (VRT) is that the voice register suddenly switches from chest to falsetto because of discontinuous voice pitch jumping when the voice pitch is raised gradually from a lower pitch. Besides, when voice register changes from chest to falsetto or falsetto to chest, the voice pitch jumps discontinuously irrespective of how smoothly the vocal fold tension changes.

Two mechanisms may cause voice register transition; one is changing in the tension and the effective vibratory mass of the vocal folds, another is the acoustic interaction between the voice-source system in the larynx and the acoustic filter of the vocal tract. The source-filter interaction (SFI) is interpreted as an extension and generalization of Fant's source-filter theory [2]. The voice-source system and vocal tract filter *in vivo* are not independent; they influence each other and the voice-source system in the larynx is influenced by the

acoustic load of the vocal tract. This acoustic interaction can then make the vocal fold motion unstable, that is, acoustically induced vocal fold instabilities. Ishizaka and Flanagan [4] showed the effects of SFI by using a two-mass model of the vocal folds and speech-generation simulation. Titze [6] studied the SFI during phonation by simulating belting and high-pitched operatic male singing using a speech-production model. In other studies, vocal fold motion and voice production were simulated where the fundamental frequency was changed in time such that modal and falsetto registers were connected under the influence of the SFI. Tokuda, *et al.* [7] used a four-mass model of the vocal folds in the simulation. Kaburagi [5] performed a computer simulation study using a voice-production model that integrated a boundary layer analysis of glottal flow and the mechanism of SFI. Results from these studies suggest that the SFI can cause voice register transition and unstable phonation when the fundamental frequency approaches the first-formant frequency.

Thus, it is suggested that source-induced and acoustically induced instabilities of *in vivo* vocal fold cause voice register transition. Zañartu, *et al.* [8] showed acoustically induced instabilities for vowel /i/ and source-induced instabilities for vowel /ae/ by using one subject performing upward and downward pitch glides. Moreover, it was showed that acoustically induced instabilities appeared abruptly and caused greater frequency jump than source-induced. This suggestion, however, has not been confirmed sufficiently because such experiments did not conducted when different subjects produced glissandos with variety of vowels. Furthermore, such measurement is difficult because high fundamental frequency during voice register transition hinders the accurate measurement of formant frequencies using speech-signal processing such as linear predictive coding analysis because the harmonic components of speech are sparse.

In this study, we investigate the relationship between the fundamental and first-formant frequencies to study the influence of SFI on the voice register transition. We measure simultaneously vocal fold motions and vocal tract resonances while subjects

perform glissandos with vowel /a/ and /i/. In addition, we statistically analyze fundamental and first-formant frequencies in voice register transition and pitch jump width.

## 2. EXPERIMENT

### 2.1. Subject and task

Five Japanese male speakers who untrained singing techniques participated in this study. Table 1 shows each subject's overlap range, the pitch range where he can phonate both chest and falsetto registers.

Measurement experiments were performed in a soundproof booth. Each subject was instructed to produce a rising glissando from chest to falsetto register following the chirp signal fed into the subject's ear as a guide sound. This chirp signal was designed so that its instantaneous frequency rose from 100 Hz to 500 Hz in two seconds. Each subject repeated such glissando trials more than twenty times with Japanese vowels /a/ and /i/, and then the vibratory state of vocal folds and the vocal tract acoustic characteristics were measured simultaneously to get the fundamental frequency and the first formant frequency.

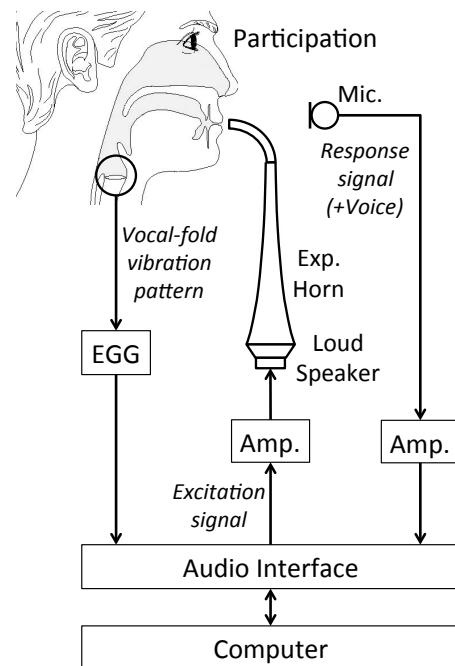
### 2.2. Measurement method

Fig. 1 shows the block diagram of the measurement system used in this study. The vocal tract acoustic characteristics was measured by using the external acoustic vocal-tract excitation (EAVE) method as described by Epps, *et al.* [1]. The vocal tract has specific acoustic characteristics that comprise formants. In the EAVE method, the vocal tract is excited by an external excitation signal such as broadband white noise. The excitation signal is input from the mouth to the vocal tract while the subject is uttering sounds. Then, the acoustic response to the excitation signal is output from the vocal tract together with the subject's own speech and these signals are recorded by a microphone placed in front of the subject's mouth. Formant frequencies are derived by analyzing the

**Table 1:** Subject number, one's age and overlap range (the pitch range where he can phonate both chest and falsetto registers)

Subject	Age	Overlap range
S1	27	B3 — C5
S2	26	C4 — F4
S3	24	A3 — F#4
S4	23	C4 — E5
S5	23	D4 — E4

**Figure 1:** Block diagram of the measurement system used in this study.

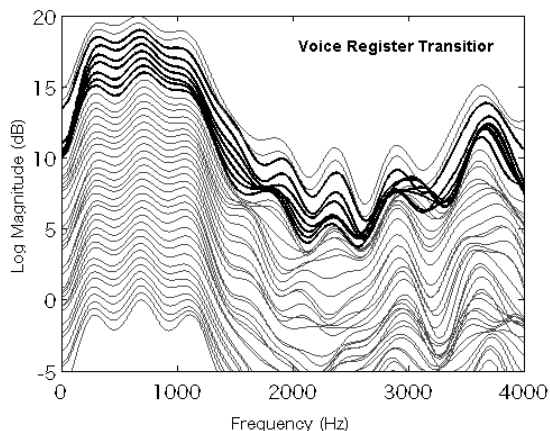


frequency characteristics of the response signal.

The EAVE device in this study was built from a speaker unit (FF165WK; Fostex) and an exponential horn of 1095 mm length connected to a flexible tube of 300 mm length and 7 mm inner radius. An excitation signal was amplified by a power amplifier (TA-V55ES; Sony) and fed to the EAVE 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; Brüel & Kjær), a preamplifier (Type 2669; Brüel & Kjær), and a conditioning amplifier (Nexus 2690; Brüel & Kjær) were used to record the output acoustic signals.

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

**Figure 2:** The temporal variation of the vocal-tract acoustic characteristics from 200 ms before VRT to 50 ms after VRT when subject S3 performed a rising glissando with the vowel /a/.



the above frequency characteristics. Finally, the external excitation signal was generated by filtering the M-sequence signal with the inverse linear filter.

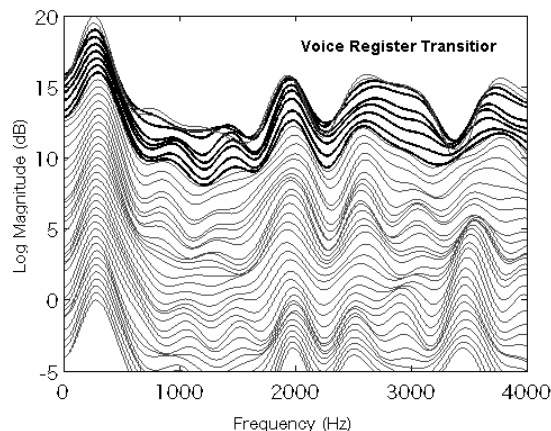
In the experiment, the microphone was set 10 cm away from the outlet flexible tube. Approximately 3 cm of the flexible tube was inserted in subject's mouth. While the subject performed the tasks, EGG and acoustic signals were recorded simultaneously and stored in the computer. The acoustic signal contained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

A vocal fold motion was measured as an electric EGG signal by means of an EGG device (Model EG-2; Glottal Enterprises) with a couple of EGG electrodes fixed on both sides 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 EAVE device.

### 2.3. Analysis of the fundamental frequency

The fundamental frequency  $f_0$  was obtained by applying the DECOM method to DEGG signals as described by Henrich, *et al.* [3]. First, DEGG signal was generated by filtering EGG signal with differentiator filter which attenuated frequency components more than the stopband frequency of 700 Hz. The glottal closure instant (GCI) was detected from the positive peaks of DEGG signals. An interval of adjacent GCI corresponds to a fundamental period  $T_0$ . Next, DEGG signal was separated into positive and negative parts and then  $T_0$  was estimated by calculating the autocorrelation of positive part. Finally,

**Figure 3:** The temporal variation of the vocal-tract acoustic characteristics from 200 ms before VRT to 50 ms after VRT when subject S4 performed a rising glissando with the vowel /i/.



$f_0$  was calculated from the inverse number of estimated  $T_0$ . The length of the hamming window was set adaptively to the quadruple of the  $T_0$  estimated from the previous analysis frame. The shift width of the analysis frame was set to the double of  $T_0$ . If  $T_0$  couldn't be estimated in the previous frame, window length and shift width were set to 40 ms and 5 ms each other.

### 2.4. Analysis of the vocal tract acoustic characteristics and the first formant frequency

Vocal tract acoustic characteristics was analyzed from the measured acoustic signal, however, it contained the subject's own speech that was the undesired signal component to be eliminated. Here, cepstrum analysis and liftering process were applied to the acoustic signal so as to remove such signal component. First, 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 was calculated from lower quefrequency components less than a threshold value. Here, the length of the hamming window was 30 ms, the shift width was 5 ms and liftering threshold value was 2.5 ms. Finally, temporal pattern of the first-formant frequency was estimated from the vocal tract acoustic characteristics for each frame by using a peak-picking method.

## 3. RESULTS AND DISCUSSION

Fig. 2 and Fig. 3 show the temporal variation of vocal-tract acoustic characteristics from 200 ms before VRT to 50 ms after VRT. Fig. 2 shows the result

**Table 2:** The analysis results of mean and standard deviation of  $F_1$  just before VRT, mean and standard deviation of pitch  $f_0^{pre}$  just before VRT and  $f_0^{post}$  just after VRT, and  $f_0$  frequency jump width in all combinations of subjects and vowels.

Subject	Vowel	Number of Data	$F_1$ (Hz)		$f_0^{pre}$ (Hz)		$f_0^{post}$ (Hz)		$f_0$ Jump Width (Cent)
			Mean	S.D.	Mean	S.D.	Mean	S.D.	
1	a	9	673.5	27.4	330.5	18.6	415.0	10.3	394.2
2	a	7	721.3	15.3	281.1	15.7	333.8	16.7	297.5
3	a	15	698.2	27.6	269.0	13.0	331.8	13.6	363.2
4	a	15	688.8	30.7	288.7	33.0	375.1	27.3	453.2
5	a	4	645.1	27.2	229.7	19.1	279.6	14.1	340.3
1	i	10	270.4	16.4	301.4	18.8	378.0	21.0	392.0
2	i	9	261.4	17.9	281.3	19.3	340.8	19.7	332.2
3	i	15	268.1	15.0	278.3	17.9	366.0	19.9	474.2
4	i	15	256.2	19.4	239.4	10.9	319.0	11.5	497.0
5	i	15	267.5	7.0	272.7	8.7	329.5	19.3	327.6

when subject S3 performed a rising glissando with the vowel /a/, and Fig. 3 shows the result when subject S4 performed a rising glissando with the vowel /i/. In Fig. 2, it was found that peaks near 700 Hz shifted continuously along time, which means that these were  $F_1$  of the vowel /a/. On the other hand, in Fig. 3, it was found that peaks near 300 Hz shifted continuously along time, that is, these were  $F_1$  of the vowel /i/.

Table 2 shows the results of mean and standard deviation of  $F_1$  just before VRT, mean and standard deviation of pitch  $f_0^{pre}$  just before VRT and  $f_0^{post}$  just after VRT, and  $f_0$  frequency jump width in all combinations of subjects and vowels. Here,  $f_0$  jump width were worked out in Cent as:

$$(1) \quad 1200 \log_2 \left( \frac{f_0^{post}}{f_0^{pre}} \right).$$

It was found that  $F_1$  for vowel /a/ were from 640 Hz to 720 Hz and for vowel /i/ were from 250 Hz to 270 Hz in each subject. It was also found that frequency range where  $f_0$  jump occurred was from 200 Hz to 400 Hz, and frequency margin of  $f_0$  jump were from 50 Hz to 90 Hz.

It was evident that there were two different types of the relationship between  $f_0$  and  $F_1$  in voice register transition. In one case,  $f_0$  was obviously lower than  $F_1$ . From Table 2, data for the vowel /a/ were considered to correspond with this case. In another case,  $f_0$  was very adjacent to  $F_1$ . Such tendencies were found for the vowel /i/. In addition, it was found that  $f_0$  jump width for /i/ were from 40 to 100 Cent larger than those for /a/ except the cases of subjects S1 and S5.

From previous studies [5, 6, 7, 8], it is known that the influence of SFI is particularly strong and

causes vocal fold instabilities when  $f_0$  is very close to  $F_1$ . Such instabilities bring about greater frequency jump than the instabilities caused by variation of vocal fold tension. From the results, it is certainly that  $f_0$  jump width tended to be larger for vowel /i/ than for vowel /a/ in most subjects. Thus, it is considered that the effect of SFI depends on the type of vowels.

These experimental results suggest that voice register transition is caused by not only source-induced instability but also acoustically induced instability by SFI, which intensify frequency jump. Hence, it was revealed that the SFI causes voice register transition in real speech, which supports previous studies [5, 6, 7, 8].

#### 4. CONCLUSIONS

In this study, we investigated the relationship between the fundamental frequency  $f_0$  and the first-formant frequency  $F_1$  in voice register transition through vocal fold and acoustic measurements.  $f_0$  was analyzed using the DECOM method from EGG signal, and  $F_1$  was analyzed using the EAVE method. The cepstral analysis was also used to eliminate the subject's own speech. The relationship between  $f_0$  and  $F_1$  values were then analyzed to determine the cause of voice register transition and abrupt  $f_0$  jump. As a result, Two patterns of the relationship between  $f_0$  and  $F_1$  in voice register transition were found. Furthermore,  $f_0$  was very close to  $F_1$  and  $f_0$  jump width tended to be larger when voice register transition took place for vowel /i/, indicating the influence of the SFI as a cause of voice register transition.

## 5. REFERENCES

- [1] Epps, J., Smith, J., Wolfe, J. 1997. A novel instrument to measure acoustic resonances of the vocal tract during phonation. *Measurement Science and Technology* 8(10), 1112.
- [2] Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- [3] Henrich, N., d'Alessandro, C., Doval, B., Castellengo, M. 2004. On the use of the derivative of electroglottographic signals for characterization of non-pathological phonation. *J. Acoust. Soc. Am.* 115(3), 1321–1332.
- [4] Ishizaka, K., Flanagan, J. L. 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell system technical journal* 51(6), 1233–1268.
- [5] Kaburagi, T. 2011. Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling. *J. Acoust. Soc. Am.* 129(3), 1554–1567.
- [6] Titze, I. R., Worley, A. S. 2009. Modeling source-filter interaction in belting and high-pitched operatic male singing. *J. Acoust. Soc. Am.* 126(3), 1530–1540.
- [7] Tokuda, I. T., Zemke, M., Kob, M., Herzel, H. 2010. Biomechanical modeling of register transitions and the role of vocal tract resonators. *J. Acoust. Soc. Am.* 127(3), 1528–1536.
- [8] Zañartu, M., Mehta, D. D., Ho, J. C., Wodicka, G. R., Hillman, R. E. 2011. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study a). *J. Acoust. Soc. Am.* 129(1), 326–339.