Variation of phonation types due to paralinguistic information: An analysis of high-speed video images

Masako FUJIMOTO† and Kikuo MAEKAWA†‡

†National Institute for Japanese Language, 3-9-14, Nishigaoka, Kita-ku, Tokyo, 115-8620 Japan, ‡CREST/JST

E-mail：mfuji@kokken.go.jp, kikuo@kokken.go.jp

ABSTRACT

As part of an effort to examine the transmission mechanism of paralinguistic information as well as phonation variation due to paralinguistic information, laryngeal observation was conducted using high-speed digital video image. The results showed that ‘breathy’ and ‘creaky’ phonation was observed for ‘disappointment’ and ‘suspicion’ utterances, respectively, while ‘modal’ phonation was observed for ‘neutral’ utterances. The phonation types characterizing each paralinguistic meaning were maintained throughout the vowel and consonant portions of the utterance. This fact suggests that the domain of phonation control due to paralinguistic information is the whole utterance. This parallels our previous findings about articulatory control, suggesting that paralinguistic information involves control of voice-quality rather than of individual segments. Acoustic comparison of more extended data is also discussed.

1. INTRODUCTION

Although paralinguistic information (like ‘suspicion’ and ‘disappointment’) occurs frequently in everyday speech, the mechanism for producing paralinguistic information is not well understood. Paralinguistic information is defined as the information controlled by the speaker but not distinguished by written text [1]. In recent years, the authors have tried to clarify what is involved in producing paralinguistic information. The results showed that intonation and rhythm are crucial [2,3,4]. Moreover, the influence on articulation affected the whole utterance, not just the vowels and consonants. This suggests that changes in phonation type, including voice quality changes, may be involved [5,6].

To examine how paralinguistic information affects phonation type, which is strongly related to the voice source, the present paper examines glottal parameters using high-speed digital video recordings [7]. Also, acoustic analysis of speech recorded simultaneously with high-speed video imaging was done.

2.METHODS

2.1 Data

An adult male speaker of standard (Tokyo) Japanese served as a subject. He produced a one-word utterance /e’ki/ ("station") with three types of paralinguistic renditions, ‘neutral’, ‘suspicion’, and ‘disappointment’, and the isolated vowel /e/ uttered with different phonation types including ‘modal’, ‘breathy’, and ‘creaky’. The glottal image via fiberscope or tele-endoscope was recorded by a high-speed digital video imaging system at the rate of 4500 frames per second [7,8]. Actual recording time for each utterance was 0.68second (3072frames). Table 1 shows the utterance list.

Table 1 Utterance list

/e/ (sustained vowel)	: modal, breathy, very breathy, creaky
/e’ki/ （“station”）	: ‘neutral’, ‘disappointment’, ‘suspicion’

2.2 Method of analysis

Measurements of the glottal opening area and distance between the two vocal folds were done. Figure 1 shows an example of how the measurements were made. The procedure for the measurement of the glottal area is shown in the left two panels: (1) rotate the images so that the view of the glottis is in the upright position, (2) set the threshold value of the brightness pixel intensity to match the glottal shape, and (3) measure the subliminal（darker）area. The rectangle in the figure shows the area in which measurements were done. Method of the distance measurement is shown in the right panel of the figure. Three lines (L1～L3) which are orthogonal to the length of the vocal folds were drawn and the distance between the two vocal folds at each line were measured. Three lines are located according to the following criteria: L1-- immediately posterior to the vocal processes (on the side of the arytenoid cartilages); L2-- immediately anterior to the vocal processes (on the side of the thyroid cartilage); L3-- the point that is more anterior than L2 and has large glottal movement. In those cases in which the cartilaginous glottis was hidden by the arytenoid cartilage, lines were drawn on the arytenoid cartilage corresponding to the above-mentioned part. Measurements of multiple lines were employed since a previous study suggested that the characteristics of the glottal movement differ depending on the glottal position－the ligamental glottis and the cartilaginous glottis [9].

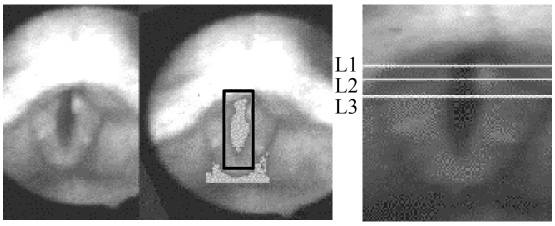


Figure 1. Example of the area measurement (left two panels) and distance measurement (right).

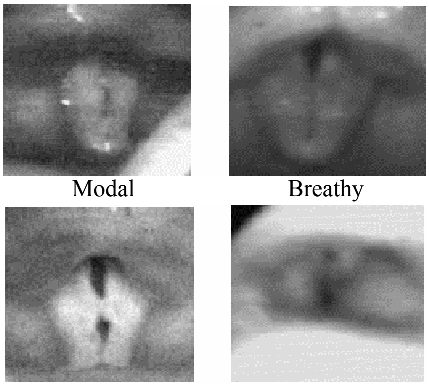
3. RESULTS AND DISCUSSION

3.1 Sustained vowel /e/

Figure 2 shows the characteristics of the vocal fold vibration for different phonation types. Each panel shows the closed phase of the glottis. Figure 3 shows the result of the distance measurement. In modal phonation, periodical glottal vibrations were observed in the ligamental glottis. The cartilaginous glottis was covered by the vocal processes, but it was assumed to be closed all the time. These observations are in agreement with previous studies [10]. As a result, the time course of the glottal distance in figure 3 shows regular opening and closing movements for L2 and L3 and not for L1 at the cartilaginous glottis. In breathy phonation, glottal vibration was observed in both the ligamental and cartilaginous glottis. However, the cartilaginous glottis was not fully closed at the closed

ISBN 1-876346-48-5 © 2003 UAB

phase. The vocal processes remain separated throughout the utterance. As a result, the time course of the glottal distance at L1 does not show full closure. Although opening and closing movements were observed at L2 and L3, the closed phases at L2 and L3 was relatively short. This results in larger OQ compared to modal phonation. Such a tendency was more prominent in L2 which is closer to the vocal processes. In very breathy phonation, a gap in the closed phase was at the ligamental glottis as well as the cartilaginous glottis. The vocal processes remained separated throughout the utterance, as in the breathy phonation. As a result, during the time course of the glottal distance, there were cases without full closure at L2 as well as at L1. As for creaky phonation, both adduction of the false vocal folds and constriction of the pharyngeal cavity were observed throughout the utterance. Consequently, the cartilaginous glottis was unable to be observed, but we conjecture it is closed all the time. The time course of the glottal distance was irregular in L2 and L3. L1 was supposed to be closed all the time.


Modal Breathy

Very breathy Creaky

Figure 2. Comparison of the vocal fold vibration patterns of the sustained vowel /e/. Closed phases are shown. (Images via fiberscope for breathy and creaky phonation and via tele-endoscpoe for the others.)

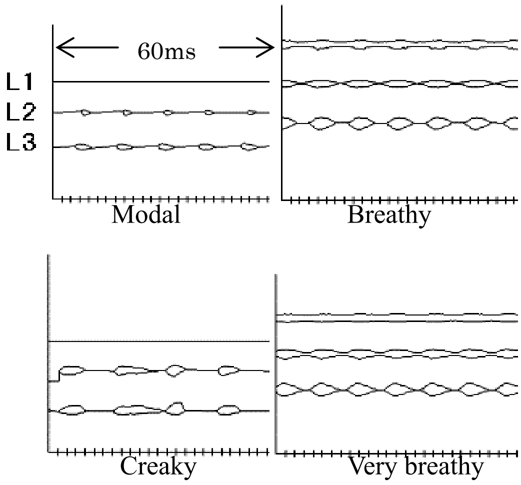


Figure 3 Variation of glottal vibration due to the phonation type. See text.

Figure 4 shows the variation of opening area due to different phonation types. The scale of the y-axis of each panel cannot be compared, since the distance between the fiberscope and the glottis differs from utterance to utterance. While the change of glottal area during modal phonation showed a regular repetition of increasing/ decreasing distance, the change of glottal area during breathy phonation did not reach 0 at any time. This is because the cartilaginous glottis did not show full closure. Such a tendency is more apparent in very breathy phonation. This is because both the ligamental and the cartilaginous glottis did not show full closure. In creaky phonation, the closing phase takes a notably longer time than for the opening phase, and the decrease of the area is irregular.

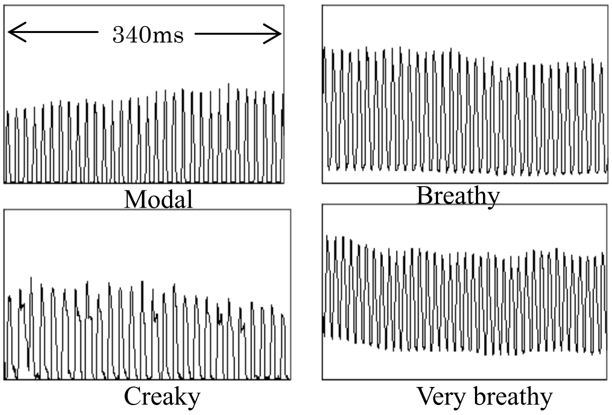


Figure4. Variation of glottal area due to the different phonation types.

3.2 The word /e'ki/

Figure 5 shows the closed phase of the vocal fold vibration during /e/ in /e'ki/ under three different paralinguistic renditions. For the 'neutral' utterance, the cartilaginous glottis remains closed throughout the utterance and the opening/closing movements were limited to the ligamental glottis, similar to what was seen for modal phonation. In the 'disappointment' utterance, the cartilaginous glottis was not fully closed at the closed phase, while there was full closure at the ligamental glottis. The vocal processes remained apart throughout the utterance, similar to breathy phonation. In 'suspicion,' adduction of the false vocal folds and constriction of pharyngeal cavity were observed, similar to creaky phonation.

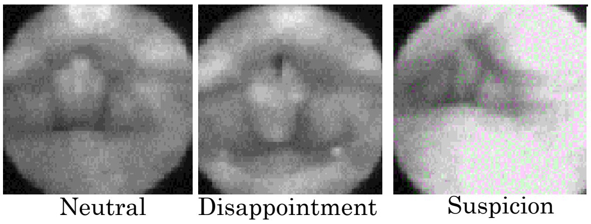
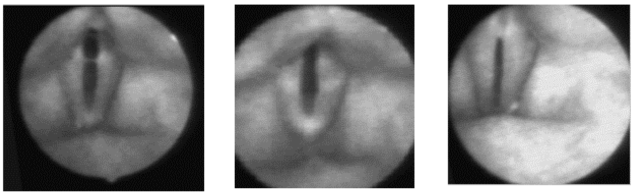

Neutral Disappointment Suspicion

Figure 5. Comparison of the vocal fold vibration pattern during /e/ of /e'ki/. Closed phases are shown. (Images are via fiberscope)

Figure 6 shows the glottal image during /k/ in /e'ki/. For the 'neutral' utterance, the arytenoid cartilages and vocal processes were both abducted. This is the usual setting of the glottis for voiceless consonants [10]. In contrast, for 'discouragement,' the arytenoid cartilages were approximated, while the vocal processes were abducted similar to that for the neutral utterance. For 'suspicion,' although adduction of the false vocal folds and constriction of the pharyngeal cavity disappeared during /k/, the vocal processes were still approximated and the arytenoid cartilages seemed to be approximated as well.



Neutral Disappointment Suspicion

Figure 6. Comparison of vocal fold posture in /k/ of /e'ki/. (Images are via fiber scope)

Table 2. Summary of glottal characteristics of /e/ and /k/ in /e'ki/ with different paralinguistic information.

/e/	Neutral	Disappointment	Suspicion
Arytenoid cartilages	Approximated	Approximated	Approximated
Vocal processes	Approximated	Separated	Approximated
Resemble to	Modal phonation	Breathy phonation	Creaky phonation

/k/	Neutral	Disappointment	Suspicion
Arytenoid cartilagees	Separated	Approximated	Approximated
Vocal processes	Separated	Separated	Approximated

Table 2 summarizes the characteristics of vocal folds due to the paralinguistic information. It is natural that the arytenoid cartilages and vocal processes approximate for the voiced sound /e/ and separate for the voiceless sound /k/, as is seen for the 'neutral' utterance in table 2 [9]. For 'discouragement,' however, the arytenoid cartilages approximated and the vocal processes separated for both /e/ and /k/. And in 'suspicion,' the arytenoid cartilages separated and the vocal processes approximated for both /e/ and /k/. These facts suggest influence of paralinguistic information on glottal adjustment. They also suggest that the domain of the influence covers both vowel and consonant.

Figure 7 shows the time course of the glottal area for /e'ki/. For convenience of comparison, the areas are normalized to the value of the mid point of /e/. The scale of the y-axis of each panel cannot be compared. The time course of the glottal area differed considerably depending on paralinguistic information. For the 'neutral' utterance, the maximum glottal area during /k/ was larger than that during the vowels /e/ and /i/. This corresponds to the observation that the cartilaginous glottis is closed during /e/, opened during /k/ and closed again during /i/. This agrees with the previous study on the glottal area change during /VCV/ [11]. In contrast, the glottal area during /k/ in 'disappointment' is rather smaller than that of the maximum area during /e/. This is because the cartilaginous glottis opened during the vowel but the extent of glottal opening did not increase during /k/. Consequently, the maximum opening of the ligamental glottis during the vowel was larger than its opening for /k/. As for 'suspicion,' the glottal opening for /k/ was smaller than the maximum opening for the vowels. This is because the cartilaginous glottis remained closed at /k/, and the maximum opening of the ligamental glottis during the vowel was larger than that of /k/.

It is possible that the measurement of glottal area during the vowel may be somewhat underestimated for 'suspicion' due to the constriction of the pharyngeal cavity. Also, the larynx moved upward from /e/ to /k/ and downward from k/ to /i/, which may have affected the measurements. However, we believe this does not effect our conclusion seriously, because the movement pattern was similar across utterances and type of paralinguistic information.

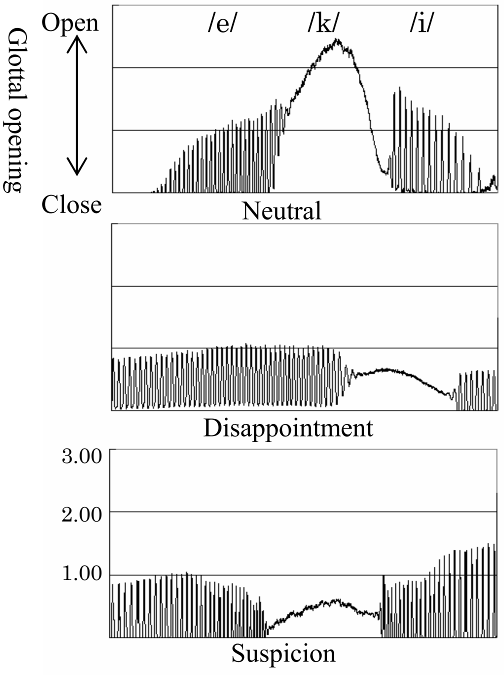


Figure 7. Variation of time course of the glottal area. Utterance duration of each utterance is 680ms.

4. ACOUSTIC ANALYSIS

4.1 Data and method.

The speech sound was recorded during the experiment and the trial sessions by DAT with a sampling rate of 48kHz. 2048 point FFT analysis was carried out at the point of maximum amplitude in /e/ of /e'ki/. Spectral tilt of H1 (f0) to H2 was examined. Since the main purpose of this experiment was to get data of high-speed video images, the number of utterances for each type of paralinguistic information varied depending on the difficulty of video-imaging. Number of samples was 33 for 'neutral', 7 for 'disappointment' and 57 for 'suspicion.'

4.2 Results and discussion

Figure 8 shows an example of FFT analysis. A previous study showed that spectral tilt is one of the major acoustic parameters that characterize phonation types. Spectral tilt is steeply negative for breathy vowels and steeply positive for creaky vowels [9].

Figure 9 shows the distribution of samples on the H1-H2 plane. Ellipses in the figure show the 95% confidence area of each utterance group. As for 'suspicion,' samples were separated into two sub-groups corresponding to different recording sessions. 'Suspicion2' was recorded earlier than 'suspicion1'. Figure 9 shows that each utterance group was well separated.

The distribution of 'disappointment' utterances clusters relatively below that of 'neutral.' This is similar to the characteristic pattern for breathy phonation reported in [9]. In contrast, the distribution for 'suspicion1' and 'suspicion2' clusters relatively above that of 'neutral.' This is similar to the characteristic pattern for creaky phonation [9]. Note that 'suspicion1' and 'suspicion2' run parallel in terms of H1-H2 ratio except that both H1 and H2 were smaller in intensity in ‘suspicion2’.

The averaged ratio of H2 to H1 was calculated in order to examine the characteristic variation due to the paralinguistic information. Results were 0.96(s.d.0.03) for ‘neutral’, 0.85(s.d.0.05) for ‘disappointment,’ 1.06 (s.d.0.05) for ‘suspicion1,’ and, 1.10(s.d.0.05) for ‘suspicion2.’ In figure 10, distribution of the ratio is shown according to the utterance groups. Distribution is well separated among paralinguistic types. A significant difference was found by ANOVA (p<0.0001) and for all the pairs (p<0.05) by Fisher’s PLSD test.

To sum up, acoustic analysis supports the results obtained from video-imaging.

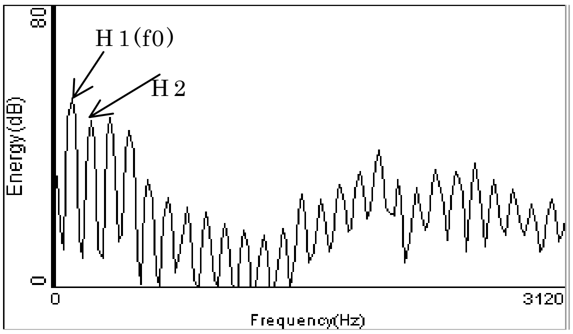


Figure 8. Example of FFT analysis. Speech sample is ‘disappointment.’

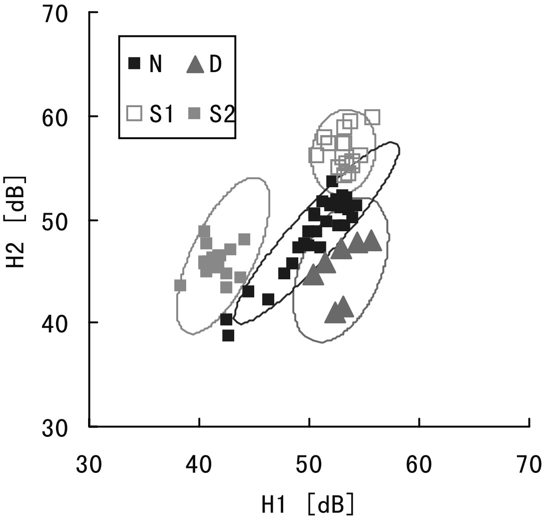


Figure 9. Distribution of H1 and H2 of utterances on /e/ in /e’ki/ under three different situations of paralinguistic information. N, D, S1 and S2 stand for ‘neutral,’ ‘disappointment,’ ‘suspicion1’ and ‘suspicion2,’ respectively. Ellipses show the 95% confidence area.

5. CONCLUSION

This study showed the influence of paralinguistic information upon phonation. ‘Disappointment’ and ‘suspicion’ utterances were characterized by breathy and creaky phonation respectively. It was also shown that the influence was not limited to a single segment; it stretches over several segments including both vowel and consonant. This finding accords with the finding of [5] that examined the influence of paralinguistic information on articulatory gestures.

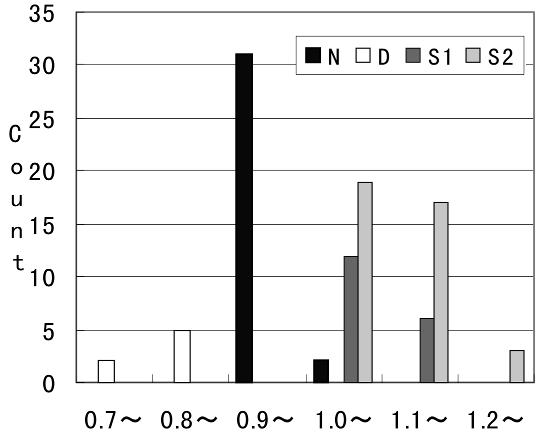


Figure 10. Ratio of H2 to H1 in /e/ in /e’ki/ under three different situations of paralinguistic information.

REFERENCES

[1] H. Fujisaki,”Prosody, Models, and Spontaneous Speech,” Sagisaka et al. (ed.), *Computing Prosody*, Springer, 1997.

[2]K. Maekawa, “Phonetic and Phonological Characteristics of Paralinguistic Information in Spoken Japanese,” *Proc. 5th International Conf. on Spoken Language Processing, vol. 2,* pp. 635-638, Sydney, Australia, Dec. 1998.

[3]K. Maekawa and N. Kitagawa, “How does speech transmit paralinguistic information?” *Cognitive studies* 9, no.2, pp.46-66, 2002.(In Japanese)

[4] H. Kasuya, K. Maewaka and S. Kiritani, “Joint estimation of voice source and vocal tract parameters as applied to the study of voice source dynamics,” *Proc. 14th International Congress of Phonetic Sciences, vol3*, pp. 2505-2512, San Francisco, USA, Aug. 1999.

[5] K. Maekawa and T. Kagomiya, “Influence of paralinguistic information on segmental articulation,” *Proc. 6th International Conf. on Spoken Language Processing, vol. 2*, pp.349-352, Beijing, China, Oct. 2000.

[6] H. Kasuya, M. Yoshizawa and K. Maewaka, “Role of voice source dynamics as a conveyer of paralinguistic features,” *Proc. 6th International Conf. on Spoken Language Processing, vol. 2*, pp.345-348, Beijing, China, Oct. 2000.

[7] S. Kiritani, H. Imagawa and H. Hirose, “Vocal cord vibration in the production of consonants —Observation by means of high-speed digital imaging using a fiberscope,” *JASJ (E)*, 17, no. 1, pp.1-8, 1996.

[8] S. Kkiritani, “High-speed digital image recording for observing vocal fold vibration,” in *Voice quality measurements*, R. D. Kent and M. J. Ball Eds,, Singular, san-diego, pp. 269-283, 2000.

[9] A.N Chasaide and C. Gobl. “Voice source variation, in *Handbook of phonetic sciences*, W.J. Hardcastle and J. Laver Eds., pp.427-461, Oxfprd Blackwell 1997.

[10] T. Chiba and M. Kajiyama, “*The Vowel : Its Nature and Structure*,” Tokyo-Kaiseikan, Tokyo, 1942. (In Japanese)

[11] H. Yoshioka, “Glottal area vibration and supraglottal pressure change in voicing control,” *Ann. Bull. Research Institute of Logopedics and Phoneatrics*, Univ. of Tokyo no. 18, pp. 45-49, 1984.