Paralinguistic information affects phonation types:  
A case study using high-speed video images

Masako Fujimoto and Kikuo Maekawa

Center for Corpus Development, National Institute for Japanese Language and Linguistics,
10–2, Midori-cho, Tachikawa, 190–8561 Japan

(Received 24 May 2012, Accepted for publication 10 August 2012)

Abstract: A case study on the correlation between phonation type and paralinguistic information in Japanese has carried out using a high-speed digital video imaging system. The results showed that “breathy” and “creaky” phonations corresponded to “disappointment” and “suspicion”-related utterances, respectively. Such influence of paralinguistic information stretches over segments including voiceless consonants. This means the alteration due to paralinguistic information is not limited to voice quality but to whole settings of the larynx. These findings are in accord with those of our articulatory study. They suggest that the domain of the phonatory and articulatory setting due to paralinguistic information is the whole utterances, rather than individual segments.

Keywords: Paralanguage, Phonation, Breathy, Creaky, High-speed video images

PACS number: 43.70.+i [doi:10.1250/ast.34.89]

1. INTRODUCTION

We express our feelings in everyday speech not only by linguistic means, but also by paralinguistic ones. Paralinguistic expression can be either unintentional, denoting a speaker’s emotions [1,2], or voluntary, conveying a speaker’s attitude [3]. The inventory and manifestation of voluntary paralinguistic expression may vary, cross-linguistically, more than emotional realization [4] and hence deserves detailed investigation. Previous acoustic studies using typical paralinguistic inventories, such as “admiration,” “disappointment,” “suspicion,” “indifference,” and “focus,” suggest that Japanese speakers display both segmental and prosodic properties in their transmission information types [4–6]. Vowel quality is contrastive between disappointment and suspicion: i.e., \( F_1 \) is low for disappointment and high for suspicion [5]. The results of a study using electromagnetic articulography (EMMA) suggested that the articulatory setting is altered by paralinguistic information types and the scope of the alteration is not a segment but whole utterances. Perceptual studies have shown that Japanese listeners are able to distinguish such information types [7].

It also appears that phonation type plays a crucial role in some paralinguistic information: i.e., breathy voice for disappointment, and laryngealized or pressed voice for suspicion [5,7]. The aim of our study is to clarify the correlation between voice quality and paralinguistic information types. In this respect, we compare the glottal characteristics of disappointment and suspicion with those of breathy and creaky phonation using a high-speed video recording system. Hereafter, the term creaky, rather than laryngealized, phonation is used because it is less ambiguous and more extensively studied than the other.

Among the range of available voice qualities, breathy and creaky phonations are most commonly adopted both linguistically and paralinguistically. They bear different roles in different cultures. A breathy voice bears a phonological role in Hindi, Newar, Tsonga, and Gujarati [8], while a creaky one bears such a role in Arabic, Chadic, Nilotic, and North American Indian languages, such as Kwakw’ala [8–10]. In White Hmong, breathy and creaky voices are both phonological [11]. Paralinguistically, a breathy voice in the English language expresses intimacy [9,12]. Creaky voice signals indicate turn-taking, boredom [9], reluctance, suppressed rage, unwilling concession [10], commiseration, and complaint in Tzeltal, a Mayan language [9]. In the balance of this paper, the term “paralinguistic information” will be used with reference to the voluntary or deliberate expression of a speaker’s attitudes [3].

2. METHODS

2.1. Data Collection

An adult male speaker of standard (Tokyo) Japanese served as the subject. He is one of the authors and was in...
his mid-40s at the time of the recordings. The subject produced a one-word utterance /e’ki/, “train station,” where ‘/’ denotes the accent nucleus. The utterance was produced with three different types of paralinguistic rendition: “suspicion,” “disappointment,” and “neutral,” as well as a sustained vowel /e/ uttered with “modal,” “breathy” and “creaky” phonation.

An otolaryngologist inserted a fiberscope in the subject’s nostril and positioned it above the larynx. Glottal images were recorded using a high-speed digital video recording system at the rate of 4,500 frames per second [12–15]. Note that this frame rate enables us to investigate glottal vibration, which is generally 100 Hz and higher and, hence, conventional videos (30 frames per second) are inadequate. A specially designed fiberscope with a thicker light guide (FNL-T15, Asahi Optical Company) as well as an image intensifier (C6276-01, Hamamatsu Photonics K.K.) is used to maximize the quality of images [16]. Nevertheless, owing to the trade-off of numerous frame rates, the resolution of individual frames may be less fine especially when the center part is enlarged. The maximum recording time for a single sample was 0.68 s (i.e., 3,072 frames), which barely covers the test word. The data transfer to a computer took 15 min per sample. Because of these time-consuming tasks, one sample for each task was recorded. An endoscope was used in some cases of the sustained vowel. Speech sound alone was recorded for additional acoustic analyses during the data transfer. An analytic result of these speech signals has been presented elsewhere [17]. These data were collected in 1999 at the Research Institute of Logopedics and Phoniatries (RILP), University of Tokyo.

2.2. Data Analysis

We measured the glottal opening area and the distance between the vocal folds using computer programs developed at RILP. Figure 1 shows an example of the measurements. As for the area measurements (i.e., the left two panels), we set the brightness threshold value of the pixel intensity in the right panel so as to match the glottal shape in the left panel, and then measured the subliminal (darker) area. The rectangle in the figure shows the bounds of the measurement area. Distances were measured at three points (L1 to L3): immediately posterior to the vocal processes at the arytenoids (L1), immediately anterior to the vocal processes of the thyroid cartilage (L2), and a point more anterior to the thyroid cartilage (L3). We used a multiple line measurements, because glottal movement differs significantly between the ligamental and cartilaginous glottises [18].

3. RESULTS AND DISCUSSION

3.1. Phonation Types in Sustained Vowel /e/

Figure 2 shows the closed phase of the glottis for modal, breathy and creaky phonation types during the sustained vowel /e/. Figures 3 and 4 show the results of the distance measurements and the opening area measurement, respectively. Note that the scale of the y-axes of each panel in the Fig. 4 differs, because the distance between the fiberscope and the glottis may vary.

In modal phonation, the ligamental parts of the vocal folds vibrate periodically. The cartilaginous glottis is unseen, but is supposedly closed during phonation, since
the arytenoids are fully closed. These observations are in strong agreement with those of a previous study [19]. As a result, the time course of the glottal distance in Fig. 3 shows regular opening and closing movements at L2 and L3. Also, the glottal area in Fig. 4 shows a repetition of smooth increase and decrease.

In breathy phonation, both the ligamental and cartilaginous parts of the vocal folds vibrate. The cartilaginous glottis is not fully closed during the closed phase, and the vocal processes remain abducted throughout the utterance. As a result, the glottal distance at L1 in Fig. 3 does not reach zero. Also, the closed phases at L2 and L3 are relatively short and show larger open quotient (OQ) values than in the case of modal phonation; this tendency is more prominent in L2 than in L3, because L2 is physically closer to the vocal process. The glottal area in the Fig. 4 does not reach zero at any time.

In creaky phonation, the false vocal folds adduct and the pharyngeal cavity constricts, which hides the cartilaginous glottis. However, like the laryngealized sounds reported by Ladefoged [20], arytenoids are assumed to be pressed inward; i.e., the cartilaginous glottis is closed throughout. The dotted line of L1 in Fig. 3 denotes this assumption. The displacement of the glottal distance in Fig. 3 shows irregularity at both L2 and L3. In Fig. 4, the duration of the closing phase is notably longer than that of the opening phase, and the decrease of the area is irregular.

These observations are consistent with general descriptions, wherein a breathy voice involves minimal adductive tension and weak medial compression, whereas a creaky voice is characterized by strong adductive tension and medial compression with vigorous ventricular involvement [9].

3.2. Paralinguistic Renditions for /e’ki/

Figure 5 shows the selected frames of a cycle of the vocal-fold vibration of the /e/ in /e’ki/, during paralinguistic renditions of neutral, disappointment and suspicion, respectively. The interval of each frame is around 10 ms. For the neutral utterance, the cartilaginous glottis remains closed throughout the vowel part and the opening/closing movements are limited to the ligamental glottis. This is similar to the modal phonation described in Sect. 3.1. In the disappointment utterance, the cartilaginous glottis does not reach full closure, but the ligamental part does at the closed phase. The vocal processes remain abducted throughout the utterance. This is similar to the breathy phonation described in Sect. 3.1. In the suspicion utterance, the adduction of the false vocal folds and constriction of the pharyngeal cavity are characteristic; this is similar to the creaky phonation described in Sect. 3.1.

Figure 6 shows the glottal images of /k/ in /e’ki/. For the neutral utterance, the arytenoids and vocal processes are both abducted; this is the usual setting of the glottis for voiceless consonants [18]. For the discouragement utterance, the arytenoids and the vocal processes are abducted as well. However, the state of vocal folds differs critically: i.e., less tension seems to be involved in disappointment. In contrast, for suspicion, the vocal processes and the arytenoids are kept adducted, although the adduction of the false vocal folds and constriction of the pharyngeal cavity become lesser than that in the preceding vowel.

Table 1 summarizes the state of the larynx. It is quite common that the arytenoids and vocal processes become adducted for the voiced sound /e/ and abducted for the voiceless sound /k/, as seen in neutral utterance [17]. In discouragement, however, the arytenoids and the vocal processes become abducted for both the vowel and voiceless consonant. By contrast, in suspicion, the arytenoids and the vocal processes were adducted for both vowels and voiceless consonants. These findings revealed that the influence of paralinguistic information on glottal adjustment covers both the vowel and the consonant.
Figure 7 shows the displacement of the glottal area for /e’ki/. For the sake of comparison, the values are normalized at the midpoint of /e/. The time course of the glottal area differs considerably depending on the paralinguistic meaning. For the neutral utterance, the maximum glottal area during /k/ was larger than that during the vowels /e/ and /i/. This agrees with the findings of a previous study of glottal-area change during /VCV/ [21]. In contrast, the glottal area during /k/ in disappointment is smaller than that of /e/. This reflects the fact that the cartilaginous glottis does not fully close during the vowel and the extent of glottal opening does not increase during /k/. Consequently, the maximum opening of the ligamental glottis during the preceding vowel was larger than that for /k/. In suspicion as well, the glottal opening for /k/ was smaller than the maximum opening for the vowels; this reflects the fact that the cartilaginous glottis remained closed during /k/. It is possible that, in suspicion, the glottal area during the vowels may have underestimated owing to the constriction of the pharyngeal cavity. However, these effects seem to be smaller than those of paralinguistic information.

4. GENERAL DISCUSSION

Although limited to one subject, this study revealed that, at least in Japanese, some paralinguistic information apparently influences the phonation type. A disappointment utterance is characterized by breathy phonation, and suspicion, by creaky phonation. These results agree with those of the previous acoustic analysis which involved more subjects [5,7]. Among the two sets of paralinguistic information examined in this study, the correlation between disappointment and breathiness is presumably stronger than that between suspicion and creaky. When disappointed, body tension becomes weaker, as does the tension of glottal muscles. If so, the closure of the glottis may be less tight than that in neutral utterance, which, in turn, increases noise during phonation, i.e., speech becomes breathy. In contrast, the physical consequence of suspicion is less straightforward. Variations found in listening impressions, pressed [5] and laryngealized [7], may be derived from this. Further study to evaluate the findings is expected.

The present results clearly showed that the influence of paralinguistic information stretches over segments including voiceless consonants. This means such alteration due to paralinguistic information is not limited to voice quality.
but to whole settings of the larynx. These findings are in accord with the articulatory investigation that revealed that the paralinguistic information alters the articulatory gestures during utterances [4]. These results, in combination, strongly suggest that speakers of Japanese manipulate both phonation and articulation in order to convey a set of paralinguistic information.

Acknowledgement

An earlier version of this paper was presented at ICPhS 2003 [22]. We thank John Ohala and two anonymous reviewers for their comments and advice. This work is partly supported by JSPS KAKENHI (C) No. 23520539.

REFERENCES


