Influence of Paralinguistic Information on Segmental Articulation

Kikuo Maekawa and Takayuki Kagomiya
Dept. of Language Behavior
The National Language Research Institute
{kikuo/kagomiya}@kokken.go.jp

ABSTRACT
Influence of paralinguistic information upon articulatory gesture was examined using EMMA record of tongue, jaw and lips. The focus was upon the contrast between “admiration” and “suspicion.” It was found that previously reported contrast in F2 is related to the horizontal displacement of the tongue body and lip aperture control. These articulatory characteristics were observed not only in vowels but also in consonant portions. This suggests that the effect of paralanguage upon speech production is a manipulation of voice-quality rather than the control of individual segments.

1. INTRODUCTION
Speech transmits paralinguistic and non-linguistic information in addition to linguistic information. Non-linguistic information, like gender and physiologically motivated emotion, is the information that the speaker can not control at his/her own will. On the other hand, linguistic and paralinguistic information are manifested under the speaker’s deliberate control [1]. The crucial differences between the linguistic and paralinguistic information can be found in that 1) paralinguistic information is rarely expressed in the written counterpart of a given utterance (note this is also true with non-linguistic information), and 2) the magnitude of paralinguistic information can be changing within a given category. For an example, the magnitude of ‘disbelief’ associated with an utterance can be either slight or deep depending on the speaker’s mental state.

2. PROBLEM
We have examined phonetic characteristics of paralinguistic information in Japanese mainly from an acoustic point of view, and found that the effect of paralinguistic information can be seen in all phonetic parameters examined, i.e. duration, pitch, and spectrum [2,3]. At this point we would like to attract readers’ attention to the fact that paralinguistic information influences not only so-called prosodic features but also segmental characteristics of speech.

Maekawa (1998) examined the F1 and F2 (the first and second formant frequencies) of the /a/ vowels at the end of three sentences and found systematic difference of F2 between ‘suspicion’ and ‘admiration’ utterances. F2 was higher in the rendition of ‘suspicion’ than in that of ‘admiration’.

The aim of the present paper consists in the examination of the effect of paralinguistic information upon articulatory gesture rather than formants which are prone to measurement error.

3. DATA ACQUISITION
Measurement of articulatory gestures was obtained using an EMMA system operating at the NTT Communication Science Laboratory in Atsugi, Japan (courtesy, Masaaki Honda and his colleagues).

Figure 1 shows the placement of coils used in the experiment. Four coils (T1-T4) were placed on the surface of tongue, two on upper and lower lips (UL and LL), and the last one on the surface of lower incisor (MN).

Two male subjects took part in the experiment, but in this paper only one subject was analyzed. The subject is a middle-aged linguist who is the first author of the present paper. He uttered four one word utterances (/e'ki/="station", /a'ki/="autumn", /sa'ke/ ="salmon", /sa'ga/= name of a Japanese prefecture; where apostrophe stands for lexical accent location), and one short phrase (/sasadaga/; a Japanese surname /sasada/ followed by a nominal particle /ga/) under four different paralinguistic renditions: ‘Neutral’, ‘Disappointment’, ‘Admiration’, and ‘Suspicion.’ The instructions given to the subjects about the four paralinguistic information types can be found in the CD-ROM version of Maekawa (1998).

Each of the resulting 20 utterances (5 texts * 4 renditions) were repeated ten times. Speech and EMMA data were digitized under 16000Hz-16bit and 250Hz-
16bit conditions respectively. EMMA data was smoothed by 12.5Hz low-pass filtering.

4. ACOUSTIC CHARACTERISTICS

4.1 Perception test
In order to show the validity of the recorded utterances, an identification test was conducted. All 200 recorded utterances were played back in a random order, and 12 naïve subjects were asked to identify the paralinguistic meaning intended by the speaker. The overall correct perception rate was higher than 98%.

4.2 Prosodic characteristics
Figure 2 shows F0 contours of typical utterances of /sasadaga/ that differ in paralinguistic meaning. Duration of A(dmiration) and D(isappointment) are nearly the same and the longest. N(utral) is the shortest, and S(uspicion) is intermediate. D has very narrow pitch range, while S has very wide one. Only S ends in rising pitch. These are congruent with our previous studies.

4.3 Formant distribution of /a/
Figure 3 shows the F2 (second formant) vs. F1 (first formant) scatter plots made separately for the first and the last vowels in /sasadaga/; both F1 and F2 axes are inverted for the sake of the comparison with articulatory data. LPC-based formant measurement with filter order of 20 was done at the time points of local minima of MN_Y coil, i.e. the time when the jaw opened most widely for the vowels. They show clearly that Suspicion samples have higher F2 values compared to Admiration. The same distributional difference was observed in the other two /a/ vowels in the same sentence. This is, again, congruent with our previous observations [2,3].

Based on these agreements, we assume that the current speaker’s manners of phonetic manifestation of paralinguistic information are essentially the same as those observed in our previous experiments.

5. ARTICULATORY CHARACTERISTICS

5.1 Tongue position of /a/
Figure 4 shows the distribution of T3 (tongue dorsum) coil measured at the same time point as in figure 3. The same vowels as in figure 3 were analyzed. T3_X and T3_Y stand respectively for the horizontal and vertical positions of the coil.

It is clear that the tongue dorsum position differs systematically according to paralinguistic information; the tongue takes a relatively more frontal position in Suspicion than in Admiration. Essentially the same tendency was observed for T1 and T2. This finding suggests that the tongue as a whole takes a different position depending on paralinguistic information. Analyses of other two /a/ vowels revealed the same tendency.

5.2 Vowels other than /a/
It is necessary to examine vowels other than /a/, because /a/ is the only open vowel in Japanese vowel system, that presumably has no phonological specification of horizontal position. It is, therefore conceivable that the observed difference in horizontal position is a phenomenon limited to the vowel which has greater degree of freedom in horizontal variability.

Figure 5 shows the acoustic and articulatory characteristics of the /e/ vowel in /e ki/ uttered by the same speaker. Although /e/ has phonological specification of [FRONT] (there is a [BACK] counterpart, i.e. /o/), the same distributional difference as in figures 3 and 4 was observed for this vowel. The analysis of /i/ in /e ki/ revealed the same distributional difference both in formant and articulation.

5.3 Consonant articulation
From a point of view of speech production theory, an important problem to be examined here is whether the observed articulatory difference is limited to vowels or not. To examine this problem, coil positions were compared through the whole sentence.

Figure 6 shows the average and standard error values of T3_X and T1_X computed for each (acoustically determined) segment of /sasadaga/. In this figure, the average of Admiration is always higher, i.e. more back, than that of Suspicion regardless of the segment.
It suggests that the effect of paralanguage applies both for vowels and consonants as far as the current speaker is concerned. See the discussion at the end of this paper.

5.4 Labial articulation

The foregoing articulatory analyses suggest strongly that the observed F2 rising/lowering is the acoustic consequences of tongue fronting/backing. However, there is at least one more articulator that has significant influence on F2 values, namely lips. Acoustic theory of

Figure 3. F2-F1 scatter plot of the first (/sa/) and last (/ga/) vowels of /sasadaga/. Both axes are inverted. Plot symbols ‘S’ and ‘A’ stand respectively for Suspicion and Admiration.

Figure 4. T3_X-T3_Y scatter plot of the first (/sa/) and last (/ga/) vowels of /sasadaga/. Plot symbols ‘S’ and ‘A’ stand respectively for Suspicion and Admiration. Unit of measurement is 0.01 mm.

Figure 5. Acoustic (F2-F1) and articulatory (T2_X-T2_Y) distribution of /e/ vowel in /e’ki/ (“station”) uttered by the same speaker. Plot symbols ‘S’ and ‘A’ stand respectively for Suspicion and Admiration.
speech production tells us that increase/decrease in lip aperture results in rising/lowering of F2 [4]. Figure 7 shows the averaged euclidian distance between the UL and LL coils as a function of phonological segment. Clearly, lip distance is greater in Suspicion than in Admiration, with the only exception of the first segment of /s1/. It is to be noted that the difference was maintained for most of the consonants.

6. CONCLUSION AND FURTHER PROBLEMS
The present experiment revealed that systematic influence of paralanguage can be observed at the physiological level of speech production.

It also revealed the prosodic nature of the influence, in the sense that the effect of paralanguage covers the whole utterance including both vowels and consonants.

These findings suggest strongly that the articulatory displacement caused by paralanguage is to be considered the consequence of manipulation of a voice-quality feature rather than segmental feature. Clarification of the manipulation mechanism is an indispensable task of a general theory of speech production.

 Needless to say, there remain many problems to be examined. One of them is the existence of inter-speaker difference of voice-quality manipulation. Preliminary examination of the data taken from the second subject showed exactly the same displacement of tongue in the case of /a/ vowel, but not for /e/ and /i/.

Moreover, the contribution of articulatory or spectral changes to the perception of paralinguistic information remains to be examined in a quantitative way. Informal perception test of spectrum-pitch converted stimuli made out of Admiration and Suspicion utterances suggests significant contribution of spectral information.

ACKNOWLEDGMENT
This work has been supported by CREST of JST (Japan Science and Technology) and the grant-in-aid for scientific research from the Ministry of Education, Science and Culture (No. 10410110).

REFERENCES

Figure 6. Mean value of T1_X (top) and T3_X (bottom) as a function of phonological segment and paralinguistic information type (A, S, and N). Short vertical lines show the standard error for each segment. Values for Neutral utterance was overlaid for comparison purpose.

Figure 7. Mean euclidian distance between the UL (upper lip) and LL (lower lip) coils as a function of segment and paralinguistic information type. Short vertical lines show the standard error for each segment. ‘A’ and ‘S’ stand respectively for Admiration and Suspicion.