TR-IT-0034

The role of speaking style, duration, and f0 in signalling affect: anger and politeness.

Etsuko Ofuka
(Leeds University, U.K.)

Nick Campbell
(ATR-ITL)

1994.1

ABSTRACT

This report describes research carried out at ATR-ITL on the prosodic manipulation of speech to emulate politeness. It is concerned particularly with the effects of changing fundamental frequency and segmental duration, and examines their inteaction with the underlying voice quality of the original speech. It was found that whereas appropriate settings of both duration and fundamental frequency can change the perception of politeness (kindness/anger) in the resultant speech, neither cue alone is strong enough to override the effect of the original voice source and trigger a percept of anger or politeness.

1. Introduction

People generally use multiple channels such as visual and auditory cues to encode and decode affect in everyday life. The interaction between visual, audio, and verbal cues may be substantial, and should be taken into consideration in order to understand the process of communicating affect. However, there are many factors involved , and it is necessary for practical reasons to restrict factors to be studied.

This study will focus on the audio channels, which can be categorised into two types: verbal content (i.e., what you say) and acoustic variables (i.e., how you say it) (Ladd et al., 1985 p. 435). The acoustic variables include duration or time structure, fundamental frequency (f0), amplitude, and voice quality (e.g., formant characteristics, energy distribution, etc.) (Reviews are in Kramer, 1963; Crystal, 1969, pp. 62-82; Scherer, 1979a, 1979b, 1982, pp. 150-169; Frick, 1985; Murray & Arnott, 1993; etc.). Among those, duration and f0 have been most commonly studied, because firstly they are robust features in the sense that they survive even if the quality of sounds is not very good (e.g., in telephone conversations), and secondly they are easy to measure in analysis and easy to manipulate in synthesis.

Time structure is reported to be the most powerful cue to indicate emotions. For example, Brown et al. (1974) studied f0 mean, f0 variance, and speech rate on benevolence and competence dimensions, and reported that speech rate was the most influential cue (accounted for 48% variance). Scherer & Oshinsky (1977) studied f0 mean, f0 variance, contour, amplitude variance, tempo, amplitude envelope, and filter cut-off frequency on

evaluation, activity, and potency dimensions, and found tempo was the most noticeable cue on all scales studied. Kitahara & Tohkura (1988) also reported that time structure was the most important for recognising anger, while pitch for joy and sadness.

The findings are, however, hard to generalise, as pointed out by Van der Bas (1992: 6-10); the researchers have used different utterances and different sets and levels of prosodic features, which were manipulated by using different techniques and software packages, and asked different subjects to rate them on different scales in different rating forms in different settings....
The only thing which can be said with certainty may be that time structure, f0, amplitude, and voice quality are all important to some extent depending on various factors such as the types of affect, speakers, languages, etc.

In this study, three types of acoustic factors, duration, f0 and speaking styles (e.g., a 'voice' spoken angrily or kindly) were studied in relation to affect in Japanese. The attitudinal meanings studied here are anger, kindness, and politeness. The utterance used in this experiment is 'Nimotsu wa koredake desuka?' meaning 'Is this all the luggage you have?', which is commonly used by a customs officer to passengers, spoken by a trained male speaker in an angry/irritated way, a kind/considerate way, and a normal/ordinary way. Politeness was examined together with anger and kindness, because it is almost always involved in social interaction, especially in the Japanese society.

2. Objectives

The experiment was designed to examine

(1) whether or not the factors function independently in signalling affect, as suggested in Ladd et al. (1985),

(2) the relative importance of the factors, i.e., speaking styles, duration, and f0 contour, on the scales of anger, kindness, and politeness, and

(3) the difference between judgements of anger, kindness, and politeness, in terms of the effects of the factors

3. Method

3.1 Design

A factorial 2 x 2 x 3 design was used with two speaking styles (angry, kind), two types of duration (angry, kind) and three types of F0 contours (angry, kind, normal), as within-subjects factors. The styles were produced by one speaker's simulating different types of affect. The variables duration and F0 were manipulated through digital resysnthesis based on a PSOLA technique (Moulines & Charpentier, 1990).

3.2 Speech material

3.2.1 Utterance

The utterance used is

'Nimotsu-wa(,) koredake desuka'

luggage-as for, this is all?

(Is this all the luggage you have? --- by a customs inspector to a passenger)

A professional male narrator was instructed to speak this utterance in a short conversation between a customs officer and a passenger in several different speaking styles (for the speech material, see Miyatake & Sagisaka, 1990). Among those, utterances

spoken in an angry/ irritated way, a kind/considerate way, and a normal/ordinary way were used in this experiment.
Although this sentence is mostly used as a routine question at the customes and people may not listen to the utterance carefully, we sometimes experience that some officers sound very kind or polite, while others very rude.

3.2.2 Prosodic parameters

Three variables were used:

(1) speaking style (STYLE)

This includes amplitude, formant characteristics, energy distribution, etc. other than the duration and F0 factors.
Two types of styles were used:

(1.1) angry or irritated style

(1.2) kind or considerate style

The auditory characteristics of these styles are the angry style voice is loud and 'tense' while the kind version soft and 'relaxed', especially the final mora (see Appendix 1).

(2) duration (DUR)

Two types were used:

(2.1) angry type

Phoneme duration extracted from the angry utterance

the rate of articulation: 105 ms / mora

(2.2) kind type

Linearly compressed phoneme duration of the kind utterance by 20%

the rate of articulation: 128 ms / mora, and

160 ms / mora for the original kind utterance

The reason why the compression was done is that the original

kind utterance sounded unnaturally slow to most subjects participated in an informal listening test. Although linear compression across different types of phonemes does not take place in human speech, the utterances which were linearly compressed by 20% sounded quite natural in terms of both quality and speed to the subjects and the experimenter.

The angry, kind and normal durations are shown in Table 3.1.

Table 3.1 Phoneme durations for the angry, kind, and normal utterances in ms

Phoneme	angry	kind	normal
n	25	15	45
i	60	85	55
m	35	50	45
o	90	110	90
ts(u)	65	13	105
w	40	110	85
a	85	115	140
PAUSE	0	0	60
k	45	105	45
o	60	60	40
r	10	15	10
e	65	85	70
d	30	43	35
a	85	108	75
k	10	70	35
e	55	95	50
d	15	40	35
e	75	80	55
s(u)	60	12	75
k	45	90	50
a	200	235	130

(3) F0 contour (F0)
Three types were used:
(3.1) angry type
F0 values extracted from the angry utterance
(3.2) kind type
F0 values extracted from the kind utterance
(3.3) normal type
F0 values extracted from the normal utterance
The three contours are shown in Appendix 2.
Table 3.2 shows the range and mean values for the contours.

Table 3.2 The range and mean values for the contours in Hz

Type	Range (min-max)	Mean (1st half, 2nd half)
angry	71 (113-185)	157 (159, 157)
kind	124 (107-231)	141 (137, 144)
normal	95 (74-169)	112 (123, 99)

where 1st half is 'Nimotsu-wa', and
2nd half is 'koredake desuka'

The angry F0 type is characterised by narrow range, which sounded 'flat', and higher mean f0 than the normal type, while the kind F0 type, wider range and higher mean f0 than the normal type.

3.3 Stimulus preparation

3.3.1 Stimuli

Table 3.3 shows twelve patterns (2 STYLEs x 2 DURs x 3 F0s) produced from the two natural utterances (i.e., the angry and kind utterances) by using a PSOLA technique.

Table 3.3 Stimuli patterns

Pattern	STYLE	DUR	F0
1 (*1)	angry	angry	angry
2	angry	kind	angry
3	angry	angry	kind
4	angry	kind	kind
5	angry	angry	normal
6	angry	kind	normal
7 (*2)	kind	kind	kind
8	kind	angry	kind
9	kind	kind	angry
10	kind	angry	angry
11	kind	kind	normal
12	kind	angry	normal

where
*1: the original angry utterance
*2: the original kind utterance compressed by 20%

The manipulation of the angry and kind types of DUR and F0 was done automatically by creating a mapping table between the original and target values. Imposition of the normal type of F0 on the angry and kind types of voice source was done manually; designating f0 values on the boundaries of vowels and interpolating them with straight lines (see Appendix 3).

The reason why the normal STYLE was not used is that generated utterances sounded quite unnatural because of the big difference between the mean f0 for the normal utterance and that for the angry/kind utterances.

3.3.2 Presentation of the stimuli

A total of 60 stimuli (12 patterns x 5 occurrences) were

produced. These stimuli were preceded and followed by two dummies in order to avoid the first and the last order effect. Five sets of presentation sequence of patterns (see 'ATR8RAWD.DAT') were used in order to minimise the order effect. Each item was presented twice, preceded by a warning signal, and followed by silence:

Item(i) consisting of a beep sound, stimulus(i), 6 seconds of silence, two beep sounds, stimulus(i), and 8 seconds of silence.

3.4 Subjects

Nineteen subjects (10 males and 9 females) participated in the listening test. All of them were native speakers of Japanese in their 20's and 30's, and members of staff at ATR. Most of them were from the western part of Japan (Kansai area). Some of them had experience of hearing synthetic speech.

3.5 Rating sessions

Subjects were instructed to listen to the stimuli through headphones and rate them on three scales:

politeness, on a 7-point bipolar scale, ranging from -3: the least polite to +3: the most polite, and anger and kindness on a 4-point scale, ranging from 0: does not sound angry/kind to +3: very angry/kind. They were also asked to tick the box in the rating form if the stimulus sounded unnatural. The rating form is in Appendix 4.

Set A, B, and C were presented to two male and two female subjects, Set D to two male and three female subjects, and Set E to two male subjects.

Each session began with an explanation of the task by the experimenter, and four practice stimuli consisting of the

original normal, kind, and angry utterances, and the compressed version of the kind utterance (Pattern 7). Then the subjects rated 32 items, followed by 5-minute break, and the rest of the stimuli (32 items). A session took about 30 minutes and was conducted independently in a small private room.

4. Results and analysis

Scores from 18 subjects (10 males and 8 females) were used for the analyses, except Quantification Method type 1, because one subject (FXD) seemed to regard 'politeness' as 'precision of articulation' while the others as global impression.

4.1 Mean values

Tables 4.1, 4.2, and 4.3 show the mean values and standard deviations for the three rating scale scores for each pattern.

Table 4.1 Mean values and standard deviations for the politeness scores

Pattern (SDF)	Mean [S.D.] All (N=18)	Male (N=10)	Female (N=8)
1 (aaa)	-1.36 [0.845]	-1.24 [0.729]	-1.50 [1.003]
2 (aka)	-0.82 [0.651]	-0.68 [0.834]	-1.03 [0.668]
3 (aak)	-0.48 [0.651]	-0.26 [0.525]	-0.75 [0.723]
4 (akk)	0.06 [0.799]	-0.04 [0.951]	0.20 [0.554]
5 (aan)	-0.28 [0.579]	-0.28 [0.391]	-0.28 [0.785]
6 (akn)	0.18 [0.735]	0.07 [0.808]	0.34 [0.640]
7 (kkk)	1.67 [0.729]	1.34 [0.626]	2.08 [0.667]
8 (kak)	0.32 [0.764]	0.28 [0.694]	0.38 [0.891]
9 (kka)	0.35 [1.100]	0.02 [1.133]	0.71 [1.011]
10 (kaa)	-1.06 [0.699]	-0.78 [0.466]	-1.40 [0.814]
11 (kkn)	1.84 [0.813]	1.74 [0.789]	1.98 [0.878]
12 (kan)	0.10 [1.000]	0.10 [0.949]	0.10 [1.136]
Mean S.D.	[0.790]	[0.741]	[0.814]

where
(SDF) is S: STYLE, D: DUR, F: F0,
a: angry, k: kind, and n: normal.

Table 4.2 Mean values and standard deviations for the anger scores

Pattern (SDF)	Mean [S.D.] All (N=18)	Male (N=10)	Female (N=8)
1 (aaa)	1.91 [0.554]	1.72 [0.527]	2.15 [0.521]
2 (aka)	1.05 [0.716]	0.82 [0.614]	1.37 [0.770]
3 (aak)	0.74 [0.459]	0.54 [0.366]	1.00 [0.454]
4 (akk)	0.49 [0.453]	0.50 [0.501]	0.49 [0.414]
5 (aan)	0.61 [0.379]	0.54 [0.327]	0.70 [0.441]
6 (akn)	0.12 [0.159]	0.12 [0.140]	0.11 [0.195]
7 (kkk)	0.07 [0.168]	0.10 [0.216]	0.03 [0.071]
8 (kak)	0.20 [0.257]	0.26 [0.284]	0.13 [0.212]
9 (kka)	0.39 [0.487]	0.33 [0.424]	0.45 [0.573]
10 (kaa)	1.13 [0.557]	1.00 [0.499]	1.30 [0.614]
11 (kkn)	0.04 [0.189]	0.08 [0.253]	0.00 [0.000]
12 (kan)	0.02 [0.094]	0.00 [0.000]	0.05 [0.141]
Mean S.D.	[0.373]	[0.367]	[0.346]

where
(SDF) is S: STYLE, D: DUR, F: F0,
a: angry, k: kind, and n: normal.

Table 4.3 Mean values and standard deviations for the kindness scores

Pattern (SDF)	Mean [S.D.] All (N=18)	Male (N=10)	Female (N=8)
1 (aaa)	0.03 [0.077]	0.02 [0.063]	0.05 [0.093]
2 (aka)	0.09 [0.189]	0.04 [0.127]	0.17 [0.243]
3 (aak)	0.16 [0.243]	0.16 [0.227]	0.15 [0.278]
4 (akk)	0.31 [0.407]	0.22 [0.274]	0.43 [0.547]
5 (aan)	0.19 [0.211]	0.18 [0.148]	0.20 [0.283]
6 (akn)	0.39 [0.541]	0.34 [0.582]	0.46 [0.513]
7 (kkk)	1.66 [0.716]	1.30 [0.424]	2.10 [0.778]
8 (kak)	0.69 [0.603]	0.58 [0.503]	0.83 [0.721]
9 (kka)	0.49 [0.448]	0.33 [0.346]	0.68 [0.501]
10 (kaa)	0.09 [0.184]	0.10 [0.216]	0.08 [0.149]
11 (kkn)	1.56 [0.787]	1.32 [0.801]	1.85 [0.707]
12 (kan)	0.67 [0.686]	0.64 [0.776]	0.70 [0.605]
Mean S.D.	[0.424]	[0.374]	[0.452]

where
(SDF) is S: STYLE, D: DUR, F: F0,
a: angry, k: kind, and n: normal.

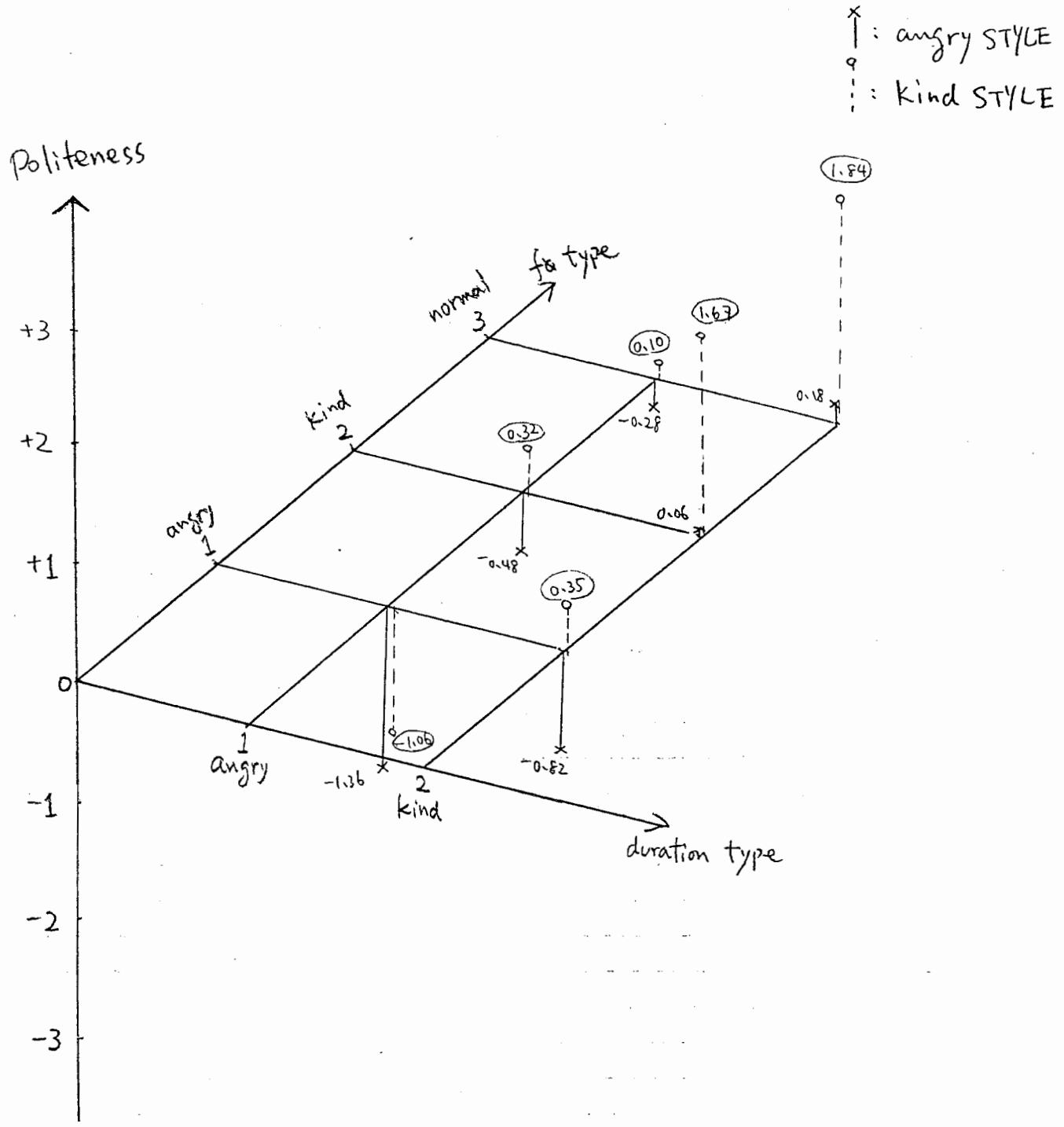


Figure 4.1. The mean values on the scale of politeness

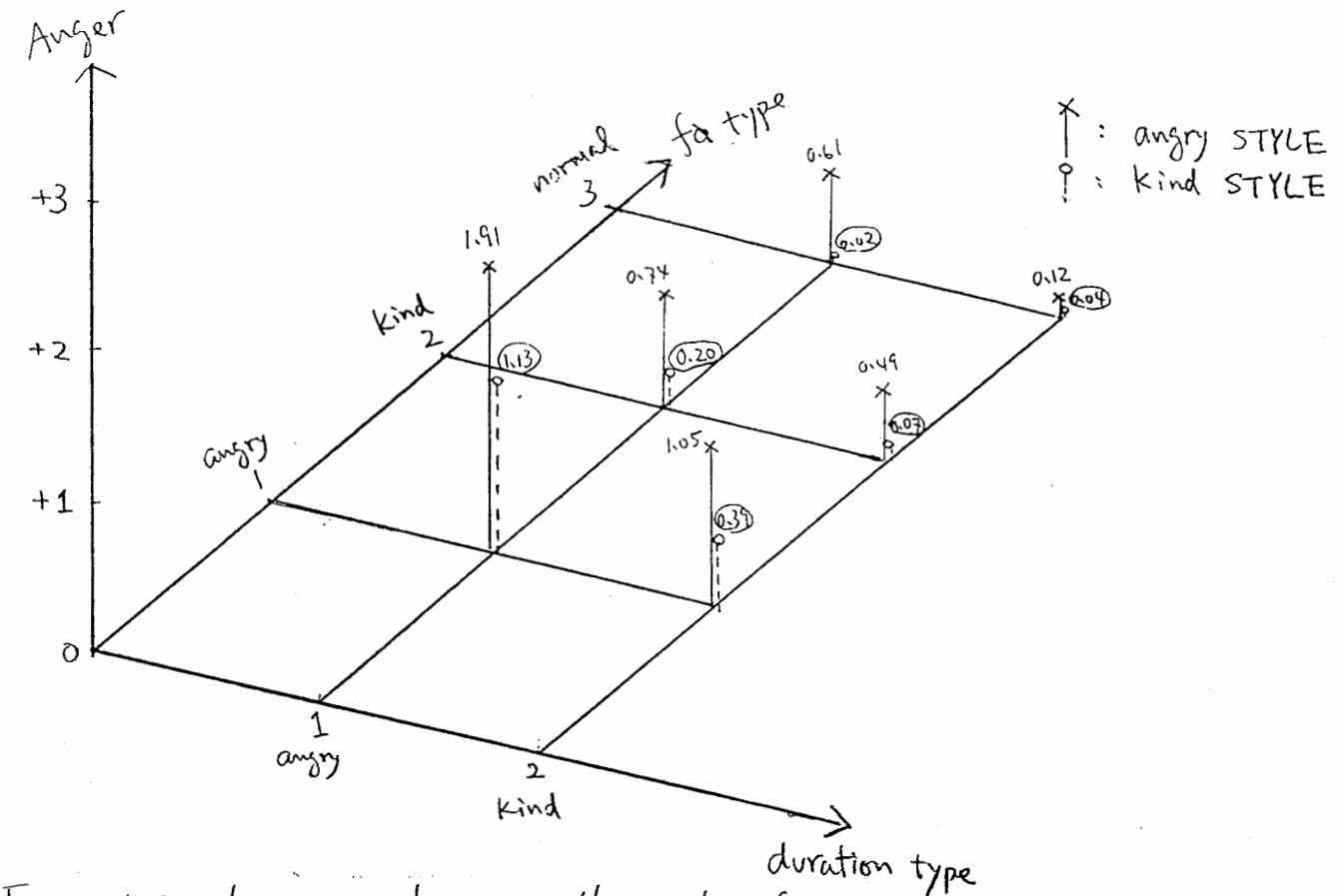


Figure 4.2 The mean values on the scale of anger

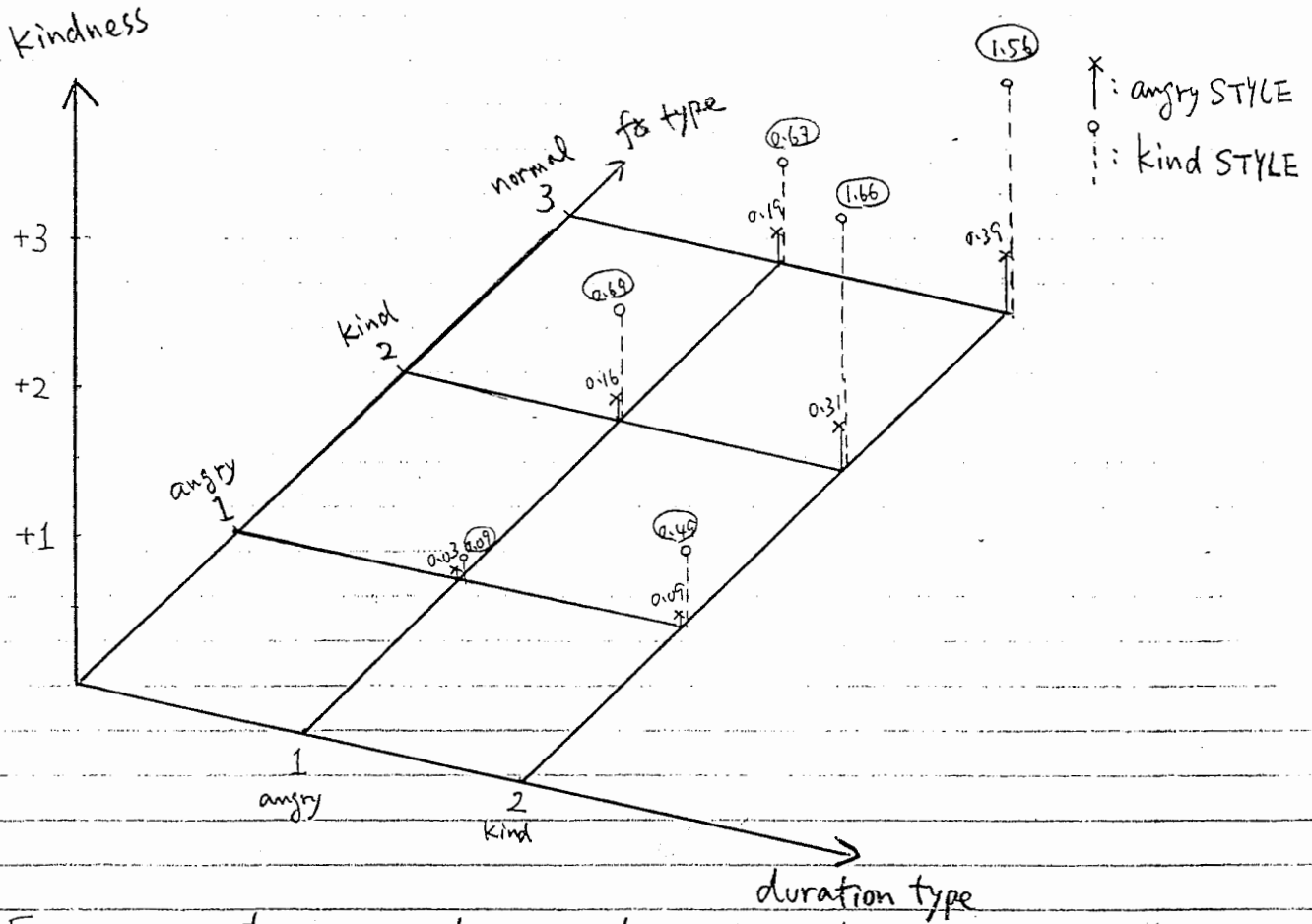


Figure 4.3 The mean values on the scale of kindness

The Figures show that F0 and DUR seem to be quite powerful cues, especially in judgement of anger and politeness. Although the kind and normal F0 types have similar effects on the judgements, the utterance (STYLE: kind, DUR: kind and F0: kind) was rated as the kindest while the utterance (STYLE: kind, DUR: kind and F0: normal) was rated as the politest. In this sense, there is a difference between the F0 types depending on the types of affect.

4.2 Correlation between the scales

A significant positive correlation was found between politeness and kindness, while high negative correlation between politeness or kindness and anger, as expected (Table 4.4).

Table 4.4 The correlation between politeness, kindness and anger (N = 12)

Politeness - Kindness:	0.95 (**)
Politeness - Anger:	-0.83 (**)
Kindness - Anger :	-0.70 (*)

where *: 1-tailed significance level = 0.01
**: 0.001

4.3 Agreement among judges

4.3.1 Kendall's Coefficient of Concordance (W)

Kendall's coefficient of concordance (W) was calculated to assess the agreement among k judges' rankings of N patterns. Since it is recommended to convert W to the average Spearman rank-order correlation (ave(rs)) for interpretation (e.g., Howell, 1992, p. 281), ave(rs) is also shown in Table 4.5.

The results show that quite high agreement among the judges on the scales of anger and politeness, and fairly high agreement on the kindness scale. Anger seems to be the easiest to judge, while kindness the most difficult.

Table 4.5 Kendall's Coefficient of Concordance (W) and the mean Spearman correlation (rs) for politeness, kindness, and anger

Scale	W	ave(rs)
Politeness	0.6875 (****)	0.669
Kindness	0.5168 (****)	0.488
Anger	0.7427 (****)	0.728

where
****: significance level < 0.0005,
W = (sum (i=1 to N) ((ave(Ri)-ave(R))squared) divided by (N(NxN-1)/12)
ave(rs) = (kW-1) / (k-1) (k is the number of judges)
where k = number of judges
N = number of objects being ranked
ave(Ri) = average of the ranks assigned to the i-th object
ave(R) = average of the ranks assigned across all objects
N(NxN-1)/12 = maximum possible sum of the squared deviations

Table 4.6 Mean rank for each pattern on the politeness, kindness, and anger scales

Pattern	(STYLE DUR F0)	Mean Rank (1 to 12) Politeness	Kindness	Anger
1	(aaa)	1.57	3.60	11.83
2	(aka)	3.50	4.00	9.53
3	(aak)	4.87	4.73	7.93
4	(akk)	6.87	6.60	6.67
5	(aan)	5.70	5.00	7.87
6	(akn)	7.67	6.27	3.93
7	(kkk)	11.10	11.03	3.53
8	(kak)	7.77	7.90	4.97
9	(kka)	7.87	7.17	5.60
10	(kaa)	3.03	4.07	10.03
11	(kkn)	11.23	10.53	3.27
12	(kan)	6.83	7.10	2.83

where
a: angry, k: kind, and n: normal

4.3.2 Inconsistency in politeness judgement and the order effect

The five scores for the same pattern by the same judge in politeness judgement, the difference of which (between the highest and the lowest) is greater than three, are shown in Table 4.7. Some of them, but not all, can be explained by the order effect. For example, subject M8C rated the pattern 1 stimuli, -2, -2, -1, +1, and -2 (see Table 4.7), and the forth score (+1), which is the highest among the five scores, was preceded by the stimulus (Pattern 10, the mean politeness score by the subject: -1.2) which sounded quite impolite while other stimuli preceded by the other four scores sounded rather polite, compared to pattern 1. The sets of inconsistent scores, which might be explained by the order effect was shown by '*' before the judge-id in the table.

Table 4.7 Inconsistency scores in politeness judgement and the presentation order
(Five scores for the same pattern by the same judge, the difference of which is greater than 3)

P	Judge	previous pattern [mean for the pattern]/ score				
1	*M8C	7[2.2]/-2	11[2.2]/-2	2[0.9]/-1	10[-1.2]/1	8[0.6]/-2
	*F3C	7[2.8]/-1	11[2.8]/-1	2[-1.8]/2	10[-1.6]/1	8[1.4]/0
2	M2B	9[0.0]/-1	9[0.0]/1	4[0.4]/0	10[-0.8]/-1	9[0.0]/-2
	M8C	10[-1.2]/2	9[1.4]/2	7[2.2]/1	1[-1.2]/1	10[-1.2]/-2
	F1A	4[1.0]/-3	6[1.2]/-1	10[-1.0]/1	8[0.0]/1	4[1.0]/1
3	M8C	4[1.2]/0	6[1.4]/-1	2[0.8]/1	7[2.2]/-1	3[-0.6]/-2
	F2B	4[-0.2]/-1	8[-0.4]/-1	4[-0.2]/1	6[0.0]/-2	12[-0.2]/-1
4	M5E	15[1.5]/-1	12[-1.2]/1	8[-1.0]/-2	10[-1.0]/-1	2[-1.2]/-1
	*F3C	8[1.4]/1	8[1.4]/-1	1[0.2]/-1	2[-1.8]/2	3[0.4]/-1
5	M2B	9[0.0]/-1	7[1.2]/-1	12[-0.2]/1	6[-0.6]/-2	2[-0.6]/0
	F6B	9[0.0]/1	7[2.0]/1	12[1.2]/-1	6[?]/-2	2[-1.5]/1
	M5E	6[-0.6]/-1	1[-2.2]/-1	11[2.0]/-1	2[-1.2]/-2	10[-1.0]/1
6	M2B	1[-0.2]/1	7[1.2]/0	12[-0.2]/-2	12[-0.2]/-1	7[1.2]/-1
	F3C	8[1.4]/1	10[-1.6]/1	9[0.8]/-1	3[0.4]/-2	6[-0.2]/0
7	M3C	6[-0.5]/-2	2[-1.4]/1	6[-0.5]/1	10[-1.2]/1	8[-0.2]/-1
8	M3C	15[0.5]/-3	1[-2.0]/1	5[-0.2]/1	11[2.2]/-1	1[-2.0]/1
	M8C	15[0.0]/0	1[-1.2]/1	5[-0.2]/1	11[2.2]/-1	1[-1.2]/2
	F1A	10[-1.0]/-3	5[-1.6]/1	10[-1.0]/1	8[0.0]/1	6[1.2]/0
	F8D	2[-1.6]/0	2[-1.6]/-1	7[2.6]/-2	6[0.2]/1	5[-0.2]/0
9	*F3C	1[0.2]/2	4[0.0]/0	9[0.8]/-1	12[0.2]/1	12[0.2]/2
10	*M1A	7[1.6]/-2	1[-1.4]/-1	3[0.0]/1	6[0.6]/-1	3[0.0]/0
	*F1A	7[2.6]/-3	1[-2.0]/1	3[-1.0]/-1	6[1.2]/-1	3[-1.0]/-1
12	M8C	4[1.2]/-2	7[2.2]/-2	4[1.2]/-1	7[2.2]/1	9[1.4]/1
	F1A	11[2.8]/-1	9[2.2]/1	8[0.0]/-2	3[-1.0]/1	3[-1.0]/-1
	*F2B	15[-0.5]/0	11[2.2]/-1	7[2.4]/-1	10[-2.0]/2	6[0.0]/-1

where P: pattern,
*: the order effect explains the inconsistency

4.4 Unnaturalness

The angry STYLE with the kind DUR were rated unnatural (Table 4.8), maybe because the big difference between the speaking rates. No systematic relation between unnaturalness and scores (i.e., unnaturalness increased/decreased scores) was found.

Table 4.8 Unnaturalness

Pattern (SDF)	Unnaturalness count (*1) subject # = 18 (10M, 8F)
1 (aaa)	0
2 (aka)	8 (5M, 3F)
3 (aak)	0
4 (akk)	9 (2M, 6F)
5 (aan)	0
6 (akn)	9 (4M, 5F)
7 (kkk)	0
8 (kak)	0
9 (kka)	4 (3M, 1F)
10 (kaa)	0
11 (kkn)	1 (1M)
12 (kan)	2 (2M)

where
*1: the number of judges who rated the stimulus unnatural, more than three times out of 5 occurrences
SDF: STYLE, DUR, and F0,
a: angry, k: kind, n: normal,
M: male subject, F: female subject

4.5 Analysis of variance (ANOVA) with repeated measures

The mean values of five scores for each pattern rated by each judge was used as input data for the ANOVA tests. Less than three scores for each pattern per judge were treated as missing values. Since the kind and normal F0 types seem to have a similar tendency on all the scales studied in terms of the mean values, the two analyses, one using only two F0 types (angry and kind) and the other using three F0 types (angry, kind and normal), were done.

Tables 4.9, 4.10, 4.11 show that

(a) the main effects of STYLE, DUR, and F0 are more significant than the interactions

This could mean that each factor contributes to perception rather independently.

(b) no interaction between DUR and F0 was found on the scale of politeness, while it was found on the anger and kindness scales

It could mean that DUR and F0 work independently for perception of politeness

(c) the effect of SEX (the sex of the judges) was found only on the scale of kindness

(d) the relative weight of contribution by using eta-squared, i.e., SS(factor) divided by SS(grand total), as a rough indicator is:

Politeness --> F0 > DUR > STYLE

Anger --> F0 > STYLE > DUR

Kindness --> STYLE > F0 > DUR

Although the eta-squared may not be the best indicator, it was used because it reflects how much of the overall variability can be attributed to each factor effect.

(e) the comparison between the two analyses, 2F0 and 3F0, shows that the effect of DUR increased in politeness judgements (e.g., SS(2F0)[DUR] vs SS(2F0)[F0] = 28.32 : 34.32, while SS(3F0)[DUR] vs SS(3F0)[F0] = 48.39 : 48.99), while the relative importance of F0 seems to be increased in anger judgements

In the case of politeness, one explanation might be that F0

is the most important cue and when it becomes confusing due to the variability of F0 types, the second important cue, i.e., duration in this case, comes to play an important role. In the case of anger judgement, since the added f0 type (the normal type) is quite different from the angry type, which is 'flat', including the normal F0 type may not have made the judging task difficult.

Table 4.9 Significant effects at the level of 0.01 on the politeness scale

Factor	STYLE	DUR	F0	STYLE by DUR	STYLE by F0
SS(2f0)	27.53	28.32	34.32	5.07	1.76
F	119.01	27.50	66.40	10.44	12.59
Sig. F	***	***	***	0.007	0.004
SS(3f0)	40.14	48.39	48.99	10.60	(1.78)
F	63.86	32.17	53.20	16.89	(4.80)
Sig. F	***	***	***	0.001	(0.017)

where *** means less than 0.001,

SS(2f0) is sums of squares (SS) obtained by the analysis using 2 types of STYLE (angry, kind), 2 types of DUR (angry, kind), and 2 types of F0 (angry, kind) as within-subjects factors, and 2 types of SEX (male, female) of judges as a between-subjects factor, and

(3f0) is the analysis using 2 STYLEs, 2 DURs, and 3 F0s (angry, kind, and normal) as within-subjects factors, and 2 SEXs as between-subjects factor.

Table 4.10 Significant effects at the level of 0.01 on the scale of anger

Factor	STYLE	DUR	F0	DUR by F0	STYLE by F0	STYLE by DUR
SS(2f0)	11.00	7.65	17.36	2.43	0.31	-
F	71.32	33.52	50.62	20.59	12.16	
Sig. F	***	***	***	0.001	0.004	-
SS(3f0)	12.46	7.91	30.52	3.09	0.88	0.88
F	76.14	53.59	55.75	12.72	11.21	15.37
Sig. F	***	***	***	***	***	0.002

where ***, SS(2f0), SS(3f0) are the same as in Table 4.9.

Table 4.11 Significant effects at the level of 0.01 on the scale of kindness

Factor	STYLE	DUR	F0	STYLE by DUR	STYLE by F0	DUR by F0	SEX by DUR
SS(2f0)	9.66	5.10	7.65	2.29	3.07	0.66	1.01
F	38.01	55.99	26.01	18.43	32.46	14.26	11.12
Sig. F	***	***	***	0.001	***	0.002	0.005
SS(3f0)	17.89	9.75	9.53	3.58	3.43	0.91	(1.68)
F	35.59	48.93	18.46	9.87	11.76	6.13	(8.44)
Sig. F	***	***	***	0.008	***	0.007	(0.012)

where ***, SS(2f0), SS(3f0) are the same as in Table 4.9.

4.6 Quantification method type 1

Scores from 19 subjects were used for this analysis. Partial correlation coefficient and multiple correlation coefficient in Quantification method type 1 were calculated (Table 4.12), Although the fitting of the model is not very satisfactory (the multiple correlation coefficients are around 0.6).

Table 4.12 Results of Quantification method type 1

	Partial correlation coefficient		
	Politeness	Anger	Kindness
STYLE	0.36	0.38	0.44
DUR	0.38	0.30	0.34
F0	0.41	0.55	0.37
Multiple correlation coefficient	0.59	0.65	0.59

5. Conclusions

The mean values show that F0 combined with DUR seem to be a quite powerful cue on all three scales. Appropriate levels of F0 and DUR can make utterances sound polite or angry or kind regardless of STYLE, while only one type of cue is not powerful enough to override the effect of voice sources.

Although the kind and normal types of F0 have a similar effect on judgement of politeness and kindness, the kind F0 was rated as the kindest and the normal F0 the politest.

The consistency among 18 judges' scores was quite high. Anger and politeness judgement were more consistent than kindness judgement. Politeness is a complex concept, but it may be easier to judge from speech than kindness, because it is to be learnt in the society, although people usually do not study what sound

polite or impolite in a form of specific rules.

The ANOVA tests show the following points:

(a) the main effects of STYLE, DUR, and F0 are more significant than the interactions

Each factor seems to contribute to perception of politeness, anger, and kindness rather independently, which agrees with the findings of Ladd et al. (1985).

(b) no interaction between DUR and F0 was found on the scale of politeness, while it was found on the anger and kindness scales

It could mean that DUR and F0 work independently for perception of politeness.

(c) no significant difference between male and female judges was found on the three scales studied, especially on the scales of politeness and anger

(d) the relative weight of contribution by using eta-squared as a rough indicator is:

Politeness --> F0 > DUR > STYLE

Anger --> F0 > STYLE > DUR

Kindness --> STYLE > F0 > DUR

It agrees with the findings of Ladd et al. (1985), in which they studied voice quality, F0 range and F0 contour in relation to several types of affect, and they suggested that F0 range seems to be related to arousal, while voice quality is related to the speaker's positive-negative evaluation of the interlocutor or semantic content. However, this disagrees with the findings of Brown et al. (1974), Scherer & Oshinsky (1977), and Kitahara & Tohkura (1988), etc., in which

duration was found to be more influential than f0 for some attitudinal meanings including anger and benevolence axes (kindness, politeness, etc.) It may be because of the selection of types for the duration and f0 factors. Further experimentation is needed.

(e) the comparison between the two analyses, 2F0 and 3F0, might show that people use a hierarchy of cues, starting with the easiest or the most powerful cue(s), and only use the other cue(s) when the first cue becomes confusing. Further analysis of utterances by different speakers is necessary to confirm this point.

6. Limitations of the study

In this experiment, a single text spoken by a single speaker was used, which is obviously the main limitation of the experiment. In fact, Ladd et al. (1985) found strong main effects for the factors of 'speaker' and 'text' (the verbal content) on almost all judgement scales they used (e.g., arousal, annoyance, control, etc.), although they reported that there were very few sizeable interactions between the speaker or text factors and such vocal variables as voice quality, f0 contour and f0 range.

Since the subjects had to rate the utterances which had the same verbal content spoken by a single speaker 64 times in a 30 minute session, it is very likely that they focus on the prosodic aspects more than they usually do in more natural settings, and consequently, minor differences in prosody might be unnecessarily magnified.

Another problem is the great variability of speakers' ability

in expressing attitudinal meanings (e.g., Pakosz, 1983, p. 316). The relative importance of each acoustic variable depends on the selection of levels or types of the factors studied. For example, if the experimenter selected two types from factor A, which happen to belong the same category, and two types from factor B, which belong to different groups in terms of perception of some affect, it is very likely for subjects to respond with factor B, although whether perception of affect is categorical or not is debatable (e.g., Ladd et al., 1985).

Rosenthal et al. (1979: 21) states that the available literature on nonverbal communication suggests that "good decoders are likely to be more accurate than poor decoders across different encoders", and the speaker used in this experiment may not be so idiosyncratic. However, one can never be certain to what extent the findings can be generalised unless analysis of more data from different speakers has been done.

7. Acknowledgement

We would like to thank Dr. Yoshinori Sagisaka and all members of Department 2, ITL/ATR for providing us an opportunity to do this experiment at ATR. We especially thank Dr. Norio Higuchi and Katsuhiko Mimura for their useful comments and advice on designing listening tests and analysing data, and Dr. Helene Valbret for her technical support by providing programs based on a PSOLA technique and useful discussions. We also thank Dr. Minoru Tsuzaki and Hiroaki Kato of HIP/ATR for their advice on designing perception tests, and Professor Peter Roach and Mitch Waterman of Leeds University for their advice, discussions and

comments. We finally thank all the subjects who participated in the listening tests in which they had to concentrate on 64 similar-sounding stimuli for half an hour.

References

Bas, M. J. van der (1992). "The recognizability of emotions in British English diphones and natural speech." Master's Thesis, Department of Linguistics/Phonetic Laboratory, Leyden University, The Netherlands.

Brown, B. L., Strong, N. J., and Rencher, A. C. (1974). Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech, Journal of the Acoustical Society of America, 55, 313-318.

Crystal, D. (1969). Prosodic Systems and Intonation in English, Cambridge: Cambridge University Press.

Frick, R. W. (1985). Communicating Emotion: The Role of Prosodic Feature, Psychological Bulletin, Vol. 97, No. 3, pp. 412-429.

Howell, D. C. (1992). Statistical methods for psychology. (Third edition) Belmont, California: Duxbury Press.

Kitahara, Y. and Tohkura, Y. (1988). Prosodic Components of Speech in the Expression of Emotions, (in Japanese), Proceedings of the autumn convention of Acoustical Society of Japan, 1-2-15, pp. 131-132.

Kramer, E. (1963). Judgement of personal characteristics and emotions from nonverbal properties of speech, Psychological Bulletin, Vol. 60, No. 4, pp. 408-420.

Ladd, D. R.; Silverman, K. E. A.; Tolkmitt, F.; Bergmann, G.; and Scherer, K.R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect, Journal of the Acoustical Society of America, 78 (2), 435-444.

Miyatake, M. and Sagisaka, Y. (1990). Prosodic Characteristics and Their Control in Japanese Speech with Various Speaking Styles (in Japanese), The Institute of Electronics, Information, and Communication Engineering Japan, Vol. J73-D-II, No. 12, pp. 1929-1935.

Moulines, E. and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones, Speech Communication, Vol. 9, Nos 5/6, 453-467.

Murray, I. R. and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion, Journal of the Acoustical Society of America, 98 (2), 1097-1108.

Pacosz, M. (1983). Attitudinal Judgments in Intonation: Some Evidence for a Theory, Journal of Psycholinguistic Research, Vol. 12, No. 3, pp. 311-326.

Rosenthal, R. (1982). "Conducting judgment studies." In Handbook of Methods in Nonverbal Behavior Research, pp. 287-361. Edited by Klaus R. Scherer and Paul Ekman. Cambridge, etc. : Cambridge University Press.

Rosenthal, R.; Hall, J. A.; DiMatteo, M. R.; Rogers, P. L.; and Archer, D. (1979). Sensitivity to Nonverbal Communication: The PONS Test. Baltimore and London: The Johns Hopkins University Press.

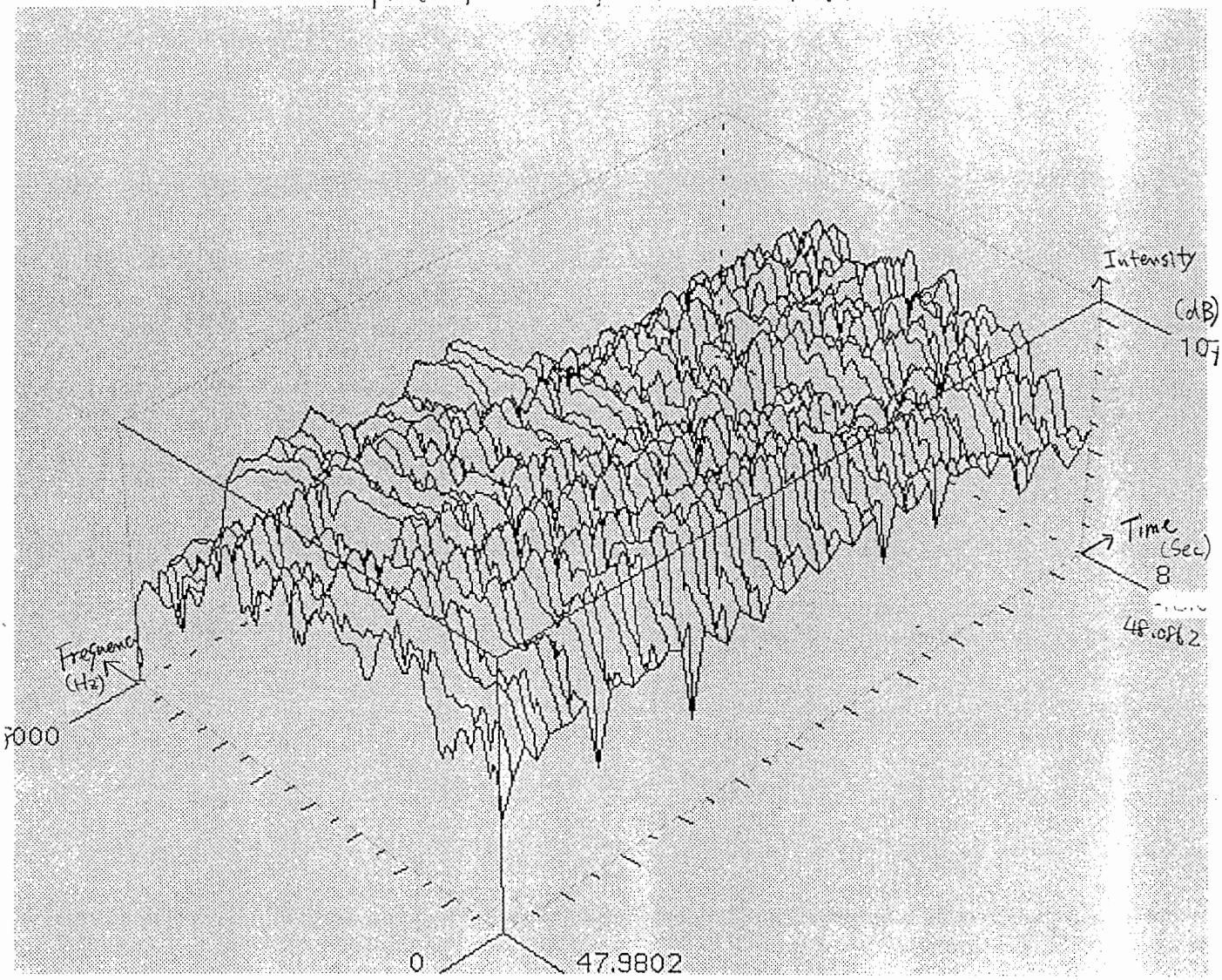
Scherer, K. R. (1979a). "Personality markers in speech." In Social markers in speech, pp. 147-209. Edited by K. R. Scherer and H. Giles. Cambridge, etc.: Cambridge University Press.

---------. (1979b). "Nonlinguistic Vocal Indicators of Emotion and Psychopathology." In Emotions in Personality and Psychopathology, pp. 495-529. Edited by C. E. Izard. New York and London: Plenum Press.

---------. (1982). "Methods of research on vocal communication: paradigms and parameters." In Handbook of Methods in Nonverbal Behavior Research, pp. 136-198. Edited by Klaus R. Scherer and Paul Ekman. Cambridge, etc. : Cambridge University Press.

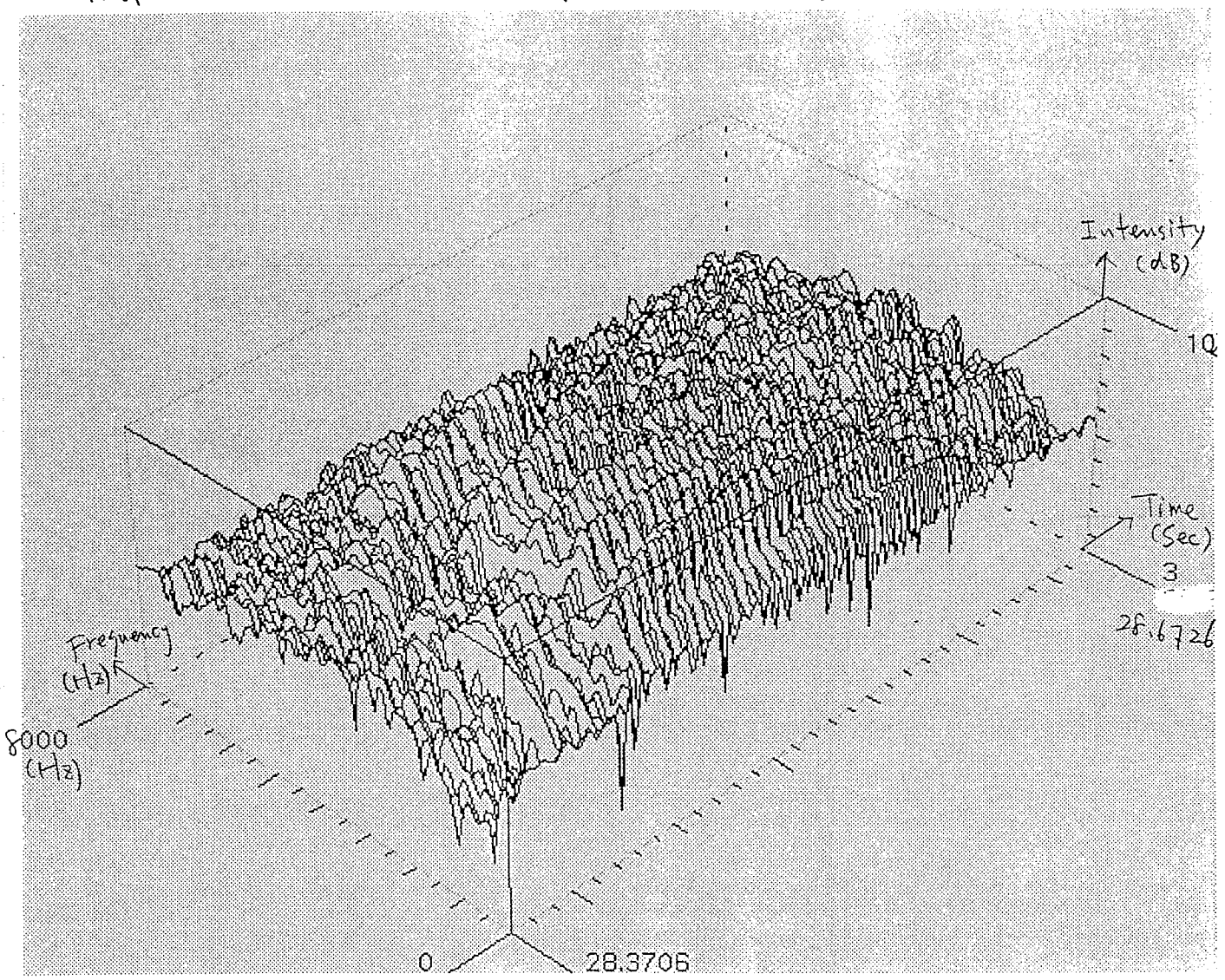
Scherer, K.R., and Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli, Motivation and Emotion, Vol. 1, No. 4, pp. 331-346.

angry 3D plot for the final mora ('ka')



Kind

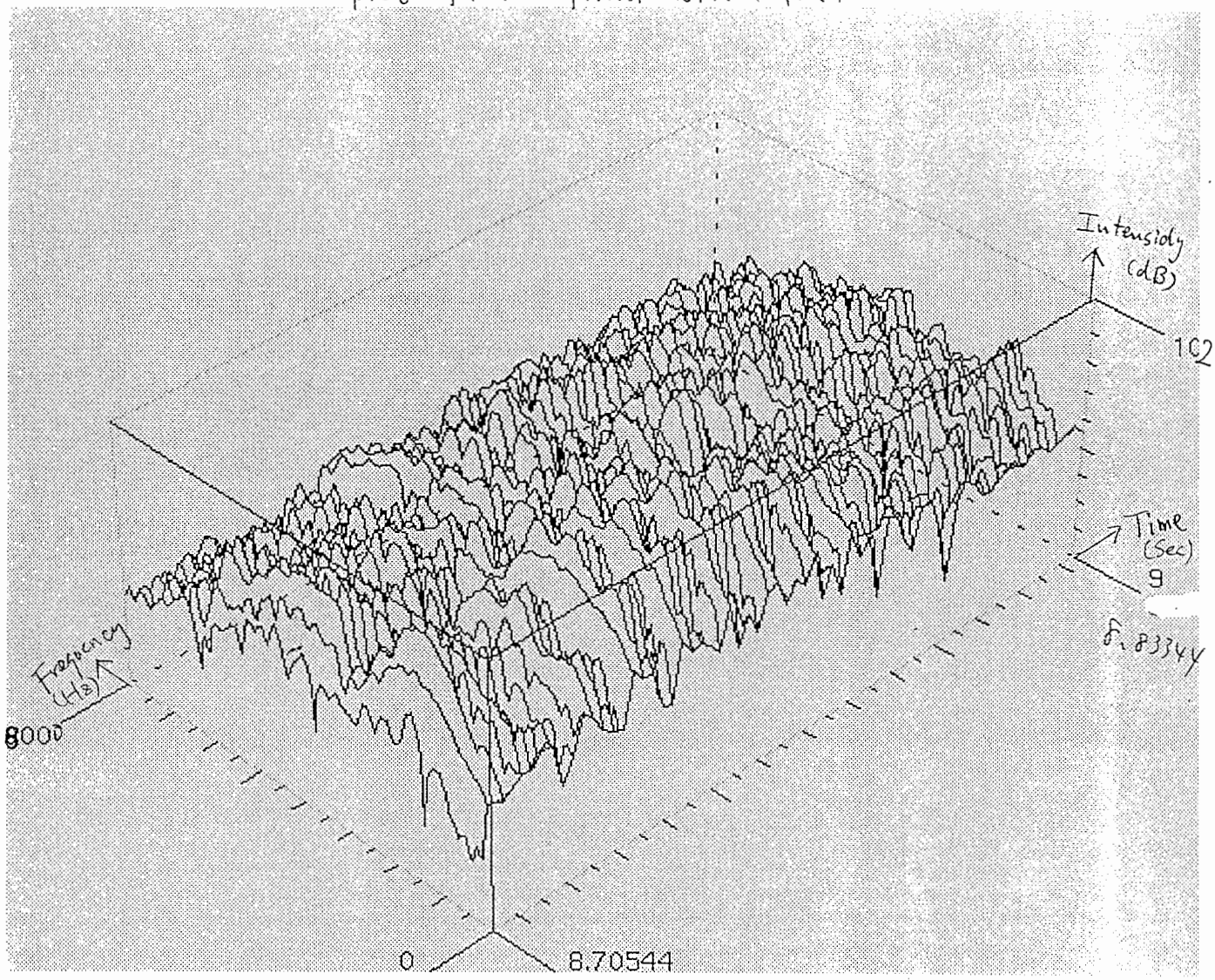
3D plot for the final mora ('ka')



Appendix 1-2

normal

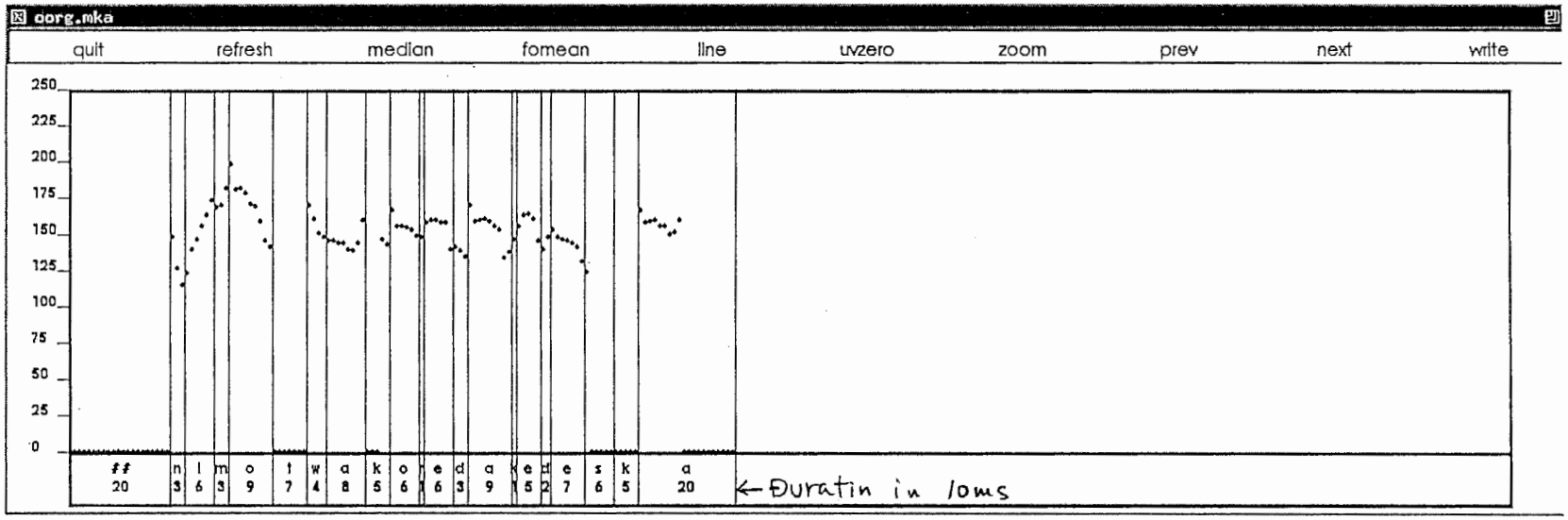
3D plot for the final mora ('ka')



Appendix 1-3

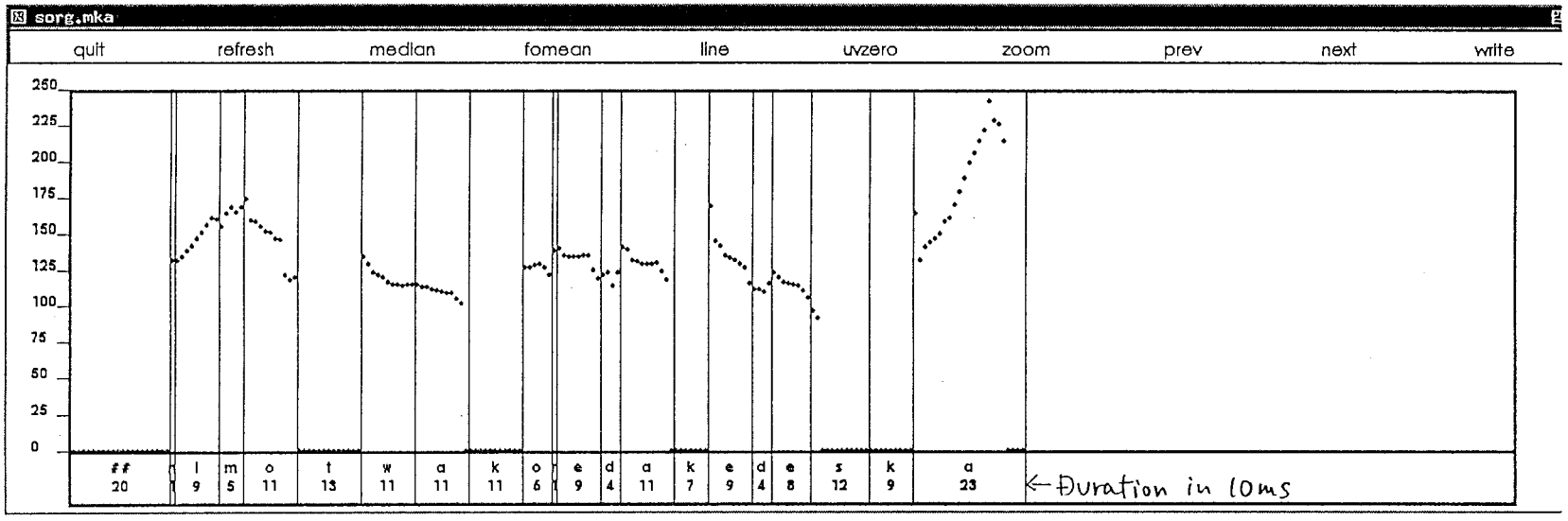
Appendix 2-1

F0 contour of the angry utterance

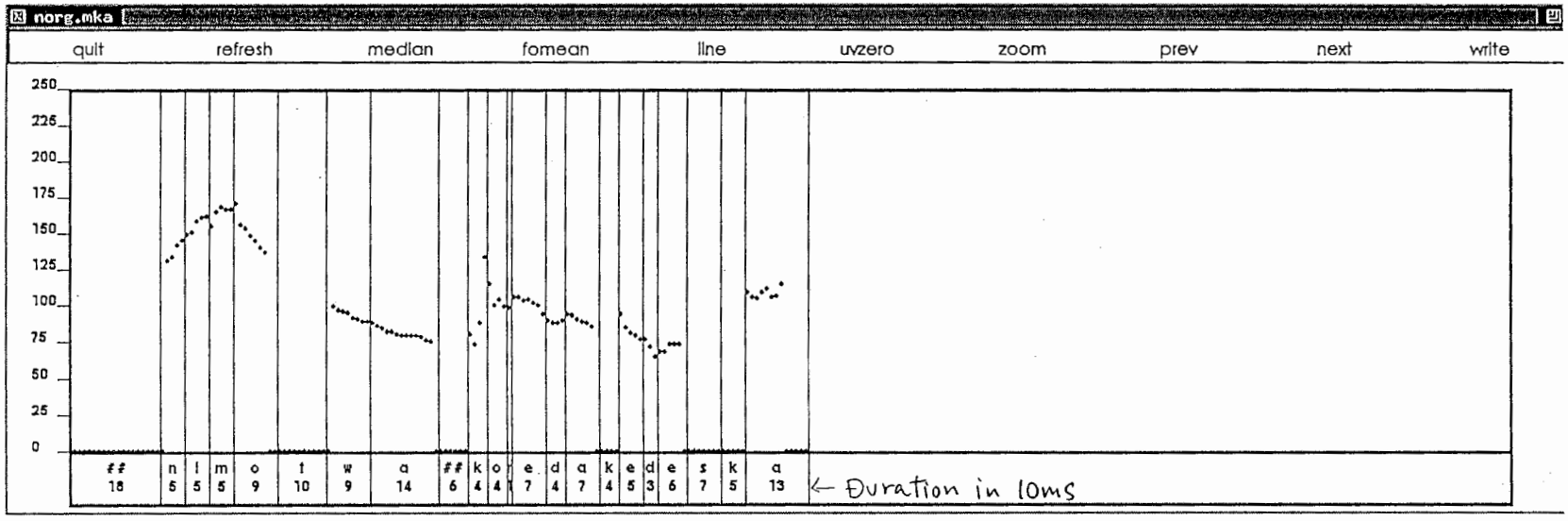


Appendix 2-2

F0 Contour of the kind utterance



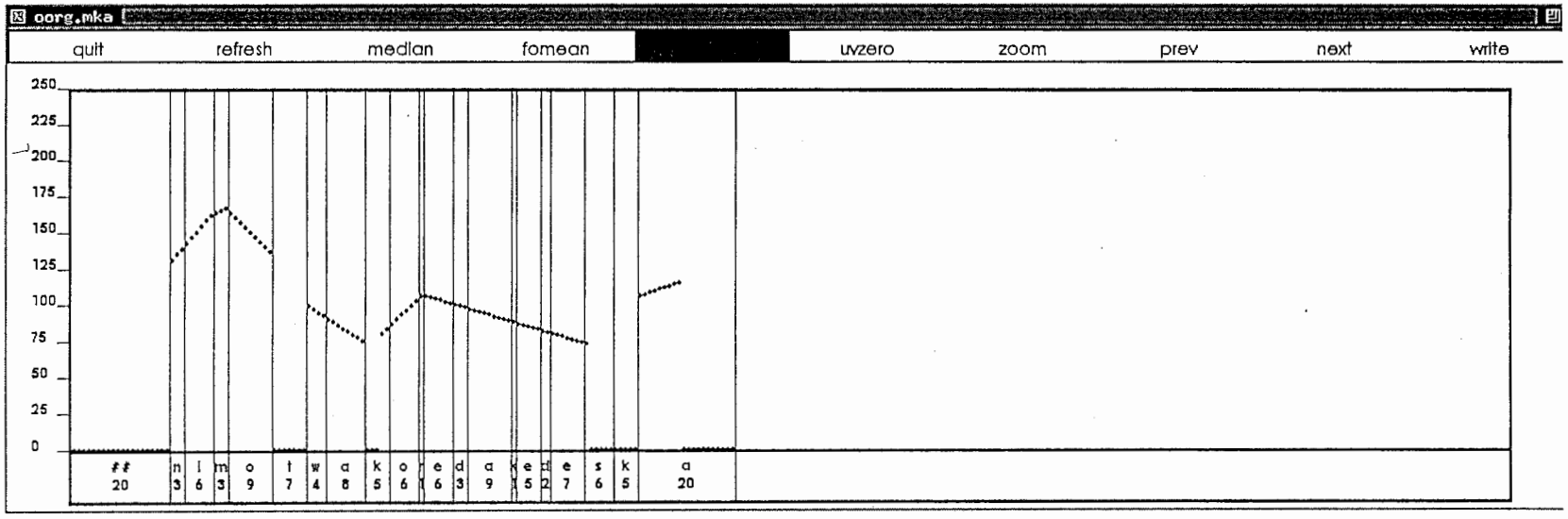
F0 Contour of the normal utterance



Appendix 2-3

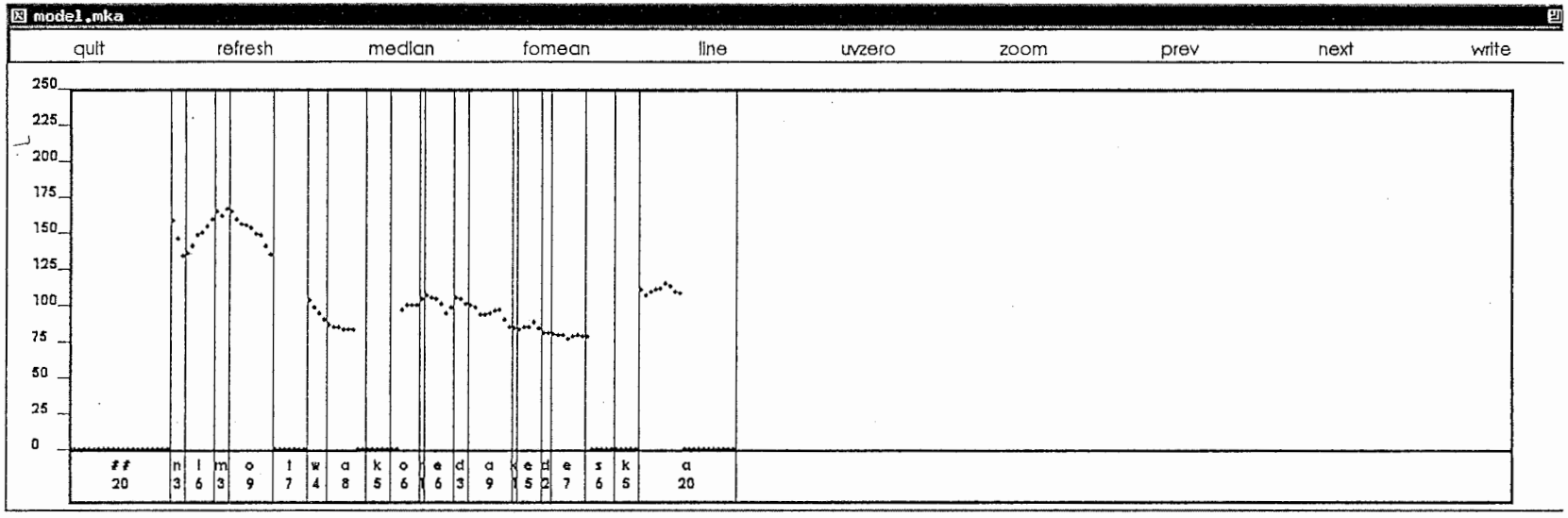
Appendix 3-1

Imposition of the normal Fo type on the angry utterance



Appendix 3-2

F0 Contour of the angry utterance with the normal F0 (Result of Appendix 3-1)



Rating form

(very impolite) 非常にない　　(very polite) 非常にある

□ 丁寧である (polite)　-3 -2 -1 0 1 2 3

□ 怒っている (angry)　0 1 2 3

□ 親切である (kind)　0 1 2 3

□ どことなく言い方が自然でない (unnatural)

非常にない　　非常にある

□ 丁寧である　-3 -2 -1 0 1 2 3

□ 怒っている　0 1 2 3

□ 親切である　0 1 2 3

□ どことなく言い方が自然でない

非常にない　　非常にある

□ 丁寧である　-3 -2 -1 0 1 2 3

□ 怒っている　0 1 2 3

□ 親切である　0 1 2 3

□ どことなく言い方が自然でない

非常にない　　非常にある

□ 丁寧である　-3 -2 -1 0 1 2 3

□ 怒っている　0 1 2 3

□ 親切である　0 1 2 3

□ どことなく言い方が自然でない