THE ROLE OF F0 AND DURATION IN SIGNALLING AFFECT IN JAPANESE: ANGER, KINDNESS AND POLITENESS

Etsuko Ofuka[*a)], Helène Valbret[*b)], Mitch Waterman[*a)],
Nick Campbell[*b)] and Peter Roach[*a)]

*a) Speech Laboratory, Department of Psychology,
University of Leeds, Leeds, LS2 9JT, ENGLAND
*b) ATR Interpreting Tele-communications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, JAPAN

ABSTRACT

This paper describes a study which is concerned with the effects of changing fundamental frequency (f0) and segmental duration, examining their interaction with different speaking styles ("angry" and "kind"). It was found that whereas appropriate settings of both duration and f0 can change the perception of these affects in the resultant speech, neither cue alone is strong enough to override the effect of the original voice source. The relative importance of these prosodic cues varied depending on the kind of affect studied.

1. INTRODUCTION

The importance of studying the communicative function of prosody has increased recently from the point of views of foreign language education and man-machine interface design using synthetic speech. Current speech synthesis technology has come to attain reasonably intelligible output for dialogue systems (e.g., telephone inquiry systems). However, synthetic speech often sounds machine-like, which gives listeners difficulty in concentrating on the speech. It has, therefore, become increasingly important to add a more human-like flavour to the synthetic speech.

Various acoustic variables have been studied in relation to signalling affect. The variables include duration, fundamental frequency (f0), amplitude, and voice quality (see for example, [1][2]). Among them, f0 and duration have been most commonly studied, because they are major acoustic correlates of pitch and tempo, robust in the sense that they survive even in very noisy environments, and rather easy to measure and manipulate compared to the other features.

This study focuses on these features, f0 and duration, in relation to three different kinds of affect in Japanese: anger, kindness and politeness. Anger is an emotion, and said to be closely related to physical arousal, while kindness and politeness are not usually categorised as emotions, but certainly a kind of attitude towards the addressee. Politeness appears to be more culture-specific than kindness, and is very important in smooth social interaction.

An experiment was conducted to examine how these acoustic factors, f0 and duration, and speaking styles work in signalling affect, the relative importance of the factors, and the difference between judgements of these affects.

2. METHOD

2.1 Design

A factorial 2 x 2 x 3 x 2 design was used with two types of speaking STYLE (angry and kind) , two types of duration (DUR) (angry and kind), and three types of F0 contours (F0) (angry, kind and neutral) as within-subjects factors, and two sexes of the subjects (male and female) as between-subjects factors.

2.2 Speech material

A single Japanese sentence was chosen from a dialogue between a customs officer talking to a passenger:

Nimotsu-wa koredake desuka

"Is this all the luggage you have?"

A trained male speaker was instructed to speak this utterance in several different speaking styles [3]. Among those, utterances spoken in an angry/irritated way, a kind/considerate way and a neutral way were used in this experiment. Although this sentence is mostly used as a routine question at the customs and people may not listen to the utterance carefully, we sometimes experience that some officers sound kind or polite, while others sound angry or rude.

2.3 Stimulus preparation

Twelve patterns for resynthesis were produced by copying f0 contours (of angry, kind and neutral) and duration (of angry and kind) extracted from the original utterances, onto the original voice source of two different speaking styles (angry and kind). Since the original "kind" utterance sounded unnaturally slow to most subjects who participated in an informal listening test, the duration was linearly compressed by 20% and was used as "kind" duration in this experiment. This utterance sounded quite natural in terms of both quality and speed to the subjects and the experimenter.

The variables duration and f0 were manipulated through digital resynthesis based on a time-domain PSOLA algorithm [4]. The manipulation of the "angry" and "kind" types of f0 and duration was done automatically by creating a mapping table between the original and target values. Imposition of the "neutral" 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.

The styles were produced by one speaker's simulating different types of affect. Table 1 shows some acoustic characteristics of the styles.

2.4 Rating session

Eighteen subjects (10 male, 8 female) participated in the listening test. All of them were native speakers of Japanese in their 20's and 30's, and were members of staff at ATR. They had a practice session consisting of 4 stimuli, including the original utterances. Then they were presented with 64 stimuli consisting of 5 occurrences of the twelve patterns, preceded and followed by two dummy stimuli, in random order, with a 5-minute break halfway. They listened to each stimulus twice preceded by a beep sound and followed by a 6 and 8 second silence respectively through headphones, and rated it on 4-point unipolar scales (0: not angry/kind, to +3: very angry/kind) for anger and kindness, and on a 7-point bipolar scale (-3: very impolite, to +3: very polite) for politeness. They were also asked to judge the naturalness of the stimuli. A session was conducted individually in a small private room and took about 30 minutes.

Table 1 Acoustic characteristics of different styles

	F0 (in Hz)	Duration	
STYLE	range (min-max)	mean	rate (ms/mora)
angry	71 (113-185)	157	100
kind	124 (107-231)	141	120
neutral	95 (74-169)	112	110

3. RESULTS AND DISCUSSION

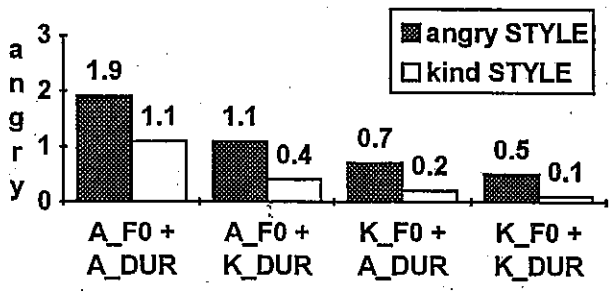
3.1 Unnaturalness

The "angry" style with the "kind" duration was rated unnatural by nearly half of the subjects. No systematic relation between unnaturalness and judgement scores was found.

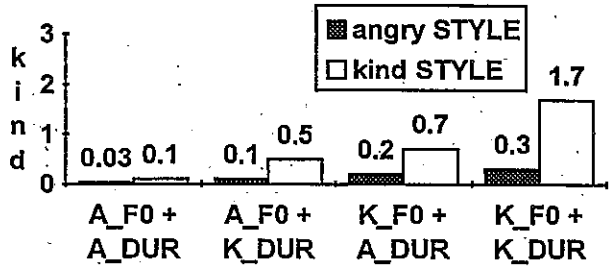
3.2 Mean values

There was no significant difference between the "kind" and "neutral" f0 types. Figure 1 shows that the "kind" style with both "angry" f0 and "angry" duration obtained higher scores for anger, lower scores for kindness, and negative scores for politeness. Similarly the "angry" style with both "kind" f0 and "kind" duration was rated as lower on the anger scale, slightly higher on the kindness scale, and positive on the politeness scale, compared to the "angry" style on which the other combinations of the f0 and duration types are imposed.

(a) Anger



(b) Kindness



(c) Politeness

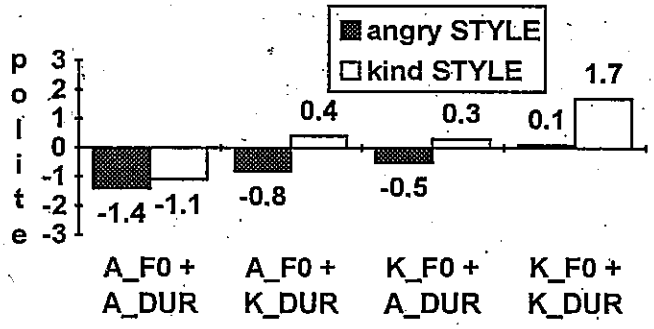


Fig. 1 Mean ratings for the affects by style, f0 and duration: (a) anger, (b) kindness and (c) politeness
Note: A: angry and K: kind

3.3 Agreement among judges

Kendall's coefficient of concordance was calculated to assess the agreement among 18 judges' rankings of the twelve patterns: 0.74 for anger, 0.69 for politeness, and 0.52 for kindness ($p<0.0005$ for all three). A very high level of consistency was found for the judgement of anger and politeness, while that of kindness was found less consistent than the others.

3.4 Analysis of Variance (ANOVA)

The mean values of five scores for each pattern rated by each judge on each scale were used as input for ANOVA. Since no significant mean difference between the "neutral" and "kind" f0 types was found, the results of an analysis using only two f0 types ("angry" and "kind") is reported in Table 2.

Table 2 Significant effects at the level of 0.01 on the scales studied

	Anger	Kindness	Politeness
Main effects	F (eta^2)	F (eta^2)	F (eta^2)
F0	50.6 (0.32)	26.0 (0.18)	66.4 (0.25)
DUR	33.5 (0.14)	56.0 (0.12)	27.5 (0.20)
STYLE	71.3 (0.20)	38.0 (0.23)	119.0 (0.20)
Interactions	F (eta^2)	F (eta^2)	F (eta^2)
F0 x DUR	20.6 (0.04)	14.3 (0.02)	NS
F0 x STYLE	12.2 (0.01)	32.5 (0.07)	12.6 (0.01)
DUR x STYLE	NS	18.4 (0.05)	10.4 (0.04)
SEX x DUR	NS	11.1	...

Note: $df_{effect}=1$, $df_{error\ within\ cells}=13$,
NS: not significant ($p > 0.05$)

Table 2 shows that

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

(b) no interaction between DUR and F0 was found on the politeness scale, but significant interactions were found on anger and kindness scales, though effect size estimates showed these to be less important than those for the main effects,

(c) an effect of the sex of the subjects was found only on the scale of kindness,

(d) the relative weight of contribution by eta-squared, (i.e., the ratio of the sums of squares of the

effect to the sums of squares of the total of within-subject effects), as an indicator, is:

Anger --> (1) F0 > (2) STYLE > (3) DUR,
Kindness --> (1) STYLE > (2) F0 > (3) DUR,
Politeness --> (1) F0 > (2) DUR >= (3) STYLE

This agrees with the findings of Ladd et al. [5], in which they studied voice quality, f0 range and f0 contour in relation to several types of affect, and they suggest 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 result disagrees with some other findings [6][7], in which duration was found to be more influential than f0 for anger and benevolence axes (kindness, politeness, etc.). The inconsistency may be due to the selection of types or values for the durations and f0 factors used in experiments in this area. In fact, the great variability of speakers' encoding ability is reported [8]. Further experimentation using more speakers and texts is necessary to examine the hierarchy of the factors.

4. CONCLUSION

The experiment was conducted to examine how duration, f0 and style function in signalling affect. It was found that appropriate settings of both f0 and duration can change the perception of affect. F0 was found to be the most influential indicator for anger and politeness, and style the most powerful cue for kindness in the utterances used.

The consistency among 18 judges' scores was quite high on the affects studied. Anger and politeness judgement were more consistent than kindness judgement.

ACKNOWLEDGEMENT

We thank Dr. Yoshinori Sagisaka and all members of Department 2, ITL/ATR for useful comments and discussions.

REFERENCES

[1] K. R. Scherer, "Methods of research on vocal communication: paradigms and parameters". In Handbook of Methods in Nonverbal Behavior Research, pp. 136-198. Edited by K. R. Scherer and P. Ekman. Cambridge: Cambridge University Press, 1982.

[2] I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", J. Acoust. Soc. Am., Vol. 93, No. 2, pp. 1097-1108, 1993.

[3] M. Miyatake and Y. Sagisaka, "Prosodic Characteristics and Their Control in Japanese Speech with Various Speaking Styles", (in Japanese), IEICE Japan, Vol. J73-D-II, No. 12, pp. 1929-1935, 1990.

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

[5] D. R. Ladd; K. E. A. Silverman; F. Tolkmitt; G. Bergmann and K. R. Scherer, "Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling affect", J. Acoust. Soc. Am., Vol. 78, No. 2, pp. 435-444, 1985.

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

[7] Y. Kitahara and Y. Tohkura, "Prosodic control to express emotions for man-machine speech interaction", IEICE Trans. Fundamentals, Vol. E75-A, No. 2, pp. 155-163, 1992.

[8] H. G. Wallbott and K. R. Scherer, "Cues and Channels in Emotion Recognition", J. Pers. Soc. Psychol., Vol. 51, No. 4, pp. 690-699, 1986.