

Title	怒りの感情音声における音響特徴量と感情知覚との関係に関する研究
Author(s)	平館, 郁雄
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1535
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

A STUDY ON THE RELATIONS BETWEEN THE ACOUSTIC PARAMETER AND EMOTIONAL PERCEPTION IN ANGER CONDITION.

Ikuo Hiradate (010095)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2002

Keywords: emotional speech, features contained in speech, accent, Neutral, Cold Anger, Hot Anger.

1 INTRODUCTION

There is a voice as one of the human most fundamental communication means. Recently, it is put to practical use that a voice is used as the interface with the human being and the computer. It can think that a smoother communication is made if the transmission of not only the linguistic information but also the paralinguistic information, non-linguistic information is done when a communication is taken with the computer by the voice. It can think that emotional information is especially indispensable to the synthesis sound like a human voice.

There is no study that a local part like an accent in the sentence was fully analyzed. So, the aim of this study is to make clear the relations between the specific in speech and the emotional perception in the accent part. The emotion of "anger" was divided into two in this study. The first emotion is "Hot Anger" that fury is expressed. It is "Cold Anger" that restrained anger is expressed. These two emotional speech were compared toward the emotional speech of "Neutral". A fundamental frequency, power, duration, formant frequency, spectrum was handled, and it paid attention to the change in the accent, and an analysis was done.

2 ACCENT

An accent is important in high synthetic speech quality, and the fundamental frequency influences it in Japanese. Japanese accent have two level (high level and low level). When an accent level changes from high level to low level, high level accent mora is called an accent nucleus. Three kinds of accent types are decided by the position of the accent nucleus. It adapted it to all the sentences based on this accent type in this study.

3 EMOTIONAL SPEECH DATA

3.1 EMOTIONAL SPEECH DATA

A voice actor and a drama experienced person know that it is exactly aware of the expression technique of the emotion by the voice in comparison with the general person[1]. Therefore, the emotional voice data handled by this study used the total 179 samples (9 patterns, 20 sentences) that it was collected from the professional woman voice actor. As for the emotion of 9 patterns, one kind of "Neutral" and "Joy" "Sadness" "Cold Anger" "Hot Anger" are two kinds respectively.

3.2 HEARING EXPERIMENT

The target of emotion was which exist in the listener by this study. Therefore, a listening experiment was done to judge for voice datas whether to be useful to analyze it in this research. It thought that it was sufficient for the analysis which aimed at the extraction of the emotion information as a result of the listening experiment.

4 ANALYSIS

It paid attention to the accent part in this study, and the analysis of the amount of acoustic parameters was done. Then, STRAIGHT was adopted as a model that the amount of acoustic parameters was extracted, changed and composed[2].

4.1 FUNDAMENTAL FREQUENCY

The averages of the fundamental frequencies in the whole of the sentences was analyzed. The fundamental frequencies of "Cold Anger" were lower than it of "Neutral", and the fundamental frequencies of "Hot Anger" were higher than it of "Neutral". When accent level rise and fall, the change rates in phoneme and the change rates in time were analyzed. The change rates of "HotAnger" were bigger than it of "Neutral", and the change rates of "ColdAnger" were smaller than it of "Neutral".

4.2 POWER

Power change rates in phoneme and time were analyzed when the accent level rise and fall. When an accent level rose, the power change rates of "Cold Anger" were smaller than it of "Neutral". When an accent level rose, the power change rates of "Hot Anger" were much bigger than it of "Neutral". When accent level fall, the power change rates of "Cold Anger" and "Hot Anger" were bigger than it of "Neutral".

4.3 DURATION

The durations of the whole of the sentences, the accent parts, the vowels and consonants in accent parts, non-accent parts and the vowels and consonants in non-accent parts were measured respectively. Then, durations of "Cold Anger" and "Hot Anger" toward durations of "Neutral" were analyzed.

The durations of "Cold Anger" and "Hot Anger" were shorter than durations of "Neutral" in sentences. The shortening rates of the accent parts were bigger than the shortening rates of the non-accent parts. A difference weren't shown in consonantal durations between each emotion. The durations of "Cold Anger" were much shorter than the durations of "Hot Anger" in the durations of the whole of sentences, the accent parts, the vowels of accents, the non-accent parts and the vowels of non-accents.

4.4 FORMANT FREQUENCY

The Formant frequencies were estimated from the prediction parameter by the LPC. The analyses examined formant frequencies in high level mora when an accent level rises from the low level to the high level. And, formant tracking by LPC went by using SpeechTools[3].

A figure1 showed the relations of the formant frequencies in "Neutral" and "Cold Anger". (A vertical axis and horizontal axis indicated formant frequency [kHz] of "Neutral" and "Cold Anger" with the common logarithm respectively.) The figure shows the tendency that generally the formant frequencies of "Cold Anger" were lower than the formant frequencies of "Neutral". The difference in F1 of "Cold Anger" and "Neutral" is big. As the order of formant rose, the differences in formant frequencies of "Neutral" and "Cold Anger" became small, and the differences was hardly seen with F4.

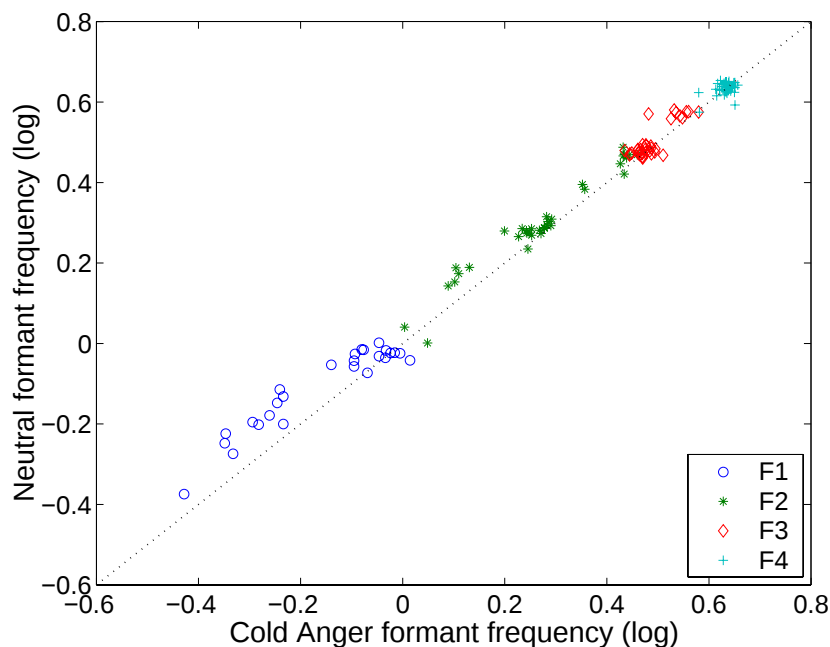


Figure 1: relations of between the formant frequency in "Neutral" and "Cold Anger"

In the same way, a figure2 showed the relations of the formant frequencies in "Neutral" and "Hot Anger". The figure showed the tendency that generally the formant frequencies of "Hot Anger" were higher than the formant frequencies of "Neutral". It was shown by the figure2 that the

differences in frequencies of "Neutral" and "Hot Anger" were the biggest in F1. As order of formant was raised, the differences in formant frequencies of "Neutral" and "Hot Anger" became small, and the differences was hardly seen with F4. Therefore, it pays attention to F1 ~ F3 by this study. An approximate expressions were led by the method of least square from the analysis result. This approximate expressions were decided to be applied in the case of the synthesis sound preparation.

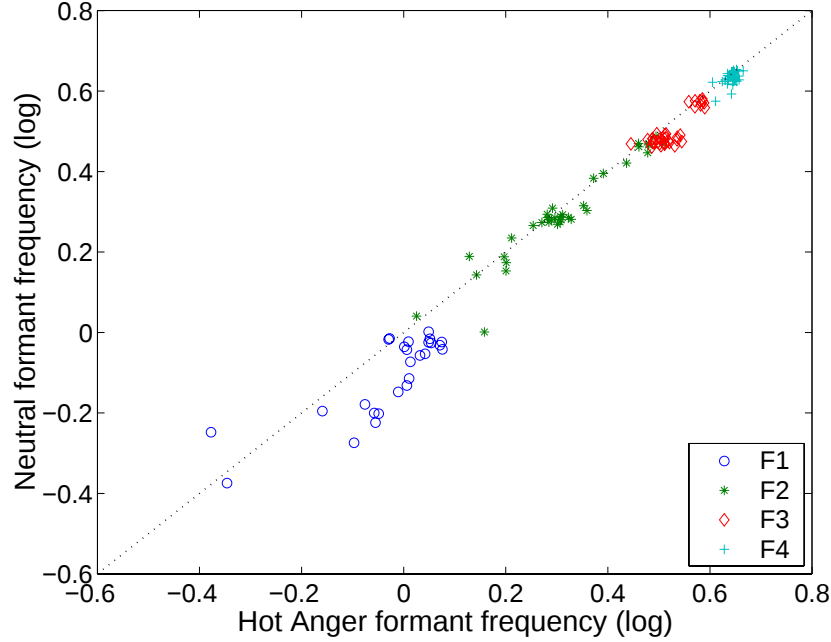


Figure 2: relations of between the formant frequency in "Neutral" and "Hot Anger"

4.5 SPECTRAL

The power spectra of moras in the low accent level and the high accent level were analyzed. Frequency axis was changed into ERB rate axis in consideration of the frequency character of basilar membrane of human inner ear[4].

The band was divided into 5 ERB rate each, and the power was analyzed about each band. The analysis result of the spectra in the accent low level were shown in the table1. (The inside of bracket were a difference in power of "Cold Anger" toward the power of "Neutral" in the same band.) The table showed that the form of the spectra in "Neutral" were the same as it

of "Cold Anger". The powers of "Hot Anger" were emphasized more than the powers of "Neutral" in the band from 25 to 30 ERB rate.

Table 1: The analysis result of the spectrum in the accent low level[dB]

	0 ~ 5 ERB rate	5 ~ 10 ERB rate	10 ~ 15 ERB rate	15 ~ 20 ERB rate	20 ~ 25 ERB rate	25 ~ 30 ERB rate	30 ~ 35 ERB rate
Neutral	13.6	25.4	20.1	10.3	2.8	0.1	-6.8
Cold Anger	17.3 (+3.7)	26.2 (+0.8)	19.8 (-0.3)	12.4 (+2.1)	3.5 (+0.7)	1.6 (+1.5)	-8.6 (-1.8)
Hot Anger	1.7 (-11.9)	15.1 (-10.3)	16.1 (-4.0)	11.6 (+1.3)	2.7 (-0.1)	4.2 (+4.1)	-7.5 (-0.7)

In the same way, the result of the spectra analysis in the accent high level were shown in the table2. The table showed the emphasis of the power of "Cold Anger" and "Hot Anger" in the band of from 25 to 30 ERB rate .

The rules to make synthesis sound of "Cold Anger" and "Hot Anger" from "Neutral" were led from the results of the spectra analyses in the accent low level, high level.

Table 2: The analysis result of the spectrum in the accent low level[dB]

	0 ~ 5 ERB rate	5 ~ 10 ERB rate	10 ~ 15 ERB rate	15 ~ 20 ERB rate	20 ~ 25 ERB rate	25 ~ 30 ERB rate	30 ~ 35 ERB rate
Neutral	9.5	22.8	19.9	11.5	4.7	-1.0	-6.6
Cold Anger	14.2 (+4.7)	24.7 (+1.9)	19.2 (-0.7)	12.4 (+0.9)	7.6 (+2.9)	2.2 (+3.2)	-9.9 (-3.3)
Hot Anger	-7.8 (-17.3)	3.4 (-19.4)	12.5 (-7.4)	10.8 (-0.7)	6.4 (+1.7)	3.7 (+4.7)	-6.8 (-0.2)

5 CONCLUSION

Remarkable characteristics were seen when an accent level rose about the fundamental frequency and the power. The changes in durations were

seen in the accent parts more than the non-accent parts. The changes in durations were seen with the vowels more than the consonants. The remarkable differences of the frequencies between the emotions were seen in F1. As orders of formant were raised, the differences in frequencies between the emotions became small. The powers of "Cold Anger" and "Hot Anger" were emphasized in the band of from 25 to 30 ERB rate.

The rules that synthesis sounds of "Cold Anger" "Hot Anger" were made from "Neutral" were led from the analysis results. From now on, synthesis sound will be made by using these rules, and it will be necessary to do a listening experiment.

References

- [1] Hiroshi Hiraga, Yoshiyuki Saitou, Shigeo Morishima, Hiroshi Harashima, "Extraction of Emotional States in Speech", IEICE, HC93-66, pp.1-8, Jan.1994.
- [2] H.Kawahara, I.Masuda Katsuse, K.Toyama, "Compensatory time window for speech analysis, modification and synthesis using STRAIGHT", ASJ Tech.Report, H97-47,1997.
- [3] Seiichi Tenpaku, Tatsuya Hirahara, "UNIX speech research tool - SpeechTools-", ASJ Tech.Report, pp.331-332, Mar.1991.
- [4] B.R.Glasberg, B.C.J.Moore, "Derivation of auditory filter shapes from notched-noise data", Hearing Research, 47, pp.103-138, 1990.