

Title	音声中の感情表現に関連する物理量とその制御に関する研究
Author(s)	杉本, 隆
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1357
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

A study on physical correlates to emotions in speech and their control

Takashi Sugimoto

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

February 15, 2000

Keywords: emotional information, synthetic speech, STRAIGHT, Lombard effect.

1 Introduction

Not only linguistic information but also emotional information, which play important roles in human communication, must be handled to synthesize speech like a human listens to easily.

F0, amplitude, and duration of utterance were discussed as physical parameters about emotional information in speech, and researches on F0 contours and other values had been done [1, 2]. However, as for speech synthesized by these researches, emotion can not be expressed well. This is because synthetic speech disregarded emotional correlates and physical correlates which originated in the generative mechanism of speech.

This paper investigates the relationship between physical values in speech and cognition of emotion. Transformation rules of physical values for controlling emotion are made using the results combining physical correlates which originated in the generative mechanism of speech. Additionally, the influence of rule-based transferred speech on emotion perception is investigated.

2 Analysis

2.1 Speech data

Speakers are two males and three females in the age of 20, and they are either actors, going to a vocal school or going to actor training place. Such speakers were chosen because they are exactly aware of techniques to express emotional condition by speech compared with a general person.

Linguistic information considerably influences on emotion. Fundamental frequency (F0) is handled in this paper. Therefore, a sentence ”いいじゃない” which composed of vowels and some voiced consonants is used as speech data.

167 samples in total were collected. Recording was done in a soundproof room. The uttered speech waves were recorded on DAT, outside the soundproof room at a sampling rate of 48kHz. There were down sampled into 20kHz, and stored in Work Station(WS).

2.2 Experiment for emotion classification

The target of emotion is which exist in the listener by this reseach. Then, listening test were done for emotion classification.

Subjects were tried to judge emotion, neutral, joy, anger, sadness. Subjects are eleven graduate school students and all have normality hearing. Subjects listened to stimuli by headphone in the soundproof room. Speech data were stored in WS outside the soundproof room, and it is shown corresponding to the listener’s answer.

Since samples of high distinction rates¹ existed the result that it was sufficient could get it as a result of the listening experiment in each emotion.

2.3 Analysis Results

Physical values related to emotion were extracted by STRAIGHT analysis-synthesis system[3], and differences between emotions classified in section 2.2 are investigated.

The results show the same tendencies as the past research[1] of the analysis toward the expression of each emotion. However, there was one difference from the past research result, that duration of utterance for anger was longer than that of neutral.

3 Transformation rule of physical values for emotional control

Transformation rule of physical values were constracted using analyzed results. Duration of utterance on consonant portions is not stretched. So, stretch of the consonant part isn’t done by time axis change map which was made in the clue of spectra.

Furthermore, the physical correlation that it originates in the formation mechanism of speech is integrated in rule of physical parameters’ transformation by Lombard effect is used.

3.1 Lombard effect

Speech wave is transformed as utterance style changes this is called as Lombard effect[4]. This effect is other physical parameters change speech becomes louder. Then, modeling of

¹The rate of the listener who replied when the same emotion was expressed is decided to be called distinction rate.

Lombard effect is effective to control anger emotion whose power is larger than that neutral. Rise in F0, shift of formant frequency, and increase of spectra in the high frequency region are occurred in Lombard effect.

Transformation rule of physical values transformation for emotional control with Lombard effect are shown as follows.

joy : duration of utterance, long time average of duration of power and F0 are the same as neutral.

increases the change rates of F0 in two times.

only the end of a word raises F0.

anger : duration of utterance is expanded in 1.2 times. (The stretch of the consonant part is not done.)

increases the long time averages of F0 in 1.2 times.

increases the long time averages of power in 1.4 times.

increases the change rates of F0 in 1.4 times.

formant frequency is shifted.

in high frequency region one increased 5dB.

sadness : duration of utterance is expanded in 1.5 times. (There is no consonant part stretch.)

The long time averages of F0 are decreased to 0.9 times.

The long time averages of power are made to decrease to 0.8 times.

The change rates of F0 are decreased to 0.5 times.

4 Experiment

Listening experiments were done to confirm whatever the synthesized speech which controlled an emotion whether to contain an emotion.

A listening experiment was done by using speech stimuli of applying rule of physical parameters' transformation the neutral speech of speech used for the analysis and making it as close test. Listening experiments were done by using speech stimulus of applying a rule it about "けっこうです", "そうですか" in ATR speech data base much in the same way as open test. Experiment conditions are the same as the emotion classified experiment.

As results, it could get a high recognition rate in the close test in all the emotions. Cognition rate was low in the open test.

However, many common clauses are found through the close test and the open test. Change rate of F0 was strongly concerned in each emotion expression, and it was found out specially that an important part as sadness.

Anger was fully expressed when duration of utterance was long. This is thought to originate in the difference from cold anger with hot anger, not that physical parameters of the duration of utterance is not important toward the emotion expression of anger.

5 Conclusion

The influence which a change in physical parameters in speech gave to the emotion distinction was examined by this research. It almost agreed with the past research result as that result.

Transformation rule of physical value for an emotional control was constructed using the analyzed results with physical correlates. Rise in F0, shift of formant frequency, and increase of spectra in the high were related to increase of power.

Emotion controlling a neutral speech by transformation rule of physical values, and influence on emotion perception was examined. The change rate of F0 was strongly concerned in each emotion expression, and it was found out specially that an important part as for sadness. As for anger, it could get the improvement of the recognition rate when physical parameters was made to relate by Lombard effect.

References

- [1] Yoshinori Kitahara, Yoh'ichi TOHKURA: "Prosodic Control to Express Emotions for Man-Machine Speech Interaction", IEICE TRANS. FUNDAMENTALS, Vol. E75-A, No.2, Feb.1992.
- [2] J.E.H.Noad, S.P.Whiteside and P.D. Green: "A MACROSCOPIC ANALYSIS OF AN EMOTIONAL SPEECH CORPUS", Proc. of Eurospeech, pp.517-520, 1997.
- [3] H.Kawahara, I.Masuda-Katsuse, K.Toyama: "Compensatory time window for speech analysis, modification and synthesis using STRAIGHT", ASJ Tech. Report, H97-47,1997.
- [4] B.J.Stanton: "Acoustic-Phonetic Analysis of Loud and Lombard Speech in Simulated Cockpit Conditions", Proc.of ICASSP, pp.331-334, Apr.1988.