International Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 1, March 2006 pp. 27-32

Design of Model to Recognize Emotional States in a Speech

Kim, Yi-Gon*
YoungChul Bae*

*Division of Electronic communication and Electrical Engineering, Chonnam National University

Abstract

Verbal communication is the most commonly used mean of communication. A spoken word carries a lot of informations about speakers and their emotional states.

In this paper we designed a model to recognize emotional states in a speech, a first phase of two phases in developing a toy machine that recognizes emotional states in a speech. We conducted an experiment to extract and analyse the emotional state of a speaker in relation with speech. To analyse the signal output we referred to three characteristics of sound as vector inputs and they are the followings: frequency, intensity, and period of tones. Also we made use of eight basic emotional parameters: surprise, anger, sadness, expectancy, acceptance, joy, hate, and fear which were portrayed by five selected students. In order to facilitate the differentiation of each spectrum features, we used the wavelet transform analysis. We applied ANFIS (Adaptive Neuro Fuzzy Inference System) in designing an emotion recognition model from a speech. In our findings, inference error was about 10%. The result of our experiment reveals that about 85% of the model applied is effective and reliable.

Key Words : ANFIS, emotional parameters, emotional state, spectrum features, wavelet, FFT(Fast Fourier Transform).

1. Introduction

There have never been a time in the history of mankind when a machine would recognize and interpret human emotion in a speech. It is only in recent decades that scientists and researchers got interested in Artificial Intelligence, which made it realistic.

Emotion has always been recognized as an important feature underlying behavior. Psychology recognizes emotion as an essential feature in all aspects of human behavior. For instance an individual behave according to his/her state of mood. A child cries when he/she is in unpleasant condition. Other sciences also are interested in the study of cognition aspect of beings, such as cybernetics, which is interested both in biological and artificial beings.

A speech contains information about the speaker's regional accent, sex, approximate age, state of health, personal identity, and about speaker's attitude or emotional state. Moreover, in face to face communication, some of the same "messages" are conveyed visually through facial expression and other movements of the body (Kinesics). Therefore, the human machine interaction through spoken language should be able to identify those features in order to convey attitude and emotion at the same time. In automatic speech recognition these features can help in decoding of a spoken language by indicating where the speaker is placing emphasis, whether the speaker is cooperating or not or, indeed, perhaps, whether the speaker is lying or not. [1,4]

In this paper the aim is to design emotion recognition model, which recognize emotion from a speech.

II. Why an Emotion Recognition Toy Machine

Technology never stop challenging the future in all careers. For instance NASA developed and launch two intelligence robots to the red planet (Mars) to explore and trace the existence of water on mars. Those machines send the data and pictures back on earth thousand miles away just in minutes. The Japanese car giant Honda made a robot capable of shaking hands with people, can identify voices, distinguish between sounds and the spoken word and respond to certain instructions. It can exchange simple sentences and greetings as well as 30 different commands. There are quite number of inventions in Artificial Intelligence to respond to human needs. [7]

In developed countries toy industries have grown dramatically in recent years, and that sector is attracting more investors. This study is in compliance with such rapid growth of toy market with a more meaningful role in the society. Toys are being used to occupy and entertain children of busy parents. In addition toys give a sense of imagination and creativity as a tool of play to the children. Toys help children to experience, explore, discover and express own unique and unlimited potential. Adults may use toys or dolls as substitute when they live a lonely life or simply as decoration in the house, in office or in the car.

In designing this emotion recognition toy machine the emphasis will be not mere possession tool of play for children, but a tool to help parents monitoring their children daily emotional status, a tool to help investigator in criminal cases to realize the emotion of suspects. It can be also a tool of emotional self evaluation.

Manuscript received Nov. 7, 2005; revised Mar. 15, 2006

III. Emotional Concepts and Parameters in Speech

The emotion belongs to the speaker. However the emotion that the speaker intended is not always communicated to the listeners due to many factors such as speaker's personality and speaking conditions (temporary psychological state). Therefore to capture and to know the emotional concepts and parameters is the basis to recognize emotion contains in a speech. What we get from a speech is a relative information between the emotional content and physical parameters. The relative information is provided not between emotional words and speech parameters themselves but between orthogonal bases extracted from them respectively. Using orthogonal bases, relative information is independent of the way and the kind of choose emotional words and speech parameters.

Emotional parameters can be classified into two categories: (a) pleasant parameters such as joy, content, relief, calm, admiration, happy, love etc..., and (b) unpleasant such as anger, disgust, sad, fear, hateful, outrage, etc.. [5]

IV. Characteristics of Sound to be considered in Speech.

It is worth to understand the concept and characteristics of sound in order to understand and incorporate these characteristics to speech. Human perceive sound by three means:

- **Loudness:** which is the strength of a sound. Its synonym is volume.
- **Pitch:** the highness or lowness of a sound.
- **Quality:** the characteristic distinguishing sounds from different sources. Its synonym is timbre.

In this study we are going to consider three characteristics of sound or speech: frequency, intensity and period.

1. Frequency.

Aside from the definition of frequency, which is the number of cycles or oscillations, a sound wave completes in a given time, we defined the importance of frequency in emotion recognition. First of all in emotion recognition, frequency gives the mutual information between energies and the corresponding phonetic label. This is very useful to our study since we are going to analyze every detail of each sound produced. Second, speech signals whether voiced or unvoiced are obtained in time domain and are analysed in frequency domain. Voiced sounds consist of fundamental frequency. let's take a look an example of three vowels /a/, /i/, and /u/ in time and frequency domain with a fundamental frequency of 100 Hz in all cases as presented in Figure 1.

In fig1., the harmonic structure of the excitation can be perceived by mean of frequency domain.

Last, with the use of fundamental frequency we can distinguish the speaker gender. Male sound produces a big fundamental frequency range for almost all emotions, while the female sound produces a narrow range of fundamental frequency. Children fundamental frequency is associated with that of females. [6]

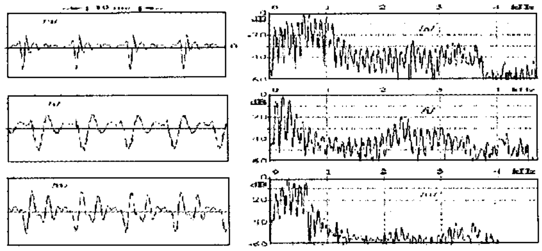

Fig.1 look an example of three vowels /a/, /i/, and /u/ in time and frequency domain with a fundamental frequency of 100 Hz

2. Intensity.

The intensity or loudness is the average rate of flow or energy per unit area perpendicular to the direction of propagation. It is measured by amplitude. Intensity or loudness depends upon the amplitude and is directly proportional to the square of the amplitude.

Speech signal is a series of sounds rapidly varying from instant to instant in frequency and intensity. A dynamic range coupled with fine intensity resolution and fine spectral resolution allows normal-hearing listeners to maintain high speech intelligibility. In another words acoustic stimulation depends on intensity level. So for the mentioned reasons, we select intensity as an equal important parameter as frequency in emotion recognition from the speech.

In the sound intensity signal is the sum of the squares of the sound amplitude signal that reaches the ear at any time. Figure 2 shows the transfer of sound amplitude into sound intensity.

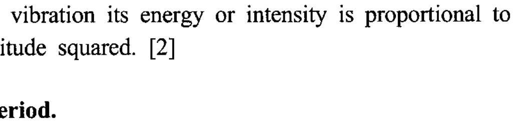

Fig.2 shows the transfer of sound amplitude into sound intensity

Actually there is always a static pressure signal in the total sound intensity frequency spectrum, because for every sinusoidal vibration its energy or intensity is proportional to sound amplitude squared. [2]

3. Period.

Period of tone is another important parameter in emotion recognition signal analysis. Its importance lies on the fact that period of tone facilitates to compare each sound duration in harmonical sinusoids and to eliminate noisy.

Mathematically, the periodic component can be given by the following formula:

Let $h[n]$ be the periodic component as the sum of the harmonically sinusoids, then

$$h[n] = \sum_{k=-L(ni)}^{L(ni)} A_k[n]\, e^{j2\pi kfo(ni)(n-ni)}$$

Where $L(ni)$ denotes the number of harmonics included in the harmonic part at $n = n_i$, $f_o(n_i)$ denotes the fundamental frequency at $n = n$. $A_k[n]$ can take on one of the following forms:

$$A_k[n] = a_k(n_i)$$

$$A_k[n] = a_k(n_i) + (n - n_i) b_k(n_i)$$

$$A_k[n] = a_k(n_i) + (n - n_i) + c_k(n_i) + (n - n_i)^2 d_k$$

Where $a_k(n_i)$, $b_k(n_i)$, $c_k(n_i)$, and $d_k(n_i)$ are assumed to be complex numbers with $arg\{a_k(n_i)\} = arg\{c_k(n_i)\} = arg\{d_k(n_i)\}$ (assumption of constant phase). These complex numbers denote the amplitude of the k^{th} harmonic. [6]

VI. Experiment set up and Data acquisition.

In gathering the data, we used Lab-View version 5.0.1, a microphone, an A/D converter, a loud speaker, and a desk-top computer. To acquire voice signal, five selected students were asked to say and act on a korean sentence featuring the eight basic emotional parameters: acceptance, anger, expectance, fear, hate, joy, sadness and surprise. The sentence is: "A, G-RUB-SSUM-NI-GGA?" Fig.3 shows the set up of our experiment. The signal flows from the speaker's speech into PC via a microphone which change the voice signal into electric signal, then the electric signal is amplified by a pre-amplifier in analog mode, and before it proceeds to the PC an A/D converter is applied to the signal in order to obtain a digitalized signal which is the computer language.

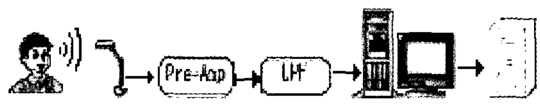

Fig.3 The experiment set

Fig.3. The experiment set up.

The sample data gathered are shown in the following figure with sampling frequency of 40KHz.

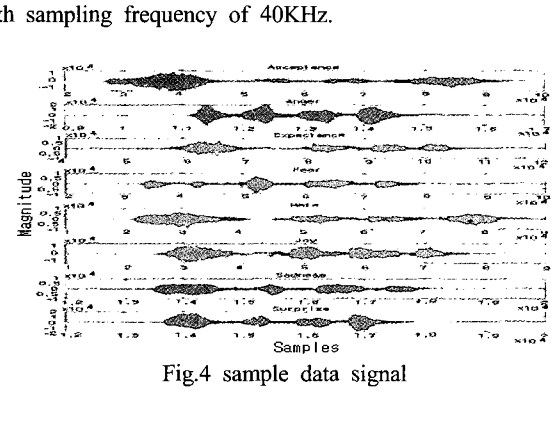

Fig.4 sample data signal

VII. Data Analysis and Algorithm.

To analyse signals obtained, we formulated a new algorithm which is characterized by seven steps:

1. Start algorithm
2. Measure the duration of voice which refers to the determination of period of tone in a voice signal.
3. Measure the energy distribution of voice signal. In another word "intensity".
4. Measure the frequency out put of each speech. To retrieve emotion from a speech, different vocal structure of each individual has to be considered.
5. Analyse signals with wavelet.
6. Infer emotion with neuro-fuzzy model.
7. End the algorithm.

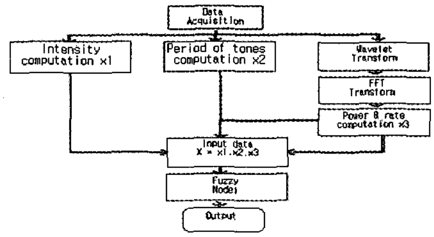

Fig.5 represents the diagram of recognition algorithms.

After setting algorithm model, the sample data signal in fig.4 were obtained in time domain were transformed into frequency domain by use of FFT and the result of frequency spectrum of sample signals is shown in fig.6

To analyse and distinguish characteristics carried out by each signal, we used wavelet transform analysis method in Matlab. We made use of one dimensional signal

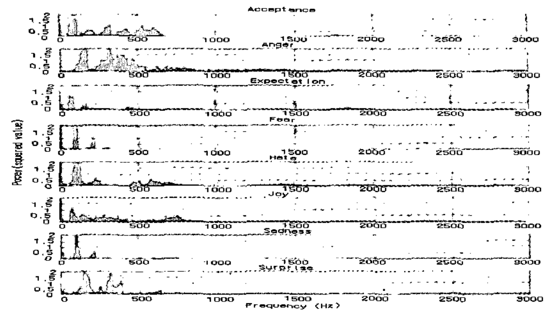

Fig.6 FFT spectrum of sample signals.

(wavelet 1-D), with analyzing wavelet = db3, and decomposition levels = 6.

Result of surprise is shown in fig.7

In original signals, there are much noise signals in overall shape. In this level-six analysis, we realized that the trend becomes more and more clear with each approximation, d1 to d6. In another words successive approximations possess less high-frequency information. With the high-frequency removed, what's left is the overall information carried by the signal.

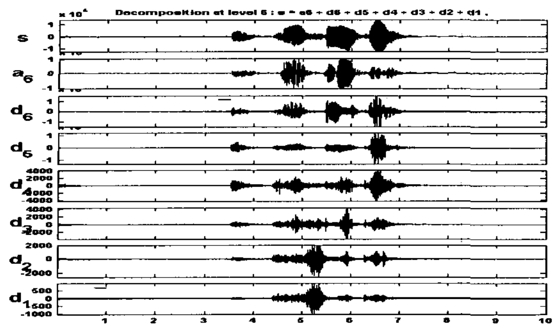

Fig.7 Surprise Signals

We stressed previously that frequency plays a major role in signal analysis. In order to figure out the emotional features each sound carries, we analysed the spectrum distribution of wavelet transform in FFT and we obtained power spectral density data as presented in figure 8.

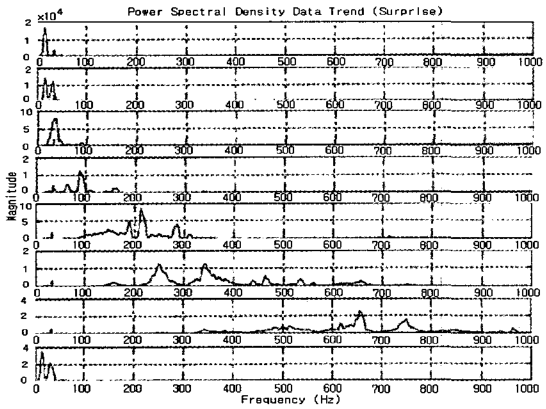

Fig.8 obtained power spectral density data as present

The results of frequency, intensity and period of tones derived from the frequency spectrums analysis are presented in table 1.

Table 1. The results of frequency, intensity and period of tones

Emot. \ Char.	Intensity (x1)	Period of Tones (Samples, 40KHz)(x2)	Frequency (Hz)	
Acceptance	270	6700	160	290
Anger	420	5700	150	310
Expectance	193	10500	90	200
Fear	158	9500	60	150
Hate	293	16000	70	510
Joy	297	12000	100	210
Sadness	190	11000	70	710
Surprise	154	7600	90	210

Table 2 and table 3 present the characteristic data result of analysis of eight basic emotion parameters in wavelet.

Table 2. characteristic data result of analysis of eight basic emotion parameters in wavelet.

Emot. \ Chan.	S	a6	d6	d5	d4	d3	d2	d1
Accept.	2.5e1	37.0	24.0	40.1	5.20	0.74	0.20	0.05
Anger	5.9e1	0.07	16.29	40.95	7.96	1.10	0.25	0.08
Expect.	6.5e0	5.60	48.3	22.1	1.10	0.40	0.10	0.05
Fear	3.4e0	10.5	31.95	8.96	5.95	3.89	1.21	0.21
Hate	1.7e1	0.75	31.0	15.1	10.0	1.97	0.33	0.04
Joy	1.9e1	4.45	48.4	19.38	4.81	1.70	0.46	0.11
Sad.	6.0e0	1.73	25.83	19.72	13.34	3.23	0.50	0.06
Surp.	4.8e0	3.75	47.45	5.05	0.94	0.25	0.09	0.02

Table 3 Characteristic data (x3)

Emo.	Acc.	Ang.	Exp.	Fear	Hate	Joy	Sad.1	Surp
Lev.	11000	23000	6000	3200	6000	11000	000	4000
d6	2200	400	3100	700	1100	2100	400	1100
d5	6100	18100	2100	2100	2200	5100	500	2100
d4	9900	21100	2000	1900	4600	4100	400	2600
d3	1300	6100	2200	360	1900	2100	400	900
d2	1100	1100	1200	280	500	1100	200	200
d1	300	700	500	140	140	300	150	1200

For more details we analysed the result of table 1 and 2 in a radiation graph and with the analysis we find out that fear, joy, and sadness are similar in emotion pattern as shown in figure 9

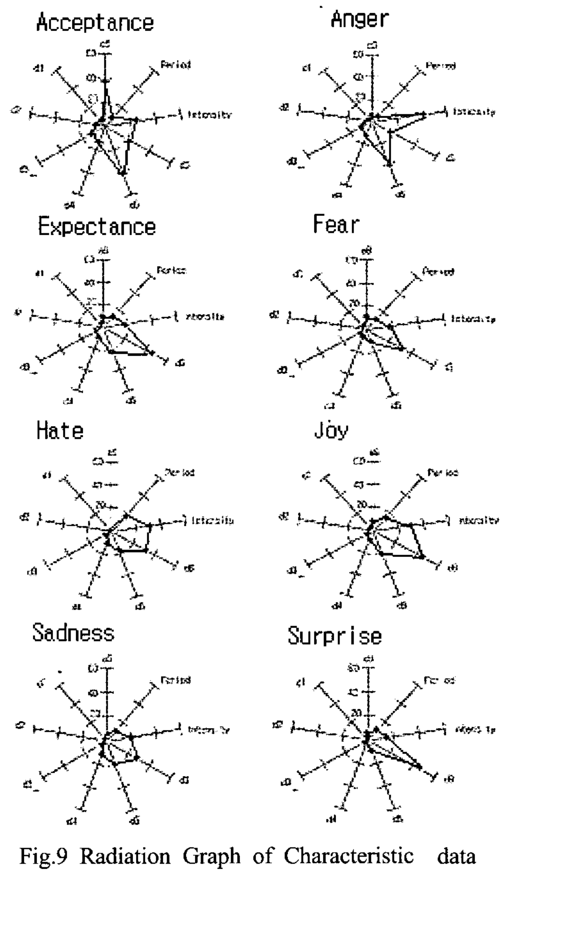

Fig.9 Radiation Graph of Characteristic data

VIII. Emotion Recognition Model Design

ANFIS (Adaptive Neuro-Fuzzy Inference System) was used in designing the emotion recognition model. Fig.10 shows the structure result of a neuro-fuzzy model.

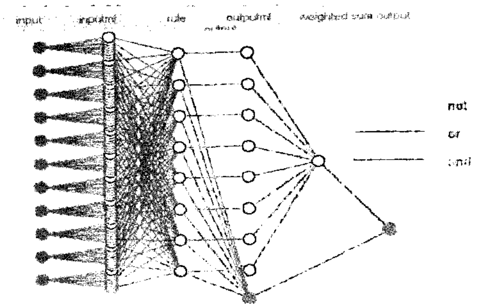

Fig 10 Analysed the spectrum distribution of wavelet transform in FFT

With the structure in fig.10, we came up with characteristics of emotion recognition model as it is presented in a surface window fig.11.

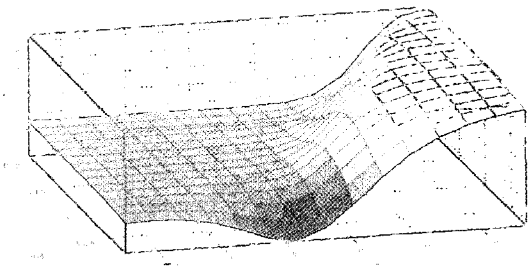

Fig.11 presented in a surface window

We evaluated the designed model with our experimental data and a percentage of error of inference set to 10%, and by computer simulation, the result is shown in fig.12

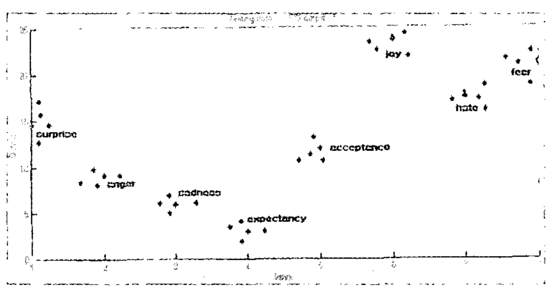

Fig.12 the result of computer simulation

IX. Conclusion

In this paper, we conducted an experiment aims to design an emotion recognition model, which would be suitable to recognize emotion of a speaker in a speech, with three input vectors: intensity, period of tones, and frequency. Based on our findings with a winning rate of about 85% from the computer simulation result, we conclude that the emotion recognition model design is effective and reliable to extract emotional states of a speaker in a speech. To materialize it's application, we are going to apply the model in designing an intelligent toy machine.

References

[1] Eibl-Eibesfeldt I.,"Ethology. The biology of behavior", 2nd ed. Holt, Rinehert, and Winston, New York, 1975.

[2] Heerens Chr. Willem "Sound intensity on the basilar membrane as square of the amplitude on the ear drum", 2003 (http://www.slechthored-plus.nl/fysica/en/heerens-02 expl.htm).

[3] I. R. Murray and J. L. Arnott,"Toward the simulation in synthetic speech: A review of the literature on human vocal emotion",J. Acoust. Soc. Am., Vol.93, No.2, pp1097-1109, February 1993.

[4] I. R. Murray and J. L. Arnott, "Synthesizing Emotions in Speech: Is it time to get excited?", The MicroCenter, Applied Computer Studies Division, University of Dundee, Dundee DDi 4HN, U.K.

[5] Moriyama T. ; Saito H. and Ozawa S.,"Evaluation of the relationship between emotional concepts and emotional parameters on speech", Department of Electrical Engineering, Kei Univ., Japan.

[6] Phonetics and Theory of Speech Production. http://www.acoustics.hut.fi/~slemmett/dippa/chap11.html

[7] Woodman Peter, PA News, "Human Robot science Museum Debut" Sunday 18, January 2004. 10:26 am U.K. (http://www.news.scotsman.com).

Yigon Kim
received the MS degree in avionic electric engineering from Hankuk aviation university, Seoul Korea, in 1986 and 1988 respectively. He received Ph.D. degree in electrical engineering from Chonnam National university in 1993. He performed research at Tokyo Institute of Technology by research member in 1991 and at Iowa State University by visiting professor in 2000-2001. He is an associate professor in the School of ECC at Yosu National University. His interests fuzzy-Neuro Modeling and its application to diagnosis and control industrial systems.

Young-Chul Bae
received his B.S degree, M.S and Ph. D. degrees in Electrical Engineering from Kwangwoon University in 1984, 1986 and 1997, respectively. From 1986 to 1991, he joined at KEPCO, where he worked as Technical Staff. From 1991 to1997, he joined Korea Institute of Science and Technology Information (KISTI), where he worked as Senior Research. In 1997, he joined the Division of Electron Communication and Electrical Engineering, Yosu National University, Korea, where he is presently a professor. His research interest is in the area of Chaos Nonlinear Dynamics that includes Chaos Synchronization, Chaos Secure Communication, Chaos Crypto Communication, Chaos Control and Chaos Robot etc.