정규논문 (Regular Paper)
방송공학회논문지 제18권 제5호, 2013년 9월 (JBE Vol. 18, No. 5, September 2013)
http://dx.doi.org/10.5909/JBE.2013.18.5.713
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

음성 감정인식에서의 톤 정보의 중요성 연구

이 정 인[a)‡], 강 홍 구[a)]

On the Importance of Tonal Features for Speech Emotion Recognition

Jung-In Lee[a)‡] and Hong-Goo Kang[a)]

요 약

본 연구는 음성의 감정인식에 있어서 크로마 피쳐를 기반으로 한 음성 토널 특성에 대하여 기술하였다. 토널 정보가 갖는 장조와 단조와 같은 정보가 음악의 분위기에 미치는 영향과 유사하게 음성의 감정을 인지하는 데에도 토널 정보의 영향이 존재한다. 감정과 토널 정보의 관계를 분석하기 위해서, 본 연구에서는 크로마 피쳐로부터 재합성된 신호를 이용하여 청각 실험을 수행하였고, 인지실험 결과 긍정과 부정적 감정에 대한 구분이 가능한 것으로 확인되었다. 인지 실험을 바탕으로 음성에 적합한 토널 피쳐를 적용하여 감정인식 실험을 진행하였고, 토널 피쳐를 사용하였을 경우 감정인식 성능이 향상되는 것을 확인 할 수 있다.

Abstract

This paper describes an efficiency of chroma based tonal features for speech emotion recognition. As the tonality caused by major or minor keys affects to the perception of musical mood, so the speech tonality affects the perception of the emotional states of spoken utterances. In order to justify this assertion with respect to tonality and emotion, subjective hearing tests are carried out by using synthesized signals generated from chroma features, and consequently show that the tonality contributes especially to the perception of the negative emotion such as anger and sad. In automatic emotion recognition tests, the modified chroma-based tonal features are shown to produce noticeable improvement of accuracy when they are supplemented to the conventional log-frequency power coefficient (LFPC)-based spectral features.

Keyword : Emotion recognition, tonality, tonal features, chroma feature

I. Introduction

Recognizing emotion with speech signal provides naturalness for human-and-machine communication. Various studies utilize the fact that spoken utterances contain emotion information with a form of linguistic and paralinguistic one. Since the emotional expression of human's voice is not consistent, i.e. emotion is not as objective as linguistic information, however, designing an efficient emotion recognition system is very difficult compared to other signal processing systems such as speech recognition or speaker

a) 연세대학교 전기전자공학과 (Dept. of EE at Yonsei university)
‡ Corresponding Author : 이정인(Jung-In Lee)
E-mail: junginida@dsp.yonsei.ac.kr
Tel: +82-2-2123-4534
· Manuscript received 25, June 2013 Revised 22, August 2013 Accepted 22, August 2013

recognition [1].

To automatically analyze and detect these ambiguous human emotions, many studies utilize a variety of features, feature measurements, and classifiers. Typically, prosody, voice quality, and spectral related features have been used [1-3]. Characteristics of prosody features including fundamental frequency (F0), intensity, and duration are investigated in Murray's study [4]. In general, the arousal of emotions is easily identified by F0 and intensity, however, the valence of emotions that is related to positive and negative feeling is relatively difficult to detect using speech signal only. Some studies argue that voice quality features slightly improve the classification accuracy in the valence dimension [5-7]. However, spectral features such as mel-frequency cepstral coefficients (MFCC) [8], log-frequency power coefficients (LFPC) [9] have been popularly used in emotion recognition systems. Though MFCC are most widely used to represent spectral characteristics in speech signals, Nwe's study argued that LFPC are more appropriate for emotion recognition [9]. LFPC represents energy distribution of spectrum, and also provides information on the fundamental frequency of speech, because log-frequency scale has higher resolution for low frequency range than mel-frequency scale,

In this study, we focus on the tonal feature that has drawn less attention than other types of features in speech emotion recognition. While listening music, human perceives the mood by the dominant music key; whether the dominant key of the music is major or minor. Our approach starts from the assumption that speech emotion is also affected by tonality. In the applications of music information retrieval (MIR), various attempts are carried out to extract the tonal features from audio signal, and the extracted features are applied to music genre recognition or music mood classification [10, 11].

A chroma feature is one of the widely used features for representing tonality [12], but it needs to be modified for speech signal applications, because the intensity of speech signal is mainly concentrated in low frequency region. By analyzing the feature characteristics of all the octave frequency range, an appropriate frequency range for the chroma feature is determined in this study. Interestingly, the effective frequency range is highly related to the fundamental frequency of human voice. Experimental results show that the combination of the proposed chroma feature with LFPC increases the recognition rates compared to the combination of LFPC+F0 or LFPC only. Especially, the proposed feature combination significantly improves overall accuracy for the gender dependent experiment.

This paper is organized as follows. Section II describes the characteristic of features used in the experiments. System setup and experimental results on tonal features and reference features are summarized in section III. Section IV follows conclusion and discussion.

II. Sub-band and tonal features

1. Sub-band power features

In the field of emotion recognition with speech, short-time subband power related features are typically used. Considering the characteristic of human perception, the sub-band power related feature is computed by:

$$S_n(m) = \sum_{k=f_m-(b_m/2)}^{f_m+(b_m/2)} (X_n(k)W_m(k))^2, \quad m = 1,2,...,M \tag{1}$$

where, $X_n(k)$ is the kth spectrum of windowed signal, $W_m(k)$ is filterbank window, n is the frame index, $S_n(m)$ is the power coefficient of mth filter, and f_m, b_m is the center frequency, bandwidth of mth filterbank, respectively. The sub-band power coefficients $P_n(m)$ is computed by:

$$P_n(m) = \frac{10\log_{10}(S_n(m))}{N_m}, \quad (2)$$

where N_m is the number of samples in the mth filterbank.

In this paper, four feature vectors having various frequency resolution such as log-frequency (LFPC), mel-frequency (MFPC), linear-frequency (LinFPC), and bark-scale (BFPC) are extracted in each frame for comparison [13]. Relationship between frequency and auditory frequency scale are shown in Fig. 1. The number of auditory filterbank, M, is set to 12 to make a fair comparison with the proposed chroma features explained in the next subsection.

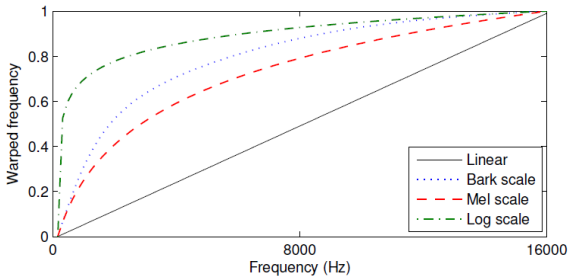

그림 1. 청각필터뱅크 종류에 따른 주파수 와핑
Fig. 1. Frequency warping for various types of auditory filterbank

2. Chroma features

Chroma features obtained by combining the normalized sub-band energies of 12 pitch classes to represent the tonal information of audio signals [10]. The 12 pitch classes are mapped into 12 semi-tones defined in western music theory such as {C, C#, D,..., B}, which are referred to as chroma [14]. In order to calculate the features, sub-band energies are computed by 88 filters centered at the pitches of A0 to C8 (MIDI pitches, p=21 to p=108), and they are decomposed into corresponding classes. The chroma features are computed by taking the following three steps [12]:

Decompose the audio signal into 88 frequency bands corresponding to the musical notes from A0 to C8.

Compute the short-time mean-square power (STMSP) for each of the 88 sub-bands.

Compute the element of chroma feature by adding up the STMSPs of all the corresponding pitches belonging to the respective class.

In the first step, log scale triangular filterbanks are computed for center frequency $f_c(k)$ described in equation (3).

$$f_c(k) = f_{\min} \times 2^{\frac{k}{\beta}}, \quad k \in [0, (\beta \times Z) - 1] \quad (3)$$

where $f_{\min}$ is minimum frequency of the analysis, k is integer filter index. Z and β denote the number of octaves and bins per octave, respectively. Short time mean square power of each filterbank $S_n(m)$ is computed by equation (1). In the final step, chroma feature of chroma pitch class b is computed by summing corresponding $S_n(m)$ for octave z, as described in equation (4).

$$C_f(b) = \sum_{z=0}^{Z-1} |S_n(b + z\beta)|, \quad z \in [0, Z-1], \quad (4)$$

where b is index associated with 12 chroma.

This yields a real 12-dimensional vectors. A 64 msec analysis window (1024 samples) is applied to speech signal in order to maintain the tonal information and to increase the frequency resolution. The frame shift length is set to 10 msec (160 samples) to make a fair comparison with other features. Note that the proposed chroma feature for speech emotion recognition is different from the conventional one used for music applications. The conventional chroma feature contains spectral information of all the 8 octaves.

Since the spectral envelope which are dominantly affected by first to third formants are varies depending on the type of uttered vowels, it is difficult to detect emotional variability when we use octave bands corresponds to the frequency range between 1kHz and 4kHz. Considering the

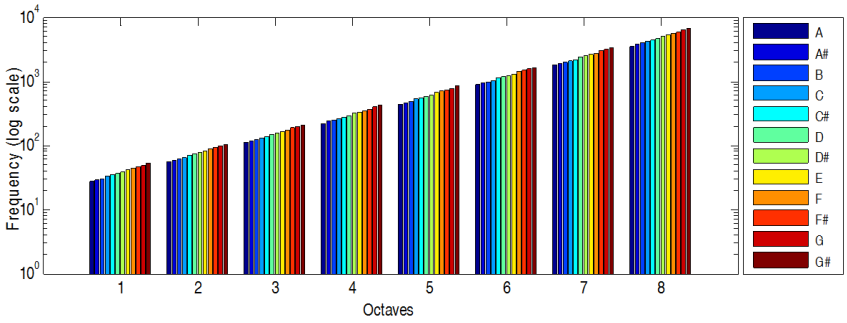

그림 2. 선형 주파수를 8개의 크로마 피치로 변형하기 위한 주파수 와핑
Fig. 2. Frequency warping between linear frequency (Hz) and chroma pitch for 8 octaves.

frequency range of fundamental frequency, F0, of speech signal, we only consider the part of frequency bands while computing chroma features. According to the analysis given in the following section, only the third and fourth octave frequency bands are very important, which are highly related to the F0 range of human voice, i.e. 110 to 415.3Hz (MIDI pitch, p=45 to p=68).

3. Feature analysis

Subjective hearing tests with artificially generated signals by chroma features are also performed to verify the relation between human perception and tonal information. Synthesized signals are computed by inverse step of the computing chroma features. In order to reconstruct the harmonic component of spectrum, center frequencies are determined from 12 chroma pitches based on the Fig. 2. Spectrum of chroma feature $V=[v_1, v_2, v_3, \ldots, v_{12}]$, is computed by adding the reconstructed harmonic spectrum of v_c. Synthesized spectrum $S_{synth}(f)$ is obtained for each frame, and inverse Fourier transform is performed within overlap add scheme to convert the frequency domain signals into time domain. Computation of each component is described in equation (5) and (6).

$$H(c,f)=\sum_{z=1}^{Z} v_c \delta(f-f_{c,z}) * W_g) \quad (5)$$

$$S_{synth}(f)=\sum_{c=1}^{12} H(c,f), \quad (6)$$

where $H(c,f)$ is harmonic spectrum computed from c^{th} chroma feature v_c, $f_{c,z}$ is center frequency which corresponds to chroma pitch c and octave z, W_g is gaussian weight, respectively.

Features are normalized to reduce the impact caused by loudness that is related to the arousal of emotions. Pause, silence, and unvoiced frames are excluded from the synthesized signal to minimize the influence caused by durational cues. Signals belonging to the neutral emotion class are used as references, and test materials are blinded and trials are randomly selected. Ten listeners participated in the test composed of 30 sentences.

표 1. 크로마를 이용하여 합성된 신호를 이용한 청각실험 결과
Table 1. The result of subjective hearing test with artificially re-synthesized signal using chroma features on emotion valence

emotion	valence	positive	negative
neutral	reference	-	-
joy	positive	0.62	0.38
sad	negative	0.29	0.71
anger	negative	0.21	0.79

The result is tabulated in Table 1. The 62% of positive emotion is perceived as positive, and 75% of negative emo-

tion is perceived as negative. Especially, about 79% of anger is selected as negative emotion. Although test materials do not include the loudness and duration, the subjective test with only the tonality of utterance derives meaningful results. Previous studies indicate that prosodic features are efficient for arousal classification, but are inefficient for valence [7]. However, according to our results, there are valence cues in the tonal features.

Subband features compute the power of sub-band frequency independently, while chroma features combine the power of pitch related multiple bands. To investigate the relevance of features for emotion recognition, the amount of mutual information (MI) between the features and emotion classes is compared.

The MI between feature vector x and class c is defined as equation (7)-(9).

$$I_s(x:c) = H_s(x) - \sum_c p_c H_s(x \mid c) \tag{7}$$

$$H_s(x) = -\int p(x) \log p(x) dx \tag{8}$$

$$H_s(x \mid c) = -\int p(x \mid c) \log p(x \mid c) dx \tag{9}$$

where p_c is the prior class probabilities, $p(x|c)$ is conditional distribution given class c, and overall data distribution $p(x)$ is computed by equation (10).

$$p(x) = \sum_c p_c p(x \mid c) \tag{10}$$

However, it is difficult to compute the distributions such as $p(x)$ and $p(x|c)$ from the collected data, the MI is calculated by utilizing the Gaussian mixture model (GMM) [15]. Estimated $p(x)$ using GMM components are written as:

$$p(x) = \sum_{m=1}^{M} \alpha_m G_m(x) \tag{11}$$

$$I(x;c) = -\frac{1}{N}\sum_{i=1}^{N} \log(\sum_{m=1}^{M} \alpha_m G_m(x_i)) + \sum_c p_c (\frac{1}{N_c}\sum_{m=1}^{N_c} \alpha_m G_m(x_i^c))) \tag{12}$$

where N is the number of overall data samples and N_c is the number of data samples belonging to class C. For GMM estimation, 800 utterances per each emotion and 32 mixtures GMM are used in this experiment.

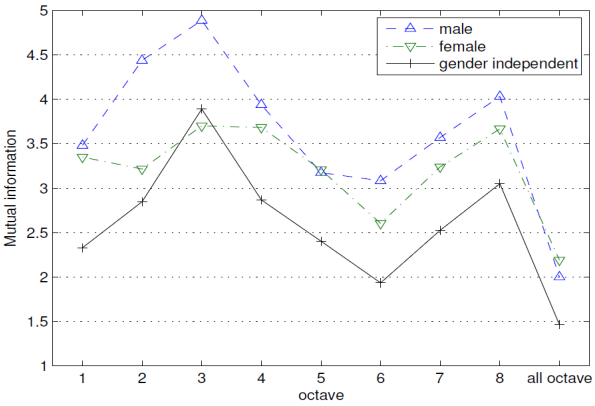

그림 3. 8개 옥타브에서 추출된 크로마피쳐에 대한 뮤추얼 인포메이션 결과
Fig. 3. Mutual information of chroma features for 8 octaves

Fig. 3 shows the MI between emotion classes and 12 semi-tone pitch classes with each of eight octaves. MI between 2nd and 4th octave which are 55Hz to 415Hz are relatively higher than other octaves and conventional chroma feature which is denoted as 'all octaves'. These frequency ranges are related to the average F0 ranges of male and female speakers. It proves that chroma features generated by the F0 related octave frequency contain relevant emotional information. Features extracted from octaves 5th to 7th which correspond to formant frequency range have lower information than the results of 2nd to 4th fourth octaves.

Based on the MI results, chroma features are extracted for 2nd to 4th octaves which are related to F0. The examples of chroma features are shown in Fig. 2. The frequency range of 2nd octave is 55Hz to 103Hz. As shown in Fig.

4, octave 2, the range is too small to be analyzed by 12 channel filters, and it contains both pitch and consonant cues. Thus, chroma feature of 2nd octave have excluded in proposed features.

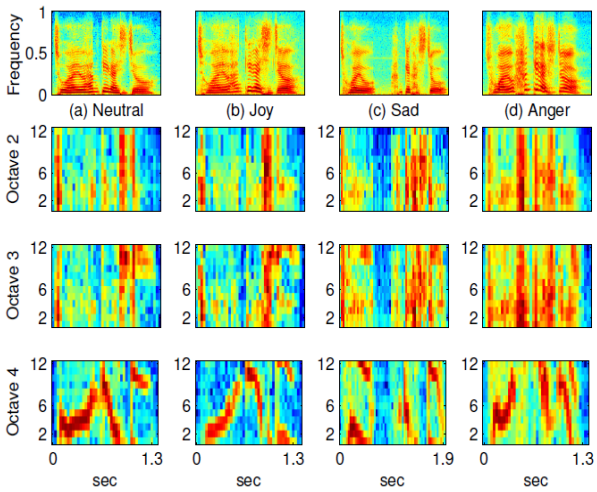

그림 4. 여성화자로부터 추출한 2-4번째 옥타브 크로마 피쳐 결과
Fig. 4 Chroma features of 2nd to 4th octaves from one female speaker.

III. Experimental results

1. Corpus

The speech material used in this study is the Korean Emotion Corpus developed by Kang et al [16]. It includes four emotional states, which are joy, sad, angry, and neutral, recorded by 15 actresses and 15 actors. Forty five scripts for each emotion were recorded 3 times with the sampling rate of 16kHz. Sentences include 22 long declarative sentences, 9 short declarative sentences, and 14 interrogative sentence. Averaged length of 45 sentences are 1.3 sec. Every utterance was examined for phonetic balance and ease of pronunciation. Corpus also contains subjective test scores (1 to 10) to represent the quality of emotion to each sentence.

2. Experimental setup

An automatic emotion recognition system using tonality features is implemented using a support-vector machine (SVM) technique with a radial basis kernel. To implement a multi-class emotion recognition system, livsvm tools [17] are used for training and test. The optimal cost parameter is obtained by using a 10-fold cross-validation technique with a parameter range from 2 to 1024. The training data set contains 1000 utterances for each emotion, and the utterances are not duplicated with test materials.

In the feature extraction step, the first and second moment (mean and std) for spectral and chroma features are computed for each utterance. Computed features are combined with one supervector per one utterance. In order to compare the spectral features and chroma features, feature orders are equivalently set up 12. Totally, utterance-level features computed from first and second moments are 24. Supervectors are collected by feature level fusion, and trained by SVM for neutral, joy, sad, and anger emotion classes [18]. 10 types of global statistics such as mean, std, kurtosis, skewness, median, percentile (5, 25, 50, 75, 95%) are used to extract F0 related features, mean and std of each coefficients are used for spectral power coefficient and tonal features.

For the performance measure, 10-fold cross validation with the total of 4452 sentences is used to reduce the speaker and sentence dependency.

In order to verify the efficiency of tonal features in speech emotion recognition, two experiments are carried out. Experiment 1 compares the modified chroma features and conventional features. Experiment 2 shows the recognition accuracy of the revised system that the proposed tonal features are combined with LFPC features.

Based on the experiment 1 and 2, combination of tonal and pitch related features with LFPC are experimented. Both tonal features and F0 related features represent glottal

source information, however, combination of different types of features can overcome the drawback of each features.

표 2. 실험 12차 크로마피쳐와 스펙트럴 피쳐를 이용한 감정인식 성능 비교
Table 2. Experiment 1: recognition results of 12th order chroma features (CHR) and spectral features

feature	neutral	joy	sad	anger	total
CHR	59.0	44.50	63.23	72.73	59.86
LFPC	68.87	57.71	67.97	75.77	67.58
Bark-FPC	64.96	59.60	62.75	75.62	65.74
MFPC	63.51	58.60	60.33	75.61	64.51
Linear-FPC	52.41	57.72	56.63	73.32	60.02

3. Recognition results

Table 2 shows the recognition accuracy of the system using chroma features and sub-band power coefficient features. The order of each feature is set to 12. LFPC has the highest accuracy, and BFPC, MFPC, and LinFPC are followed. The results given in Table 2 show that the features which have high resolution in low-frequency range show higher recognition rates. This result follows the argument of LFPC efficiency in emotion recognition proposed by Nwe et al [9]. Recognition rate of the proposed chroma features is 59.9%, which is somewhat lower than that of other sub-band energy features, especially, the detection rate of neutral and joy is significantly lower than sad and anger emotion.

As mentioned in the summary of previous researches [2][3], emotional cues are contained in both glottal source and vocal tract information. It means that both pitch related features and spectral features are important to represent characteristics of emotions in speech signal. Thus, the glottal source or spectral biased features such as CHR and Linear-FPC have relatively lower performance than LFPC.

Experiment 2 is designed to verify the efficiency of tonal features when it is combined with LFPC which shows the highest accuracy in experiment 1. Optimal order of LFPC are determined from the preliminary experiments. As shown in Fig. 5, 20th order LFPC and their △ and △△ coefficients achieve the highest recognition rates. This result shows similar tendency with the result of Nwe's study.

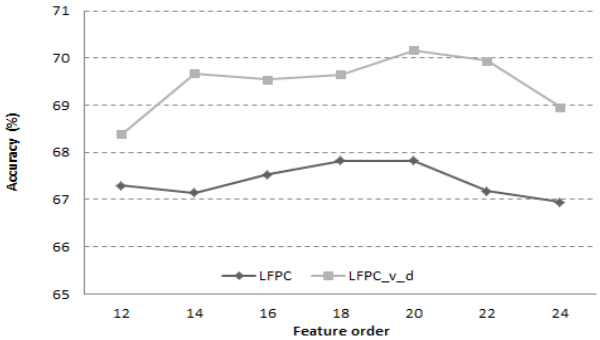

표 5. 12차부터 24차에 따른 LFPC 성능 변화
Fig. 5. Recognition results for LFPC order from 12 to 24

To extract utterance-level features, first and second moments of 20th LFPC and △and △△ coefficients for overall frames are computed. Totally 120 features of LFPC are extracted from one utterance In order to verify the efficiency of proposed features, static measurements of F0 with LFPC is also compared. The utterance-level measurements of F0 and △F0 are mean, std, kurtosis, skewness, median, percentile (5, 25, 50, 75, 95%). Totally, 20 features related F0 are extracted from one utterance. Dimension of reference feature set is 140.

Tonal features are computed first and second moments of 12th chroma features from one utterance. Dimension of proposed feature is set to 144. The result of experiment 2 is summarized in Table 3. Chroma features denoted as CHR3 and CHR4 denote the features extracted from the third and fourth octave frequency range. In the gender independent case, LFPC+CHR4 shows 71.69% accuracy which is the best result in these experiments. In gender dependent experiments, LFPC+CHR4 also shows the highest accuracy, 73.81% for female speakers. Though, LFPC+

CHR4 achieves noticeable improvement for female speaker, LFPC+F0 has slightly higher accuracy than proposed chroma features for male speaker. F0 of male speakers are varied around 100Hz and dynamic range of F0 is relatively smaller than female speakers. Chroma feature analyzes the F0 frequency range by decomposing 12 frequency bands. Thus, it should be less efficient to analyze the tonal information of male speakers which have narrow F0 ranges.

표 3. 실험 2: LFPC와 크로마 피쳐, F0의 조합 실험 결과
Table 3. Experiment 2: recognition results using LFPC with chroma features and F0. CHR3 and CHR4 denote chroma features for 3rd and 4th octaves

Feature (dimension)	accuracy (%)		
	Gender independent	female speaker	male speaker
LFPC (120)	69.89	71.79	69.27
LFPC+F0 (140)	70.73	71.31	72.53
LFPC+CHR3 (144)	70.57	72.37	70.16
LFPC+CHR4 (144)	71.69	73.81	72.01

In order to overcome the drawback of tonal and F0 related feature, combination of tonal and F0 related features are applied with LFPC. Features used in combined features are 164 features which consist of LFPC (120), CHR4 (24), and F0 (20). Experimental results are tabulated in Table 4. Combined features provide improved performance both male and female speakers. Relative improvements of female and male speaker are approximately 1% and 3% compared to the best performance of experiment 2. Because the feature dimensions used in experiment 2 and 3 are different, it does not seem to be a fair comparison. However, increased feature dimension does not always improve the performance as shown in Fig. 3. According to the recognition results, tonal features and F0 related features provide reliable cues of emotional states in glottal source of speech.

표 4. 실험 3:164차 LFPC+CHR4+F0를 이용한 SVM 인식 결과
Table 4. Experiment 3: recognition results of LFPC+CHR4+F0 using SVM. Feature dimension is 164

Gender	neutral	joy	sad	anger	total
gender ind.	80.16	60.83	80.07	68.42	72.37
female	85.86	54.49	80.02	76.73	74.53
male	90.89	62.91	82.68	61.70	74.54

IV. Conclusion

This paper described the efficiency of tonal features in speech emotion. We were motivated by the concept of tonality used for perceiving mood or emotion in music applications. From the feature analysis and subjective hearing tests, we concluded that F0 related octave frequency was suitable to extract tonal features for speech, and they were reliable to detect neutral and sad emotions.

Recently, many studies in emotion recognition have a tendency to focus on increasing the type and size of features and finding out optimal features combinations. The proposed method is meaningful in the sense that human perception is involved in the feature set selection process and relationship between tonality and emotion is analyzed. From the analysis using mutual information, tonal features related to F0 frequency range contain reliable information about emotional states. We observe the improvement of emotion recognition accuracy using chroma based tonal features with spectral features. Furthermore, combined features of tonal features and F0 related features show the noticeable improvement for female and male speaker experiments. Increasing the features associated with glottal source information provides the improvement of recognition accuracy. It means that sufficient features for glottal source and spectral features are required to represent the emotional cues contained in speech signals.

References

[1] R. Cowie, E. Douglas-Cowei, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion Recognition in Human Computer Interaction,” IEEE Signal Processing Magazine, pp. 32-80,

2001.
[2] D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” Speech Communication, vol. 48(9), pp. 1162-1181, 2006.
[3] M. E. Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases”, Pattern Recognition, vol. 44, pp. 572-587, 2011.
[4] I. Murray, J. Arnott, “Toward the simulation of emotion in synthetic speech: A review of the literature of human vocal emotion,” J. Acoust. Soc. Am, vol. 93 (2), pp. 1097-1108, 1993.
[5] C. E. Williams and K. N. Stevens, “Emotion and speech: Some acoustical correlates”, J. Acoust. Soc. Am, vol. 52(4), pp. 1238-1250, 1972.
[6] C. Gobl and A. N. Chasaide, “The role of voice quality in communicating emotion, mood and attitude”, Speech Communication, vol. 40, pp. 189-212, 2003.
[7] M. Goudbeek and K. Scherer, “Beyond arousal: Valence and potency/control cues in the vocal expression of emotion”, J. Acoust. Soc. Am, vol. 128, pp. 1322-1336, 2010.
[8] S. Yacoub, S. Simske, X. Lin, J. Burns, ”Recognition of Emotionsin Interactive Voice Response System,” Proceedings of the Eurospeech 2003, Geneva, 2003.
[9] T. L. Nwe, S. W. Foo, and et al, “Speech emotion recognition using hidden markov models”, Speech Communication, vol. 41(4), pp. 603-623, 2003.
[10] M. A. Bartsch and G. H. Wakefield, “Audio thumbnailing of popular music using chroma-based representation,” IEEE Transactions on Multimedia, vol. 7(1), pp. 96-104, 2005.
[11] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, “Music emotion recognition: A state of the art review”, Proc. 11th Int. Soc. Music Information Retrieval Conf.(ISMIR), pp. 255-266, 2010.
[12] M. Muller, F. Kurth, and M. Clausen, “Audio matching via chroma-based statistical features”, Proc. 5th Int. Soc. Music Information Retrieval Conf.(ISMIR), pp. 288-295, 2005.
[13] T. F. Quatieri, Discrete-Time Speech Signal Processing, Prentice-Hall, NJ, 2002.
[14] H. Purwins, “Profiles of Pitch Classes: Circularity of Relative Pitch and Key: Experiments, Models, Computational Music Analysis, and Perspectives,” Ph. D. dissertation, Berlin Univ. of Technol., Berlin, Germany, 2005.
[15] T. Lan, D. Erdogmus, U. Ozertem, and Y. Huang, “Estimating mutual information using Gaussian mixture model for feature ranking and selection”, Proc. Int. Joint Conf. on Neural Networks, pp. 5034-5039, 2006.
[16] B.-S. Kang, “Text-independent emotion recognition algorithm using speech signal,” M. S. Thesis, Yonsei university, Electrical and Electronic Engineering Department, 2000.
[17] C.-C. Chang and C.-J. Lin, “Libsvm: a library for support vector machines”, ACM Transactions on Intelligent Systems and Technology, vol. 2, 27:1-26, 2011.
[18] P. Shen, Z. Changjun, X. CHen, “Automatic Speech Emotion Recognition using Support Vector Machine,” Int. Conf. on Electronic and Mechanical Engineering and Information Technology (EMEIT), vol. 2, pp. 621-625, 2011.

저 자 소 개

이 정 인

- 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 음성 신호처리, 음성인식, 감정인식

강 홍 구

- 연세대학교 전기전자공학과
- 주관심분야 : 음성 신호처리