Perception of Emotions Using Constructive Learning through Speech

K. Sai Sruthi, Amarjot Singh, N. Sumanth Kumar, and Arundathy Reddy

***Abstract*—Constructive learning is an important research area having wide impact on teaching methods in education, learning theories, and plays a major role in many education reform movements. Teachers play a major role in improving the learning skill of the students. It is observed that constructive learning advocates the interconnection between emotions and learning. When students fail to get the excepted results they tend to feel that they are not good at the subject/task. Teachers should make them realize that failure is also a part of learning process and improve their learning rates. Human teachers identify the emotions of students with varying degrees of accuracy. In learning with computers, computers also should be given the capability to recognize emotions so as to optimize the learning process. Literature survey indicates the wide use of image processing to understand the constructive learning theory. The paper presents a novel system which can be used by computer to access the emotional state of the learner further presenting a corrective measure to improve their learning states. This is the first paper which analyses constructive learning using speech analysis. It is the primary paper which analyses the effect of emotions on the learning rate using pitch tracking. A database consisting of acoustic waveforms produced by an amateur musician is taken and the learning rates are analyzed. The pitch contours of waveforms are compared with the standard waveform and the error graphs are plotted. Analysis of the emotion of the subject is also made by observing the error plots.**

***Index Terms*—Learning rate, autocorrelation, constructive learning, pitch tracking.**

I. Introduction

Emotions and learning are interdependent of each other [2]. Learning means acquiring knowledge and skills which requires thinking. Emotions impact how we think which in turn influences the learning process. Fear, anger, sadness, joy are the most common emotions. Fig 1 shows the transitions between different emotions representing the emotion transition of a person from positive emotion to negative emotion.

The channel for interpersonal communication is human body, which conveys information related to interpersonal attitudes and emotions. The learning process has been rarely modeled by educators, who give importance to conveying information and facts. Because of this, when students don't get desired results, they tend to believe that they are not good at their task/subject. Exceptional teachers make them realize that failure is also a part of learning and recognize their emotional states and take some action to impact their learning positively [2].

In intelligent and affective computing, computers need to recognize and also express emotions to humans. In learning with computers, the learner's emotions are to be identified and accordingly necessary action is to be taken by the computer for optimized learning. Literature survey indicates that many image processing techniques were used to analyze constructive learning and also the interplay between emotions and learning. Face occupies a crucial position in the human perception of emotion. In face to face interactions, facial expressions are main channels for conveying emotions. Many methods like tracking of the facial feature points [4], observing the subjects upper body and eyes [2] were used for the measurement and perception of emotions so as to aid the human machine interactive learning process. By analyzing the emotions, inferences can be drawn about the affective state of the learner which can be used to optimize the learning.

During interaction with a computer tutor, nonverbal behavior from facial expressions, head posture, eye gaze can be used to infer about learner's emotions. The existing databases consist of emotions which are expressed deliberately, so their application to a context like 'learning with a computer' is irrelevant. But nonverbal data is rich, hard to validate, and is also ambiguous. So the collection of data relevant to the context is a time consuming process [7]. Usually the emotional expressions overlap, co-occur or blend subtly into a background expression, this makes demarcating the beginning and end of an emotional expression very difficult. Because of these reasons we make an effort to analyze constructive learning using speech.

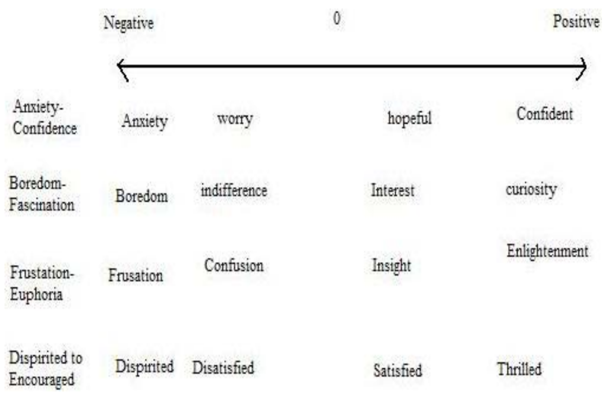

Fig. 1. Indicating the transitions between emotions

In this paper we analyze learning process using pitch tracking. Pitch detection plays a vital role in many speech processing systems. Valuable information about the nature of

Manuscript received April 2, 2011; revised May 15, 2012.

K. Sai Sruthi, is with the Electronics and Communication Department of National Institute of Technology, Warangal, India

N. S. Kumar, A. Singh, and A. Reddy are with the Electrical Department of National Institute of Technology, Warangal, India. (e-mail: amarjotsingh@ieee.org).

the excitation is given by pitch. It is difficult to estimate accurate and reliable pitch due to many reasons [1]. One reason is during voiced speech segments, due to co articulation determining the exact beginning and ending of the pitch periods is difficult. The glottal waveform's (excitation waveform) pitch period is not constant and also within the pitch periods the structure is different, this comprises the second reason. Third difficulty is that the vocal tract formants alter the glottal excitation waveform making it difficult to estimate the pitch.

Wide variety of algorithms have been proposed in the speech processing literature to estimate the pitch period because of its importance and also due the problems involved. The problem associated with the methodology can be easily resolved using the robust and effective algorithms. One of the most robust and reliable technique of them all is the time domain autocorrelation method [3], which operates on the time domain signal itself.

The remaining paper is classified into the following sections. The next section explains about constructive learning and section 3 gives the algorithm and code in detail

and also the application of the algorithm to the learning process. Section 4 presents the results obtained by using the algorithm and section 5 concludes the results.

II. Constructive Learning

Before describing the details of constructive learning, the emotions involved are to be analyzed. Two to twenty basic or prototype emotions have been proposed by previous emotion theories. Fear, anger, sadness and joy are the most common emotions appearing on many theorists' list. Eight basic emotions: fear, anger, sorrow, joy, disgust, acceptance, anticipation, and surprise were distinguished by Plutchik [1980]. Ekman [1992] has focused on a set of six to eight basic emotions that have been associated with facial expressions. However, none of the existing theories seem to address emotions commonly seen in learning experiences, some of which have been noted in Fig. 1[2].

Fig. 2 indicates the interconnection of cognitive dynamics of learning process and emotion axes of Fig.1 [2]. Positive emotions are to the right side of the horizontal axis, whereas the negative emotions are to the left side. Vertical axis above the horizontal axis represents the constructive learning axis while below the horizontal axis symbolizes unconstructive learning. Learning is seen as addition or negation of knowledge to facilitate its pictorial representation. Constructive learning is indicated by the upward vertical axis whereas the negative vertical axis indicates destructive learning.

In the beginning, the player is highly motivated to learn the guitar and song. The state of mind of the player is in the first quadrant of the learning cycle as he is motivated to learn and hence his learning will be constructive. The initial attempts of the player will have large error and will be far away from perfection. With the increase in attempts of the player the error will decrease, as the player learns to play, as shown in the table. Now, if the attempts of the player are not sufficient enough to improve the error, the player becomes frustrated, but still he/she attempts to play well and learn. In this case the learning is constructive but the emotion is negative because of which the player enters into the second quadrant of learning cycle.

Depending upon the future attempts of the player, the learning rate of the player may fall into the third quadrant or may shift back into the first quadrant motivated from the fourth quadrant. If the player is still not able to improve his error, he/she moves into the third quadrant due to depression (negative emotion) and unconstructive learning as in this state the person doesn't add anything to his knowledge. On the other hand, if the player is able to improve the error, the quadrant is shifted back to the first quadrant. Further decrement in the learning rate will lead the person to the fourth quadrant which motivates the person to start again.

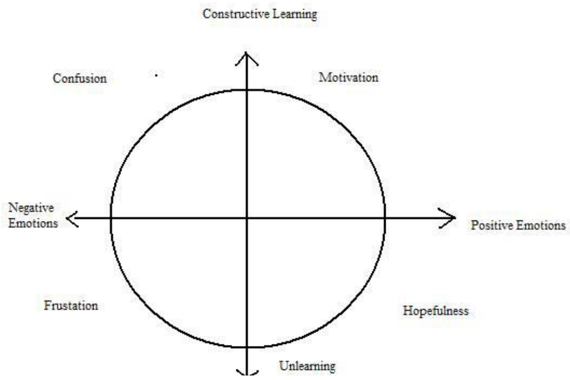

Fig. 2. Indicating the relation between emotions and learning

III. Time Domain Autocorrelation

The most important feature of periodic signals is the similarity of the waveforms at different times. The aim of the algorithm is to find the pitch period by comparing the original signal with its shifted version. When the distance by which the signal is shifted, equals the pitch period, then the autocorrelation function will be maximum. Autocorrelation $R(t)$ is given by

$$R(t) = \sum_{0}^{N-1} s(n)s(n+t)$$

where s(n) is the sampled version of the signal s(t). Analysis frames are taken to calculate autocorrelation between two speech signals. Each frame comprises of 400 samples. The spacing between the starting points of these windows is 10 ms i.e. an analysis frame of length 400 samples and frame shift of 10 ms is taken. The criteria being that the size of our window (frame) should be greater than two pitch periods. The next step is the

computation of autocorrelation for the frame.

Two thresholds T1 and T2 are considered to differentiate between voiced and unvoiced regions. If the maximum value of autocorrelation function is less than T1 (taken as 0.1) or if the number of peaks in the autocorrelation function exceeds T2 (taken as 130), then pitch is taken to be nonexistent (i.e. pitch frequency of that frame is taken zero) for that analysis frame. The resulting values (frequency=1/ (pitch period)) for all the frames is stored in an array and are plotted.

This code shown in figure 3 is used to analyze the learning rate. A standard waveform is taken and pitch contour is computed using the above code. Waveforms of an amateur singer, who tries to sing a song and simultaneously play the tune on guitar, are recorded. Pitch contours are computed using the above code. The errors between the pitch contours of the standard waveform and the recorded waveforms are plotted. The error magnitudes will indicate the learning rate of the person. The learning rates can be used to infer the emotions of the person.

```
[a,Fs,nbits]=wavread('singgui1.wav');
 % reading the wavefile
 k3=400/Fs;j=0;
 for k1=(1/Fs):0.01:(length(a)/Fs)-k3
  j=j+1;k2=k1+k3;n(j)=(k1+k2)/2;
  %taking a frame of 400 samples
  a400=a((k1*Fs):(k2*Fs));
   %computing the autocorrelation of the frame
  lac=xcorr(a400);r=1;
  %calculating the number of peaks in the autocorrelation
  for i=2:length(lac)-1
    if(lac(i+1)<lac(i))
      if(lac(i-1)<lac(i))p(r)=i;r=r+1;
      end
    end
  end
  k4=1;
  while(p(k4)>0)
    if(k4<length(p))k4=k4+1;
    else break;end
  end
  if(k4>=length(p))k5=k4;
  else k5=k4-1;end
  %if the number of peaks exceeds 130 or if maximum magnitude
  %of autocorrelation is <0.01,means the frame indicates unvoiced region
  if(k5>=130)
    disp('autocorrelation erratic-so no pitch');t=0;b(j)=t;
  elseif (max(lac)<0.01)
      disp('autocorrelation almost zero-so no pitch');t=0;b(j)=t;
  else
  %voiced region
 [X,I]=max(lac);k=I;
   while(k<length(lac))
         if(lac(k+1)<lac(k+2))
            if(lac(k)>lac(k+1))
               z=k;break;end
         else k=k+1;end
   end
   l=1;
  for i=z:length(lac)
  q(l)=lac(i);l=l+1;end
  [Y,J]=max(q);t=(z-I+J+1)/Fs;b(j)=1/t;end %pitch is stored in array b
 a400(:)=0;p(:)=0;end
plot(n,b);axis([0 1.5 0 500])
```

Fig. 3. Code for pitch contour extraction.

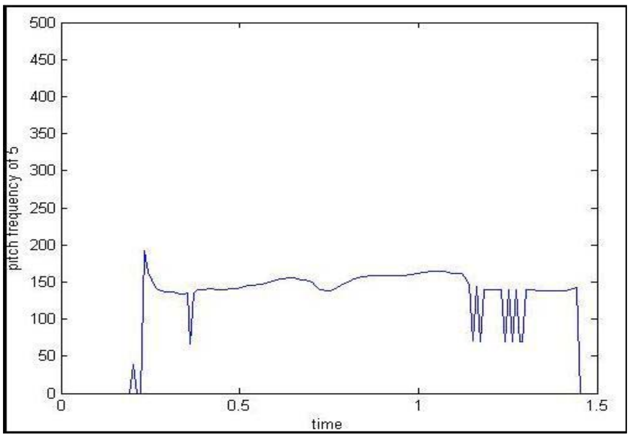

Fig. 4. Reference wave pitch contour.

IV. Results

The aim is of the study is to analyze the state of mind of the player for a continuously improving learning rate. The corpus of the speech data required is formed by the audio recording of first 2 words of the song “khuda jaane” with the instrument (guitar). An amateur musician (male) tries to sing the song and play the guitar simultaneously and his attempts were recorded. The recordings are recorded with a sampling rate of 16KHz encoded in a single channel 16 bit PCM using wave surfer and Logitech microphone. All simulations were carried out on an Intel dual core 1.6 GHz machine. The results obtained from the simulations enable us to investigate the learning rates and the emotions associated with them. Waveforms consisting of both the words and instrumental were recorded. Pitch contour was plotted as shown in fig. 5(a) for the learning rate waveforms using the algorithm explained above. The resultant pitch contours are used to analyze the learning rates. By comparing the pitch contours, learning rates can be understood. The last waveform i.e. waveform-5 was taken as the standard waveform.

The player in the starting is highly motivated to learn the guitar and song. The state of mind of the player is in the first quadrant of the learning cycle show in Fig. 2, as he is motivated to learn hence his learning will be constructive. The initial attempts of the player will have large error and will be far away from perfection. We have taken a threshold error ‘ŋ’ (1.5671e+004) and waveforms with errors equal to and less than this threshold have been displayed in the table. As the person tries to produce perfect (expected) waveform and is unable to do so, he moves into the 2nd quadrant due to frustration. If his waveforms are persistently not good then the player feels depressed and thinks that he can’t play the guitar at all because of which he will be pushed into the third quadrant. The player analyses the ideas which do not work and starts afresh from first quadrant after moving to fourth quadrant. It is observed that during this process the error values will not decrease. With the increase in attempts of the player, the error will decrease as the accuracy of the player towards guitar playing and singing increases. The learning rate of the player improves which leads to the decrease in error incurred by the player with the increase in his attempts. As the attempts of the player increase, the error decreases and it finally becomes minimum when the player achieves perfection. The number of times a person will traverse the cycle varies from person to person and it depends on the person’s motivation levels. We further analyze the waveforms obtained after the playing error of the person moves below the specified threshold ŋ.

We used direct subtraction technique to plot the error rates. The pitch contours of the remaining waveforms were subtracted from the pitch contour of the standard waveform. The subtraction performed was an array element to element subtraction (i.e.; an element in one array is subtracted from the element with the same index in other array). The square of this is plotted against the average value (time) of the analysis frame taken. These comprise the error graphs.

We have used cross correlation technique to compare the similarity between the standard waveform and the other waveforms as shown in fig. 5(b). As the similarity increases the correlation function’s magnitude increases. These are the two techniques used to indicate the learning rates. The average values of the error and cross correlation functions have been indicated in the table. It can be seen that the average values of the error decrease as we go down the table i.e. as the number of attempts towards perfection increases. The decrease in error is not uniform and difference in errors decreases as we go down the table. This decrease in error also indicates that the person is in 1st quadrant of learning cycle, as his error rates are decreasing. So his emotions are positive.

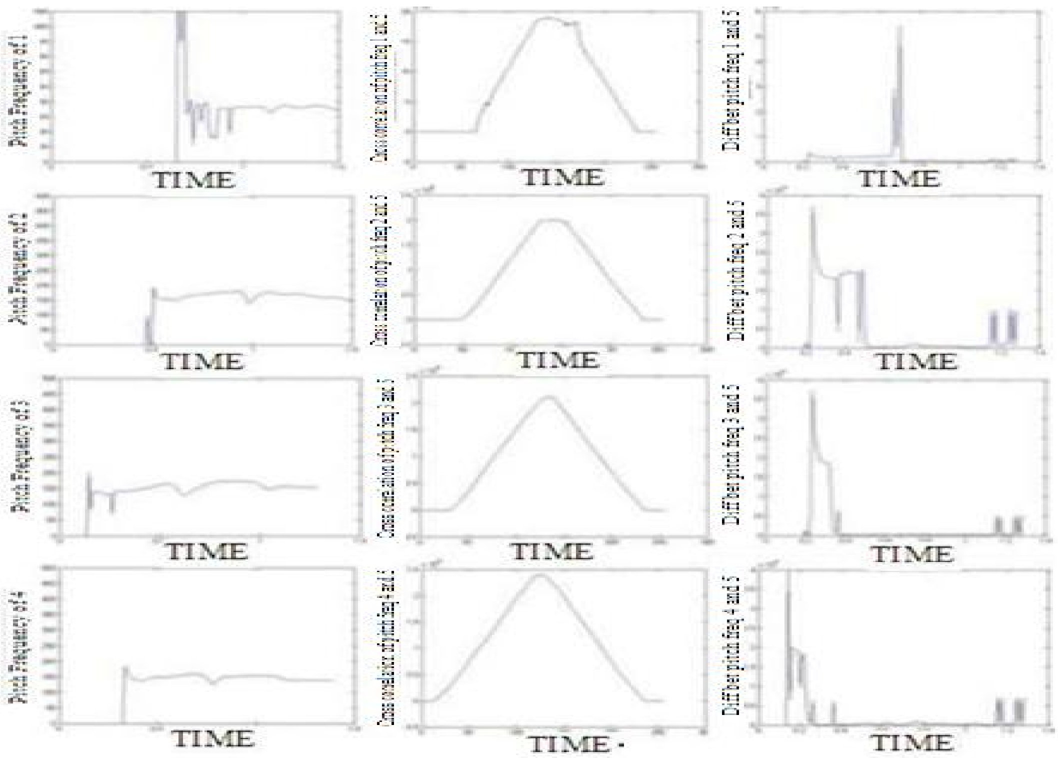

Fig. 5. (a) Pitch contour of recorded waveforms (b) Crosscorrelation of recorded waveforms (c) Error graph for recorded waveform

So based on the patterns of error s shown in fig 5 (c), the quadrant the person is in can be assessed. Based on this his emotions can be recognized and the computer's can give the necessary interventions to improve the learning rate. The average values of the cross correlation functions increase as we go down the table i.e. as the number of attempts towards perfection increases. This means that the similarity between the standard waveform and the waveforms taken into account is increasing. The difference between two consecutive cross correlations increases as we go down the table as shown in fig. 6.

Waveforms(compared to the standard waveform)	Average value of errors	Average values of cross correlations
Waveform-1	1.5671e+004	7.5088e+005
Waveform-2	4.6653e+003	8.0459e+005
Waveform-3	2.0592e+003	8.6318e+005
Waveform-4	1.9155e+003	1.0675e+006

Fig. 6. Average error and correlation value with respect to every waveform

The code takes 0.735 seconds to produce each pitch contour. The runtime of error calculation process is 4.26 seconds whereas for cross correlation is 4.35seconds (these are calculated taking the time taken to produce pitch contours of five waveforms also into account).

V. Conclusion

This section presents a concise summary of the paper. This paper presents a novel approach to analyze the learning rate from the speech waveform. The paper analyses the learning rate of the musician playing the guitar as well as singing a song simultaneously. The error graphs of the player with relation to the emotion are studied which can be used to improve the learning rate of the player. Depending upon the improvement or decrement in the learning of the person the future quadrant is decided. If after the first attempt, the error of the player increases, the person gets frustrated but simultaneously attempts to learn because of which the person moves into the 2nd quadrant. Now in order to improve the learning rate of the person or to stop him from moving into the third quadrant, the person needs motivation or his mood has to be changed (lightened) which is highly influenced by the instructor. The instructor should analyze the emotional state of the person carefully and should try to move the person back to first quadrant because if he/she moves to the third quadrant (state of 'depression') it is highly likely that the person will move to the fourth quadrant which will lead to a fresh start of the whole process. This can position the person in a state of depression and the person may start doubting his capabilities.

This paper can be used to study and improve the learning rate of students. It is an effective methodology by which students can be helped to learn effectively in a better way which will improve their leaning rate. Excellent human teachers identify the student's emotion and take actions to impact the learning rate positively. In situations like learning with computers, computers need to have emotion identification abilities so as to assist the learner in a better way. By observing the error trends, necessary actions can be taken by the instructor so as to optimize the learning. The decrease in error indicates that the person is in 1st quadrant of learning cycle. So his emotions are positive. If the error is not decreasing with the number of attempts, it indicates the person might move into the quadrants of negative emotion. So based on the patterns of error, the quadrant the person is in can be assessed. Based on this, emotions can be recognized and the computer's can give the necessary interventions to improve the learning rate.

VI. Application

In professions like SMET (science, math, engineering and technology) failure is a natural part of learning process. To forbid the students from going into the negative quadrants of emotion, the above technique can be used in cases where repeated attempts are being made to master a task.

References

[1] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. M. Gonegal, "A Comparative Study of Several Pitch Detection Algorithms," *IEEE transaction on Acoustics , Speech and signal processing*, vol. ASSP 24, no.5, October 1976.

[2] B. Kort, R. Reilly, and R. W. Picard, "An effective model of interplay between emotions and learning: Reengineering Educational Pedagogy-Building a Learning Companion".

[3] L. R. Rabiner "On the use of autocorrelation analysis for pitch detection," *IEEE transactions on acoustics, speech and signal processi*ng, vol. ASSP 24, no. 1, 1977.

[4] S. Afzal, T. M. Sezgin, Y. Gao, and P. Robinson "Perception of emotional expressions in different representations Using facial feature points," Affective computing and Intelligent Interaction and workshops, 2009.

[5] K. Leidelmeijer, Emotions: An Experimental Approach Tilburg University Press, 1991

[6] R. Plutchik, "A general psycho evolutionary theory of Emotion," *in Emotion Theory, Research, and Experience* (R. Plutchik and H. Keller man, eds), vol. 1, Theories o Emotion, Academic Press, 1980.

[7] S. Afzal and P. Robinson "Natural affect data-Collection and annotation in learning context," *Affective computing and Intelligent Interaction and Workshops*, 2009.