Journal of Theoretical and Applied Information Technology

www.jatit.org

LAUGHTER INQUISITION IN AFFECT RECOGNITION

[1]Nachamai. M, [2]T.Santhanam
[1]Research Scholar, Mother Teresa Women's University, Kodaikanal, India.

[2]Head, PG & Research Dept. of Computer Applications, DG Vaishnav College, Chennai, India.

E-mail: subha_muthu@yahoo.com, santhanam_dgvc@yahoo.com

ABSTRACT

Laughter and humor are major ingredients of humanity but does not have any black and white authenticate. Laughter is a physiological process, which activates facial, respiratory and laryngeal muscles. Laughter may occur instinctively - in response to humor or to appropriate emotional or sociological stimuli and can also be elicited upon command – voluntary, contrived or faked laughter. Exploring the pattern of laughter is an intricate and arduous errand. Speech and laughter are quite disparate in the vocalization, duration and in regularity of its occurrence. This work is a convincing attempt to catalog the different types of laughter as positive laughter, negative laughter, laughed speech and vague category. Feature selection was done using Radial Basis Boltzmann Machine Network, categorization was done using an Ergodic- Hidden Markov Model. The corpus used for experimental study was the International Computer Science Institute (ICSI) Meeting Corpus from which 30 meetings were taken as the data set. The technique used has yielded a persuasive result in categorizing the types of laughter.

Keywords: Speech *recognition, Laughter, Neural Network (NN) , Hidden Markov Model(HMM).*

1. INTRODUCTION

Laughter is one of the renowned features of human to human communication. Laughter has a conscientious culture which adds color to speech. Negative forms of laughter are quite common in today's life. The laughter pattern is a mixture of contextual/ semantic impulse that puts the speaker in a laughing state. While laughing there are bursts of air exhalation (along with audible voicing) and aspiration (unvoiced segment) that each last for a petite period.

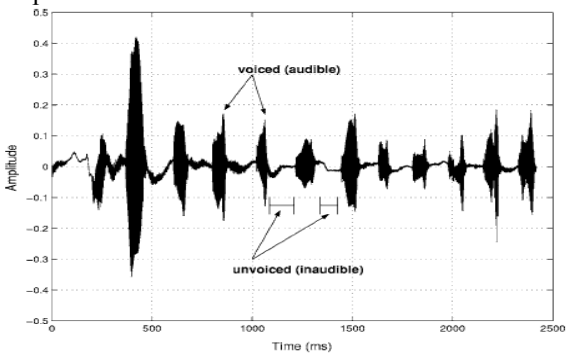

Fig. 1 Laugh cycle with intermittent laugh pulses

The intermittent voicing pattern can be seen as an "oscillatory behavior" that can be monitored in most laughter spells [1]. An episode of laughter begins with an inhalation to its end is known as a "Laughter Bout". Each bout comprises of alternating voiced and unvoiced sections. This alternating texture is termed as "Laughter Cycle" with intermediate "Laughter Pulses" along with aspiration sounds in between [2]. A laugh cycle is depicted below.

The types of laughter categorized in this paper are given below in the tabulation.

Positive laughter	Elation, Ecstasy, Joy, Delight, Benevolent, Compassionate, Concerned, Poised, Sophisticated, Respite.
Negative laughter	Hassle, jittery, Harassed, Worried, Humiliated, Discomfited.
Vague	Ironical, Criticism, Mock, Poignant, Weird, incongruous
Laughed Speech	Simultaneous talking and Laughing

2. METHODOLOGY

The methodology uses an amalgam of neural network with a mathematical model and proved to be very profiicient [3].

2.1 Laughter segmentation

The first step includes seceding the laughter occurrences from the incoming speech signals. Start and stop labels are prefixed at each episode of laughter. Laughter occurrence in speech does not have a fastidious or customary pattern, so it is intricate to mark these random signals [4]. The experimental corpus is placed with .stm files which comprises the sound instances are adjacent on both the left and right on a time stamped uttering boundary, so the end points are marked.

2.2 Feature Extraction

The feature vector is extricated from the input signals segmented only with laughter episodes [5]. The features that were construed are

2.2.1 Spectral entropy

The spectral entropy is a useful feature for end-point detection and is superior to energy [5]. Entropy is the measure of disorganization used to measure the peak in a distribution. Entropy values are lower when speech occurs in the signal due to clear formants, and entropy values are higher in the case of silence. The entropy feature improves performance and robustness when additional noise is encountered. The improved feature of spectral entropy that was used to increase robustness against various noises was

$$\Delta P_i = P_i^{'} - P_i$$

$$= \frac{1}{\frac{\sum_{k=0}^{N-1} Y(f_k)}{k} + N}.(1 - N.P_i)$$

2.2.2 Kurtosis

It is the 4th order moment of a signal divided by the square of its 2nd order moment [6]. The value of kurtosis is high for isolated speech and low when overlapping occurs. The frequency domain kurtosis was reckoned with the magnitude spectrum.

2.2.3 Pitch prediction feature

The listener's ability to discern two different pitches is dependent on frequency, type and loudness of the tone. Pitch finding is grueling in quiet signals [7]. If a frame consists of two speakers the difference between them can be seen by computing the residual error with Linear Predictive Coding values and the standard deviation of the inter-peak differences will be high. The method of autocorrelation was adopted for calculating the PPF. $R_n(k) = \sum_m x_n(m) x_n(m-k)$

$R_n(k)$ - An even function, is a projection of $x_n(m)$ onto $x_n(m-k)$ so it is maximum at k=0.

For resolving Fo- the pitch the main lobe width of the window length was taken ($2*\omega_c$) not exceed Fo.

2.2.4 Cross – Correlation Ratio (CCR)

It is the measure of similarity of two signals commonly used to acquire features in an unknown signal by comparing it with the renowned one. It is a function of relative time between the signals. It is analogous in nature to the convolution of two functions, the only amend in cross- correlation is signals are shifted and multiplied by another signal and not reversed.

The features were cliqued using a RBBM Network [8]. The Radial Basis Function Networks are "Universal Approximators" [9]; with enough hidden neurons the network can approximate any continuous function with arbitrary precision. The Boltzmann Machine espousing the simulated annealing technique can be entitled as a stochastic recurrent neural network. They are adroit of learning internal portrayals and solve difficult combinatorial predicaments. The network is compiled of units with "energy" demarcated for the each. The global energy E is defined as

$$E = -\sum_{i<j} W_{ij} S_i S_j + \sum_i \theta_i S_i$$

$$S_i - State\ S_i \in \{0,1\}$$

The Boltzmann machine is made up of stochastic units. The probability P_i of the i^{th} unit is given by

$$P_i = \frac{1}{1+\exp(\frac{-1}{T}\Delta E_i)}$$

For high value that is for large variance the generalization error proliferates linearly. Below the variance symmetry breaking phenomenon depends critically on the value of the 4^{th} cumulant. If the 4^{th} cumulant is not equal to zero the generalization bias dwindles. Since, training error is an approximate constant in the vicinity of the symmetry breaking point this model gives slightly enhanced generalization error. For selection of optimal network edifice the predictive minimum description length principle was employed - to ascertain optimal number of input nodes. So, the over and under-fitting problem in the network is taken care automatically. The network works best with noise attenuation and generalization capabilities.

2.3 Laughter Classification

An exceptional noise modeling HMM was used for classification [10]. The selected feature set was noshed into a 4-state E-HMM [11]. The 4 states commune to the four categories to be made. Ergodic embrace the perception of randomness. Meeting speech is considered as an ergodic series of data, there is no memory, and no correlation need to be present with past data so each new data point adds the same amount of new information. For every frame the accumulated log-probability is evaluated by Viterbi search. Miscellany of log-likelihood and log-probability gives a new score- called ergodic score which is used for the categorization.

3. RESULTS AND CONCLUSION

The International Computer Science Institute (ICSI) meeting corpus [12] deployed in this study encloses 75 different meetings, of which 30 meetings were used that contained annotated laughter .stm files along with the input files totaling around 20 hrs of time. The experimental results show that the spectral entropy is high for positive laughter and low for vague catalog. The CCR value ascertains clearly the laughed speech category as it finds more than one patterned signal coming in from the input. . Despite the factual that the patterns of speech and laughter are different, Laughed speech category is the most grueling to identify as speech and laughter occur at the same time frame. Cataloging the mood of the laughter to be identified in such a category is a Herculean task. The lower the PPF value laughter falls into the negative category.

Category	Samples	Correctly classified	% of correct classification
Positive	42	41	97.62
Negative	31	25	80.65
Laughed	19	12	63.16
Vague	08	6	75

The suggested approach has produced an accuracy rate of 79.11% in detecting the category of laughter when compared with the built in annotation given along with the corpus.

REFERENCES

[1]. *Shiva Sundaram and Shrikanth Narayanan, "Automatic acoustic synthesis of human like laughter"*, Journal of Acoustical society of America, 2007, Vol. 121, No.1, pp. 527-535.

[2]. *Bachorowski et al., "The Acoustic features of human like laughter "*,Journal of Acoustical society of America , 2001, Vol. 110, No. 3, pp.1581-1597.

[3]. *Kiet.P.Truong and David A.Van Leeuwena, "Automatic discrimination between laughter and speech "*,Speech Communication, 2007, Vol. 49, No. 2, pp.144-158.

[4]. *Kornel Laskowski and Sussane Burger, "Analysis of the occurrence of laughter in meetings"* Proc. of X ICSA International Conference on Spoken LanguageProcessing, 2007, pp. 1258-1261.

[5]. *Chuan JIA and Bo XU, "An improved entropy-based endpoint detection algorithm",* ISCSLP, 2002, pp. 96.

[6]. *M.A.Lewis and R.P.Ramachandran, "Cochannel speaker count labeling based on the use of cepstral and pitch prediction derived features*", Pattern Recognition, 2001, Vol. 34, pp. 499-507.

[7]. *Laurence Devillers and Laurence Vidrascu, "Positive and negative emotional states behind the laughs in spontaneous spoken dialogs"*, Interdisciplinary Workshop on The Phonetics of Laughter , 2007, pp. 37-40.

[8]. *Shotaro Akaho and Hilbert J.Kappen, " Nonmonotonic generalization bias of Gaussian mixture models"*, Neural Computation, 2000, Vol. 12, No.6, pp. 1411-1428.

[9]. *Mark A. Kon, "Neural Networks, Radial basis functions and complexity*", in Statistical Physics Proceedings, Bialowieza, 1997, pp. 322-335.

[10]. *Kazuhiko Ozeki, "Likelihood normalization using an Ergodic HMM for continuous speech recognition"*, ICSLP, 1996, pp. 2301-2304.

[11]. *Laskowski .k and Schultz, "Modeling vocal interaction for segmentation in meeting recognition"*, Proc. MLMI, 2007.

[12]. *A. Janin, D.Baron, J. Edwards, D.Ellis, D.Gelbart, N.Morgan, B.Peskin, T.Pfau, E.Shriberg, A.Stolcke, and C. Wooters, "The ICSI meeting corpus",* in Proc. ICASSP 2003, pp. 364-367.