INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at:
Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.
as well as inOpen J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)]
Registered & Listed at: Index Copernicus Publishers Panel, Poland
Circulated all over the world & Google has verified that scholars of more than 1388 Cities in 138 countries/territories are visiting our journal on regular basis.
Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA
www.ijrcm.org.in

CONTENTS

CHIEF PATRON

PROF. K. K. AGGARWAL
Chancellor, Lingaya's University, Delhi
Founder Vice-Chancellor, GuruGobindSinghIndraprasthaUniversity, Delhi
Ex. Pro Vice-Chancellor, GuruJambheshwarUniversity, Hisar

PATRON

SH. RAM BHAJAN AGGARWAL
Ex.State Minister for Home & Tourism, Government of Haryana
Vice-President, Dadri Education Society, Charkhi Dadri
President, Chinar Syntex Ltd. (Textile Mills), Bhiwani

CO-ORDINATOR

MOHITA
Faculty, Yamuna Institute of Engineering & Technology, Village Gadholi, P. O. Gadhola, Yamunanagar

ADVISORS

DR. PRIYA RANJAN TRIVEDI
Chancellor, The Global Open University, Nagaland
PROF. M. S. SENAM RAJU
Director A. C. D., School of Management Studies, I.G.N.O.U., New Delhi
PROF. S. L. MAHANDRU
Principal (Retd.), MaharajaAgrasenCollege, Jagadhri

EDITOR

PROF. R. K. SHARMA
Professor, Bharti Vidyapeeth University Institute of Management & Research, New Delhi

CO-EDITOR

MOHITA
Faculty, Yamuna Institute of Engineering & Technology, Village Gadholi, P. O. Gadhola, Yamunanagar

EDITORIAL ADVISORY BOARD

DR. RAJESH MODI
Faculty, YanbuIndustrialCollege, Kingdom of Saudi Arabia
PROF. PARVEEN KUMAR
Director, M.C.A., Meerut Institute of Engineering & Technology, Meerut, U. P.
PROF. H. R. SHARMA
Director, Chhatarpati Shivaji Institute of Technology, Durg, C.G.
PROF. MANOHAR LAL
Director & Chairman, School of Information & Computer Sciences, I.G.N.O.U., New Delhi
PROF. ANIL K. SAINI
Chairperson (CRC), GuruGobindSinghI. P. University, Delhi
PROF. R. K. CHOUDHARY
Director, Asia Pacific Institute of Information Technology, Panipat
DR. ASHWANI KUSH
Head, Computer Science, UniversityCollege, KurukshetraUniversity, Kurukshetra

DR. BHARAT BHUSHAN
Head, Department of Computer Science & Applications, GuruNanakKhalsaCollege, Yamunanagar
DR. VIJAYPAL SINGH DHAKA
Dean (Academics), Rajasthan Institute of Engineering & Technology, Jaipur
DR. SAMBHAVNA
Faculty, I.I.T.M., Delhi
DR. MOHINDER CHAND
Associate Professor, KurukshetraUniversity, Kurukshetra
DR. MOHENDER KUMAR GUPTA
Associate Professor, P.J.L.N.GovernmentCollege, Faridabad
DR. SAMBHAV GARG
Faculty, M. M. Institute of Management, MaharishiMarkandeshwarUniversity, Mullana
DR. SHIVAKUMAR DEENE
Asst. Professor, Dept. of Commerce, School of Business Studies, Central University of Karnataka, Gulbarga
DR. BHAVET
Faculty, M. M. Institute of Management, MaharishiMarkandeshwarUniversity, Mullana

ASSOCIATE EDITORS

PROF. ABHAY BANSAL
Head, Department of Information Technology, Amity School of Engineering & Technology, Amity University, Noida
PROF. NAWAB ALI KHAN
Department of Commerce, AligarhMuslimUniversity, Aligarh, U.P.
DR. ASHOK KUMAR
Head, Department of Electronics, D. A. V. College (Lahore), AmbalaCity
ASHISH CHOPRA
Sr. Lecturer, Doon Valley Institute of Engineering & Technology, Karnal
SAKET BHARDWAJ
Lecturer, HaryanaEngineeringCollege, Jagadhri

TECHNICAL ADVISORS

AMITA
Faculty, Government M. S., Mohali
MOHITA
Faculty, Yamuna Institute of Engineering & Technology, Village Gadholi, P. O. Gadhola, Yamunanagar

FINANCIAL ADVISORS

DICKIN GOYAL
Advocate & Tax Adviser, Panchkula
NEENA
Investment Consultant, Chambaghat, Solan, Himachal Pradesh

LEGAL ADVISORS

JITENDER S. CHAHAL
Advocate, Punjab & Haryana High Court, Chandigarh U.T.
CHANDER BHUSHAN SHARMA
Advocate & Consultant, District Courts, Yamunanagar at Jagadhri

SUPERINTENDENT

SURENDER KUMAR POONIA

CALL FOR MANUSCRIPTS

Weinvite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email addresses: infoijrcm@gmail.comorinfo@ijrcm.org.in.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION**:

DATED: _____________

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE**: The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS**: The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT**: Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS**: Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT**: Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS**: All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS**: All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT**: The main text should follow the following sequence:

   INTRODUCTION

   REVIEW OF LITERATURE

   NEED/IMPORTANCE OF THE STUDY

   STATEMENT OF THE PROBLEM

   OBJECTIVES

   HYPOTHESES

   RESEARCH METHODOLOGY

   RESULTS & DISCUSSION

   FINDINGS

   RECOMMENDATIONS/SUGGESTIONS

   CONCLUSIONS

   SCOPE FOR FURTHER RESEARCH

   ACKNOWLEDGMENTS

   REFERENCES

   APPENDIX/ANNEXURE

   It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES**: These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS**:These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES**: The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITE

Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

A SYSTEM FOR EMBEDDING FIVE TYPES OF EMOTIONS IN SPEECH: USING TIME DOMAIN PITCH SYNCHRONIZATION OVERLAP AND ADD (TDPSOLA)

MAMTA SHARMA
LECTURER
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA

MADHU BALA
LECTURER
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA

ABSTRACT

Speech is the primary means of communication between people .To interact with computer, more naturally embedding emotions in speech is a step in this direction. The aim of the paper is to make an attempt to shed light on the way the system for embedding different emotions in the speech using the Time domain pitch synchronization overlap and Add (TD-PSOLA), that is used to enhance the neutral vocal expressions impact through the use of this system, with some required target emotions by selecting the modified parameters that will make humans perceive a targeted emotion in the way the system wants them to understand. An experiment was performed using the Punjabi voices both male and female which led to clear and synthesized emotional speech by using the Time domain pitch synchronization overlap and Add (TD-PSOLA). The experiment concludes with the findings that the system developed with the use of TD-PSOLA can be successfully used to embed emotions in the input neutral speech leading to better human understanding and evaluation of emotional speech.

KEYWORDS

Embedding Emotion, Speech, Speech Modeling, Time Domain Pitch Synchronization.

1. INTRODUCTION

This Paper concerns the process of embedding an emotion in the speech using Time Domain Pitch Synchronization Overlap and Add (TDPSOLA). The goal was to modify the pitch of the speech that can portray emotion with different levels of intensity. To achieve this, the system was based on theoretic frameworks developed by Psychologists to describe emotions.

The basic goal is to perform synthesize the speech so that it sounds naturally. To increase the naturalness of the synthesized speech, the synthesized speech should deliver certain content in right emotion therefore making the speech and content more believable. Emotions can make the interaction with the computer more natural because the system reacts in ways that the user expects. Embedding emotion to the speech is a step in this direction. In this paper it is described that recognition for synthesis system is to automatically select a set of possible parameter values that can be used to resynthesize emotional speech. The system deals with the synthesis using TD-PSOLA. The modification of the input speech is only prosody modifications.

2. LITERATURE REVIEW

2.1 SPEECH PRODUCTION

Speech sounds are produced by causing modulation of the airflow through constrictions in the airways between the larynx and the lips. This modulation of the flow gives rise to the generation of sound. The acoustic process involved in production of speech sounds can be modeled as in Fig-1 shown under the figures section.

When the vocal folds are appropriately positioned and the pressure is raised in the airways below the glottis, the folds are set into vibration and the airflow through the glottis is modulated periodically. The spectrum of this modulated flow is rich in harmonics. The frequency of vibration of the vocal folds during normal speech production is usually in the range 80-160 Hz for adult males, 170-340 Hz for adult females, and 250-500 Hz for younger children [14].

The transfer function of the vocal tract for vowels has a relatively simple form for the special case in which the area function is uniform, the acoustic losses are neglected, and the radiation impedance at the mouth opening is assumed to be small. For a cylinder tube of length x, this transfer function is expressed as equation (Eq-1).The velocity of the sound, v, at body temperature is 354 m/s, and the length of a typical adult male vocal tract is 0.17 m [14]. For male speech the normal range of variation is F1 = 180-800 Hz, F2 = 600-2500 Hz, F3 = 1200- 3500 Hz, and F4 = 2300-4000 Hz. The average distance between formants is 1000 Hz. Females have on the average 20% higher formant frequencies than males, but the relation between male and female formant frequencies is nonuniform and deviates from a simple scale factor [3].

2.2 SPEECH MODELING

2.2.1 SOURCE FILTER SPEECH MODELING

A rather useful model of speech production consists of a filter that is excited by either a quasiperiodic train of impulses (for voiced sounds) or a random noise source (for unvoiced sounds). The source-filter model realized by electrical circuits was first proposed by H. Dudley at Bell Laboratories in the 1930s [4]. At present, two types of the source-filter models are useful for speech processing: the all-pole model known as the autoregressive (AR) model, and the pole-zero model known as the autoregressive moving average (ARMA) model [5]. The AR model of a vocal tract is well known in speech processing as a linear predictive coding (LPC) model [6]. Another type of the source-filter speech model is the cepstral model using homomorphic signal processing based on the idea of the log magnitude approximation filter.

2.2.2 SINUSOIDAL SPEECH MODELING

2.2.2 (a) SINUSOIDAL MODEL IN SPEECH CODING AND SYNTHESIS

When compared with the source-filter model, a rather different approach represents that a sinusoidal speech model is simple because it models the speech signal as a sum of sine waves with defined frequencies, amplitudes, and phases. Perhaps, the first most detailed description of speech analysis/synthesis based on a sinusoidal model was presented in 1986 by R. J. McAulay, and T. F. Quatieri [7], [10], although some information about sinusoidal and harmonic coding and synthesis had been published a few years before also by other authors. In this model, first in every frame the amplitudes are computed from the local maxima of the magnitude spectrum and the phases are determined from the phase spectrum at the corresponding frequencies. However, this model cannot be used for speech synthesis or speech coding at low rates because of a very high number of sinusoidal parameters.

In [8] the authors propose the overlap-and-add (OLA) method with triangular, Hanning, or trapezoidal window instead of a computationally expensive matching algorithm with linear interpolation of amplitudes and cubic interpolation of phases. The sinusoidal model is suitable for prosodic modifications that are necessary in the text-to-speech (TTS) systems.A time-scale and pitch modification system that preserves shape-invariance property during voicing is done using a version of the sinusoidal analysis/synthesis system modeling and independently modifying the phase contributions of the vocal tract and the vocal chord excitation [11]. In the sinusoidal model was compared with a code-excited linear prediction (CELP) concluding with complementarity of the two methods. The authors state that an ideal coder should be able to combine the noise-free quality of sinusoidal models with the robust analysis-by- synthesis procedures of CELP coders [9].The harmonic sinusoidal analysis is used to encode the periodic part of speech and to split the speech spectrum into two frequency regions of harmonic and random components.

Harmonic speech model is a multiband excitation (MBE) model. It uses an analysis-by-synthesis method, in which Voiced speech is synthesized from the voiced envelope samples by summing the outputs of a band of sinusoidal oscillators running at the harmonics of the fundamental frequency. Unvoiced speech is synthesized from the unvoiced envelope samples by first synthesizing a white noise sequence. Its normalized Fourier transform is multiplied by the spectral envelope and then synthesized using the weighted OLA method.

Synthesis is also performed in a pitch-synchronous way using an OLA process. If the frame is voiced, the noise part is filtered by a high-pass filter with cutoff frequency equal to the maximum voiced frequency. At AT&T Labs-Research, HNM was compared with a time-domain pitch-synchronous OLA (TD-PSOLA) method [13]. TD-PSOLA relies on the speech production model described by the sinusoidal framework, although the parameters of this model are not estimated explicitly.

2.3 INTRODUCTION TO SPEECH SYNTHESIS

Speech is the primary means of communication between people. Speech synthesis, automatic generation of speech waveforms, has been under development for several decades. The text-to-speech (TTS) synthesis procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from this phonetic and prosodic information as in Fig-2.

2.4 EMOTIONS IN SPEECH

There are four basic traditions in emotion research in Psychology [1]. Each theory focusing on different components and making different assumptions on what is important for describing an emotion.

2.4.1 THE DARWINIAN PERSPECTIVE

Charles Darwin in his book *The Expression of Emotion in Man and Animals* laid the groundwork for much of modern psychology and also for emotion research. He describes emotions as reaction patterns that were shaped by evolution. This implies that emotions are common in all human beings and also that some emotions might be shared with other animals.

2.4.2 THE COGNITIVE PERSPECTIVE

Cognitive emotion theories relate emotions to appraisal, which is the automatic evaluation of stimuli by low level cognitive processes. Scherer‟s component process model [15] makes physiological predictions relevant to speech from such appraisal processes.

2.4.3 THE SOCIAL CONSTRUCTIVIST PERSPECTIVE

In the social constructivist perspective, emotions are seen as socially constructed patterns that are learned and culturally shared [1]. Emotions have a social purpose that regulates the interaction between people.

2.5 DESCRIPTIVE SYSTEMS OF EMOTIONS

This section is only supposed to give a quick overview over the two main description systems used by most researchers.

2.5.1 EMOTION CATEGORIES

Different emotion theories have different methods for selecting such basic emotion words. In a Darwinian sense the basic emotions have been evolutionarily shaped and therefore can be universally found in all humans. There is a Jamesian extension that expects to find specific patterns for the basic emotions in the central nervous system.

2.5.2 CIRCUMPLEX MODEL OF AFFECT

Instead of independent emotion categories, several researchers have described an affective space [12]. Russell has developed a circular ordering of emotion categories that makes it straightforward to classify an emotion as close or distant from another one. By having subject’s rate similarity of different emotion words and converting the ratings into angles, a circular ordering emerged as in Fig-3.

2.6 EMOTION IN SPEECH

Vocal expression has been recognised as one of the primary carriers of affective signals for centuries. Darwin in his pioneering monograph on the expression of emotions in man and animals underlined also this importance of the voice as an affective channel. Notably, Scherer has done important work in his studies on acoustic profiles. But many other studies have been undertaken that examine the relationship of vocal expression and emotion. To examine the vocal correlates of emotions, one has to analyse a speech database. The source of the content of such a database has been widely debated.

2.6.1 SOURCES OF EMOTIONAL SPEECH

To obtain authentic emotional speech data is one the biggest challenges in speech and emotion research. The most frequently used and oldest method is to have an actor portray certain emotions. This method has the advantage of control over verbal, phonetic, and prosodic speech content. Because only the emotion is varied in the actors‟ portrayal, direct comparisons of voice quality and prosody between different affective states are feasible. Another advantage is that it is easier to obtain expressions of full blown and extreme emotions.

The elicitation of authentic emotions in participants in a laboratory has been tried by number of researchers. There are a few techniques that are termed mood induction procedures which can be used in different settings.

2.6.2 ACOUSTIC CORRELATES OF EMOTIONS

During evolution speech was added to a ”primitive analog vocal signaling system”. This means that the study of speech parameters expressing emotions is very complex. Acoustic parameters vary in their function of linguistic information carriers and non-verbal information carriers. Therefore it is not clear which parameters should be measured. Parameters like voice quality are important carriers of emotion in speech but are very difficult to measure.

3. IMPORTANCE OF STUDY

This has many applications especially in the field of assisting blind and deaf people, speaker verification and security.

4. PROBLEM DEFINITION

Considering the wide range of possible modifications that can be applied on a speech signal to synthesize emotional speech, there is a need for a system that can select the parameters that are expected to perform well, thus narrowing down the sample set that needs to be evaluated by human raters. Considering the significantly different performances for different emotions, and the differences observed between human and machine perception of emotions, however, at this stage we prefer to view the proposed automated evaluation more as a preprocessing step than a replacement to human evaluations.

One of the biggest challenges in embedding emotion in speech is the selection of modification parameters that will make humans perceive a targeted emotion. The best selection method is by using human raters. However, for large evaluation sets this process can be very costly.

5. OBJECTIVE

The objective is to embed emotion in speech by using the time domain pitch synchronization Overlap and Add (TD-PSOLA). To do so we also require reading the recorded speech wave and its different parameters and then create an environment in MATLAB to implement the complete process of embedding emotion in speech using TD-PSOLA.

6. HYPOTHESES

6.1 EMBEDDING EMOTION IN SPEECH

Different synthesis techniques allow control over voice parameters in varying degrees. In most previous systems, only three to nine full blown emotions were modeled. These systems are:

6.1.1 RULE BASED SYNTHESIS

Rule based synthesis, also known as formant synthesis, creates speech through rules of acoustic correlates of speech sounds. Although the resulting speech sounds quite unnatural and metallic, it has the advantage that many voice parameters can be varied freely.

6.1.2 DIPHONE SYNTHESIS

Diphone synthesis uses recorded speech that is concatenated. The Diphone recording is usually made in a monotonous pitch and an F0 contour is generated through signal processing at synthesis time. This technique allows only limited control over voice parameters.

6.2 SPEECH SYNTHESIS TECHNIQUES AND ALGORITHMS

Synthesized speech can be produced by several different methods. The methods are usually classified into three groups. Articulatory synthesis, which attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of vocal tract based on source-filter-model. Concatenative synthesis, which uses different length prerecorded samples derived from natural speech.

6.3 PSOLA METHODS

It is actually not a synthesis method itself but allows prerecorded speech samples smoothly concatenated and provides good controlling for pitch and duration. There are several versions of the PSOLA algorithm.

Time-domain version, TD-PSOLA, The basic algorithm consist of three steps (Charpentier et al. 1989, Valbret et. al 1991). The analysis step where the original speech signal is first divided into separate but often overlapping short-term analysis signals (ST), the modification of each analysis signal to synthesis signal, and the synthesis step where these segments are recombined by means of overlap-adding. Short term signals *xm(n)* are obtained from digital speech waveform *x(n)* by multiplying the signal by a sequence of pitch-synchronous analysis window *hm(n)* as expressed in equation (Eq-8). The windows, which are usually Hanning type, are centered around the successive instants *tm*, called pitch-marks. These marks are set at a pitch-synchronous rate on the voiced parts of the signal and at a constant rate on the unvoiced parts. The used window length is proportional to local pitch period and the window factor is usually from 2 to 4 (Charpentier 1989, Kleijn et al. 1998). The pitch markers are determined either by manually inspection of speech signal or automatically by some pitch estimation methods (Kortekaas et al. 1997). The segment recombination in synthesis step is performed after defining a new pitch-mark sequence. Manipulation of fundamental frequency is achieved by changing the time intervals between pitch markers as in Fig-4. The modification of duration is achieved by either repeating or omitting speech segments. In principle, modification of fundamental frequency also implies a modification of duration (Kortekaas et al. 1997).

Another variations of PSOLA, Frequency Domain PSOLA (FD-PSOLA) and the Linear-Predictive PSOLA (LP-PSOLA), are theoretically more appropriate approaches for pitch-scale modifications because they provide independent control over the spectral envelope of the synthesis signal (Moulines et al. 1995). FD-PSOLA is used only for pitch-scale modifications and LP-PSOLA is used with residual excited vocoders. Some drawbacks with PSOLA method exists. The pitch can be determined only for voiced sounds and applied to unvoiced signal parts it might generate a tonal noise (Moulines et al. 1990).

6.4 Linear Prediction based Methods

Linear predictive methods are originally designed for speech coding systems, but may be also used in speech synthesis. In fact, the first speech synthesizers were developed from speech coders. Like formant synthesis, the basic LPC is based on the source-filter-model of speech. The digital filter coefficients are estimated automatically from a frame of natural speech.

The main deficiency of the ordinary LP method is that it represents an all-pole model, which means that phonemes that contain antiformants such as nasals and nasalized vowels are poorly modeled. The quality is also poor with short plosives because the time-scale events may be shorter than the frame size used for analysis. With these deficiencies the speech synthesis quality with standard LPC method is generally considered poor, but with some modifications and extensions for the basic model the quality may be increased.

7. RESEARCH METHODOLOGY

7.0 OVERVIEW OF TIME-SCALING TECHNIQUES

7.1 TIME DOMAIN TECHNIQUES: TIME-SEGMENT PROCESSING

In time-segment processing, the basic idea behind the time-stretching technique is to divide the input sound into segments, then if the sound is to be shortened, some segments are discarded, or if the sound is to be lengthened, some segments are repeated. In general, all time-segment processing techniques are based on overlapping and adding adjacent segments extracted from the input signal.

7.1.1 OVERLAP-ADD (OLA)

Overlap-add (OLA) techniques are generally the most computationally inexpensive of all the time-scaling techniques since the basic algorithm requires only simple read/write pointer manipulation and accumulate instructions. For basic time-scale compression, small windowed segments are extracted at time ti and added to the output at time $t'_i = \alpha t_i$ where α is the time scale factor. The main artifact from the OLA technique comes from the amplitude and phase discontinuity at the boundary of the segments which causes pitch period discontinuities and consequent distortions that are detrimental to signal quality.

7.1.2 SYNCHRONOUS OVERLAP-ADD (SOLA)

One strategy for reducing the artifacts associated with the OLA technique is to modify the offset for placing each time-segment within a small range around the time-scale factor offset so that the cross-correlation between the overlapping samples is maximized for each pair of overlapping segments. This technique is referred to as Synchronous Overlap-Add (SOLA). The SOLA technique does a much better job at preserving the pitch, magnitude and phase relationships of the time-scaled signal.

7.2 SIGNAL MODEL ANALYSIS/SYNTHESIS

Signal modeling techniques model sounds as a sum of elementary sinusoidal components called partials. These techniques start by extracting partials, in the form of time dependent magnitude and instantaneous phase data, from a signal via time-frequency analysis, usually a Fourier transform based analysis, such as the short-time Fourier transform (STFT). The decomposition into individual partials is expressed in the additive synthesis equation (Eq-2). The instantaneous phase and instantaneous frequency are related by equation (Eq-3).

As its name implies, the additive synthesis expression is used to synthesize a signal from the time dependent magnitude and instantaneous phase data. Time-scaling is then achieved by modifying the phase and magnitude data before synthesizing. The modification is a time mapping function . To time-stretch a signal, we set >1, or to compress in time we set <1. The time-scaled signal is thus expressed as equation (Eq-4).

7.3 TIME-DOMAIN PITCH-SYNCHRONOS OVERLAP ADD (TD-PSOLA)

Time-Domain Pitch-Synchronous Overlap and Add (TD-PSOLA) is a variation of the SOLA technique in which the signal is first analyzed to identify local pitch across the signal, and local pitch information is used to adjust a variable segment size parameter, and a variable segment offset parameter to preserve the local pitch (fundamental frequency) while achieving a desired time-scale change.

Considering a periodic signal s(n), it is possible to get pitch modified version of s(n) by summing OLA frames si(n), extracted from the weighted window w(n), and changing the time-shift between frames from the original pitch period T0 to the desired one T expressed as equation (Eq-5).

If T≠T0, the operation results, again according to the Poisson Formula, in a re-harmonization of the spectrum of si(n) with the fundamental frequency 1/T can be described as equation (Eq-6).

Consequently, w (n) can be chosen so that the spectrum of si(n) closely matches the envelope spectrum of s(n). So the stated process provides a simple and efficient way to change the pitch of a periodic signal. Synthesizing speech is still not perfect, but the simplicity and efficiency has been increased with PSOLA. For

T=T0 the equation simply results in reconstructed signal which is proportional to the original signal as equation (Eq-7). The TD-PSOLA algorithm block diagram is under the section figures (Fig-5).

8. RESULTS & DISCUSSIONS

The experiment is performed on Punjabi voices both for male and female. We achieved very clear synthesized emotional speech through TD-PSOLA for many cases. The TD-PSOLA method is not effective for stochastic speech signal as the frequency domain peak-picking cannot estimate a modulation rate in the aspiration noise source. The threshold value is still not perfect as we do more testing on different voices and utterances using TD-PSOLA. The pitch marking of the wave is very limited to the testing voices as figures are shown under the section figures.

9. FINDINGS

The results show that the proposed system is promising for selecting parameters for embedding emotional in speech. Considering the significantly different performances for different emotions, and the differences observed between human and machine perception of emotions, however, at this stage we prefer to view the proposed automated evaluation more as a preprocessing step than a replacement to human evaluations.

10. RECOMMENDATIONS AND SUGGESTIONS

With the application of neural network we can improve the artificial sound of the synthesized speech by adding emotion to it. The emotional synthesis speech will sound more natural with the neural network trained speech signals. The parametric behavior of the speech will get more accurately tuned for embedding emotion in the speech resynthesis.

11. CONCLUSIONS

The system, developed with supervised training, consists of synthesis (TD-PSOLA). The experimental results show evidence that the parameter sets selected by the system can be successfully used to embed emotion in the input neutral speech, demonstrating that the system can assist in the human evaluation of emotional speech.

12. SCOPE FOR FURTHER RESEARCH

Our future research will be directed towards the design of more robust systems, more sophisticated parameter modifications, and experimenting with different parameter selection techniques and additional emotions and embedding emotion in a stochastic speech.

13. ACKNOWLEDGEMENT

It is our pleasure to acknowledge to our debt to many people involved directly or indirectly in our project. No work will be done by a single person. Every work should be completed with co-operation and co-ordination of many people. We do not find words to express our sentiments about all those people who involve with us in our project.First of all, I must thank to Ms. Simrat Kaur Sandhu who have encouraged me to do research on the topic of "Embedding Emotion in Speech". She helped me in almost every field of this project.

14. APPENDIX

MATLAB IMPLEMENTATION

MATLAB FILES:

1. detect_vuv.m : it perform s the detection of voiced and unvoiced samples in speech wave.
2. energy.m : it calculates the energy of the input frames of speech wave.
3. find_pmarks.m : it calculates and marks the pitches at the peaks in the short time energy function.
4. plot_pmarks.m : it plots the pitch marked by the „find_pmarks.m‟.
5. tdpsola.m : it performs the time domain pitch synchronous overlap add function for embedding emotion in speech.
6. test.m : integrated file to perform the whole process with an ease.
7. window.m : it calculates the windowed coefficients of the speech frames passed through window.
8. zerocr_frm.m : it calculates the average zero crossing rate of the input speech frames.

15. FIGURES & TABLES

FIG-1: REPRESENTATION OF SPEECH SOUND PRODUCTION

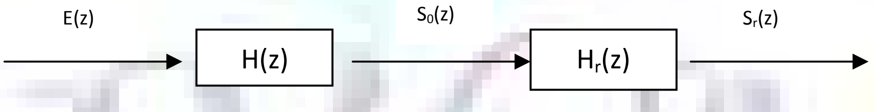


FIG-2: SIMPLE TEXT-TO-SPEECH SYNTHESIS PROCEDURE

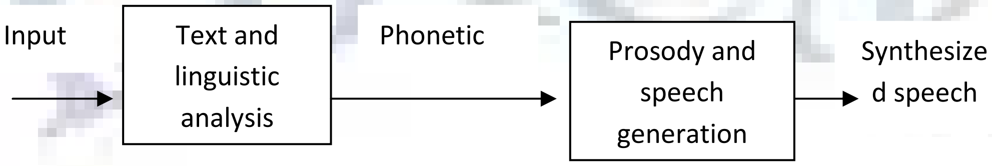


FIG-3: CIRCUMPLEX MODEL OF AFFECT AS DESCRIBED BY RUSSELL (1980)

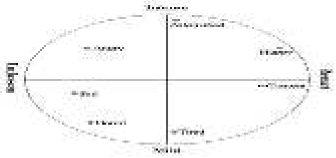


FIG-4: PITCH MODIFICATION OF A VOICED SPEECH SEGMENT

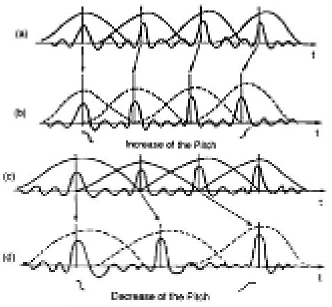


FIG-5: BLOCK DIAGRAM OF TD-PSOLA

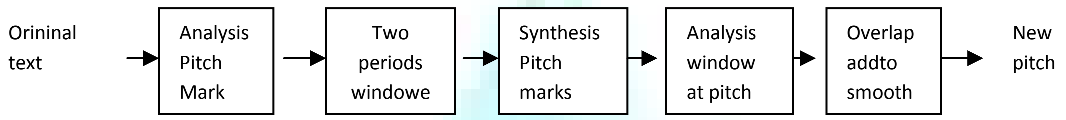


FIG-6: SPEECH WAVEFORMS USED FOR TESTING

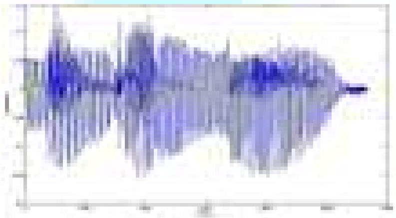


FIG-7: PITCH MARKS ON THE ORIGINAL SPEECH WITH THE THRESHOLD VALUE AS

a. Tscale: 1.5

b. Pscale: 0.7

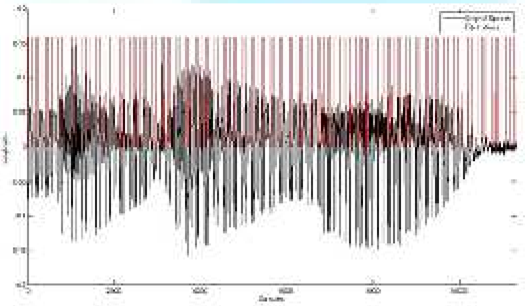


FIG-8: VOICED AND UNVOICED DETECTED MARKING IN SPEECH FRAMES

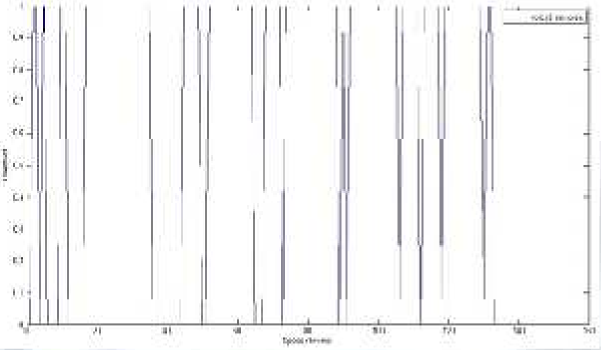
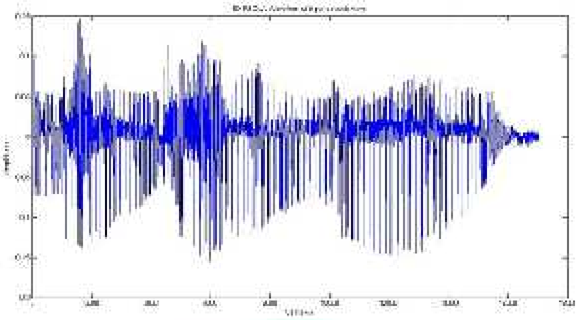


FIG-9: TD-PSOLA WAVEFORM OF THE INPUT WAVE



16. EQUATIONS

$$H(\omega) = \frac{1}{\cos(\omega \cdot x / v)} \quad \text{(Eq-1)}$$

$$s(t) = \sum_{k=1}^{K} a_k(t)\cos\phi_k(t) \quad \text{(Eq-2)}$$

$$\phi_k(t) = \int_0^t \omega_k(\tau)\,d\tau \quad \text{(Eq-3)}$$

$$s'(t') = \sum_{k=1}^{K} a_k(t')\cos(\alpha\phi_k(t')) \quad \text{(Eq-4)}$$

$$s_i(n) = s(n)\,w(n-iT_0) \qquad \bar{s}(n) = \sum_{i=-\infty}^{\infty} s_i(n - i(T - T_0)) \quad \text{(Eq-5)}$$

$$\text{if } s_i(n) \xleftrightarrow{f} S_i(\omega) \qquad \text{then } \bar{s}(n) \xleftrightarrow{f} \frac{2\pi}{T}\sum_{i=-\infty}^{\infty} S_i\left(i\frac{2\pi}{T}\right) \quad \text{(Eq-6)}$$

$$\begin{aligned} \bar{s}(n) &= \sum_{i=-\infty}^{\infty} s(n)w(n-iT) \\ &= s(n)\bar{w}(n) \\ &\approx K s(n) \end{aligned} \quad \text{(Eq-7)}$$

xm (*n*) = *hm* (*tm* - *n*)*x*(*n*) (Eq-8)

17. REFERENCES

1. Cornelius, R.R. (2000). Theoretical approaches to emotion. In Proceedings of the ISCA Workshop on Speech and Emotion, Northern Ireland.
2. Deller, J., R., Proakis, J., G., Hansen, J., H., L.: Discrete-Time Processing of Speech Signals, Macmillan Publishing Company, 1993.
3. Fant, G.: Acoustical Analysis of Speech. In: Crocker, M., J., Ed., Encyclopedia of Acoustics, pp. 1589-1598, John Wiley & Sons, Inc., 1997.
4. Juang, B., H., Ed.: The Past, Present, and Future of Speech Processing. In: IEEE Signal Processing Magazine, vol. 15, no. 3., pp. 24-48, May 1998.
5. Makhoul, J.: Linear Prediction: A Tutorial Review. In: Proceedings of the IEEE, vol. 63, no. 4, pp. 561-580, April 1975.
6. Markel, J., D., Gray, A., H.: A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method. In: IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-22, no. 2, pp. 124-134, April 1974.
7. McAulay, R., J., Quatieri, T., F.: Speech Analysis/Synthesis Based on a Sinusoidal Representation. In: IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-34, no. 4, pp. 744-754, August 1986.
8. McAulay, R., J., Quatieri, T., F.: Sinusoidal Coding. In: Kleijn, W. B., Paliwal, K., K., Eds.: Speech Coding and Synthesis, pp. 121-173, Elsevier Science, Amsterdam, 1995.
9. Nakhai, M., Marvasti, F.: A Hybrid Speech Coder Based on CELP and Sinusoidal Coding. In: IEICE Transactions on Information and Systems, vol. E83-D, no. 8, pp. 1190-1199, August 1999.
10. Quatieri, T., F., McAulay, R., J.: Speech Transformations Based on a Sinusoidal Representation. In: IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-34, no. 6, pp. 1449-1464, December 1986.
11. Quatieri, T., F., McAulay, R., J.: Shape Invariant Time-Scale and Pitch Modification of Speech. In: IEEE Transactions on Signal Processing, vol. 40, no. 3, pp. 497-510, March 1992.
12. Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology. 39:1161-1178.
13. Schere, K. R (1986) Vocal affect expression: A review and a model for future research. Psychological Bulletin. 99:143-165.
14. Stevens, K. : Models of speech production. In: Crocker, M., J., Ed., Encyclopedia of Acoustics, pp. 1565-1578, John Wiley & Sons. Inc., 1997.
15. Syrdal, A., et al.: TD-PSOLA Versus Harmonic Plus Noise Model in Diphone Based Speech Synthesis. In: IEEE International Conference on Acoustics, Speech, and Signal Processing .98, Seattle, Washington, 1998.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mails i.e. **infoijrcm@gmail.com** or **info@ijrcm.org.in** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail infoijrcm@gmail.com.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

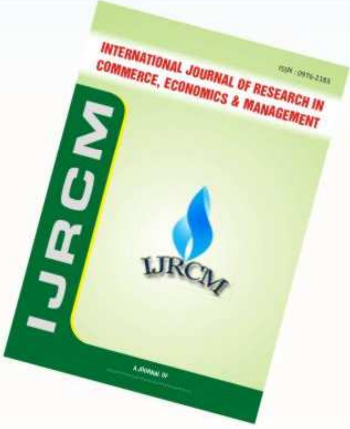
Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals



ISSN : 0976-2183
INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, ECONOMICS & MANAGEMENT
IJRCM

ISSN : 0976-2183
INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT & MANAGEMENT
IJRCM

ISSN : 0976-2183
INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE & MANAGEMENT
IJRCM