2011

Automated Semantic Understanding of Human Emotions in Writing and Speech

Ricardo A. Calix
Louisiana State University and Agricultural and Mechanical College

AUTOMATED SEMANTIC UNDERSTANDING OF HUMAN EMOTIONS IN WRITING AND SPEECH

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in
Engineering Science

by
Ricardo A. Calix
B.S., Industrial and Systems Engineering, Universidad Tecnologica Centroamericana, 2001
M.B.A, Louisiana State University, 2006
M.S., Engineering Science, Louisiana State University, 2010
August, 2011

DEDICATION

To my parents Georgina and Ricardo Calix.

ACKNOWLEDGEMENTS

First and foremost, I express my sincere appreciation to my advisor Dr. Gerald M. Knapp for his support and guidance during the past four years of graduate study in the college of engineering at LSU. His vision of human centered engineering helped me to learn and grow in an emerging area of research at the cross section between industrial engineering and computer science which I truly enjoy. I am grateful for having him as an advisor. I also thank my committee members, Dr.'s Craig Harvey, David Constant, Andrea Houston, Boryung Yu, and Bahadir Gunturk, for their guidance and support during my studies. Their patience is much appreciated.

This work would not be possible without the help and support of the people in the department of industrial engineering and our great semantic analysis research group in the interdisciplinary program of engineering science. To my fellow graduate students, Sri Abhishikth Mallepudi, Anuradha Maha, Manthan Malde, Bin Chen, Mahdi Khazaeli, Leili Javadpour, Brandon Pitts, and Heather Dylla, I extend my appreciation. I wish you all the best in your careers. I would also like to thank Denis Daly who provided his talents, free of charge, to help record the speech extension of the Affect Corpus 2.0.

Additionally, I thank my good friends Sylvester Tang, Amra Durmisevich, and Frank White who encouraged me to pursue the PhD when I was still undecided. Finally, special thanks go to Dr.'s Jing Wang, Isabelina Nahmens, Arthur Shr, and Peter Chen for their help with references, conferences, and financial support during the summers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Affective Human Computer Interaction (A-HCI) will be critical for the success of new technologies that will prevalent in the 21st century. If cell phones and the internet are any indication, there will be continued rapid development of automated assistive systems that help humans to live better, more productive lives. These will not be just passive systems such as cell phones, but active assistive systems like robot aides in use in hospitals, homes, entertainment room, office, and other work environments. Such systems will need to be able to properly deduce human emotional state before they determine how to best interact with people.

This dissertation explores and extends the body of knowledge related to Affective HCI. New semantic methodologies are developed and studied for reliable and accurate detection of human emotional states and magnitudes in written and spoken speech; and for mapping emotional states and magnitudes to 3-D facial expression outputs. The automatic detection of affect in language is based on natural language processing and machine learning approaches. Two affect corpora were developed to perform this analysis.

Emotion classification is performed at the sentence level using a step-wise approach which incorporates sentiment flow and sentiment composition features. For emotion magnitude estimation, a regression model was developed to predict evolving emotional magnitude of actors. Emotional magnitudes at any point during a story or conversation are determined by 1) previous emotional state magnitude; 2) new text and speech inputs that might act upon that state; and 3) information about the context the actors are in. Acoustic features are also used to capture additional information from the speech signal. Evaluation of the automatic understanding of affect is performed by testing the model on a testing subset of the newly extended corpus. To visualize actor emotions as perceived by the system, a methodology was also developed to map predicted emotion class magnitudes to 3-D facial parameters using vertex-level mesh morphing.

The developed sentence level emotion state detection approach achieved classification accuracies as high as 71% for the neutral vs. emotion classification task in a test corpus of children's stories. After class re-sampling, the results of the step-wise classification methodology on a test sub-set of a medical drama corpus achieved accuracies in the 56% to 84% range for each emotion class and polarity. For emotion magnitude prediction, the developed recurrent (prior-state feedback) regression model using both text-based and acoustic based features achieved correlation coefficients in the range of 0.69 to 0.80. This prediction function was modeled using a non-linear approach based on Support Vector Regression (SVR) and performed better than other approaches based on Linear Regression or Artificial Neural Networks.

CHAPTER 1: INTRODUCTION

Affective Human Computer Interaction (A-HCI) will be critical for the success of future interactions between humans and computer systems. Such systems will need to be able to properly deduce the emotional implications related to human interaction before conveying information. If current trends are any indication, there will be continued rapid development of automated semantic understanding systems that help humans to live better, more productive lives.

In healthcare, for instance, high costs and the increase in the size of the elderly population have put a strain on the supply of care giving medical personnel. Many tasks that require medical support staff such as rehabilitation and elderly care are expensive and are not being performed adequately. Additionally, it has been suggested that caring for people needs to have a human touch (Tapus et al. 2007). Therefore, automated systems could be used to fill this need but they must be able to detect human behavior and emotion to perform these tasks, especially when dealing with elderly patients and children. Automated systems that understand human emotions could serve as companions providing entertainment or challenging interactions that keep the elderly more active and healthier.

For automated systems to understand emotions, first they need to learn how to detect them. Ideally, humans would be able to hand craft all required emotion detection rules into standalone machines or organize all emotional information available in the world (e.g. the web) so that they can interpret and respond to human emotions. This, however, is not possible since annotating all data in the world or hand coding all possible rules would require resources that are not available. As Gantz and Reinsel (2010) note, the digital universe is expanding faster than what people can handle manually. They estimate that by 2020 the amount of available digital information will be 40 times larger than it is today. Therefore, this rapid expansion will require new automated methodologies that can process digital information including emotional content.

This dissertation focuses on this issue of how to develop affective aware system methodologies that can effectively understand natural languages and interact with humans. A key problem in developing affective HCI is understanding the high level semantics in language that convey emotion. In this work, high level semantics in language refer to detection and estimation approaches that find concepts, relations between concepts or features, semantic representations, and entailments between concepts in the context of emotion understanding and response.

Finding ways of detecting emotion in speech or ways of developing interactive affect aware robot aides that adapt to a person's emotions will have great impact to human computer interaction, communication, automation of information technology, and overall artificial intelligence. Developing these methodologies will be very useful in many areas such as social robotics, cognitive ergonomics, text-to-scene processing, game design, movie animation, information retrieval, task automation, and healthcare.

1.1 Applications of Affective Human Computer Interaction

In the following sections, descriptions of some important application areas are provided.

1.1.1 Applications to Health Care

A current research area in healthcare is the use of socially enabled robots to treat patients with autism, rehabilitation needs, and other problems. Tapus et al. (2008) have shown that patients have better performance in their rehabilitation tasks when assisted by automated systems that appropriately respond to their behavior. They also note that autistic children who interact with robots can improve behavior and cognitive processing. Unfortunately, at this point, smart automated aides are difficult to implement because they need to interpret human emotions from language inputs. In general, it is expected that in the future socially aware assistive systems will be required to provide additional care to patients.

1.1.2 Applications to Cognitive Ergonomics

The field of cognitive ergonomics could greatly benefit from the use of machine learning techniques to automate tasks and develop environments that can adapt to human behavior and emotions. Aircraft

cockpits, hospital operating rooms, military simulators, and other work environments could automatically adapt based on a person's mood. This can have great benefits in safety and increased productivity.

1.1.3 Applications to Information Retrieval

Retrieving content from text and audio collections is an important topic in information retrieval. There are many text and audio collections on the web that may contain what a user needs but that do not contain the metadata necessary to find it. As a result, new tools will be needed to bridge the gap between semantic meaning in speech content and low level speech features. When looking for audio or video files, for instance, users usually think of ideas, emotions, or people they want to see or hear about. Therefore, developing efficient technologies that can detect these semantic concepts, retrieve them, and make them available in an enriched way could improve communication and overall user satisfaction.

Predicted emotion magnitudes (chapter 8) can be used for many applications related to content search. In content search, estimated emotion magnitudes could be used for automatic tagging or automatic rating of audio narratives based on emotion content.

1.1.4 Applications to Movie Animation and Video Game Design

Emotion detection has many applications in video game design and movie animation. In animation, a character's facial expression is critical in conveying the overall feeling for the scene. Currently, animation is a highly complex and laborious practice which can involve hundreds of people and take thousands of hours to complete. Automatic 3-D generation from text descriptions can be used to reduce this workload by automatically creating initial renderings of scenes and characters from text descriptions. This can greatly reduce the time it takes to go from production to market.

In the field of entertainment, for instance, it has been shown that realism in the 3-D graphics domain is key for the audience or user to connect with the game or movie they are watching. A good example of this is the 2009 movie Avatar by James Cameron. This movie demonstrated the importance of effective emotional facial expressions in 3-D based movies.

Emotions are an important component that must also be addressed in video game design. Callele et al. (2006) argue that in video games, developers need to address both the functionality and the emotional interaction of the user. They refer to the "player's experience" which relates to the user's emotional response to the game. Callele et al. (2008) supports the argument that emotion requirements should drive the development of the video game as much as other requirements.

1.2 Problem Statement

Much work has been done in emotion detection for document summarization, opinion mining, and sentiment polarity detection using lexical approaches and knowledge bases. However, many issues are still unresolved. Previous research has used multimodal approaches such as text, speech, images, and biometrics for emotion detection in hopes of capturing more information. However, these approaches are still limited to low level semantic understanding of emotions because they usually do not consider higher semantic aspects of language.

Emotion detection in language is hard because: (1) emotions can be very subjective and (2) the set of semantic and acoustic features that captures emotion in language is not clearly defined. Therefore, additional work needs to be done in the area of affective natural language processing for affective HCI. Affective Human Computer Interaction (HCI) systems will need to be able to properly deduce human emotional state before they can determine how to best interact with people

In emotion language understanding, vocabulary approaches alone are not enough (Calix et al. 2010) because emotion words that are usually associated with specific emotion classes can be used for other emotions. An example of this is when people use sarcasm, where they say one thing but they mean another. Additionally, acoustic features are usually related to some emotions but noise and other factors can affect the ability of a system to predict emotion. Furthermore, when dealing with language understanding, there are many higher level semantic aspects that must also be considered to obtain true understanding. Semantic understating, however, is very challenging. Important issues related to emotion detection include: How should an automated system consider the actors in a story or conversation who

experience the emotion? How does the automated system know when the emotional state of an actor has changed, and what it has changed to? How does the system represent emotional state and magnitude? And how should the automated system respond to a person? These are some of the issues this dissertation addresses.

This dissertation focuses on understanding and developing affective aware systems such as a dialogue system that can effectively understand natural language inputs and respond based on emotion content. Specifically, two main issues are addressed. The first issue is how an automated system can recognize human emotion from multimodal sources, in this case text and speech (text and speech refer to the semantic and acoustic components of speech). The second issue is how an automated system can respond after detecting human emotion. To achieve this, new methodologies for emotion and actor level emotion magnitude prediction from written and spoken speech are developed. Additionally, a methodology for the integration of emotion detection inputs to the human computer interaction cycle using real-time 3-D graphics is developed. The automatic detection of affect in written text and speech is based on natural language processing and supervised machine learning approaches.

The key contributions of this research include new algorithms, methodologies, text based features, and new corpus annotations to detect actor level emotional states and magnitudes through a story or conversation. Four emotion related methodologies are developed, and results and limitations are discussed: First, a methodology to automatically extract emotion word features from annotated corpora is developed. Second, a step-wise classification methodology for affect detection using sentiment composition and sentiment flow features in text and speech is developed and analyzed using two different communication formats (narrative vs. dialogue). Third, a methodology for nominal sentient entity detection is developed which can be used to identify actors. Fourth, a methodology for actor level emotion magnitude prediction and mapping to 3-D graphics using a prior state feedback regression model and mesh morphing is developed. Additional aspects about the challenges of affect detection in NLP such as feature selection, corpus development, unbalanced datasets, and anaphora resolution are also discussed.

The features used in the regression model for emotion magnitude prediction include: 1) previous emotional state; 2) new text and acoustic inputs that might act upon that state; and 3) information about the environment the actors are in. For speech, the main features that are considered include speech intensity, pitch, formants, and Mel Frequency Cepstral coefficients (MFFCs).

1.3 Objectives

The major objectives of this study are to:

- Automatically identify evolving emotional states and magnitudes of actors in multi-actor environments from multimodal sources (writing/speech and voice features). Specific tasks to accomplish this included:
    - Use Natural Language Processing (NLP) methodologies and corpora to identify and annotate actors participating in conversations (directly or via reference), and use them for emotion magnitude prediction.
    - Define a set of input features from text and speech processing sufficient for accurate and reliable classification of emotional state and magnitude. For text, features include features derived from sentiment flow and sentiment composition principles; descriptive adjectives; actions applied to or performed by the actor; emotion words; and environment/location descriptors that may convey emotional content. For speech, features include aspects of tone, pitch, volume, energy, speech rate, and spectral properties.
    - Develop a recurrent (prior-state feedback) predictive model to predict evolving emotional states and their magnitude based on the input feature vectors described above. Machine learning and optimization techniques are used to train the system. The model assigns emotion state to the individual actors and includes information about the scene and the environment to establish the influence of these external conditions on emotional state.

- Develop, for training and evaluation purposes, a corpus of text and speech with emotion labels (classes & change in magnitudes).
- Analyze the performance of the methods against the corpus.

- Develop a model for automated rendering of 3-D facial expressions based on emotion class and magnitude.

1.4 Organization of this Dissertation

This dissertation is organized as follows: Chapter 1 presents the introduction and motivation for the work. Chapter 2 presents the literature review and background for the methods used. Chapter 3 describes an overview of the methodology developed in this work. Chapter 4 describes a feature extraction methodology to determine key semantic emotional words for emotion detection. Chapter 5 describes the corpora that was developed to train and test the methodologies, as well as to explore the generalization of the developed scheme. Chapter 6 addresses multimodal emotion detection at the sentence level in narrative and dialogue based mediums. Chapter 7 proposes a methodology for actor and environments detection. Chapter 8 proposes the recurrent (prior-state feedback) predictive model for actor level emotion magnitude estimation using an annotated corpus. Chapter 9 demonstrates the use of these methodologies to provide a system response based on initial emotion detection. Finally, chapter 10 provides a summary of contributions and potential future work.

CHAPTER 2: LITERATURE REVIEW

This chapter describes the current state of the art in emotion detection from multimodal sources, and related applications. Relevant statistical and probabilistic machine learning methods are also discussed.

2.1 Emotion Recognition Systems

There have been several attempts to develop affective HCI systems. Important early work in this field was done by Rosalind Picard at MIT. She is credited as being one of the founders of the field of affective computing. Picard's book (1997), offers a very comprehensive overview of the topic and issues. The book provides early frameworks for emotion representation and modeling, as well as suggested approaches using pattern recognition (such as Hidden Markov Models) for affect detection and synthesis.

Most of the current systems in affective HCI address emotion detection using multimodal approaches. Sebe et al. (2005), for instance, provides an important summary of multimodal affective HCI studies of relevance to this work. In general, these approaches use feature vectors that fuse information from multiple sources for pattern recognition. The main multimodal sources used in emotion detection include text, speech, images, video, and biometrics.

Current multimodal studies using biometrics (Chang 2009; Wang 2009) and vision (Zhao et al. 2003; Wang 2008) have many applications to medicine, cyber security, and law enforcement. Studies such as Chang (2009) implemented an emotion recognition system that used visual and physiological input signals to detect emotion. Features such as facial points, skin conductivity, finger temperature, and heart rate were used to train classifiers such as Artificial Neural Networks. Studies such as this, however, do not include a critical component for emotion understanding which is language.

An important early affective HCI system that used speech information is the Kismet robot from MIT (Breazeal 2000). The main design objective of this robot was to interact with humans (socialize) using sound and images. This robot was important because it would use facial expressions to respond to

detected human inputs. However, no high level semantic language understanding and response was possible. The Kismet robot was programmed to socialize with human beings in a very simple way much like an infant. The vision system would extract features that allowed Kismet to detect people. The auditory system of the Kismet robot focused mainly on distinguishing between sounds emitted by a human and sounds emitted by things. To detect human sounds (including vocal affect), the system would extract features related to presence of sound and speech, time stamped pitch tracking, time stamped energy tracking, and time stamped phonemes. After feature extraction, the system would use supervised learning techniques to perform classification. Kismet's architecture was implemented using a multi-layer approach where a low level may extract simple easy to process features which would determine if higher more complicated processing was required (e.g.. if human skin is not detected, then no further processing is needed).

Another important system in this area which utilizes a more semantic approach to understand different high level cognitive processes such as emotions is EM-ONE (Singh 2005). EM-ONE (Emotion Machine ONE) uses multiple heuristics that encode information and common sense knowledge about the world (ConceptNet discussed below, is an example of this type of repository of common sense knowledge). This system takes new approaches to the problem of artificial intelligence which are mostly inspired by the work of Marvin Minsky. In Minsky (2007) and Singh (2005), the authors propose that there are at least six levels in which information should be represented and analyzed. Therefore, their model proposes that multiple aspects of a problem should be analyzed concurrently. These aspects include: (1) types of problems that may be solved with information, (2) goals, (3) similarity to other concepts, (4) useful cases in which information can be used, (5) narrative information about use of information or concept, (6) relevancy related contextual cues, and (7) credit assignment to important features. Their basis for this model is the fact that the brain thinks about a problem in many different ways and therefore computational learning models should also store and process information in many different ways.

This dissertation will focus on developing a multimodal emotion detection system using only two input mediums which are: text and speech. For this reason, a review of emotion detection from natural language is provided in the following sections.

2.2 Emotion Recognition from Text

Emotion detection in text is a difficult problem because of the richness and ambiguity of language. Words, combinations of words, special phrases, and grammar all play a role in formulating and conveying emotional information. There are hundreds of words in the English language that can be used to represent emotions, most of which do not represent unique emotions but instead represent varying levels of the degree in which people experience emotions. An example of this is “ecstatic”, which describes a degree of happiness that a person experiences.

Supervised learning techniques in natural language processing and semantic analysis are commonly used to detect and classify information in a document. Supervised learning has been extensively used in topic detection (Manning and Schutze 1999), entity recognition (Nguyen and Cao 2008), and extraction of temporal relations (Bethard 2007) to name a few applications. The application of these techniques in emotion or affect detection is more recent but has been successfully used by Osherenko (2007), Osherenko (2008), Strapparava and Mihalcea (2008), Minato et al. (2007), Mathieu (2004), Alm (2005), Alm (2008), Wiebe (2005), Tokuhisa (2008), Calix et al. (2010), and Pang and Lee (2008). These papers are discussed below.

Emotion detection studies in text can loosely be classified into lexical-based and high semantic-based studies.

2.2.1 Lexical Based Approaches

In general, lexical based studies for affect detection from text use a bag of words approach combined with other text features such as syntactic elements to perform classification. Osherenko (2008), for instance, used the presence or absence of negations and intensifiers as features to train and test an emotion

detection model. In Alm (2005, 2008), the authors conducted an empirical study to determine the emotional affinity of sentences in the domain of children's stories. To achieve this, they developed the UIUC affect corpus and explored different types of features and their contribution to emotion classification. This study also explored affect in speech with a special interest in expressive text-to-speech synthesis. In Tokuhisa et al. (2008), the authors propose a model for detecting the emotional state of a user that interacts with a dialog system. Tokuhisa et al. (2008) uses corpus statistics and supervised learning to detect emotion in text. They implement a two-step approach where coarse grained emotion detection is performed first followed by fine grained emotion detection. Their work found that word n-gram features are useful for polarity classification.

To select lexical text features, Calix et al. (2010) proposes a methodology to automatically extract emotion relevant words from annotated corpora. The emotion relevant words are used as features in sentence level emotion classification with Support Vector Machines (SVM) and 5 emotion classes plus the neutral class.

In general, most of these studies extract features which are used to learn models for classification. Common features include POS tags (VB, NN, JJ, RB), exclamation points, sentence position in story, thematic role types, sentence length, number of POS tags, WordNet emotion words, positive word features, negative word features, actual words in the text, syntactic parses, etc.

2.2.2 Sentiment vs. Commonsense Knowledge Approaches

For some applications, effective acquisition of meaning from text requires the use of semantic representations. The four main types of semantic representations are: semantic maps, first order logic, conceptual dependency, and frame based representations (Jurafsky and Martin 2008). Most semantic studies in the literature make use of one of these representations.

Studies that include higher semantic ideas for emotion detection can be divided into sentiment composition approaches and knowledge based approaches. Sentiment composition, as defined by Moilanen and Pulman (2007) states that emotion content in a sentence is based on the emotional content

of its constituent parts. In their work, Moilanen and Pulman (2007) propose a series of sentiment composition rules (i.e. {(+) (N) → (+)}) which can be used to infer the emotional polarity of a sentence. Neviarouskaya et al. (2009) also uses a sentiment composition and a rules-based approach to detect emotion in text. Pang and Lee et al. (2008) provides a survey of the current methodologies and issues related to opinion mining and sentiment analysis.

Common sense knowledge based approaches are those that use some form of encoded knowledge or rules to extract information that can be used to infer emotion in a document or sentence. In Liu (2003), the authors use a knowledge based approach to detect emotion in text. They argue that common sense knowledge about the world (i.e. is_a, located_at, etc.) is required to disambiguate and detect emotions. Similar arguments are proposed in Minsky (2007). In Lu et al. (2006), the authors consider the issue of who experiences an emotion by using a semantic role labeling tool which for each verb in a sentence identifies constituents with a semantic role such as patient or agent. This helps to find possible subjects and objects in a sentence. The researchers use the internet to learn adjectives that could be common to several words. For instance, in their example “A girl met a tiger”, the system would use role labeling to learn that “girl” is the subject, “met” is the verb, and “tiger” is the object. Next, the system looks up definitions for girl and tiger and extracts any adjectives associated with these words. In this case, “girl” is linked with youthful and tiger is linked with predatory. The final step infers the emotion based on the knowledge that “youthful met predatory”. The drawback of this approach is that the inference rules to determine that:

youthful & predatory => fear

have to be coded manually. Machine learning techniques could be used to learn this knowledge. In their paper, they also suggest that knowledge bases such as ConceptNet (Havasi et al. 2008) could be useful to learn these rules.

Concordance lines (Coyne and Sproat 2001) are another example of knowledge based approaches. The basic idea is that certain emotions will happen more often with certain words. To gain this knowledge, a window approach is taken of words around a target word (i.e. emotion word). This can be useful in situations like the following. The sentence "I am excited because it is my birthday but not happy about my physical therapy" has two emotions and two possible causes. On the one hand, the subject is "excited" about his birthday but on the other hand he is "not happy" about the therapy. This can be expressed using First Order Logic (Jurafsky and Martin 2008) as follows:

excited(I, birthday)

NOT happy(I, therapy)

After learning likelihoods from a corpus, the system knows that the word birthday is more likely to be associated with a positive emotion such as "excited".

2.3 Emotion Recognition from Speech

Many studies have been conducted to detect emotion from speech. Three common issues for emotion detection from speech are: (1) which features to use, (2) what segment of the signal should be used, and (3) what machine learning methods should be used to detect emotion.

On the first issue, Chuang et al. (2004) showed that emotions in speech are closely related to prosody features such as pitch and energy (formants). (Pitch and energy are described in section 2.7.2). Chuang et al. (2004) described how pitch for happiness or anger is higher than for sadness. Chuang et al. (2004) also indicated that the energy in speech associated with anger or surprise is greater than the energy associated with fear. Other authors such as Luengo et al. (2010) concluded that spectral features (MFCCs) have a higher contribution to emotion detection. Mel Frequency Cepstral Coefficients (MFCCs) are usually very useful in emotion detection because they represent a deeper view of the speech signal. They are a deeper view because they help to analyze the data on the Mel scale of the auditory system. In contrast to the cepstrum (rate of change in spectrum bands), the MFFC cepstrum frequency bands are mapped to the Mel

scale. This helps to capture speech information. The Mel Frequency Cepstral coefficients are obtained as a result of the Cosine Fourier Transform of the Mel scaled Fourier Transform of the speech signal.

With regards to the second issue of speech segment length, Klabbers et al. (2007) conclude that the length of the speech segments is important for emotion detection. Different segment lengths have been used to extract information from the speech signal. In Hansen et al. (1997), the authors used word level features to detect stress in utterances using a vocabulary of 35 words. This work concluded that "when stress is present, recognition rates decrease significantly". In contrast, Busso et al. (2009) concluded that statistics such as the mean of the pitch contour at the sentence level are a better indicator of emotion than other shorter term statistics (e.g. at the phrase or word level).

Finally, on the third issue of the learning methods used in emotion detection, many approaches have been used. Schuller et al. (2003), for instance, used continuous Hidden Markov Models (HMMs) to detect emotion in speech. Their approach incorporated temporal information and states to detect additional context about the speech signal and its relation to emotion. Grimm et al. (2007a, 2007b) used a primitives-based evaluation to study emotions in speech. In their work, the researchers analyzed the estimation of continuous values for 3 emotion class related properties or dimensions: valence, activation, and dominance. Valence represents the positive vs. negative relation of an emotion. Activation describes the level of excitation (from calm to excited) and dominance describes the influence of the person (from weak to dominant). Their study focused on speech only features at the utterance level to estimate the emotion levels using several regression approaches. The features from the input signal include pitch, energy (formants), speaking rate, and the spectral characteristics. The approach in Grimm et al. (2007a, 2007b) is different from other implementations because it predicts magnitudes for speech related primitives instead of just classes for the speech signal.

Ververidis (2005) used Gaussian Mixture Models (GMM) with a Bayes classifier to classify utterances (phrases or sentences) after feature extraction. In Shafran (2005), the authors compared the performance of different classifiers on the detection of emotion in speech. Results of their study indicate that word

content is better than speaking style when it comes to detecting emotion but that other speech features can enrich the information. Other similar techniques are described in Busso et al. (2004).

2.3.1 Automatic Speech Recognition (ASR) and Speech Characteristics

Emotion detection from speech is closely related to emotion detection in text. Speech is related to text because speech can be converted into text by many off-the-shelf Automatic Speech Recognition (ASR) systems. The issue to consider is the additional features that speech includes such as pitch, volume and intonation (Jurafsky 2008). The literature refers to this area as emotion recognition from vocal prosody (rhythm, stress, and intonation of speech). In general, features used in speech analysis include pitch, energy (formants), speech rate, intensity analysis, harmonicity analysis, linear predictive coefficient (LPC), and spectral analysis (see speech features section 2.7.2 for additional information on use of these features for affect detection from speech).

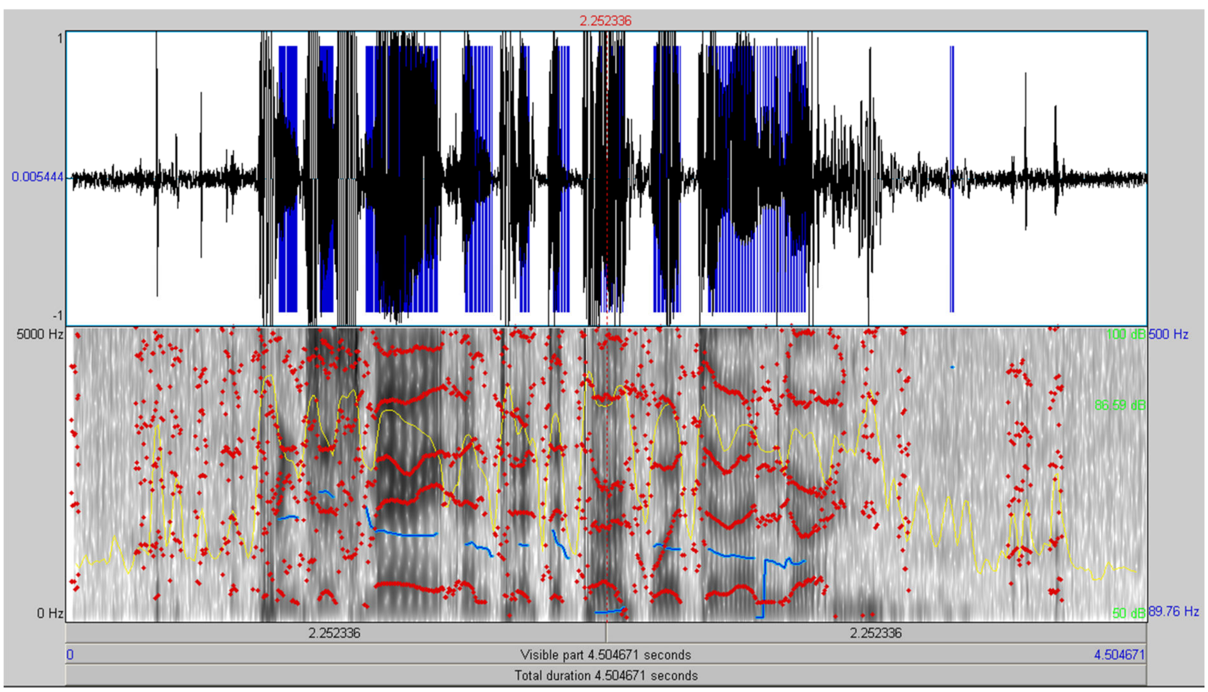


Figure 1: Praat waveform and spectrogram for "The princess Mary entered the dark woods of Engar."

Available software for speech recognition includes: CMU Sphinx (Placeway 1996), Microsoft Windows 7 SAPI Speech Recognition, Loquendo ASR (Loquendo 2002), CSLU toolkit (Sutton et al. 1998), Nuance Dragon Naturally Speaking (Nuance 2011), and Praat (Boersma and Weenink, 2005).The

CSLU toolkit and the Praat program are two of the most widely used and versatile tools for speech recognition, annotation, and analysis. The Praat program incorporates modules for pitch extraction, formant analysis, Fourier spectrum and spectrogram analysis, and scripting language capabilities.

Figure 1 illustrates the signal analysis for a speech signal using Praat. The upper panel shows the waveform of the signal. The lower panel shows the spectrogram for the speech signal. The blue solid line in the lower half of the figure represents the pitch signal. The yellow line is the intensity signal. The red dotted lines represent the formants (energy) of the speech signal. The upper half of the figure shows the pulses as vertical dark blue solid lines. Additionally, the Praat application allows for the segmentation of a signal so that different sections can be annotated with the word they represent. Figure 2 shows an example of the signal with its corresponding annotated words.

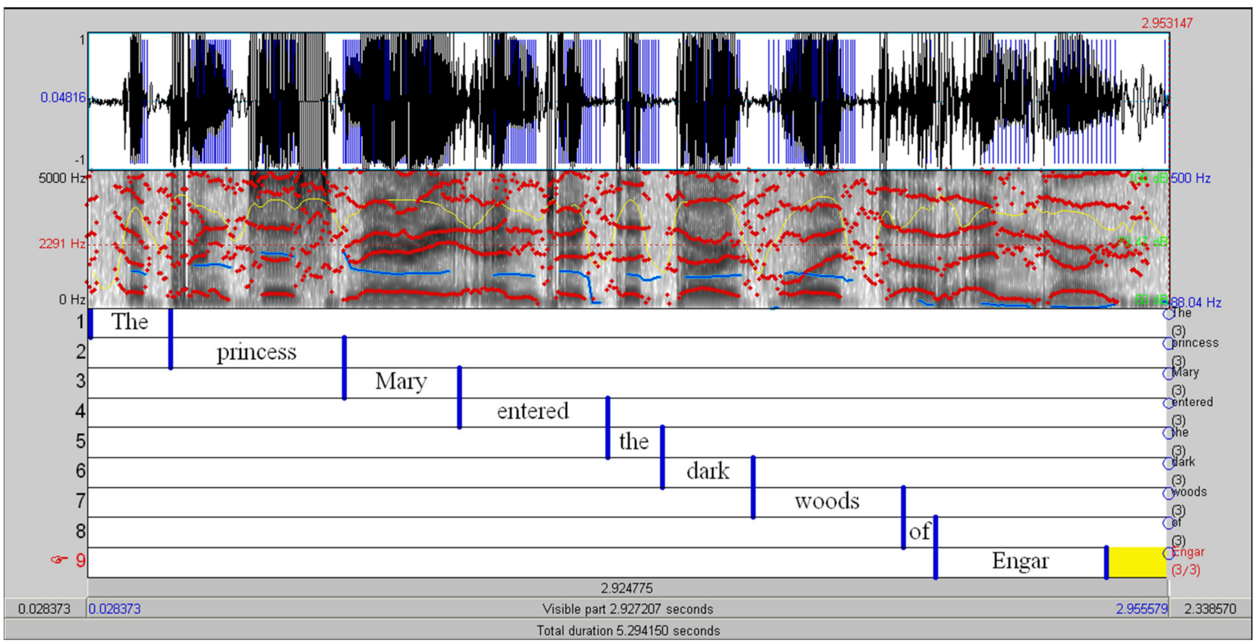


Figure 2: Speech signal with words for "The princess Mary entered the dark woods of Engar"

2.4 Multimodal Emotion Recognition

Research in multimodal emotion recognition is not as extensive as research in the individual areas. However, several studies have been conducted that integrate information from two or more sources. In general, two approaches have been used for multimodal emotion detection: separated recognition and

joint recognition (Sebe et al. 2005). Separated recognition performs classification for every modality and later integrates the results. Joint recognition, on the other hand, integrates the modality features and then performs classification or inference. Learning methods used in multimodal affective recognition have included HMM, GMM (Jurafsky 2008), ANN and Dynamic Bayesian Networks (Tawfik and Neufeld 1994).

Studies that have used multimodal approaches for affect detection include Kanluan et al. (2008), Song et al. (2008), Chuang (2004), Tawari et al. (2010), and Castellano et al. (2007). In Kanluan et al. (2008), the authors used audio-visual information to detect emotion. An emotion space concept was introduced to describe emotions using just three feature types which were valence, activation, and dominance. The system detected features individually and then combined the results using a linear function with weight adjustment. In Song et al. (2008), the researchers used a hidden Markov model for emotion recognition. Their implementation allows for state synchronization of multiple source signals and preserves the correlation of the signals over time.

Most authors concluded that multimodal approaches can be more effective for emotion recognition. In Chuang et al. (2004), the authors showed that detection of emotion from speech produces higher classification accuracies than detection of emotion from text. This result can be attributed to the fact that natural language is complex and that higher level approaches using semantic analysis are needed. Chuang et al. (2004) and Castellano et al. (2007) showed that multimodal emotion recognition (i.e. text, speech, biometrics, and vision) performs better than uni-modal emotion recognition. Sebe et al. (2005) notes that determining the optimal point at which different modes of information should be merged is crucial and still an open research problem. Studies such as Tawari et al. (2010) propose the use of context (such as speaker gender) and other sources to analyze emotion. An example of an important multimodal and machine learning based emotion detection system that incorporated sound was the Kismet robot (Breazeal, 2000). Sebe et al. (2005) provides a good overview of work and challenges related to this area.

2.5 Actor Detection

Actor detection for high semantics emotion detection, prediction, and understanding is very important. Most approaches for emotion detection in the literature analyze and detect emotions at the sentence or document level. This approach neglects the fact that emotions are experienced by actors (participants or users) in the story or conversation. Therefore, actors must be detected before their emotions can be identified and tracked.

The process of detecting and tracking actors can vary greatly depending on the input medium being analyzed. The process can be as simple as assuming whoever types information into a login screen in a dialogue system is the user, to as complicated as finding actors and referring expressions to that actor throughout narrative text. Some examples, of ways that users can be identified and the level of difficulty include:

- Assuming there is only one actor or user interacting with an affective dialogue system (easiest).
- Assuming there are multiple actors but that all possess some type of identification such as a biometric or RFID tag.
- Detecting actors based on their speech patterns in dialogue formats using speech recognition and diarization based classification. Diarization approaches partition an audio signal into segments based on an individual's characteristics. With a clustering approach, for instance, diarization techniques will group speech segments based on the characteristics of the speaker.
- Detecting actors from dialogue text where each actor has been identified such as in scripts with some type of XML tag.
- Detecting actors solely based on semantic content in narrative stories using entity detection and relation recognition.

The first 4 approaches are straightforward and can be handled with relatively low complexity. However, the last approach involves many very challenging aspects of NLP such as nominal entity recognition and

anaphora resolution. In this section, a review of some of the common approaches to this problem will be discussed.

Studies that have addressed the issue of actor detection from speech can be classified based on the features they use into studies that use acoustic features only, studies that use NLP techniques in text only, and studies that combine both approaches.

Detection of an actor from an audio file involves capturing the speech signal, transcribing the signal into text, extracting text and speech features, detecting the sentient actors, and enriching the content with the actor location information. In the following sections, a discussion of studies that have addressed the issues of actor detection are provided.

2.5.1 Actor Detection Using Low-level Features

Bigot et al. (2010) studied mainly acoustic approaches to detect actors. The study proposes a methodology for speaker role recognition in conversational speech. Their methodology tries to classify speakers into categories such as anchor, journalist or other. To achieve this, the audio files are segmented using speaker diarization algorithms to find the different speaker patterns. Once the files are segmented, temporal, acoustic, and prosodic features are extracted to perform classification. Using hierarchical supervised classification, the system achieves accuracies of 92% for speaker role detection. In contrast, a study that has tried to obtain a higher understanding of the speech content is Kucuk and Yazici (2008). In this work, the authors propose a methodology for the extraction of semantic objects from videos (using the audio segments) for automatic tagging. The methodology uses lexical resources and the identification of co-referential chains for semantic annotation of news videos. The methodology extracts salient entities from the text transcript of the videos and then performs detection of co-reference chains. The salient entities are extracted based on a lexical resource which consists of proper names of political people, political status, position, continent, country names, city names, and political institutions. This lexical resource is domain specific and consists mainly of named entities. The salient entities are extracted using a regular expression (Jurafsky and Martin 2008) approach and are then used to identify co-reference

chains. These co-reference chains are identified using heuristic rules. In the following section, approaches that address this more specific issue are discussed.

2.5.2 Sentient Actor Detection Using Higher Semantic-level Features

Sentient actor detection refers to techniques used to identify entities that perform actions and (from the perspective of this work) are subject to social characteristics such as emotions. Higher semantic information about the speech signal can be obtained from the transcribed text content. In general, sentient actors in text refer to named (Named Entities) and un-named entities (Nominal Entities). Most studies in sentient actor detection in text focus exclusively on Named Entity Recognition (NER) as the much broader problem of nominal actor detection (for un-named entities) is highly complex and is still an open research problem (Pang and Fan 2009a).

Named entities have proper names such as Lucinda, Jane, and Neo. Unnamed entities do not have a proper name but are instead referenced typically in a noun phrase such as “the wizard” or “the big bad wolf”. Named entity recognition (NER) approaches have been addressed in the literature and have relatively mature implementations available. On the other hand, un-named entity recognition methods are less common in the literature.

The main approaches for un-named entity detection use named entity recognition followed by disambiguation. McShane (2009) and Cassimatis (2009), for instance, argue that the next generation of intelligent systems will need knowledge about the world in order to effectively detect objects and events such as these from language inputs.

Detection of these un-named entities in text can be divided into 4 steps which are: (a) Pre-processing, (b) Named Entity Recognition, (c) Relation Recognition, and (d) nominal entity recognition.

Pre-processing uses tokenizing, tagging, and chunking with a chunk grammar to annotate and define units within sentences. A chunk grammar is a set of rules that dictates how a sentence should be divided or chunked. For example, a grammar can be expressed as a regular expression (Jurafsky and Martin 2008) as follows:

Grammar = {<DT>?<JJ>*<NN>}

A chunk parser using the previous grammar can then proceed to chunk sentences that contain sequences of part of speech tags in the order specified in the grammar. In this case, the grammar specifies an optional determiner (DT) followed by zero or more adjectives (JJ) and ending with a noun (NN). The important chunk in entity recognition is the noun phrase (NP). Noun phrases (e.g. "The big bad wolf") are important because they usually contain entities. Noun phrases and noun phrase chunks can be detected based on the syntactic and part-of-speech tags assigned to the words in a sentence text.

Named Entity Recognition (NER) looks for entities which are likely proper nouns. or definite noun phrases. A definite noun phrase is a noun phrase where the referent is identifiable (i.e. the girl, the cat). In contrast, an indefinite noun phrase is a noun phrase where the referent is not identifiable (i.e. a girl, a cat). Named entity recognition is an extension of chunking which has the additional function of disambiguating between words that are used as names or as other types of words. Disambiguation can be achieved by combining the regular expression grammar with other techniques such as supervised learning methods, and disambiguation through knowledge bases. Studies that have addressed this issue include Nguyen and Cao (2008), Chifu and Chifu (2008), Todorovic et al. (2008), Paab and Pilz (2009) and Mansouri et al. (2008).

In Nguyen and Cao (2008), the authors note that most approaches in NER applications use semi-automatic detection or heuristics with pattern recognition (i.e. regular expressions) for recognition and disambiguation of entities of a specific type (such as persons or locations). They argue, as Minsky (2007) does, that knowledge about the world (i.e. Is_a ("cat", "animal") is required to deal with this issue in a more significant way. In Chifu and Chifu (2008), the authors used an unsupervised learning approach based on Self Organizing Maps (SOM) to identify named entities. This method first detects noun phrases from the text sample and represents them as feature vectors. Concepts from an ontology driven framework are also represented as feature vectors. To link the concept vectors to the noun phrase vectors, the system uses a Euclidean distance function to classify or organize entities into concept categories.

Other authors like Mansouri (2008) used a fuzzy support vector machine approach to detect named entities.

In general, these works note that most approaches in NER applications use lexical resources and semi-automatic learning approaches (or heuristics) with pattern recognition to identify entities of a specific type (such as persons or locations). NER approaches, however, still miss information about the relations between named and referring expressions and do not detect nominal entities (e.g. un-named entities).

Relation recognition looks for links or relations between the entities identified in the actor detection step and referring expressions (e.g. pronouns). Two important issues to consider in relation recognition are dependency representation and co-referencing which are needed for entity disambiguation. A dependency representation is used to show word heads and their dependents such as actor and patient. Co-referencing techniques are used to resolve the issue of co-referents (Jurafsky and Martin 2008). That is, how do we know that in

"Joe went to a party. There he danced and had a great time"

that "he" refers to "Joe"? This issue has been approached in several ways. Coyne and Sproat (2001) matched pronouns such as he, she or they by scanning words within a window that had the appropriate number and gender. Harder situations where a noun co-refers to another noun such as:

"The dog was on the table. The animal slept all day."

can be handled by using WordNet to estimate if they are linked (Coyne 2001). WordNet (See corpora section below) has distance functions built in that return the distance from one synset to another. The function basically counts the number of nodes between synsets which can serve as a metric of the relation between two concepts. Diesner (2009) addressed this issue by combining corpus statistics with network analysis to formulate a non-algorithmic approach using human reviewers to analyze the problem of reference resolution. They basically build social networks of words which are used to determine relations between them. They found that anaphora resolution can be strengthened by using co-reference resolution in graph-based approaches. Soon et al. (2001) proposed one of the most popular approaches for anaphora

resolution. Implementations for co-reference resolution software include BART (Versley et al. 2008), KIM (Kiryakov et al. 2005), Guitar (Steinberger et al. 2007), Gate (Cunningham et al. 2002), and JavaRAP (Qiu et al. 2004).

All current relation resolution systems such as BART perform relatively poorly for anaphora and co-reference resolution such as: singular pronoun disambiguation (i.e. he, she), singular pronouns in dialogs (quotes), plural pronouns disambiguation ("Billy and Ann went down the hill. Then they ate an apple pie."), and when the referring expression is not a pronoun ("The Cheshire cat smiled. Then the cat disappeared."). Since these approaches perform poorly, the current state of the art relies heavily heuristics to resolve domain specific relation recognition issues.

Finally, the more complicated issue of nominal entity recognition (for un-named entities) has not been addressed as much in the literature. Some approaches include Pang and Fan (2009a) and Pang and Fan (2009b). In these works, the authors proposed a two-layer model with semantic role labeling that utilizes the results of co-reference resolution to perform nominal entity recognition. This dissertation proposes an algorithmic and supervised learning approach to detect nominal entities.

2.6 Corpora

Affect detection is the process of inferring emotion states for a test sample through automated learning approaches (see machine learning section 2.8) using multimedia inputs (e.g. text or speech). The objective of these approaches is to train a prediction model so that, given an input, it can predict some output. To perform this training and to test the model, a training set must be used. In this work, the resources used to train and test a model are referred to as corpora (for plural) or corpus (for singular).

A corpus is a resource that includes information that has been analyzed and annotated by a person. A prediction model uses a corpus to try to find a correlation between the corpus content and the given annotations such as class labels or magnitude assignments. Since this work focuses on language, the medium used is the text and acoustic content of a speech signal. Corpora have been developed for many applications and can be classified into: text corpora, speech corpora, and knowledge corpora.

2.6.1 Types of Corpora

- **Text Corpora**

Emotion annotated text corpora are not as common in the research literature as other types of general purpose corpora such as the Brown corpus or the Penn Treebank (Jurafsky and Martin 2008). Previous implementations for affect detection corpora include: MPQA (Wiebe et al. 2005), Movie Reviews (Pang and Lee 2008), ISEAR (Scherer 1997), EARL (Schroder 2006), and the UIUC affect corpus of children's stories (Alm 2008). Most of these corpora are hybrids which started with opinion annotations and which were later extended to include some emotional tags. These implementations of affective corpora have been used for machine learning methodologies such as in Calix et al. (2010).

A problem with some of these corpora is that they have focused on sentence level emotion annotation and on just emotion class labels. They include class labels such as happy, sad, angry, etc. but do not include the actual intensity of the emotion (e.g. happy level with scale from 0 to 1) which is an important aspect for emotion recognition systems. Additionally, in order to know who experiences an emotion and who produces it, annotation needs to consider the actor that experiences the emotion. This means that actors and their respective referring expressions must be identified before emotions can be annotated.

There are some corpora that have addressed the issue of entities and anaphora resolution such as Ace, MUC-6 and MUC-7 (Jurafsky and Martin 2008). These corpora are very important in the field but do not combine annotations on actors, actor presence in a sentence, and emotion magnitudes experienced by each actor. Therefore, to train semantic emotion magnitude prediction models, emotion magnitudes must be annotated at the actor level. This combination of annotations requires a corpus that includes actors per story, actor presence and location in a story, and emotion magnitudes per actor to determine evolving emotional state of actors. This dissertation developed such a corpus to train and test the proposed methodologies.

A good annotation requires using good annotation tools and methodologies. Important annotating tools include NLTK (Bird et al. 2009), BART (Versley et al. 2008), GATE, Nuance Dragon Audio Mining

SDK (Nuance), Praat and various other standalone programs. The scheme used for annotations used in this dissertation was based on Wiebe et al. (2005) annotation guidelines for emotion corpora.

- **Emotion Categories**

Emotion text corpora are annotated with emotion labels. The set of emotion labels used is directly related to the understanding and theories of how human emotions are represented and expressed. An important early work on human emotions can be traced back to Descartes' "The Passions of the Soul" (1649/1989). In this work, Descartes proposed new ideas which were revolutionary at the time about the sensations and perceptions of man. This work was important because it allowed emotions, for the first time, to be analyzed from a scientific point of view. Building on this work, researchers like Robert Plutchik and Johnson-Laird (1989) have reasoned that emotions can be categorized into groups. The main group consists of eight primary emotions. These emotions are: anger, fear, sadness, disgust, surprise, curiosity, acceptance, and joy. Plutchik, for instance, argues that these eight emotions are biological in nature and are essential for human survival. The other groups of emotions, argue researchers, consist of combinations of the emotions in the main group. Additionally, almost all emotions can be classified as either positive or negative. Some previous studies have added a third category to represent emotions such as astonishment which are neither positive nor negative (Mathieu 2004). This dissertation uses a set of 6 emotion labels to perform the analysis. It is important to note that six emotions classes are common in the field and are referred to as "The Big Six". The 6 classes used are: anger, fear, happiness, sadness, surprise, and neutral (the absence of emotion).

Currently, emotion annotated corpora use different mark-up or annotation schemes. This is due to the fact that a standard emotion mark-up language (Baggia 2009) has only recently been proposed and it is not widely used. Additionally, annotations usually depend on the user's needs and domain.

Another important aspect to consider in emotion detection from text is emotion magnitude. In Chuang (2004), the authors use an emotion scale to represent the data. In their paper, the authors argued that

emotion magnitudes can be modified by words such as "very" and "not" as in "very happy" and "not happy". This dissertation uses scale from 0 to 100% to annotate emotion magnitudes.

- **Speech Corpora**

Speech corpora include two or more modes of information such as text and acoustic information. Corpora for speech recognition and synthesis include the following: CELEX (Baayen et al. 1995), CMU pronouncing Dictionary (CMU 1993), The PRONLEX dictionary (LDC-Pronlex 1995), TIMIT corpus (NIST 1990), AT&T's HMIHY 0300 corpus (Shafran 2005), Switchboard corpus (Greenberg et al. 1996), and Buckeye corpus (Pitt et al. 2007). Most of these corpora include the text transcript of the audio recordings available in the corpus. When developing new speech corpora, Librivox (2011) is a good source to obtain audio recordings of public domain texts.

Emotion detection in speech collections requires the use of both text and speech features to better understand emotion content. For un-transcribed audio collections, an Automatic Speech Recognition (ASR) can be used to convert the signal into a text transcript. The state of the art in ASR technology has progressed to the point where good results in speech-to-text translation can be obtained. The results are usually measured in word error rate (WER) for certain domains. The commercially available Nuance Dragon software (Nuance 2011) is currently a leader in speech-to-text synthesis and achieves over 90% accuracy for one speaker speech transcriptions. These ASR systems rely on huge vocabularies to produce good transcriptions. The Dragon software can use vocabularies with more than 300,000 words.

- **Knowledge Corpora**

Knowledge corpora refer to databases that include information about words, their definitions and relations to other words. There are several implementations such as: ConceptNet (Havasi 2007), Opencyc (Opencyc 2010), WordNet and WordNet-Affect (Valitutti et al. 2004a; Valitutti 2004b), and FrameNet. The version of ConceptNet 3.0, for instance, contains over one million assertions collected by human annotators from the World Wide Web. The format of output produced by ConceptNet for a word such as "cat" is given in Figure 3.

AtLocation (cat, lap) []
AtLocation (cat, bed) []
CapableOf (cat, hunt mouse) []
CapableOf (cat, eat mouse) []
CapableOf (cat, corner mouse) []
HasA (cat, four leg) []
HasA (cat, fur)[]
IsA (cat, carnivore) []

Figure 3: ConceptNet selected output for "cat"

2.6.2 Inter-annotator Metrics

When annotating a new resource, the quality of the annotation process must be measured. Inter-annotator metrics refer to techniques used to measure the overall agreement between the annotations of two or more individuals. Important metrics used to evaluate inter-annotator agreement include: average observed agreement, Pi, alpha, S, and Kappa (Artstein and Poesio 2008). These metrics differ in how they correct for expected chance agreement. Expected chance agreement is a probability that 2 annotators will agree on their annotation for an item by chance. This probability depends on the number of classes. Formally, this probability is calculated as follows:

$$A_e = \sum_{k \in K} P(k|c_1) \cdot P(k|c_2) \qquad \textbf{(Eq. 1)}$$

where c_i is the annotator I, and k is the assigned category. Inter-annotator agreement metrics are important because they help to set theoretical boundaries on the accuracy that a given machine learning methodology can achieve using the annotated corpora (Bird et al. 2009). A brief description of each technique is provided in the following sections.

- **Average Observed Agreement (A_o)**

Averaged observed agreement is the easiest metric to compute. It is the percentage of annotations that two annotators agreed upon. The metric is formulated as follows where "samples" is the total number of annotation samples and "agreed" is the amount of samples for which both annotators agreed.

$$A_o = \frac{1}{samples} \sum agreed \qquad \textbf{(Eq. 2)}$$

- **Chance-corrected Metrics (A_{corr})**

Chance corrected metrics are those that take into account the expected chance agreement A_e. Once chance agreement is defined, the metric can be corrected. These types of metrics include: s, alpha, and kappa. Formally, the main concept in these metrics is defined as follows:

$$A_{corr} = \frac{A_o - A_e}{1 - A_e} \quad \textbf{(Eq. 3)}$$

- **Speaker Normalization**

Additionally, in speech, multiple speaker normalization (Busso 2009) can be used to reduce variability caused by different speaker voices and recording conditions. This approach scales the speech signal for each speaker according to a specific reference signal. The scaling factor calculation is formulated as follows:

$$\mathbf{ScFact_{spkr}} = \frac{\mathbf{E_s}}{\mathbf{E_{ref}}} \quad \textbf{(Eq. 4)}$$

where $\mathrm{E_s}$ is the average value, either energy or pitch, for a given speaker's speech signal and $\mathrm{E_{ref}}$ is the average value for the reference.

2.7 Features and Feature Selection

2.7.1 Text Features

In Johnson-Laird and Oatley (1989), the authors performed an analysis of English words which are related to emotions. This work argues that there are basic sets of words that relate to each of the basic emotion classes. Other works, like Ortony et al (1987, 1998), also suggest that emotion states are associated with specific sets of words. Determining what is the set of most important emotion words for particular application domain is very important. These lists of emotion words per class can be obtained manually or automatically from different sources such as WordNet and its extension WordNet-affect (Vaitutti 2004), or automatically from annotated corpora such as Calix et al. (2010).

In addition to the emotion words, other text based features can be used. The features used by empirical studies such as Alm (2008) included quotes, POS tags, exclamation points, sentence position in story, thematic role types, sentence length, number of POS tags (VB, NN, JJ, RB), and syntactic parses.

In general, text features can be binary or numeric. Binary feature include the presence or absence of words, part of speech tags, chunks, syntactic parses, semantic parses, etc. Numeric based text features can be derived from performing counts or calculating distance metrics between words or higher level semantic concepts. For example, this dissertation proposes numeric text-based features based on sentiment composition and sentiment flow.

2.7.2 Speech Features

In speech analysis, emotion-related features consist of pitch, formants (energy), intensity, harmonicity, noise energy ratio, linear prediction coefficients (LPC), duration, and Mel frequency Cepstral Coefficients (MFCCs) (Jurafsky 2008). The speech signal is a continuous evolution of the vocal tract as words are being uttered. Therefore, to capture speech information, a time series of spectra must be extracted from the signal. This is usually achieved using, for instance, a sliding window of 20 milliseconds. After an acoustic waveform is sampled, the samples are used to calculate speech characteristics about the signal. The characteristics are represented as feature vectors for each speech sample. These feature vectors produced after feature extraction can contain statistics on mean, standard deviation, min and max, and quartiles of the signal.

A fundamental technique to extract these features from the speech signal is the Fourier transform (Osgood 2009). The Fourier Transform (both continuous and discrete) is a method that allows a signal to be decomposed into simpler signals such as sines and cosines. The process involves representing the original function as a sum of weighted sine waves. The signal can be represented as follows:

$$\mathbf{F(x)} = \sum_{k=1}^{N} A_k \operatorname{Sin}(2\pi k t + \varphi_k) \qquad \textbf{(Eq. 5)}$$

where A_k represents the amplitude of the signal, k represents the frequency of the signal, and φ_k represents the phase of the signal. Therefore, the signal is transformed from a spatial domain representation to a frequency domain representation. For ease of processing, however, the sine and cosine components are converted to complex exponentials using Euler's equation. The inverse of the Fourier transform is used to convert the signal from the frequency domain back to the spatial domain. The Discrete Fourier transform is defined as follows:

$$\mathbf{F\,(u) = \sum_{n=0}^{N-1} f[n] e^{-2\pi iun}} \qquad \textbf{(Eq. 6)}$$

See Osgood (2009) or Gonzales (2008) for more details on the Fourier transform.

The main speech features used in this work are described as follows:

- **Pitch (F0)**. Pitch is a perceptual property of a signal. Jurafsky and Martin (2008) define the pitch of sound as a "mental sensation of fundamental frequency." Higher pitch correlates to higher fundamental frequency. Pitch can be detected by using the higher Cepstral values of a signal (the peaks in the higher frequency band of the cepstrum). The cepstrum is a way to de-convolve (i.e. separate) the source (glottal) from the filter (vocal tract) in a speech signal (Jurafsky and Martin 2008). By doing this, the elements of pitch and phones (speech sounds; for example, the "k" sound in skill) can be identified. As described by Jurafsky and Martin (2008), this filter correlates to the shape of the vocal tract as the speech signal was being uttered. These cepstrum properties are obtained by applying the Fourier transform to the speech signal. Formally, the cepstrum is defined as the inverse Discrete Fourier Transform (DFT) of the log magnitude of the DFT of a signal (Jurafsky 2008). A function of pitch over time is also referred to as the pitch contour. Once the pitch contour of a signal is calculated, statistical properties of it such as the mean, standard deviation, minimum and maximum values can be computed and represented as feature vectors per speech segment.

- **Intensity**. Intensity relates to the amplitude of the vocal cord vibrations. According to Jurafsky and Martin (2008), high amplitude indicates higher air pressure and the intensity is a normalized representation of the power of the amplitude. To calculate the intensity of a signal, a sampled sum of the amplitudes of the signal over time is taken and is normalized to human auditory thresholds (Jurafsky and Martin 2008). Formally, the intensity is defined as follows:

$$Intensity = 10\log_{10}\frac{1}{NP_0}\sum_{i=1}^{N}x_i^2 \qquad \textbf{(Eq. 7)}$$

where N is the number of samples, x is the amplitude measurement for the ith sample, and P_0 is the threshold pressure for the auditory system ($P_0 = 2 \times 10^{-5}$ Pascals). Intensity is represented in decibels (db). For this dissertation, intensity mean, standard deviation, minimum and maximum values from a signal are used.

- **Formants (F1-F5)**. The energy of a signal can be captured by formants. Different sounds will have different formants. These differences are in location and the magnitude of the peaks in the given spectrum. The peaks are known as formants and capture the resonance of the activity in the vocal tract. These Formants are usually grouped as concentrations of acoustic energy in a specific frequency range. They can be seen in the following figure (Figure 4) as the darker areas in the spectrogram.

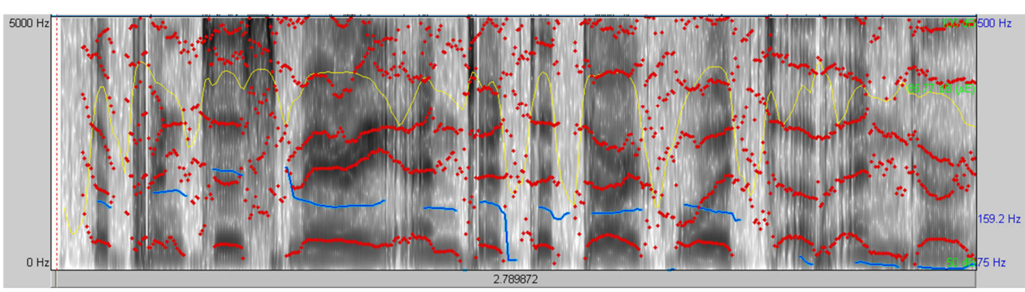


Figure 4: Spectrogram analysis (Formants appear as dark concentrations)

One formant will be assigned per 1KHz frequency band (i.e. F1 from 0 to 1000 Hz, F2 from 1000 to 2000 Hz, etc). The formant average, standard deviation, minimum and maximum for every frequency band can be extracted to represent a given speech segment.

- **Mel Frequency Cepstral Coefficients (MFCCs)**. These features are calculated by obtaining the Fourier Transform of the mel-scaled Cosine Fourier transform of the speech signal. They represent a deeper analysis of the speech signal and are very important for emotion detection. A total of 12 MFFCs are used in this work.

Other speech features that have been by other studies used include pulses, voicing, shimmer, jitter, and harmonicity.

2.7.3 Feature Selection Techniques

The use of feature selection techniques in emotion detection is critical to achieving good prediction results. There are many feature selection methods in machine learning. These are approaches which take a given set of features and try to map to a new set of transformed features (i.e. PCA) or try to reduce the number of features without transformation such as chi-square ranking (Witten and Frank 2005).

Some of the main reduction based feature selection techniques used in the literature include document frequency (DF), mutual information (MI), the chi-square statistic, and information gain (IG) (Yang and Pederson 1997). Document frequency (DF) is a method used to calculate the number of documents that a word appears in. Information Gain (IG) determines the number of bits of information contributed by a word for classification by knowing the absence or presence of a word in a given document (Yang and Pederson 1997). Mutual Information (MI) is a method which calculates the two-way contingency between words and categories. It measures the mutual dependence between two variables (Yang and Pederson 1997). Chi-square is similar to MI with a major difference being that chi-square values are normalized (Yang and Pederson 1997). Chi-square includes and additional term that counts the number of times that neither the class nor the term occurs.

Mutual Information (MI) has a tendency to favor terms that occur with less frequency (Yang and Pederson 1997) and are, therefore, appropriate in emotion detection from text since terms can occur infrequently but have high contribution to classification. Formally, MI can be defined as follows:

$$I(t,c) = \log \frac{P(t,c)}{P(t)P(c)} \qquad \textbf{(Eq. 8)}$$

where I represents the mutual dependence between t and c, the numerator P(t, c) is the joint probability between t and c, P(t) and P(c) are the individual probabilities for t and c respectively, t is the term, and c is the class.

In speech processing, Busso et al. (2009) notes important feature selection techniques such as forward or backward feature selection, sequential forward floating search, principal component analysis, genetic algorithms, evolutionary algorithms, and Linear Discriminant Analysis.

2.8 Overview of Machine Learning Approaches

Machine Learning (ML) is essential for automated systems to make decisions and to infer new knowledge about the world. This section describes some of the most important methodologies currently in use in the field of machine learning (see Table 1 below). Machine learning approaches can be divided into supervised learning (such as Support Vector Machines) and unsupervised learning (such as K-means clustering).

Within supervised approaches, the learning methodologies can be divided based on whether they predict a class or a magnitude into classifiers and regression models, respectively. An additional categorization for these methods depends on whether they use sequential and non-sequential data. The methodologies presented in this dissertation will focus on supervised learning approaches which consider both sequential and non-sequential data, and which are used to predict both classes and magnitudes. Therefore, a description of these different approaches is provided in the following sections. A comparison of different machine learning approaches is provided in Table 1.

Table 1: Machine learning techniques

Technique	Definition	Pros	Cons
Support Vector Machines	Supervised learning approach that optimizes the margin that separates data.	SLT Confidence characteristic (expected risk)	class imbalance issues
Decision Trees	This method performs classification by constructing trees where branches are separated by decision points.	Easy to understand	Not flexible
Neural Networks	Model represents the structure of the human brain with neurons and links to the neurons.	Versatile	Can obscure the underlying structure of the model
K-means clustering	Unsupervised method that forms k-means clusters to minimize distance between centroids and members of cluster.	Unsupervised – so no training needed	Needs clearly defined separations in the data in order to be effective
Linear Discriminant Analysis (LDA)	Creates linear function of features to classify data	Simple yet robust classification method	Normality assumptions of the classes
Gaussian Mixture models (GMM)	This probabilistic method represents signals as weighted sums of normal distributions	Can be used to represent non-normal distributions	Initialization is important for optimization
Naïve Bayes	Probabilistic Learning to calculate the probability of seeing a certain condition in the world selecting the most probable class given the feature vector	Fast, easy to understand the model	Bayes assumptions of independence
Maximum Likelihood Estimation (MLE)	Calculates the likelihood that an object will be seen based on its proportion in the sample data	Simple	Too simplistic for some applications
Expectation-Maximization	Similar to MLE but is used when there is missing data in the training set	Very useful when missing data	Too simplistic
Hidden Markov Models (HMM)	A Markov Chain is a weighted automaton consisting of nodes and arcs where the nodes represent states and the arcs represent the probability of going from one state to another.	Probabilistic. Good for sequence mining	Combinatorial complexity/ needs prior knowledge
Bayesian Networks	Probabilistic networks	Graphical representation improves understanding	Requires knowledge of probabilities

2.8.1 Classifiers

Classifiers are machine learning approaches that produce as an output a specific class given some input features. Important classifiers include Support Vector Machines (Burges 1998) commonly implemented using LibSVM (Chang and Lin, 2001) and Naïve Bayes, artificial neural networks, decision trees, random forests, and the k-nearest neighbor classifier (Witten and Frank, 2005). Naïve Bayes is a probabilistic classifier which usually performs worst with emotional data. Therefore, it helps to set the baseline for the classification task. Random forest classifiers consist of several decision trees and usually produce good results with emotion data. Artificial neural networks and Support Vector Machines (SVMs) are classifiers that can handle non-linearly separable data. In theory, this capability allows them to model data that may be more difficult to classify. The k-nearest neighbor classifier is also a good technique with the added advantage that it does not require parameter tuning. Because of its importance to this dissertation and to A-HCI, the Support Vector Machines methodology will be discussed in further detail.

- **Support Vector Machines (SVM)**

Support Vector Machines is a binary classification method based on statistical learning theory which maximizes the margin that separates samples from two classes (Burges 1998; Cortes 1995). This supervised learning machine provides the option of evaluating the data under different spaces through Kernels that range from simple linear to Radial Basis Functions (Chang and Lin 2001; Burges 1998; Cortes 1995). Additionally, its wide use in the field of machine learning research and ability to handle large feature spaces makes it an attractive tool for NLP studies.

Statistical Learning Theory (SLT) methods assume prediction models that can be ascribed a confidence characteristic. They are based on the fact that both structural and empirical risks are minimized (Muller et al. 2001). The expected risk can be calculated based on the empirical risk present in the data with the associated upper and lower bounds. This generalization error can be expressed as follows:

$$\mathbf{R}(\boldsymbol{\alpha}) \leq \mathbf{R}_{\mathbf{emp}}(\boldsymbol{\alpha}) + \sqrt{\left(\frac{\mathbf{h}\left(\log\left(\frac{2\mathbf{l}}{\mathbf{h}}\right)+1\right)-\log\left(\frac{\mathbf{n}}{4}\right)}{\mathbf{l}}\right)} \qquad \textbf{(Eq. 9)}$$

In SVM, the maximization of the margin is based on the training samples that are closest to the optimal line (also known as support vectors). Because the method tries to maximize the margin between the samples of two classes under a set of constraints, it ultimately becomes an optimization problem to find the maximum separation band. The function that represents the margin is quadratic and can be solved using quadratic programming techniques with Lagrange operators. The "objective function" and constrains can be represented as follows where W is the weight vector:

$$\mathbf{1/_2\ \|W\|^2} \qquad \textbf{(Eq. 10)}$$

$$\mathbf{Y_i(W \cdot X_i + b) \geq 1} \qquad \textbf{(Eq. 11)}$$

Non-linearly separable cases can be solved by mapping the initial set of features to a higher feature space by way of a Kernel Trick. This will provide higher freedom in separating the data in higher dimensional space. The Kernel trick takes advantage of the fact that SVMs do not need to know the mapping function because this is expressed as the dot product of the input data.

$$\mathbf{Xi \cdot Xj \ \rightarrow\ \emptyset(Xi)\ \cdot \emptyset(Xj)} \qquad \textbf{(Eq. 12)}$$

Since most real world data includes outliers and noise, a soft margin approach can be introduced in the model to allow for some errors to occur. This softening of the margin is achieved by introducing an error term where the cost represents the penalty for each error

$$\mathbf{C \sum Ei} \qquad \textbf{(Eq. 13)}$$

Under the soft margin approach, the objective function with constraints can be written as follows:

$$\mathbf{1/_2\ \|W\|^2 + C \sum E_i} \qquad \textbf{(Eq. 14)}$$

$$\mathbf{Y_i(W \cdot X_i + b) \geq 1 - E_i} \qquad \textbf{(Eq. 15)}$$

where i represents each training sample, W is the weight vector (normal vector) that defines the maximum margin, b is the bias, Y_i is the class for each training sample i, and X_i is the feature vector for each training sample i. The E_i term in the model represents the slack error for each sample. The cost parameter

(C) represents the penalty for each error. Although the SVM is used for binary classification, multiclass implementations can be achieved by creating different classifiers for each pair of class comparisons. The implementation used in this study, LibSVM (Chang 2001), relies on a one-against-one approach. Class imbalance problems in the data set are an important issue that needs to be addressed when using SVMs and most other machine learning techniques (Chang 2001).

Multiple kernels can be used with support vector machines to map the feature vectors from input space to higher dimensional feature space. However, during this research it was found that best results were usually obtained with the Radial Basis Function (RBF) kernel. The RBF kernel is defined in Equation 16 below.

$$K(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \qquad \textbf{(Eq. 16)}$$

2.8.2 Regression Approaches for Magnitude Prediction

In emotion understating, emotion classification is just one part required to solve the problem. Once an emotion is detected, it is also important to determine the magnitude of that emotion. For example, slightly happy and very happy convey very different states about a happy emotion. In this dissertation, each state can have an assigned magnitude which can range from 0 to 1, where 0 is the minimum value of the state and 1 represents the maximum value for the state.

To address the issue of magnitude prediction, regression approaches can be implemented. Common learning methodologies to address magnitude estimation include linear and non-linear regression models such Linear Regression, Artificial Neural Networks (ANNs), and Support Vector Regression (SVR).

- **Linear Approaches: Linear Regression Analysis**

In linear regression, multiple variables and coefficients are combined to form an equation that can be used to fit a particular data sample. The fitting process involves an optimization approach that minimizes the Least Squares Error. The coefficients or weights for each variable are useful in determining the

contribution of each variable to the fit of the data. The use of dummy variables (i.e. variables that only take 1 or 0) to indicate the presence or absence of a particular categorical effect in a sentence (such as word presence in a sentence) can be very useful to determine the contribution that a text element makes to predicting some class. Although common in the literature, linear regression models suffer from several drawbacks. Among them are the assumptions of linearity and normality of the data.

- **Non-Linear Approaches: Support Vector Regression and ANNs**

Artificial Neural Networks and Support Vector Regression (SVR) are important methods to model predictive equations. These methods have the advantage that they can be used to map non-linear data to higher dimensional spaces and do not suffer from parametric assumptions and the requirements of linear regression such as normality. ANNs are common in the literature but can over fit the data.

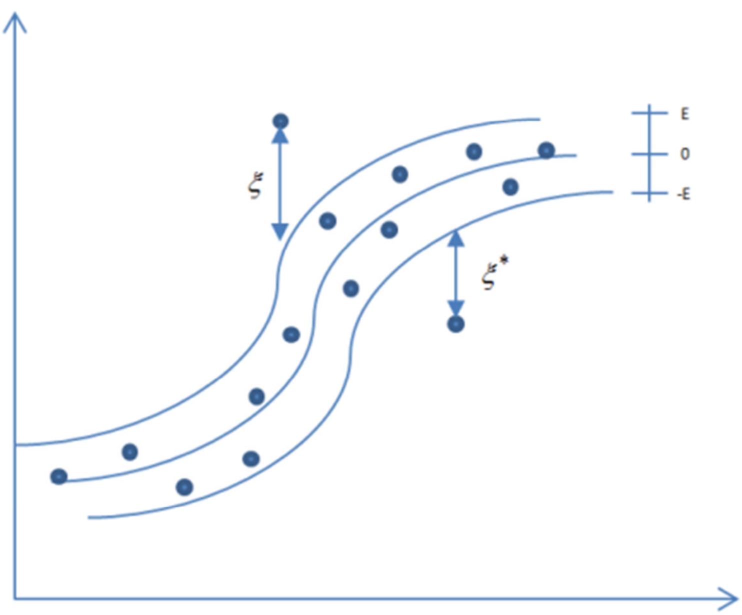


Figure 5: Regression line in input space

According to Smola and Scholkopf (2004), Support Vector Regression (SVR) modeling is a regression approach based on the frame work of Support Vector Machines which can perform better than linear regression and ANNs. SVR is better because it minimizes structural and empirical risk. The objective is to

fit a linear regression to a data set in a higher dimensional feature space after mapping the data set from a non-linear input space (Figure 5).

Similarly to regular least squares regression, a line (Equation 17) is fitted to the dataset in feature space by minimizing the sum of errors. In Equation 17, $\langle \bullet, \bullet \rangle$ denotes the dot product in the input space, x is the input vector, w is the weight vector, and b is the bias. Unlike linear regression, however, SVR allows for an Epsilon

$$f(x) = \langle w, x \rangle + b \qquad \textbf{(Eq. 17)}$$

(E) error margin. Therefore, only errors outside this margin are considered. This error margin helps to improve generalization by excluding samples inside the error margin. These errors or deviations above the accepted Epsilon (E) error margin are referred to as slack variables and are formally described as an E-intensive loss function (Smola and Scholkopf 2004). Formally, this loss function is described as follows:

$$|\xi|_E := \begin{cases} 0 & if |\xi| \le E \\ |\xi| - E & otherwise \end{cases} \qquad \textbf{(Eq. 18)}$$

Additionally to the minimization of the sum of errors, the objective function includes a component to minimize the flatness of the function by calculating the norm of the weight vector. This norm of the vector provides for the quadratic structure of the function which insures one single minimum point. The weight vector and the bias are obtained by minimizing the objective function subject to the constraints (see chapter 8 for optimization).

2.8.3 Non-Sequential vs. Sequential Methods

A non-sequential model is a method that does not consider previous information about a sample. Therefore, non-sequential models lose some information that may be important in detection and prediction. The use of sequential data in emotion detection and prediction is important because knowing the previous emotional state of an actor can be very useful to determining the current emotional state.

This is referred to as sentiment flow. Prediction of sentiment flow involves learning techniques that can iteratively store and retrieve previous state information about an actor or entity.

Studies that have addressed the issue of sentiment flow over time include: Burns (2003), Mao (2006), Wang (2008), and El-Nasr (1999). In El-Nasr (1999), the authors developed PETEEI, a pet with evolving emotional intelligence. This system used a three dimensional table to keep track of agent actions. Three actions in a sequence were used to learn probabilities about state transitions and their possible outcomes for emotion detection. This system combined a fuzzy logic model with heuristic rules to simulate emotion in agents which depended on action sequences and rewards. The system can adapt emotional state and intensity based on probabilities of actions in a sequence (sequence mining) and user feedback. Similarly, in Burns et al. (2003) the authors propose the use of hierarchical Bayesian networks for adaptive reasoning over time. Mao et al. (2006) conducted a study on the prediction of sentiment flow in documents using conditional random fields. The objective of their study was to predict a sequence of sentiments in a document based on a sequence of sentences. Results of their study indicate that sequential models are better than non-sequential models at predicting and describing sentiment. Additional general purpose methods used for sequence mining include Hidden Markov Models (HMM) and Temporal Bayesian networks based on probability approaches. Although probabilistic approaches are common, they have disadvantages related to the reliability of the probabilities, the assumptions of independence, and how to calculate the probabilities. In this dissertation, non-probabilistic alternatives to these sequence mining techniques which consider sentiment flow are proposed.

2.9 System Response to Emotions

Responses that an automated system can have to human emotions are an important aspect in affective HCI. To complete the interaction cycle, an automated system needs to provide a response based on emotion inputs. The system response refers to what the machine will do in response to the detection of a particular emotion. Responses can be in tone of speech, facial expression, type of language, physical proximity in the case of robots, and environment background. A facial response system, for instance, can

be implemented in hardware using a robot or in software using animation of computer graphics. Robotic approaches have advantages with relation to presence because they are material tangible entities which a human can relate to and physically perceive. “Embodied Conversational Agents” (ECA) are another approach in human computer interaction. They are important in virtual worlds because they provide a visual simulation of human faces and expressions (Pelachaud 2009). These interfaces can be used as dialogue systems which the system uses to respond to the user. Important studies in this area include Massaro et al. (2001), Gratch (2002), Cassell (2001), and Pelachaud (2009).

This dissertation’s main focus is in providing system responses using 3-D virtual worlds with virtual agents in them. Therefore, the next sections cover the 3 aspects used in this work to provide system responses. The 3 aspects are: environment mapping, facial expression mapping, and speech generation.

2.9.1 Environment Responses

Automatic rendering of environments or scenes from specifications and language inputs (text-to-scene processing) has been addressed in several ways. In Mukerjee et al. (2000), a methodology was proposed using a large visual database of objects and actions, and a set of domain specific constraints to generate visual scenes of 2-D urban parks. Several instances of an object are generated and the ones closest to the linguistic description are selected for display. In Coyne and Sproat (2001), the authors propose WordsEye, a system that generates scenes based on input text descriptions using frame based and an entity-position language. Other examples include: Carsim (Johansson 2004), and the Put system (Clay and Wilhelms 1996). In general these systems use heuristics and key word spotting to find relevant words that can be used to select a scene.

2.9.2 Facial Expression Responses

One of the most influential studies on how facial expression of emotions is conceptualized was done by Paul Ekman (Ekman 1998, 1978). In this study, the authors showed that certain emotions are consistent across cultures and can be expressed by a set of face muscle positions. Other important discussions on emotions can be found in Jenkins et al. (1998).

A facial response system can be implemented in hardware using a robot or in software using animation of computer graphics. Currently, robots have limitations in expressing emotions because human emotions are conveyed mostly by complex facial expressions. The human face has many muscles and skin which all need to work together to produce an emotion expression on the face. This ability can be limited at the hardware level with robots as was shown in a recent study by Beer (2009). Instead, other authors propose the use of virtual agents to communicate back to users.

This dissertation is only concerned with the methodology necessary to produce a facial expression response without worrying about the actual medium. In theory, the methodology could work with any 3D (computer graphics) model or physical (mechanized) model such as a robot. Therefore, only the techniques to define a mapping model and the actual implementation medium are discussed. In this case the implementation medium used is a 3-D mesh of a humanoid man. Blender's "Mancandy" 3D puppet (Gumster 2009) is used as the 3-D model (Figure 6) although any 3D model could be used.

Implementation of a facial response by rendering a 3-D animation of a human face is usually referred to as facial expression in virtual agents. Substantial research has gone into this area and there are many successful techniques that can be used to accurately render facial expressions using computer graphics (CG). In Noh and Neumann (1998), the authors provide a good survey of the main facial modeling and animation techniques.

Figure 6: Mancandy 3-D puppet

They divide the field of facial modeling/animation into two basic groups which are: geometry manipulations and image manipulations. Geometry manipulations are further divided into Interpolation and parameterization. Since this research is focused on automatic rendering, only interpolation and parameterization methods are discussed. For the more detailed description of the field as a whole, the reader may refer to Noh and Neumann (1998).

As the name implies parameterization generates possible facial expressions based on parameter values. Special tuning and coded rules are required to convert the parameters into deformation or transformation commands that can render the desired facial mesh topology. The Facial Action Coding System (FACS) (Noh and Neumann 1998; Ekman 1978), for instance, is an approach which is based on the movements of facial muscles and the jaw. FACS consists of 44 basic action units which represent these basic facial movements. Studies like Chen (2008) have implemented FACS approaches for face animation. Their work extracted 28 features from facial expressions in video. They used a linear model to map the features into rendering parameters for 3D facial expression modeling. The linear model is used to calculate a new vector of coordinates (i.e. x, y, z points) to render the new facial expression. The extracted features are the input to the model which is multiplied by the action units to produce the new coordinate vector.

Interpolation through mesh morphing between targets is another important technique used for facial expression rendering. Mesh morphing (Akenine-Moller et al. 2008) is the process of seamlessly generating a mesh by combining the vertices from a neutral mesh with the equivalent vertices from one or more pose meshes. This process is achieved using GLSL shader programming. The objective of mesh morphing is to perform an interpolation on a per vertex basis that allows different points to be combined. Therefore, a difference vector is calculated between the neutral mesh point and the target mesh point. These difference vectors are added to the neutral vector and can be adjusted by weights. Using weights to adjust the deformation allows for facial expressions, talking sequences, and gestures to be modified by simply changing the weights of the per vertex difference between a neutral pose and a target. This technique has advantages with regards to speed, efficiency and quality of rendering because it performs

all processing in the GPU using shader programming techniques. Wang et al. (2007) used this technique to render emotional expressions.

Rendering can be done on any type of 3-D animation software such as OpenGL, Blender, Maya, etc. The set of parameters can be conveyed to a 3-D model in Maya or Blender by way of scripting languages such as MEL for Maya and python for Blender or through C\C++ code using OpenGL. For this dissertation, examples of these techniques are provided using GLSL shaders, C\C++ with OpenGL, and python with Blender. The key aspect of how to calculate this weight based on language inputs is addressed in chapter 8.

2.9.3 Speech Response

Speech responses require the automatic generation of speech with emotion attributes. The analysis of emotions in speech has not been limited to just capturing emotion but it has also tried to generate emotion in speech synthesis. One such study was conducted by Theune et al. (2006). In their study, the researchers focused on developing a storytelling speaking style geared for children. A set of prosodic rules was explored to convert neutral speech into storytelling speech.

There are many very good speech synthesizers currently available in the market. Microsoft's operating system Windows provides MS SAPI speech tools which are very clear and effective. An application using SSML mark-up language can easily be developed to read any text provided as a text file and generate a speech signal. Parameters such as speed, pitch, and volume in the SSML mark-up can be adjusted to simulate voice speed and intonation. Parameter tuning depends on the system analysis of the multi-modal sources.

2.10 Emotion Detection and Response Implications in Affective HCI

As with all proposed automated systems that deal with specific groups of people, there are several important issues that must be considered or at least discussed. These issues relate mostly to the problems and benefits that may arise from using such systems. In this section a brief description of some of the

most important issues is presented. Tapus et al. (2007) has proposed some important issues to consider in healthcare such as:

- **Human attachment to automated systems**: This issue is important because social automated agents can become an object of affection for a human (especially for children or the elderly).
- **Erroneous learning by system**: This aspect is important as an inappropriate response by a system could upset a user which would be contrary to the purpose of the system.
- **Erroneous interpretation of system actions by patient**: Even if the system provides the pre-determined response, improper development on the interface by the developer can result in an incorrect interaction.

CHAPTER 3: APPROACH

Detection of emotion from language inputs towards affective HCI requires multiple issues to be resolved. This chapter provides an overview of the methodologies used in this dissertation and how they fit together.

Chapter 4 focuses on the task of identifying an affective feature set for reliably detecting emotions in text. For this initial work, detection was performed at the sentence rather than actor level. Results from this work also helped identify and resolve issues concerning corpora requirements and unbalanced training sets, as well as establishing a baseline evaluation for performance and the difficulty of classifying between the multiple emotion classes.

In supervised learning training and testing the models requires an annotated corpus. No existing corpus had the necessary annotations for this work, and therefore it was necessary to build one. Chapter 5 describes the process by which text/speech corpora were developed to train and test the methodologies developed in this work. Two annotated corpora were developed, one based on children's stories, and another on dialog from a popular medical TV series. The purpose of two different corpora was to provide some measure of the generalizability of the methodology in this work to different communication forms.

Chapter 6 focuses on a multimodal (speech + text features) emotion detection methodology aimed at improving performance over previous results. As in chapter 4, detection was performed at the sentence level. A step-wise was used to first classify and filter out neutral sentences, and then a second phase used to distinguish between positive, negative, and the 5 emotion classes for the remaining sentences. In addition, recurrent features based on previous story context and sentiment composition/flow were added.

Actor level emotion detection requires a reliable mechanism for determining which actors are present at any time. Chapter 7 develops a methodology for actor and environment detection in text stories. In chapter 8, all previously described chapters are brought together to develop a methodology for actor-level

emotion magnitude estimation for use in emotion expression rendering which is based on a recurrent (prior-state feedback) regression model.

Finally, chapter 9 demonstrates how the detected actors, environments, emotion classes and magnitudes, and other features can be used for automatic text-to-scene rendering in virtual worlds. Each chapter includes experimental results, discussion, and conclusion of the proposed methodologies.

- **Tools**

The following tools are used to perform the tasks required in this methodology: Python 2.6 and NLTK (Bird et al. 2009); BART: Beautiful Anaphora Resolution Toolkit (Versley et al. 2008); Machine Learning with LibSVM, Matlab optimization toolbox, and WEKA (Witten and Frank 2005); Stanford parser (Klein and Manning 2002) and NER algorithms; ConceptNet 3.0 (Havasi 2007) and WordNet (Miller 1995), and Praat (Boersma and Weenink, 2005).

3.1 Assumptions

To reduce the complexity of this task, the following assumptions were made:

- Actors are sentient entities (people and creatures) which can feel and cause emotions. In general, such entities are identified by nouns and noun phrases. Each story is analyzed independently and actors from one story to another are assumed to be different. There can be multiple actors in each story or conversation. The annotations of the Affect Corpus 2.0 (Calix and Knapp 2011) are used to identify and track actors in a story.

- Environmental elements are places and things which can provoke emotions (i.e. being in a dark cemetery can cause fear). Location and mood can influence emotion state (i.e. being in a dark forest vs. a birthday party).

- Each actor can have multiple emotions at the same time. Actors hold the following types of emotions: angry, sad, happy, surprised, afraid, and neutral. The emotional state of each actor is evolving throughout the story. Actors are represented as emotion-position vectors where each

emotion can vary in degree of magnitude from 0 to 1. Here, 0 represents absence of the emotion and 1 represents full degree of emotion.

- Evolving emotion state is tracked by sentence position in the story rather than time. The methodology does not track emotional state across flashbacks or flash forwards in the timeline.

- Emotion detection considers sequences of sentences using a window approach. All previous emotional information is considered to be stored in each actor's emotional state vectors as recorded on a sentence basis.

- Neutral (no emotion) states represent a high percentage of the total number of states in the real world and corpus. Over-sampling or weight adjustment may be necessary to deal with the class imbalance problem present in the corpus.

- The setting is a controlled environment free of background noise
- There is only one speaker reading the sentences for a story; however, it is not assumed that the same speaker will read each story.

3.2 Performance Assessment

This section discusses how the system performance was measured and analyzed to determine the accuracy of the methodology.

3.2.1 System Accuracy on Test Corpora

For purposes of evaluation, as is standard in the field of machine learning, the corpus is divided into two sections; one section for training purposes and one section for testing purposes. The training set uses 80% of the corpus and the testing set uses 20%. Both the initial (UIUC) annotation of the corpus and the new annotations produced in this dissertation were used to train and test the models.

The objective is to determine how accurate the system's estimated calculations are in comparison to those provided by the human annotators. For classification tasks, metrics such as precision, recall, and F-measure are used to evaluate the predicted results. For regression estimation, the Root Mean Squared

Error (RMSE) and correlation coefficient metrics are used. Comparison to other studies such as Alm (2008) was also performed, although modifications or generalizations of the results from other studies may be necessary in order to have a more accurate comparison.

3.2.2 Text Annotator Agreement

An evaluation of obtained annotation agreement is performed using Avg_Ao, Pi, kappa, S, and alpha metrics. This approach is used to determine if the results are similar or if there are significant differences between annotators.

3.2.3 Generalization of the Method

Stories from a medical drama corpus were used to test the accuracy of the classification model in a more generalized setting. These resulting accuracy scores were compared to the results from the UIUC corpus to determine how the model generalizes to other domains.

3.2.4 Ranking Analysis of Emotion Triggers

An important analysis in this study is to determine the ranking of the different sets of features used for evolving emotional state change detection. In this case, an analysis was performed to determine which features have the highest contribution to emotion detection. Chi-square feature ranking and other methods are used for this task.

CHAPTER 4: EMOTION FEATURE EXTRACTION AND EMOTION RECOGNITION IN TEXT[1]

Selecting the right set of features is essential in training a classification model to correctly detect emotions. This chapter develops and evaluates a methodology to automatically extract lexical features that are relevant to emotion detection. These features are used to perform emotion classification using Support Vector Machines (SVM). Therefore, the methodology presented in this chapter uses a lexical only approach to determine the contribution of words to emotion detection.

The algorithm for the extraction of emotion related words uses intuition from the Mutual Information (MI) feature selection technique. Challenges in emotion detection such as class imbalances and the limits of lexical approaches in emotion detection are discussed.

4.1 Methodology

The work presented in this chapter expands on work done by Alm et al. (2005) and Alm (2008) by exploring in more detail how word features alone affect the accuracy of classification methods in emotion recognition in text. The methodology presented in this chapter is divided into 2 parts. First, a feature selection algorithm is developed to automatically extract emotion word features from an affect corpus based on the likelihood that the words are related to emotions. Second, a methodology to detect emotion in sentences using the extracted features, an SVM model, and unbalanced data is developed.

4.1.1 Automatic Feature Extraction Approach

Before developing the feature selection approach, the effects of other feature selections techniques such as mutual information (MI) on the classification task were explored. MI was selected for this comparison because of its tendency to favor terms that occur with less frequency (Yang and Pederson 1997). Its use is

[1] Portions of this chapter reprinted by permission of IEEE from Calix, R.A., Mallepudi, S.A., Chen, B., Knapp, G.M. (2010). Emotion Recognition in Text for 3-D Facial Expression Rendering. Multimedia, IEEE Transactions on Multimedia on, Vol. 12, no. 6, pp. 544-551, 1520-9210 © 2010 IEEE

appropriate in emotion detection since terms may occur infrequently but have high contribution to classification.

Although mutual information (MI) is commonly used in modeling of word/class associations and related applications (Yang and Pederson 1997), its performance in the initial feature selection in this chapter did not show significant improvement. Therefore, a modified approach for feature selection building on the framework of MI was implemented. The approach was determined after a detailed evaluation of the characteristics of the UIUC corpus. The analysis showed several intuitions that resulted in the feature selection approach presented here. First, the analysis of the corpus showed a considerable class imbalance due to having more neutral sentences in the corpus than there are emotion sentences. Class imbalance is an issue that can affect feature selection.

In MI, a metric is calculated to determine the association between terms and classes using Equation 8. To measure a term's ranking for selection purposes, an average MI metric for each term is calculated. This average is obtained by adding all metrics for a term per given class and dividing them by the total number of classes (Yang and Pederson 1997). Based on this average, all terms above a specific cut-off are selected. This approach can have disadvantages when dealing with imbalanced data. For example, given a cut-off of 0.45, if a term has an MI of 0.40 for the emotion class and 0.60 for the neutral class, under the described MI approach, the term would be selected since the average MI score is 0.50. In this case the neutral strength of the feature helped to select it even though the emotion MI metric was below the cut-off. The algorithm presented in this chapter is used to address this issue by only extracting features that have high emotion content.

Second, for fine grained emotion classification (e.g. happy, sad, angry, etc.), if a word is detected to be strongly correlated to emotion, then it is expected that the actual word meaning will help to determine what type of emotion the sentence is conveying. Therefore, fine grained classes can be combined so that emotion counts can offset neutral classes to some extent.

- **Approach**

The objective of the feature extraction approach is to identify word features that relate to emotions. The proposed feature extraction algorithm uses words from the actual training corpus and their related classes to learn the likelihood that a word is related to an emotion. To achieve this, the algorithm generates a two way association table of words and classes. First, the algorithm counts the number of times that each word appears associated with each of the six classes. Once the counts have been performed for all words, a probability of emotion content is calculated for each word. This probability is the result of dividing the total number of times that a word appears in an emotion class by the number of times the word appears in all classes. Therefore, Equation 19 includes the associations for the 5 emotion classes and Equation 20 includes the associations for all 6 classes. Formally, the model is as follows:

$$A_i = \sum_{j=1}^{5} W_{ij} \qquad \forall j = 1..5, i = 1..N \qquad \textbf{(Eq. 19)}$$

$$B_i = \sum_{j=1}^{6} W_{ij} \qquad \forall j = 1..6, i = 1..N \qquad \textbf{(Eq. 20)}$$

where A_i is the number of times a word appears associated with emotion classes, B_i is the number of times a word appears associated with any class, W_{ij} represents the number of times the word w_i is associated with emotion class j, W is the set of all words, and N is the number of words in W.

$$P_i = \frac{A_i}{B_i} \qquad \forall i = 1..N, B_i > 0 \qquad \textbf{(Eq. 21)}$$

Finally, P_i represents (Equation 21) the probability that the word w_i is related to emotion. Therefore, P_i is the probability of emotion content as determined by the approach. Once all the probabilities (likelihoods) per word are calculated, a selection of the emotion strong words is made. To control the precision of the output features, the algorithm uses a cut-off limit. This cut-off is used to select only those features that are

above a certain probability cut-off. The cut-off for the results presented in this chapter was set at 50% after performing empirical analysis with different values. For other domains, other empirical based or optimization based techniques can be applied to calculate this cut-off such as cross-validation.

4.1.2 Classification: Support Vector Machines

After the features have been extracted, a model is trained to detect emotion in descriptive sentences. For this classification, a Support Vector Machines (SVM) model was used. An SVM is used since the data is highly subjective and the classes are highly overlapped. Therefore, the SVMs kernel mapping capabilities are used to see if the data is easier to classify in a higher dimensional space.

The SVM was trained on 80% of the UIUC corpus and then tested on the remaining 20% to determine the accuracy of the classification task for all experiments. Classification was done using LibSVM (Chang and Lin 2001). The affect corpus (Alm 2008), used in this study, consists of 176 children's stories by three authors (The Brothers Grimm, H. C. Andersen, and B. Potter). Each word in the corpus has been annotated with its corresponding part-of-speech (POS) tag, and each sentence has been assigned a pair of emotion and mood labels.

The story sentences are classified into one of a set of pre-defined emotion classes. The final set of classes used for this study consists of: anger, fear, happiness, sadness, surprise, and neutral (the absence of emotion). To perform the classification task, each sentence is represented as a feature vector and classified using Support Vector Machines (SVM). Performance was measured using precision and recall accuracy scores. Feature vector values are binary (present/not present) and were extracted using python code, with the help of the NLTK toolkit (Bird et al. 2009). After the classification task, a feature analysis was performed to identify feature contribution to accuracy and to highlight possible approaches to 3-D emotion expression rendering.

4.1.3 Assessment of the Feature Extraction and Classification Methodologies

Different word feature sets were explored to determine their influence on classification accuracy. The first set consists of a small list of 230 initial emotion only words. This small sample of words was used to

establish a baseline accuracy score from which to compare. The second set consists of a longer list of 1085 initial emotion words. Both lists were collected by the researchers using WordNet and the Internet. Other larger sets of between 5,000 and 10,000 words were used to determine if there was an optimal set of words for this particular task. The final set was collected automatically using the feature extraction algorithm discusses earlier in the chapter. Additionally, to determine whether the number of samples of the training corpus affects classification accuracy, all tests were conducted on a subset of the corpus (i.e. Grimm's sub-corpus) and on the combined corpora (consisting of Grimm's, Potter and Andersen stories). Finally, stemming and stop word removal were used to improve the SVM's performance (Bird et al. 2009).

4.2 Analysis and Results

The analysis was performed in three phases. The first phase was a preliminary exploratory overview of the data to understand the strengths and weaknesses of the UIUC corpus. The second phase was a more detailed and exhaustive review to determine optimal parameters and kernels for the SVM classifier. This phase also included identification of a set of optimal weights to be used with the SVM to deal with the class imbalance challenge present in the data. The third phase involved the evaluation of the feature selection algorithm developed from insights obtained from the previous two phases. The results of the analysis are presented below.

Finally, since one of the objectives of this study is to provide a detailed analysis of this new emotion annotated corpus (i.e. the UIUC corpus), the results of the classification task are presented using multiple kernels, lists of features and other parameters. The best parameters used for the classification task included a cost of 32 and gamma of 0.0078125 for the RBF Kernel. These parameters were estimated via cross validation using LibSVM's built in grid search. Classification accuracy was determined by comparing the SVM's predicted classes to the human annotated classes. Only one of the two annotator classes per sentence was considered.

4.2.1 Preliminary Data Analysis

The preliminary analysis was a broad scope study of the data to gain an understanding of the peculiarities of the corpus and the challenges related to emotion classification. Analysis of the data was performed on each separate sub-corpus and on the combined corpus of all three authors (Grimm's, Andersen, and Potter). All 8 classes from the corpus were used in this preliminary analysis. In general, the Grimm's sub-corpus always yielded lower classification rates than the other sub-corpora.

Table 2 shows the results using two different lists of features (see appendix A for details of word lists). The first list included 230 features and the second list included 1085 features. From Table 2, it can be observed that of the three sub-corpora, testing on the Andersen corpus yielded the highest accuracy at 77.74 % with a set of 230 initial features and 78.30 % with a set of 1085 initial features. Among the three sub-corpora, the lowest accuracies were obtained using the Grimm's Sub- Corpus. This result may be due to Grimm's story type, which

Table 2: Results of classification task using support vector machines on three sub-corpora (Transactions on Multimedia © 2010 IEEE)

	Accuracy Score (230 initial Features)	**Accuracy Score (1085 initial Features)**
Grimm's Sub Corpus 5,360 Sentences	44.87%	47.29%
Andersen's Sub Corpus 7,996 Sentences	77.74%	78.30%
Potter's Sub Corpus 1,946 Sentences	74.04%	74.04%
All Corpora 15,302 Sentences	69.22%	70.55%

in general tends to be of a darker tone. With the Grimm's corpus, the classifier obtained accuracies of 44.87% and 47.29% for the 230 features list and the 1085 features lists, respectively. For the Potter sub-corpus, the classifier achieved accuracies of 74.04% for both sets of features. Testing on the combined corpora (Andersen's, Grimm's & Potter') resulted in an overall accuracy measure of 69.22% for a list of

230 initial features and 70.55% for a list of 1085 initial features. These results appear promising but do not say anything about the models ability to discriminate between emotion classes.

To determine the impact of the feature selection technique in discriminating between emotion classes, the emotion samples were broken down into the individual emotion classes (e.g. happy, sad, angry, etc.). This break down helps to better understand how each individual sentence is being classified by the SVM.

Additionally, by separating the sentences into their respective emotion classes, the effects of neutral sentences on the overall classification score can be separated. This separation is important because a large portion of the sentences in the corpus were annotated as neutral sentences. The results of this analysis can be seen in Table 3. The results from Table 3 indicate that there is a class imbalance challenge in the data which is affecting the classification accuracy of the SVM model.

Finally, mutual information feature selection was performed to reduce the number of features in the dataset and see if reducing the features would improve classification accuracy.

Table 3: SVM classification accuracy-failure analysis per emotion class (Testing set: 20% of corpus) (Transactions on Multimedia © 2010 IEEE)

		List of Features (1085)		List of Features (5400)		List of Features (All words)	
Class	**Total**	**Sent.**	**Accu.**	**Sent.**	**Accu.**	**Sent.**	**Accu.**
Angry (1)	158	15	9.49%	12	7.59%	8	5.06%
Disgusted (2)	80	0	0%	1	1.25%	1	1.25%
Fear (3)	162	37	22.83%	31	19.13%	8	4.93%
Happy (4)	273	41	15.01%	41	15.01%	33	12.08%
Sadness (5)	106	12	11.32%	18	16.98%	15	14.15%
Surprised + (6)	87	0	0%	0	0%	0	0%
Surprised – (7)	103	0	0%	0	0%	2	1.94%
Neutral (8)	2091	2054	98.23%	2040	97.56%	1995	95.40%
Total	**3060**	**2159**	**70.55%**	**2143**	**70.03%**	**2062**	**67.39%**
All Emotions	969	105	10.83%	103	10.62%	67	6.91%

Sent. = Number of correctly classified sentences per class Accu. = Accuracy score per class
Total = Total number of sentences per class

As can be seen in Figure 7, the mutual information approach does not improve the system's per class classification accuracy. Therefore, this analysis indicates that another feature selection approach is needed to improve classification accuracy. Results of the developed feature selection approach are presented in Figure 8 and Figure 9.

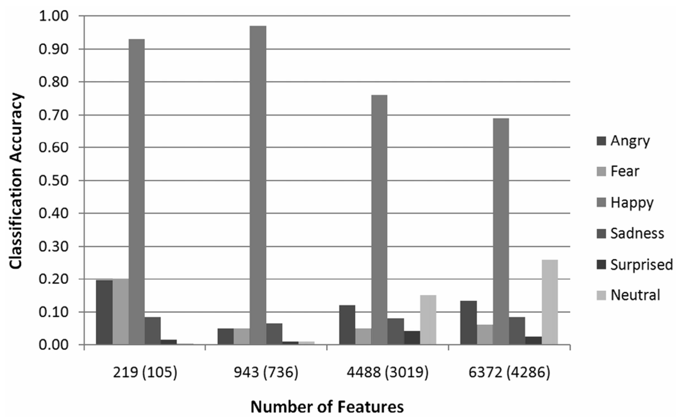


Figure 7: Linear kernel (using Mutual Information feature selection)
(Transactions on Multimedia © 2010 IEEE)

The next section presents the results of a more in-depth analysis to obtain optimal SVM parameters, weights, and kernels to deal with the class imbalance challenge present in the data and try to achieve better classification results.

4.2.2 Kernel-based Data Analysis and Class Imbalance

In this section, the analysis of the corpus was performed using the four main Kernels used in Support Vector Machines (Linear, Polynomial, RBF, and Sigmoid) (Chang and Lin 2001). The experiments were performed on a list of 234 initial features, 1090 initial features and an all words list. In general, the Linear Kernel provided higher per class accuracy scores. Linear Kernels tend to be better when the size of the feature set is high (Chang and Lin 2001). The linear Kernel was closely followed by the RBF Kernel. The other Kernels were not effective for emotion detection in this corpus. Therefore, they were discontinued for subsequent experiments. Additionally, the number of classes was condensed from 8 to 6 classes. The new reduced set was used to try to reduce the possibility of error in the classification task. Related classes were merged together (i.e. surprised – and surprised +).

Because of the class imbalance challenge present in the data, a set of weights had to be identified which could improve classification accuracy. After trying several weight combinations, a set was identified which consistently provided better results for the classification task across all feature sets. See LibSVM (Chang and Lin 2001) for details on how to apply the weights for model training in SVM. The set of weights is provided in the following table (Table 4):

Table 4: SVM weights for class imbalance challenge (Transactions on Multimedia © 2010 IEEE)

Class i	Weight i	Cost	Cost i
Anger	7	32	224
Fear	3	32	96
Happy	8	32	256
Sadness	5	32	160
Surprised	4	32	128
Neutral	1	32	32

4.2.3 Improved Results Analysis

The final phase of the analysis was performed using improved parameters only. That is, only parameters that were determined to provide best results. Results for the classification task using the linear kernel are presented in Figure 7 and Figure 8. For Figure 7, Figure 8, and Figure 9 the horizontal axis represents the number

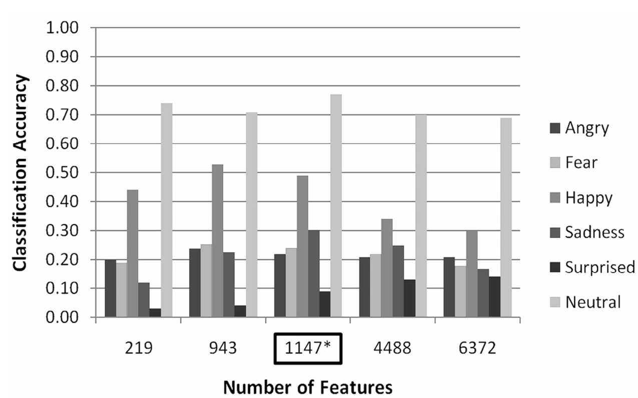


Figure 8: - Linear kernel * = emotion feature selection algorithm (Transactions on MM © 2010 IEEE)

of features used and the vertical axis represents the accuracy scores achieved by the SVM. Figure 7 presented the results of the classification task using Mutual Information Feature selection and a linear kernel. From Figure 7, it can be seen that smaller sets produced better classification results than larger feature sets under MI selection.

The analysis presented in Figure 8 and Figure 9, on the other hand, includes the results of the analysis using the proposed feature extraction algorithm. From the analysis presented in these two figures, it can

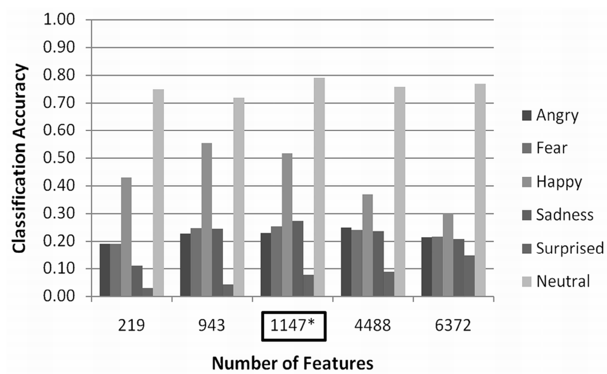


Figure 9: RBF kernel * = emotion feature selection algorithm (Transactions on MM © 2010 IEEE)

be seen that the classification accuracies are still poor using lexical features for emotion classification but that the automatically selected features perform as well as the manually selected features. This finding is

Table 5: Sample of words from the feature selection list that contributed to classification accuracy (Transactions on Multimedia © 2010 IEEE)

Angry	**Fear**	**Happy**	**Sadness**
annoyed	dreadfully	hurrah	alack
hatch	fright	seam	undone
jay	concert	thanked	sad
furious	frightful	smiling	mourning
anger	befall	dripped	mournfully
sale	uneasy	smiled	sorrowfully
grumbling	mischance	excited	wept
unicorn	terrified	perfectly	sighing
boar	army	rejoiced	despair
punish	hanged	heartily	avail
angrily	grating	friendly	

very important because it indicates that words alone are not enough for emotion detection and that the class imbalance in the training corpus must be addressed to improve classification results. All final tests were performed using the porter stemming algorithm (Porter 1980), the set of weights presented in Table 4 and using a condensed list of 5 emotion classes and neutral. Stemming causes the list to be reduced in size after removing duplicates (e.g. from 234 to 219 features).

Finally, Table 5 presents a sample of some of the words that were selected by the proposed word feature extraction algorithm. It can be seen from this list that of all the features selected automatically, many were emotion related words. This helps to corroborate the intuition that words with an emotion context have a higher contribution to classification accuracy.

4.3 Conclusions

Based on the analysis of the results, it can be concluded that the main factor that contributed to classification accuracy was the quantity of emotion content in the selected features from the training corpus.

Another interesting conclusion of the analysis presented in this chapter is that automatic word feature extraction algorithms can obtain feature sets that are as good as those obtained manually. This approach, therefore, can save time and reduce costs by obtaining lexical features automatically and can be extended to other domains such as healthcare.

Additionally, it is important to note that emotion only words (e.g. happy, sad, angry, happiest, etc.) are not enough to classify the sentences. Instead, other words termed here "emotion context" words like 'death' or 'marriage' should also be included to capture some context of the sentiment content in a sentence. These types of words are what the feature extraction algorithm helps to extract. The degree of emotion content in the features can be controlled by the cut-off parameter depending on the application.

Finally, the high presence of neutral sentences versus emotions sentences creates a class imbalance in the training set which makes fine grained emotion classification more difficult to perform. As a result,

some type of under-sampling or oversampling technique must be applied to balance the data and obtain better results.

CHAPTER 5: EMOTION CORPORA[2]

Advances in multimedia semantic analysis require rapid analysis and implementation of new tools and methodologies for information understanding. Currently, there are a relatively small number of existing annotated corpora for affect and multimodal content detection, and these corpora are limited in affect information available. This impedes the development of new ideas and implementations that require corpora for training and testing. To address this issue, affect and multimodal corpora needs to be created or extended and made available.

The annotations developed and presented in this chapter extend upon corpus work started at UIUC (University of Illinois at Urbana-Champaign). Additionally, for generalization purposes a corpus based on TV medical drama conversations which will be referred to as the LSU-MD corpus was also developed.

- **The UIUC Corpus**

The UIUC (University of Illinois at Urbana-Champaign) affect corpus of children's stories which is used as a base for this work is a relatively new corpus designed specifically for emotion classification which was developed as part of a dissertation project by Cecilia Alm at UIUC (Alm 2008). The UIUC corpus consists of 176 children's stories by three authors which are the Brothers Grimm, H. C. Andersen, and B. Potter. Each word in the corpus has been annotated with its corresponding part-of-speech tag, and each sentence has been assigned a pair of emotion and mood labels by human annotators. The corpus is annotated with eight emotion classes. They are anger, fear, happiness, sadness, positive (+) surprise, negative (-) surprise, disgust, and neutral (the absence of emotion).

Each of the 176 stories in the corpus is stored in two text files. One text file contains the annotation of the emotion classes for each sentence and the other has the part-of-speech tag for each word. The corpus is contained in a directory which is divided into three sub-directories; one for each author. Within each

[2] Portions of this chapter reprinted by permission of ACM from Calix, R.A., Knapp, G.M. (2011). Affect Corpus 2.0: An Extension of a Corpus for Actor Level Emotion Magnitude Prediction. In Proceedings of the 2nd ACM Multimedia Systems (MMSys) conference, Feb. 2011, San Jose, California, U.S.A. © ACM, Inc. http://doi.acm.org/10.1145/1943552.1943570

sub-directory there are two additional sub-folders named “emmood” and “pos”. The “emmood” directory contains all the stories in the corpus annotated at the sentence level with emotion classes. An example of annotation in this directory is as follows:

0:0 N:N N:N Once upon a time there was a village shop.

The first two numbers in the example above, indicate sentence position in the story corresponding to two human annotators. The four Ns are two sets of emotion and mood labels which were provided by the human annotators. Emotion can be any of the eight classes as described earlier. Mood is a relatively longer lasting and less specific emotion state. Mood is not considered in this work.

The “pos” directory contains part-of-speech (pos) tags for each word in a story. Each story is stored as a text file which includes pos tag for each word. An example of annotation in this directory is as follows:

(RB Once):(IN upon):(DT a):(NN time):(EX there):(AUX was):(DT a):(NN village):(NN shop):(. .)

The naming convention for the UIUC corpus is to use the name of the story for each text file that contains it. Names with spaces are separated with underscores. The file extensions are “.emmood” for stories in the “emmood” folder and “.sent.okpuncs.props.pos” for stories in the “pos” folder (Alm 2008). In the next sections, a detailed description of the new corpus annotations preformed for this dissertation is given.

5.1 Methodology

5.1.1 Affect Corpus 2.0 Annotation and Evaluation Methodology

To address the issues presented in this dissertation, the UIUC corpus has been used and extended. This new extended corpus, the Affect Corpus 2.0 (Calix and Knapp 2011) includes automatic annotations such as syntactic parses using state of the art NLP toolkits as well as manual annotations for affect magnitude detection, actors, actor presence and reference resolution. Human annotations are specific to the areas of affect detection, sentient nominal entity recognition, and actor presence resolution. Both types of annotations can be used for machine learning methodologies. The magnitudes can be used as classes or predicted values (outputs) and the automatic annotations as features for a learning methodology (inputs).

- **Automatic Annotation**

The Stanford parser and BART toolkit (Versley et al. 2007) were used to produce XML mark-up versions of each story in the UIUC corpus. The XML mark-up provides a simple data structure which can be used to automatically extract text features from the stories. Using this approach, all additional NLP information about each story is contained in the XML mark-up. This includes the sentence parse for each sentence, POS tags, enamex tags for actors, tags for semantic classes about each actor, and other useful tags. Additionally, referring expressions (pronouns) and their sentence position for each story were extracted.

- **Manual Annotation**

In this step, human annotators read, identified, and annotated actor level emotion magnitudes in text. The UIUC corpus was extended by manually extracting the actors in each story and annotating them with their evolving emotional state. Three human annotators (two males and a female) used a new annotation tool to select actors, annotate their presence in a sentence, and assign evolving emotional state per actors throughout the story. The annotators were undergraduate and graduate students from the college of engineering at Louisiana State University (LSU). The structure of the output frame for each actor-emotion annotation pair is as follows:

AE (SentenceNum, Actor, HappyMagnitude, AngerMagnitude, SadMagnitude, SurpriseMagnitude, AfraidMagnitude, Presence, sentence)

Annotation of actors was performed assuming that the stories were acted out as plays to simplify the annotation process. The annotator was free to assign the name to be used for the actor but had to use the same name throughout the story (e.g. as a unique ID).

Whenever an actor was present in a sentence, the annotator had to select the actor name to indicate presence. References to the actor can be made by name, referring expression, or implied in the context of each sentence. The ID used for each actor is the name, if available, or a description of the actor (e.g. the old miller). The annotated actors for each story are stored in a text file.

- **Annotation Tool**

For the purposes of this dissertation, a new VB.NET based annotation tool was developed which is available from LSU-NLP. The annotating tool has two main sections (Figure 10). One section displays the story and highlights the current sentence being annotated. The second section displays the actors in the story. This second area can be considered the stage where the actors will interact. In this section, each actor has five scroll bars which are used to adjust the magnitude of each of the emotion classes. The score for each emotion is from 0.0 to 1.0. Emoticons are also included to help the annotators assign the magnitudes. The emoticon expressions change as the annotators adjust the scroll bars for emotion magnitude assignment.

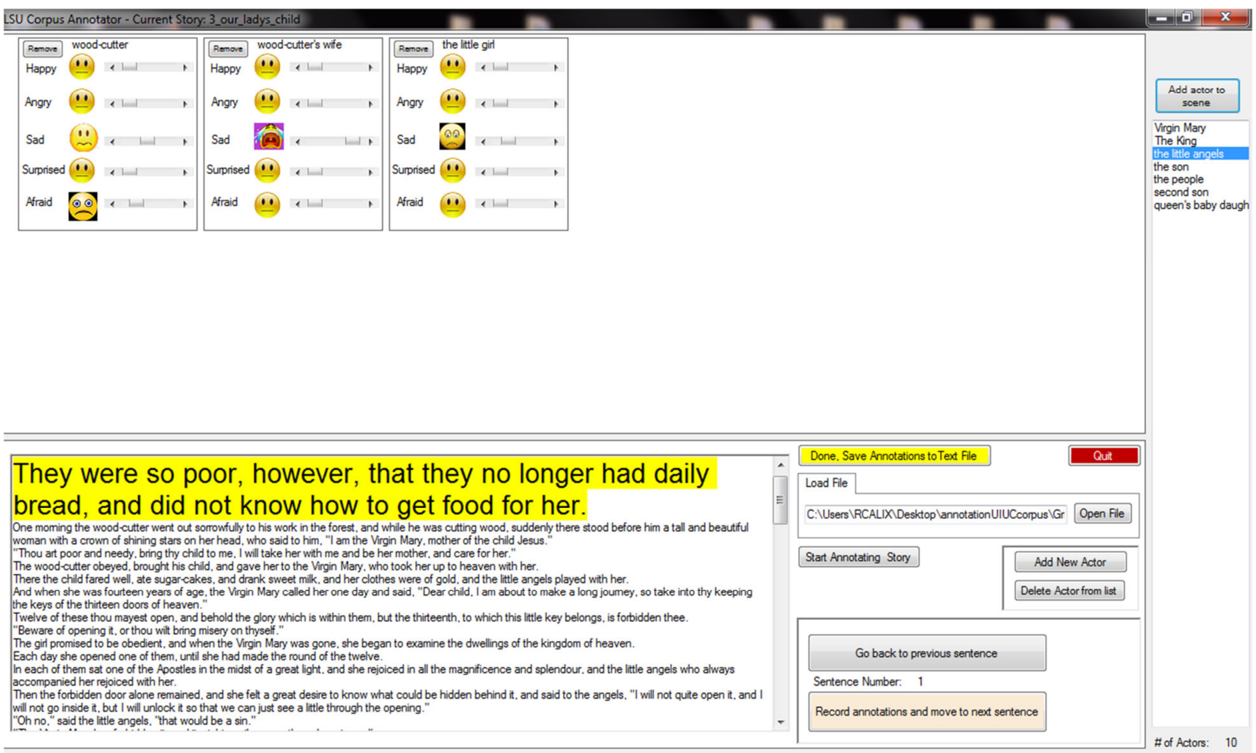


Figure 10: Annotation tool

The annotating application records the annotator's user name and relevant statistics. At the start of each session, the annotator proceeds to select one of the stories he or she is responsible for. The annotator loads the story into the system by searching for the story's name in the stories directory. When the user clicks

on the “begin annotation” button, the first sentence of the story is highlighted and the user can begin annotation.

Each sentence in the story is highlighted as the user annotates it. As the user is reading each sentence, he or she will identify actors in the story and determine if they have already been added to the list of actors. If the actor is not on the list, the annotator will click a button to add the actor to the actor’s list section. Once the actor is on the list, the annotator can select the actor and add him or her to the stage section of the annotator tool to adjust emotional state at the given sentence position in the story. This process is repeated for each sentence until the story is completed.

Each annotation produces vectors with the actor name, emotional states, and the position of the actor’s in the story. Emotional states per actor need to be adjusted only if the states change. Otherwise, the previous states are recorded for the current sentence. Once the actor leaves the scene, the emotion states are reset to zero.

- **Speech Extension**

The UIUC corpus was also extended for speech by adding audio recordings for 89 stories currently available from the corpus. Some of the stories were obtained from Librivox as public domain MP3 files. Other stories were recorded by a professional reader (see acknowledgement section). Praat was used to perform manual annotation on the speech files so that the speech segments aligned with previously annotated sentences from the UIUC corpus. The size of each speech recording per story is around 5 MB. All annotated files are stored in MP3 format.

- **Annotation Detail**

The Affect Corpus 2.0 includes annotations of the actors in each story and their presence in a given sentence. The annotated actors for each story are stored in a text file. The corpus also includes actor level emotion magnitude annotations to perform the training and testing of the prediction model. Therefore, the emotional states and magnitudes can be determined at the actor level. Magnitudes were annotated for 5

emotion classes: happy, sad, angry, afraid, and surprised. Neutral is represented when the emotion magnitude from all emotion classes is set to zero.

poor man
poor man's son
The king
King's daughter
the snake
second snake
the servant
the skipper

Figure 11: Actors for "Three snake leaves" by the brothers Grimm

The corpus includes a file of actors per story, and a file of actor-emotion-magnitude vectors for each story. The list of actors is a simple ASCII format file containing the names for each sentient actor in the story. An example is provided in Figure 11. For the actor-emotion vectors, information on sentence position, actor name (ID), and emotion magnitudes is provided. Each vector (Figure 12) includes a

[84,mrs. rebeccah puddle-duck,0,0,0,0,0,0,Jemima became much alarmed.]
[84,jemima puddle-duck,0,0,0,42,66,1,Jemima became much alarmed.]
[84,Kep,0,0,0,0,0,0,Jemima became much alarmed.]
[84,gentleman with sandy whiskers,0,0,0,0,0,1,Jemima became much alarmed.]

Figure 12: Emotion vectors with emotion magnitudes

presence feature which indicates if an actor is present in the sentence. Here, an actor could be referred to by name, a referring expression or implied by story context.

5.1.2 Multimodal Health Care Related Corpus Annotation and Methodology (LSU-MD)

For generalization purposes, a second multimodal corpus (LSU-MD) was annotated. The LSU medical drama (LSU-MD) corpus is dialogue based, multiple speaker, extraneous noises (beepers, machinery), and with shorter utterances. It consists of 6 episodes from the popular television show "Grey's Anatomy". The annotation tool was used to assign the emotion classes at the sentence level. This second corpus is used to compare the emotion classification methodology in two domains: narrative (UIUC) and dialogue (LSU-MD).

5.2 Analysis and Results

5.2.1 Affect Corpus 2.0

The new extension of the affect corpus, the Affect Corpus 2.0 (Calix and Knapp 2011), consist of 173 stories that were annotated with emotion magnitudes, actors, and actor presence information. The emotion magnitude classes included happy, sad, angry, surprised, and afraid. The neutral class was represented when all emotion class magnitudes were set to zero. From a syntactic parse analysis, it was observed that the corpus contains 45,120 noun phrases (NPs) with Nouns as constituents. Additional, statistics and metrics about the corpus can be found in Alm (2008) and Calix et al. (2010).

For the speech section, 89 audio files were collected that matched the transcript of 89 children's stories in the UIUC corpus. Therefore, a total of 89 stories in the Affect Corpus 2.0 include an audio recording, a text transcript, and emotion magnitude annotations. The audio versions are stored in MP3 format. These audio files were manually annotated and are used to extract speech features at the sentence level. Characteristics of the different speakers that recorded the stories in this Affect Corpus 2.0 can be seen in Table 6.

Table 6: Children's stories speech dataset characteristics

	Number of audio recordings	Male	Female	Number of speakers
Grimm	59	38	21	32
Potter	18	11	7	15
H.C. Andersen	12	5	7	9

The emotion vectors are very important because they can provide information about sentiment flow per actor in the story. This can allow sentiment flow to be represented as an emotion signal over time (Figure 13). Several important aspects of the stories such as correlation between actors and actor emotions can be visualized with the emotion magnitude vectors. For example, in Figure 13, the evolving emotional state of Tom Thumb in the story can be visualized. Here it can be seen, without reading the story, that Tom Thumb started out happy but then something happened that caused him sadness, surprise and a lot of fear.

These problems, however, seem to have been resolved because the final emotion state is happy. The referring expressions are provided in a text file for the entire corpus.

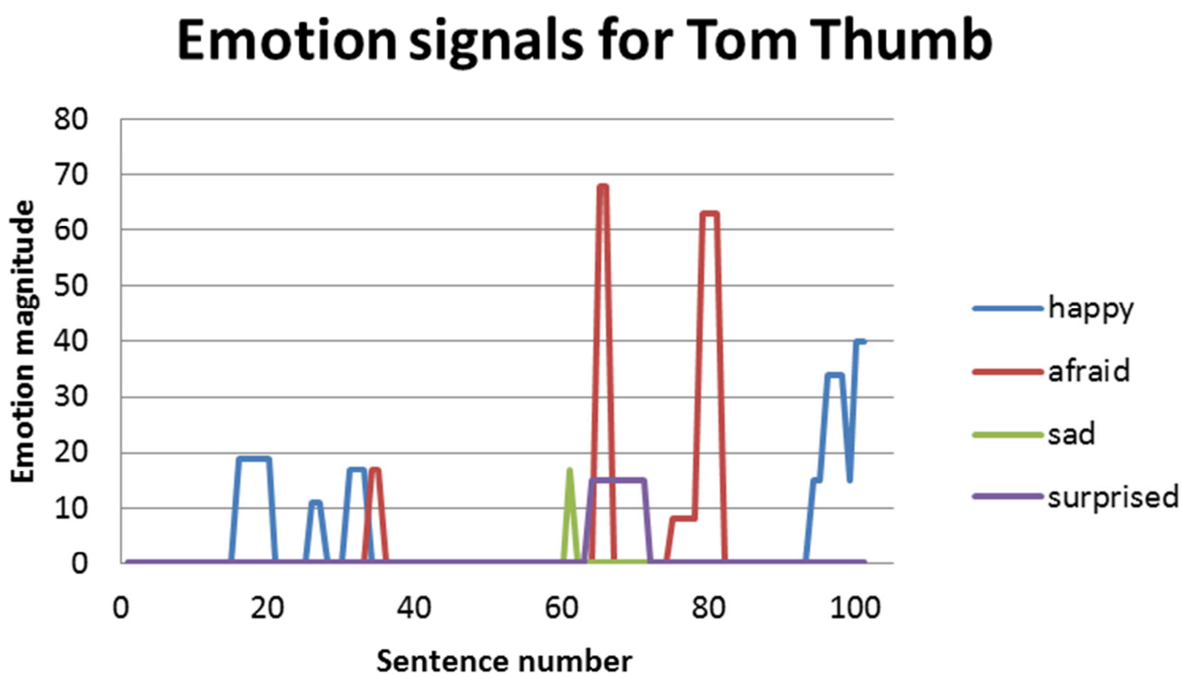


Figure 13: Emotion magnitudes for Thomas Thumb in "Tom Thumb" by the Brothers Grimm

- **Inter-annotator Agreement**

To evaluate the reliability of the annotation scheme, a subset consisting of 19 stories from the Affect Corpus 2.0 was double annotated to measure inter-annotator agreement. This subset was analyzed in two different ways. The first approach included all emotion magnitude assignments including neutral emotional states. Including all the data helped to evaluate emotion class assignment between the five classes and the neutral state.

The second approach limited the sample to annotations where the actor received at least one emotion magnitude assignment other than zero. This second approach was used to evaluate annotator agreement on emotion magnitude assignment.

The metrics used to evaluate inter-annotator agreement included average observed agreement, Pi, alpha, S, and Kappa (Artstein and Poesio 2008). In their simplest form, these metrics (e.g. Avg_Ao) serve to determine the percentage of samples that were equally annotated by two or more annotators. Other metrics like Kappa, consider expected chance agreement in the calculations. By knowing these rates, theoretical upper bounds on expected accuracy can be determined for automatic systems trained on annotated data (Bird et al. 2009).

Table 7: Inter-annotator metrics for emotion class assignment

	Happy	Angry	Sad	Surprised	Afraid
Avg_Ao	0.897	0.867	0.872	0.794	0.742
π	0.222	0.463	0.280	0.086	0.089
S	0.863	0.823	0.829	0.725	0.657
Kappa	0.223	0.464	0.289	0.128	0.129
Alpha	0.222	0.463	0.280	0.086	0.089

Since most main agreement metrics in the literature use categorical label assignments, the magnitude data presented here was categorized into four groups. Magnitudes between 0-25 were designated as low, 26-50 as medium low, 51-75 as medium high, and 76-100 as high. The inter-annotator metrics are shown in Table 7 and Table 8.

Table 8: Inter-annotator agreement for emotion magnitude

	Happy	Angry	Sad	Surprised	Afraid
Avg_Ao	0.586	0.412	0.555	0.551	0.475
π	0.090	0.164	0.318	0.126	0.139
S	0.448	0.216	0.407	0.402	0.300
Kappa	0.096	0.186	0.332	0.186	0.182
Alpha	0.091	0.166	0.321	0.128	0.141

5.2.2 LSU-MD Corpus

The LSU-MD corpus consists of utterances collected from 6 episodes of the popular television medical drama “Grey’s Anatomy”. Using conversations from this medium helps to train and test the model on a noisy environment where the speakers are performing acted tasks; in this case practicing medicine. The

conversations include medical dialogue and TV drama emotional content. This is especially useful for the task proposed in this paper of emotion detection. The LSU-MD corpus is available from LSU-NLP.

5.3 Conclusions

In this chapter, the annotation scheme for 2 different corpora was presented. These corpora are used to train and test the models for emotion detection and emotion magnitude prediction that are proposed in this dissertation.

Inter-annotator agreement metrics were calculated on a subset of the Affect Corpus 2.0 to determine the percentage of agreement between annotators. From these results, it can be seen that class assignment (Table 7) had higher inter-annotator agreement than emotion magnitude assignment (Table 8). This reflects the fact that emotion magnitude assignment is a subjective and difficult task because what is very happy to one person may just be average to another.

In conclusion, new annotations for a beloved set of children's stories are provided with the intention that future research will be able to focus on developing applications instead of resources. The corpus provides new annotations for emotion magnitudes per actor, actor detection, and actor presence detection. Inter-annotator agreement metrics compared annotation of emotion categories vs. annotation of emotion magnitudes.

CHAPTER 6: DETECTION OF AFFECTIVE STATES FROM TEXT AND SPEECH

The methodology developed in this chapter uses a step-wise approach to detect emotion from text and speech at the sentence level by first distinguishing between emotion and neutral classes, and then between positive, negative, and five emotion classes. The methodology is trained and tested on two affect corpora: (1) the children's stories corpus developed at the University of Illinois at Urbana-Champaign (UIUC) by Alm (2008) and (2) the new medical TV drama corpus developed at Louisiana State University (LSU). The UIUC affect corpus is narrative based, with one speaker per story, low noise, and longer more grammatical sentences. In contrast, the LSU medical drama (LSU-MD) corpus is dialogue based, multiple speakers, a large number of extraneous sounds, and shorter utterances.

Results of the methodology using various supervised learning techniques, feature sets, and corpora are presented and discussed. Using two corpora to train and test the model helped to identify differences and similarities in emotion detection between narrative and dialogue based communication. Finally, feature analysis using feature ranking techniques (section 2.7.3) was performed to determine which features had the highest contribution to emotion detection.

6.1 Methodology

The methodology developed in this chapter uses a step-wise approach to detect emotion in sentences by first filtering out neutral sentences, then distinguishing between positive, negative, and 5 emotion classes. Text and speech features are combined in feature vectors of 69 features (Table 9) each to train and test the emotion detection model.

The first step in the methodology compares neutral vs. emotion sentences. The second step compares positive and negative sentences, and the third step uses 5 emotion classes which are happy, sad, angry,

Table 9: Text and speech features for multimodal emotion classification

Number of features	Names	Type	Category	Heuristic rule for calculation
1	Contains NNP	Text	NNP binary (Syntactic)	If NNP in sentence: NNP = 1
5	Counts for happy, sad, angry, afraid, surprised	Text	Number of words per emotion class	If word from sentence in happy words list: counthap += 1 Note: same for other 4 emotion classes
1	Number of words in sentence	Text	Number of words in sentence	Number of words in sentence list
2	Subject actor/environment	Text	Metric for subjects	If VP not in BranchStack: IntSubj += 1
2	Object actor/environment	Text	Metric for objects	If NP not in BranchStack: IntObj += 1
1	Sentiment composition current sentence	Text	Sentiment composition	If word from sentence in happy or surprised words list: SentCompCurr += 1 Negative_list = [sad , angry, afraid] If word from sentence in Negative_list: SentCompCurr -= 1
5	Sentiment composition accumulated whole story for happy, sad, angry, surprised, and afraid	Text	Sentiment flow	SentCompAccumHappy = SentCompAccumHappy + counthap Note: same for other 4 emotion classes
1	Sentiment composition change previous current accumulated	Text	Sentiment flow (delta)	SentCompChanPrevCurrAccum = SentCompAccumCombined - SentCompAccumPrevSent
4	Sentiment composition accumulated whole story for POS, NEG, POS1, NEG1	Text	Sentiment flow	CompAccumPos = CompAccumPos + w * counthap CompAccumNeg = CompAccumNeg + w * (countsad + countang + countafr) Note: w parameter changed manually
1	Sentiment composition accumulated whole story for the previous sentence	Text	Sentiment flow	SentCompAccumPrevSent (s)= SentCompAccumCombined (s-1) Note: s stands for sentence
1	Sentiment composition accumulated whole story combined	Text	Sentiment flow	SentCompAccumCombined = SentCompAccumCombined + counthap – countsad – countang – countafr + countsup
5	Number of happy, sad, angry, afraid, and surprised phrases	Text	Emotion phrase count	If phrase from sentence in happy phrases list: PhraseCountHap += 1 Note: same for other 4 emotion classes
2	Intensify emotion, reduce emotion	Text	Intensifier or reducer features (e.g. very)	If word in sentence is in list of intensifiers: IntensifyEmt += 1 Note: same approach for reducer
5	Changes in happy, sad, angry, surprised, and afraid	Text	Change in sentiment flow (delta)	ChangeHap = counthap – PreviousHap PreviousHap = counthap Note: Same for 4 other classes
33	F0 max and avg., intensity max and avg., F1, F2, F3, F4, F5, mean and standard deviation for 12 MFCCs (MFCC1-MFCC12).	Speech		Using Praat scripts
69			**Total**	

surprised, and afraid to perform classification. The emotion classification is performed at the sentence level (UIUC corpus) or utterance level (LSU-MD corpus) rather than at the phrase or word level to take advantage of the higher amount of information available in longer communication segments. For the LSU-MD corpus, the additional step of filtering out sentences with 3 words or less is also taken. This step is helpful since most short utterances (common in dialogue based corpora) do not contribute much to semantic content and can be better handled by heuristic rules or speech only approaches. Context is used in the methodology by utilizing previous sentences' emotional intensities as features (see text features section below). All features used in this chapter are summarized in Table 9.

6.1.1 Speech Features

In this work, speech (acoustic) features were extracted at the sentence or utterance level (Busso et al., 2009) using Praat scripts (Boersma and Weenink, 2005). The features included pitch average and max, intensity average and max, formants (F1-F5), and mean and standard deviation for 12 MFCCs. After sampling the acoustic signal, the features were extracted as feature vectors containing 33 features per sentence or utterance (Table 9).

6.1.2 Text Features

A total of 36 text-based features (Table 9) were used for this analysis. These features included the counts of emotion words and phrases per class, counts of some intensifiers, syntactic features, sentiment composition features, and accumulated sentiment flow features.

Five lists of emotion words (see appendix A or LSU-NLP) were used to perform the counts per emotion class (i.e. happy, sad, angry, surprised, and afraid). These counts are used to create emotion signals which quantify the intensity of each emotion per sentence throughout the story or conversation. These lists were collected manually and automatically, and then expanded using WordNet and ConceptNet (Havasi et al. 2008). Parts of speech and syntactic annotations were also used to extract additional syntactic information such as intensifiers, subjects/objects, etc. (see Table 9). These syntactic annotations were performed using the Stanford parser and BART toolkit (Versley et al. 2008).

Sentiment composition features in this work follow a simpler approach than the one proposed in Moilanen and Pulman (2007). This simplified approach is used to avoid the time consuming task of hand coding many rules. Here, the 5 emotion lists and simple per class rules (see Table 9) are used as decision points to determine sentiment composition per class. If positive emotion words are found in the sentence, the sentiment composition feature is incremented by +1. If negative emotion words are found, the sentiment composition feature is decreased by -1. The feature is calculated for each sentence independently (with count starting at zero). Although there are no upper or lower boundaries for this metric, it is expected that within the same domain (i.e. children's stories) the sentiment composition should fall within a specific range (Table 9).

To capture story context and previous information, the accumulated sentiment flow throughout a story is also recorded. Accumulated sentiment flow features take into consideration the emotion levels of previous sentences in the story. Here, per class emotion levels refers to the accumulated sum of the individual emotion counts per sentence. Additional sentiment delta features are used to calculate the sentiment change between the current sentence and previous sentences (see sentiment flow category on Table 9 for additional detail on these features).

6.1.3 Classification Model

Classification methodologies employed include Support Vector Machines (Burges 1998) using LibSVM (Chang and Lin 2001) and Naïve Bayes, artificial neural networks, decision trees, random forests, and the k-nearest neighbor classifier (Witten and Frank 2005) using the Waikato Environment for Knowledge Analysis (WEKA). Naïve Bayes is a probabilistic classifier which usually performs worst with emotional data. Therefore, it helps to set the baseline for the classification task. Random forest classifiers consist of several decision trees and usually produce good results with emotion data. Artificial neural networks and Support Vector Machines (SVMs) are classifiers that can handle non-linearly separable data. In theory, this capability allows them to model data that may be more difficult to classify. The k-nearest neighbor

classifier is also a good technique with the added advantage that it does not require parameter tuning. Additional background on classification models is presented in section 2.8.

6.2 Analysis and Results

Results of the classification task at each step of the methodology are presented in this section. The experiments are conducted on 89 stories from the UIUC corpus and on 6 episodes from the LSU-MD corpus. Each corpus is randomly divided into 80% of the samples for training and 20% for testing.

An analysis of the methodology using speech features only from the UIUC corpus was performed first to determine the contribution of speech features to the classification problem. This analysis is followed by an analysis of the step-wise methodology using multimodal features from the UIUC corpus.

The final section shows the results of the generalization of the methodology on the LSU-MD corpus. Each section, with recall classification results, is followed by feature analysis using chi-square feature selection techniques (Witten and Frank 2005).

6.2.1 Speech Affect Detection

In this section, an analysis of the impact of speech features alone is performed. The data set uses 33 speech features with 2503 positive samples, 1177 negative samples, and 3049 neutral samples from the

Table 10: Speech-affect detection using multiple classifiers

Multiple Classifier Comparison – Training Set (5383) and Testing Set (1346)										
	SVM		**Naïve Bayes**		**Multilayer Perceptron**		**Random Forest**		**Nearest Neighbor**	
	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr
Positive (1)	**47.7%**	52.3%	37.9%	62.1%	54.4%	45.6%	**55.7%**	44.3%	49.8%	50.2%
Negative (2)	12.2%	87.8%	50.7%	49.3%	15.4%	84.6%	19%	81%	27.1%	72.9%
Neutral (3)	87.2%	12.8%	39.1%	60.9%	73%	27%	68.2%	31.8%	58%	42%
All	**59.6%**	40.4%	40.6%	59.4%	56.3%	43.7%	**55.3%**	44.7%	49.8%	50.2%

Legend: Corr. = Correct Incorr = Incorrect

UIUC corpus. The results of the classification are shown in Table 10. Classification results indicate that positive samples are easier to detect than negative samples. The speech analysis from Table 11indicated

that MFFCs have the highest contribution to emotion detection. The SVM classifier used an RBF kernel with cost of 2.5 and gamma 2.5.

Table 11: Speech feature analysis for 3 class problem

Rank	Chi	Feature	Rank	Chi	Feature
1	289	MFCC 9 mean	12	62	MFCC 6 mean
2	211	MFCC 3 mean	13	56	MFCC 9 std.
3	181	MFCC 12 mean	14	51	Intensity max
4	158	MFCC 11 mean	15	40	MFCC 8 std.
5	157	MFCC 4 mean	16	39	MFCC 12 std.
6	150	MFCC 2 mean	17	37	MFCC 4 std.
7	128	Intensity avg.	18	35	F0 avg.
8	128	MFCC 10 mean	19	33	MFCC 11 std.
9	114	MFCC 2 std.	20	31	MFCC 1 std.
10	106	MFCC 7 std.	21	31	MFCC 3 std.
11	88	MFCC 5 mean	22	21	MFCC 6 std.

6.2.2 Multimodal Affect Detection for 2 Classes (Emotion vs. Neutral)

In this section, classification results on the 2 class problem in the UIUC corpus are presented using both text and speech features. This is the first step in the methodology which helps to filter out neutral sentences. It is important to filter out neutral sentences because in general they account for half of all the samples in the corpus. The dataset contains 3,680 emotion samples and 3049 neutral samples. In machine learning, dealing with class imbalances in the dataset is a very important step that if not addressed

Table 12: Multimodal emotion classification between emotion and neutral classes

Multiple Classifier Comparison – Training Set (5383) and Testing Set (1346)												
	SVM		**Naïve Bayes**		**Multilayer Perceptron**		**Random Forest**		**Decision Trees**		**Nearest Neighbor**	
	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr
Emotion (1)	**71.2%**	28.8%	48.5%	51.5%	61.5%	38.5%	**77.7%**	22.3%	63.1%	36.9%	70.5%	29.5%
Neutral (2)	**71.6%**	28.4%	71.8%	28.2%	70.3%	29.7%	57.4%	42.6%	60.4%	39.6%	65.7%	34.3%
All	**71.4%**	28.6%	58.9%	41.1%	65.5%	34.5%	**68.6%**	31.4%	61.9%	38.1%	68.4%	31.6%

Legend: Corr. = Correct Incorr = Incorrect

properly can negatively affect the learning process of the models. The best results of the classification methodology (Table 12) were obtained using the Random Forest (68.6%) and SVM (71%) classifiers. The SVM classifier normalized the data and used an RBF kernel with cost of 4 and gamma 1. Feature

analysis (Table 13) indicates that sentiment composition text features have the highest contribution to classification accuracies but that speech features also play an important role.

Table 13: Feature analysis for emotion vs. neutral classes

Rank	Chi	Feature	Rank	Chi	Feature	Rank	Chi	Feature
1	220	Sentiment composition accumulated whole story combined (Text)	**17**	95	Angry phrases count (Text)	**33**	53	Happy words list (Text)
2	180	Sentiment composition change previous current accumulated (Text)	**18**	93	Afraid words list (Text)	**34**	51	Intensity max (Speech)
3	167	Sentiment composition accumulated whole story negative 1 (Text)	**19**	91	MFCC 11 mean (Speech)	**35**	48	NNP (Text)
4	158	Sentiment composition accumulated whole story negative (Text)	**20**	87	MFCC 5 mean (Speech)	**36**	44	MFCC 3 mean (Speech)
5	153	MFCC 2 mean (Speech)	**21**	87	Surprised phrases count (Text)	**37**	38	Afraid change (Text)
6	150	MFCC 12 mean (Speech)	**22**	87	Sad phrases count (Text)	**38**	38	MFCC 10 mean (Speech)
7	145	Sentiment composition accumulated whole story afraid (Text)	**23**	87	Happy phrases count (Text)	**39**	33	MFCC 7 std. (Speech)
8	144	Sentiment composition accumulated whole story sad (Text)	**24**	87	Afraid phrases count (Text)	**40**	30	Intensity avg. (Speech)
9	142	Number of words in sentence (Text)	**25**	80	Sad change (Text)	**41**	29	MFCC 1 std. (Speech)
10	139	Sentiment composition accum. whole story previous sentence (Text)	**26**	73	Sentiment composition accumulated whole story surprised (Text)	**42**	24	F0 max (Speech)
11	138	Sentiment composition accumulated whole story angry (Text)	**27**	71	Angry words list (Text)	**43**	24	Object environments (Text)
12	138	MFCC 9 mean (Speech)	**28**	71	MFCC 4 mean (Speech)	**44**	18	Surprised words list (Text)
13	104	Sad words list	**29**	61	MFCC 10 std. (Speech)	**45**	18	Happy change (Text)
14	103	Sentiment composition accumulated whole story positive 1 (Text)	**30**	61	Angry change (Text)	**46**	14	Surprised change (Text)
15	101	Sentiment composition accumulated whole story positive (Text)	**31**	60	MFCC 6 mean (Speech)	**47**	12	Subject actor (Text)
16	101	Sentiment composition accumulated whole story happy (Text)	**32**	53	MFCC 2 std. (Speech)	**48**	10	Sentiment composition current sentence NPs (Text)

6.2.3 Multimodal Affect Detection for 3 Classes (Positive, Negative, and Neutral)

In this section, classification results on the 3 class problem in the UIUC corpus are presented. The dataset includes: 2503 positive samples, 1177 negative samples, and 3049 neutral samples. The best results

Table 14: Multimodal emotion classification using 3 classes

Multiple Classifier Comparison – Training Set (5383) and Testing Set (1346)												
	SVM		**Naïve Bayes**		**Random Forest**		**Multilayer Perceptron**		**Decision Trees**		**Nearest Neighbor**	
	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr
Negative (1)	56.9%	43.1%	33.3%	66.7%	**61.5%**	38.5%	55.7%	44.3%	49%	51%	55.7%	44.3%
Positive (2)	31.7%	68.3%	45.7%	54.3%	28.5%	71.5%	30.3%	69.7%	34.4%	65.6%	44.8%	55.2%
Neutral (3)	**78.1%**	21.9%	65.5%	34.5%	68.2%	31.8%	67%	33%	58.5%	41.5%	67.8%	32.2%
All	**62.3%**	37.7%	49.8%	50.2%	59.1%	40.9%	56.6%	43.4%	50.9%	49.1%	59.4%	40.6%

Legend: Corr. = Correct Incorr = Incorrect

(Table 14) were obtained using the Random Forest (59%) classifier and SVM (62%) classifier with RBF kernel with cost of 4 and gamma equal to 1. These results also help to see the class imbalance between neutral, positive, and negative sentences since in the next section (after filtering out neutral) the results are better. The feature analysis (Table 15) indicated that text features have a higher contribution than speech features to classification accuracy. Here, accumulated sentiment composition features are the most useful for emotion detection. This analysis helps to illustrate that neutral classes, if not filtered out, decrease the performance of the classifier. Analysis after filtering out neutral samples is presented in the next section.

Table 15: Feature analysis for 3 class problem

Rank	Chi	Feature	Rank	Chi	Feature	Rank	Chi	Feature
1	427	Happy words list (Text)	**17**	189	Number of words in sentence (Text)	**33**	111	Sentiment composition current sentence NPs (Text)
2	399	Sentiment composition accumulated whole story combined (Text)	**18**	181	MFCC 12 mean (Speech)	**34**	106	MFCC 7 std. (Speech)
3	342	Sentiment composition accumulated whole story negative 1 (Text)	**19**	158	MFCC 11 mean (Speech)	**35**	94	Sad change (Text)
4	334	Sentiment composition accumulated whole story angry (Text)	**20**	157	MFCC 4 mean (Speech)	**36**	89	Angry change (Text)
5	315	Sentiment composition accumulated whole story negative (Text)	**21**	150	MFCC 2 mean (Speech)	**37**	88	MFCC 5 mean (Speech)
6	302	Sentiment composition accumulated whole story previous sentence (Text)	**22**	142	Afraid words list (Text)	**38**	62	MFCC 6 mean (Speech)
7	289	MFCC 9 mean (Speech)	**23**	142	Sad words list (Text)	**39**	60	Afraid change (Text)
8	279	Sentiment composition accumulated whole story sad (Text)	**24**	133	Angry words list (Text)	**40**	56	MFCC 9 std. (Speech)
9	276	Sentiment composition accumulated whole story afraid (Text)	**25**	128	Intensity avg. (Speech)	**41**	51	Intensity max (Speech)
10	254	Sentiment composition accumulated whole story positive 1 (Text)	**26**	128	MFCC 10 mean (Speech)	**42**	48	NNP (Text)
11	249	Sentiment composition accumulated whole story positive (Text)	**27**	127	Angry phrases count (Text)	**43**	40	MFCC 8 std. (Speech)
12	249	Sentiment composition accumulated whole story happy (Text)	**28**	122	Sad phrases count (Text)	**44**	39	MFCC 12 std. (Speech)
13	242	Happy change (Text)	**29**	114	Happy phrases count (Text)	**45**	38	Object environments (Text)
14	222	Sentiment composition accumulated whole story surprise (Text)	**30**	114	Surprised phrases count (Text)	**46**	37	MFCC 4 std. (Speech)
15	213	Sentiment composition change previous current accumulated (Text)	**31**	114	Afraid phrases count (Text)	**47**	35	F0 avg. (Speech)
16	211	MFCC 3 mean (Speech)	**32**	114	MFCC 2 std. (Speech)	**48**	33	MFCC 11 std. (Speech)

6.2.4 Multimodal Affect Detection for Positive vs. Negative

In this section, all neutral samples are filtered out and the classification model tries to separate between positive and negative classes. The UIUC subset includes 2503 positive samples and 1177 negative samples. All classification models performed better once neutral sentences were filtered out (Table 16). Both the nearest neighbor method (73.6%) and SVM (76.9%) achieved better accuracy scores. The SVM classifier normalized the data and used an RBF kernel with cost of 4 and gamma of 1. The feature

analysis (Table 17) indicated that both text and speech features are important in distinguishing between positive and negative sentences. The "positive and negative" and the "5 emotion class" feature vectors were mapped to a 3 dimensional space (3 Principal Components) for visualization purposes (see Figure 14 and Figure 15). This mapping, however, shows the complexity of the data and that it is not linearly separable in a low dimensional space.

Table 16: Multimodal emotion classification after filtering out the neutral class

Positive vs. negative: multiple classifier comparison – Training Set (2944) and Testing Set (736)												
	SVM		**Naïve Bayes**		**Random Forest**		**Multilayer Perceptron**		**Decision Trees**		**Nearest Neighbor**	
	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr
Negative (1)	**89.7%**	10.3%	71.7%	28.3%	**92.3%**	7.7%	76.1%	23.9%	78.1%	21.9%	79.4%	20.6%
Positive (2)	50.8%	49.2%	**57%**	43%	45%	55%	61.2%	38.8%	48.3%	51.7%	62.0%	38%
All	**76.9%**	23.1%	66.8%	33.2%	**76.8%**	23.2%	71.2%	28.8%	68.3%	31.7%	73.6%	26.4%
5 emotion classes: multiple classifier comparison – Training Set (2944) and Testing Set (736)												
	SVM		**Naïve Bayes**		**Random Forest**		**Multilayer Perceptron**		**Decision Trees**		**Nearest Neighbor**	
	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr	Corr.	Incorr
Angry - 1	**48.8%**	51.2%	14.6%	85.4%	**53.7%**	46.3%	35.8%	64.2%	43.1%	56.9%	52.8%	47.2%
Fear - 2	**66.5%**	33.5%	25.8%	74.2%	**57.7%**	42.3%	55.5%	44.5%	40.7%	59.3%	50.5%	49.5%
Happy - 3	**73.9%**	26.1%	43%	57%	**70.5%**	29.5%	58.9%	41.1%	47.3%	52.7%	65.7%	34.3%
Sad - 4	33.1%	66.9%	61.4%	38.6%	39.4%	60.6%	24.4%	75.6%	33.9%	66.1%	45.7%	54.3%
Surprised - 5	**37.1%**	62.9%	10.3%	89.7%	32%	68%	33%	67%	33%	67%	38.1%	61.9%
All	**56%**	44%	32.9%	67.1%	**54.1%**	45.9%	44.8%	55.2%	40.8%	59.2%	52.7%	47.3%

Legend: Corr. = Correct Incorr = Incorrect

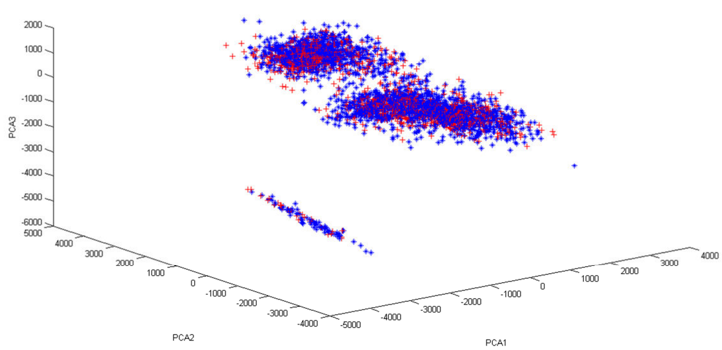


Figure 14: Positive and negative vectors mapped to 3 PCAs

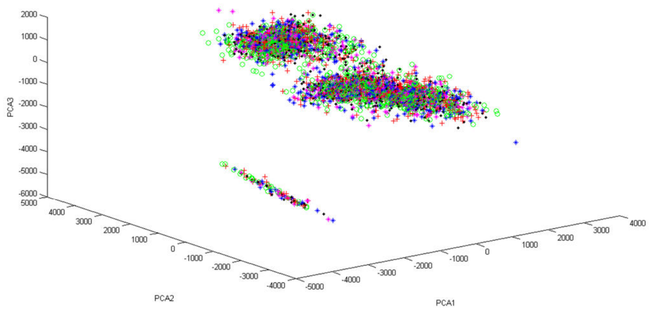


Figure 15: 5 class feature vectors mapped to 3 eigenvectors.

Table 17: Feature analysis for positive vs. negative

Rank	Chi	Feature	Rank	Chi	Feature	Rank	Chi	Feature
1	327	Happy words list (Text)	**17**	61	MFCC 10 mean (Speech)	**33**	30	MFCC 1 std. (Speech)
2	211	Happy change (Text)	**18**	56	MFCC 3 std. (Speech)	**34**	29	Number of words in sentence (Text)
3	140	MFCC 3 mean (Speech)	**19**	56	MFCC 7 std. (Speech)	**35**	28	MFCC 6 std. (Speech)
4	139	Sentiment composition accumulated whole story combined (Text)	**20**	53	Angry words list (Text)	**36**	27	Sad words list (Text)
5	138	Sentiment composition accumulated whole story previous sentence (Text)	**21**	53	MFCC 12 mean (Speech)	**37**	24	Sad phrases count (Text)
6	131	Sentiment composition accumulated whole story negative 1 (Text)	**22**	52	MFCC 9 std. (Speech)	**38**	23	Sentence composition change previous current accum. (Text)
7	127	Sentiment composition accumulated whole story surprise (Text)	**23**	44	Afraid words list (Text)	**39**	21	F0 avg. (Speech)
8	122	Sentiment composition accumulated whole story positive 1 (Text)	**24**	40	MFCC 12 std. (Speech)	**40**	20	Happy phrases count (Text)
9	117	Sentiment composition accumulated whole story positive (Text)	**25**	38	MFCC 8 std. (Speech)	**41**	20	Angry phrases count (Text)
10	117	Sentiment composition accumulated whole story happy (Text)	**26**	38	MFCC 4 std. (Speech)	**42**	20	Afraid phrases count (Text)
11	107	Sentiment composition accumulated whole story negative (Text)	**27**	36	MFCC 4 mean (Speech)	**43**	20	Surprised phrases count (Text)
12	102	Sentiment composition accumulated whole story afraid (Text)	**28**	35	MFCC 2 std. (Speech)	**44**	14	Object environments (Text)
13	91	MFCC 9 mean (Speech)	**29**	33	Angry change (Text)	**45**	14	MFCC 5 std. (Speech)
14	90	MFCC 11 mean (Speech)	**30**	32	Sentiment composition accumulated whole story sad (Text)	**46**	13	Intensify emotion (Text)
15	87	Sentiment composition accumulated whole story angry (Text)	**31**	31	MFCC 11 std. (Speech)	**47**	11	Object actor (Text)
16	73	Sentiment composition current sentence NPs (Text)	**32**	30	Afraid change (Text)	**48**	11	Sad change (Text)

6.2.5 Multimodal Affect Detection for 5 Emotion Classes

In this section, all emotion sentences are analyzed using 5 emotion classes (happy, sad, angry, surprised, and afraid). The UIUC subset includes 643 samples for angry, 907 for fear, 1013 for happy, 617 for sad, and 500 for surprised. To visualize misclassification errors, the confusion matrix for the 5 classes using the SVM classifier can be seen in Table 18.

Table 18: Confusion matrix for 5 classes

	Angry	Fear	Happy	Sad	Surprised
Angry	**60**	27	22	6	8
Fear	21	**121**	24	6	10
Happy	8	23	**153**	9	14
Sad	13	37	27	**42**	8
Surprised	16	16	22	7	**36**

Table 19: Feature analysis for 5 class problem

Rank	Chi	Feature	Rank	Chi	Feature	Rank	Chi	Feature
1	352	Happy words list (Text)	**17**	165	Sentiment composition accumulated whole story sad (Text)	**33**	57	MFCC 8 std. (Speech)
2	308	Sentiment composition accumulated whole story surprise (Text)	**18**	164	MFCC 2 mean (Speech)	**34**	56	Sad phrases count (Text)
3	296	Sentiment composition accumulated whole story angry (Text)	**19**	144	MFCC 9 mean (Speech)	**35**	55	MFCC 9 std. (Speech)
4	289	Sentiment composition accumulated whole story positive 1 (Text)	**20**	135	MFCC 10 mean (Speech)	**36**	51	Intensity avg. (Speech)
5	289	Sentiment composition accumulated whole story positive (Text)	**21**	118	MFCC 12 std. (Speech)	**37**	51	MFCC 1 mean (Speech)
6	289	Sentiment composition accumulated whole story happy (Text)	**22**	113	MFCC 4 std. (Speech)	**38**	50	F0 avg. (Speech)
7	288	Sentiment composition accumulated whole story afraid (Text)	**23**	108	MFCC 7 mean (Speech)	**39**	50	MFCC 8 mean (Speech)
8	284	MFCC 11 mean (Speech)	**24**	98	Sad words list (Text)	**40**	49	Angry change (Text)
9	277	Sentiment composition accumulated whole story combined (Text)	**25**	95	Angry words list (Text)	**41**	48	Surprised words list (Text)
10	269	Sentiment composition accumulated whole story negative (Text)	**26**	88	Sad change (Text)	**42**	48	MFCC 6 mean (Speech)
11	264	Sentiment composition accum. whole story previous sentence (Text)	**27**	83	MFCC 3 std. (Speech)	**43**	47	MFCC 11 std. (Speech)
12	263	Sentiment composition accumulated whole story negative 1 (Text)	**28**	79	Number of words in sentence (Text)	**44**	44	Afraid words list (Text)
13	231	MFCC 12 mean (Speech)	**29**	75	MFCC 7 std. (Speech)	**45**	42	Surprise change (Text)
14	226	Happy change (Text)	**30**	75	Sentiment composition current sentence NPs (Text)	**46**	42	Sentiment comp. change previous current accum. (Text)
15	218	MFCC 3 mean (Speech)	**31**	66	Angry phrases count (Text)	**47**	40	Intensify emotion (Text)
16	167	MFCC 1 std. (Speech)	**32**	64	MFCC 4 mean (Speech)	**48**	38	Surprised phrases count (Text)

From the misclassification matrix (Table 18), it can be seen that angry emotion are confused with fear the most, fear is confused the most with happy, happy is confused the most with fear, sad is confused the most with fear, and surprised is confused the most with happy. Therefore, for this analysis, fear and happy are the two classes that are hardest to identify. These two classes have the highest number of samples (Table 18) and as a result cause a class imbalance.

Results of the classification task using 5 emotion classes are presented in Table 16. Overall, the Random Forest and SVM classifiers performed best. The SVM classifier normalized the data and used an RBF kernel with cost of 4 and gamma of 1. Feature analysis is presented in Table 19. The feature analysis shows that text-based accumulated sentiment composition features have the highest contribution to emotion detection for the 5 class problem. Table 20 provides some examples of incorrectly classified samples from the UIUC corpus.

Table 20: Incorrectly classified samples from the UIUC corpus

Annotated class	Predicted class	Sentence
Surprised	Angry	Mr. Tod was mystified; he sat quite still, and listened attentively.
Happy	Angry	Yes--there was no doubt about it--it had turned out even better than he had planned; the pail had hit poor old Tommy Brock, and killed him dead!
Afraid	Happy	Oh; oh! they are coming back!
Afraid	Angry	What dreadful bad language!
Afraid	Sad	He lost one of his shoes among the cabbages, and the other shoe amongst the potatoes.
Surprised	Angry	Why on earth don't you run away? exclaimed the horrified Pigling.
Afraid	Happy	She was excited and half- frightened.
Angry	Happy	And Mr. McGregor was very angry too.
Sad	Angry	Timmy coughed and groaned, because his ribs hurted him.
Surprised	Happy	Why, I see a table spread with all kinds of good things, and robbers sitting round it making merry.
Angry	Happy	But the cat, not understanding this joke, sprang at his face, and spat, and scratched at him.
Afraid	Happy	Thereupon she took her leave of her father, and rode away with them, and rode to the court of her former betrothed, whom she loved so dearly.
Happy	Afraid	The woman who had hoped to find a good sale, gave him what he desired, but went away quite angry and grumbling.
Afraid	Angry	If we quarrel with him, and he strikes about him, seven of us will fall at every blow; not one of us can stand against him.
Afraid	Angry	But he did not venture to give him his dismissal, for he dreaded lest he should strike him and all his people dead, and place himself on the royal throne.
Afraid	Sad	It is rather dark, said he; "they forgot to build windows in this room to let the sun in; a candle would be no bad thing."
Sad	Angry	Then you may imagine how she wept over her poor children.
Sad	Afraid	She sought her children, but they were nowhere to be found.
Angry	Afraid	The miller thought to himself: "The wolf wants to deceive someone," and refused; but the wolf said: "If you will not do it, I will devour you."
Afraid	Sad	In his trouble and fear he went down into the courtyard and took thought how to help himself out of his trouble.

6.2.6 Application of the Methodology on a Medical Drama Corpus

In this section, the analysis of the methodology using 2,657 annotated samples (utterances) from a medical drama corpus (LSU-MD) is presented. The 2,657 samples consisted of 1,246 emotion samples and 1,411 neutral samples. Sentences with 3 words or less were filtered out for this analysis. Table 21 presents the recall accuracy results of the classification task between emotional and neutral samples. The SVM model which was trained using a polynomial kernel performed best overall (Recall: 61%).

Feature analysis for the classification between emotion and neutral classes in the LSU-MD corpus indicated that the speech features were slightly more useful than text features for emotion detection when compared to the UIUC corpus. This result is to be expected since each sentence or utterance in the medical drama corpus is shorter. As a result, there are fewer words for semantic analysis. The most useful text features were accumulated sentiment composition features for the emotion classes.

Table 22 presents the recall accuracy results of the classification task for positive and negative classes, and the 5 emotion classes. After filtering out neutral samples and utterances with 3 words or less, the dataset consisted of 1,246 samples. The number of samples per emotion class was as follows: 969 negative, 277 positive, 277 happy, 106 sad, 429 angry, 200 surprised, and 234 afraid. As can be seen, the data is highly imbalanced and there are fewer samples per class. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) proposed by Chawla et al. (2002) was applied to the data. Class imbalances can cause classification algorithms to underperform (Qiao and Liu 2009).

Table 21: Multimodal emotion classification between emotion and neutral classes (LSU-MD)

Multiple Classifier Comparison – Training Set (2126) and Testing Set (531)												
	SVM		**Naïve Bayes**		**Multilayer Perceptron**		**Random Forest**		**Decision Trees**		**Nearest Neighbor**	
	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.
Emotion (1)	**61%**	39%	53%	47%	58%	42%	**61%**	39%	49%	51%	57%	43%
Neutral (2)	55%	45%	66%	34%	56%	44%	49%	51%	58%	42%	59%	41%
All	**58%**	42%	60%	40%	57%	43%	**54%**	46%	54%	46%	58%	42%

Legend: Corr. = Correct Incorr. = Incorrect

The SMOTE technique oversamples minority classes of a dataset to even out the representation of the data. The number of samples per class after resampling was as follows: 969 negative, 554 positive, 426 happy, 424 sad, 429 angry, 420 surprised, and 444 afraid. As can be seen in Table 22, the results are much better once the number of samples per class is balanced.

The analysis of the methodology on this new noisy healthcare care related corpus seems promising. Therefore, this methodology may generalize to everyday speech. As expected, for the classification

between emotion classes, both text and speech features were important. Both SVM classification models for "positive vs. negative" (cost: 32; gamma: 0.1) and "five emotions" (cost: 100; gamma: 0.9) used an RBF kernel.

Table 22: Multimodal emotion classification after filtering out the neutral class (LSU-MD)

Positive vs. negative: multiple classifier comparison – Training Set (1218) and Testing Set (305)												
	SVM		Naïve Bayes		Random Forest		Multilayer Perceptron		Decision Trees		Nearest Neighbor	
	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.
Negative (1)	**83%**	17%	66%	34%	92%	8%	84%	16%	78%	22%	**71%**	29%
Positive (2)	**69%**	31%	65%	35%	60%	40%	63%	37%	59%	41%	**81%**	19%
All	**78%**	22%	65%	35%	80%	20%	76%	24%	71%	29%	**75%**	25%
5 emotion classes: multiple classifier comparison – Training Set (1714) and Testing Set (429)												
	SVM		Naïve Bayes		Random Forest		Multilayer Perceptron		Decision Trees		Nearest Neighbor	
	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.	Corr.	Incorr.
Angry - 1	**56%**	44%	34%	66%	49%	51%	32%	68%	31%	69%	33%	67%
Fear - 2	**73%**	27%	41%	59%	55%	45%	53%	47%	45%	55%	75%	25%
Happy - 3	**61%**	39%	29%	71%	38%	62%	39%	61%	40%	60%	69%	31%
Sad - 4	**84%**	16%	60%	40%	64%	36%	51%	49%	48%	52%	91%	9%
Surprised - 5	**71%**	29%	40%	60%	43%	57%	50%	50%	36%	64%	80%	20%
All	**69%**	31%	41%	59%	50%	50%	45%	55%	40%	60%	70%	30%

Legend: Corr. = Correct Incorr. = Incorrect

6.2.7 Comparison of Results between the UIUC and the LSU-MD Corpus

The lower performance of the methodology in the "emotion vs. neutral" problem using the LSU-MD corpus can be attributed to the following factors: fragmented sentences, shorter sentences, environment noise, dialogue format of the text, sarcasm, and smaller sample size for training and testing. Additionally, adults who read children's stories may be more expressive.

The healthcare data used in this study is dialogue based. Therefore, the samples are less likely to be fully formed sentences like in the children's stories corpus (UIUC corpus) and have fewer words per sentence. For example, the samples in the medical drama corpus included many utterances which are not always grammatical. In fact, the presence of many short statements like "yes", "good", or "ok" presented a challenge to the classification model which expected longer sentences. This reason influenced the

approach of filtering out samples with 3 words or less. Furthermore, the speech recordings included all kinds of background noise such as beepers, music, and other types of noise. This noise may have confused the detection system.

Another factor that may have influenced the performance is the flow of the conversations. Dialogue based conversations have a different flow than narrative based stories. In dialogue formats, the sentiment flow jumps from one emotion to another since utterances are coming from different speakers. Finally, in dialogue mediums, speakers do not have to describe the environment as in children's stories and they make substantial use of humor and sarcasm.

Although the results were less accurate for the "emotion vs. neutral" problem in the medical drama corpus, the model still performed moderately well (72% on UIUC corpus and 61% on LSU-MD corpus). Both text and speech features contributed to the detection. The analysis suggests that text features can be more discriminative on communication data with richer language (e.g. children's stories). On the other hand, for short utterances with less semantic content, acoustic features appear to be more discriminative.

6.3 Conclusions

Classification between emotion and neutral sentences achieved recall accuracies as high as 77% for emotion samples in a children's stories corpus and 61% in a medical drama corpus. Once neutral sentences were filtered out, the system achieved recall accuracy scores for detecting negative sentences as high as 92.3%. The methodology presented in this work using semantically rich and imbalanced corpora produced results which are consistent with results presented in Chuang et al. (2004), Alm (2008) and Busso et al. (2009). As can be seen in Table 23, the results obtained in this dissertation are comparable to those of recent studies on emotion detection. It is important to note that the results from Table 23 are performed on different corpora and at different levels of emotional granularity.

The results obtained in this dissertation are important because they represent a baseline for emotion detection studies using imbalanced data. In general, a system using this methodology, enough training

samples, and 100% balanced classes for training purposes can expect to achieve recall accuracies higher that the ones presented in this paper.

Table 23: Classification accuracies comparison to previous studies

Study	**Accuracy score**	**Classification type**
Busso et al. (2009)	77% (55%-96%)	Recall for Emotion vs. Neutral
Chuang et al. (2004)	62%-88%	Fine grained classification
Tokuhisa et al. (2008)	40%-80%	F-measure for fine grained classification
Alm (2008)	69%–78%	Emotion vs. Neutral (3 Authors)
Alm (2008)	78%	Emotion vs. Neutral (Grimms)
Alm (2008)	71%	Emotion vs. Neutral (Potter)
Dissertation methodology – Chapter 6	71%	Recall for Emotion vs. Neutral (3 authors)
Dissertation methodology – Chapter 6	82%	Recall for Emotion vs. Neutral (Grimms) with resampling using SMOTE (SVM – 90% split)
Dissertation methodology – Chapter 6	78%	Recall for Emotion vs. Neutral (Potter) with resampling using SMOTE (SVM – 90% split)

The feature analysis helps to corroborate recent findings by Luengo et al. (2010) that spectral features (MFCCs) are better than prosodic features (pitch) for emotion detection. The results also show promising results for text-based features that consider accumulated sentiment composition and sentiment flow in a story or conversation. The results of this study suggests that gold standard corpora can be used to develop and tune emotion recognition systems but that specific characteristics of the application domain must also be considered.

CHAPTER 7: ACTOR AND ENVIRONMENT DETECTION

In this chapter, a methodology for sentient entity or actor detection in text and speech for use in emotion detection is developed. The methodology for actor detection uses a recursive algorithm to extract NPs from sentence parses. The NPs are actor candidates from which actors will be extracted using a classification approach. The algorithm, when extracting an actor candidate, also extracts associated features for that actor candidate. The extracted features per actor candidate are used as feature vectors for classification purposes between actors and non-actors. Acoustic features are also added to the actor candidate feature vectors to help in the classification task. These acoustic features are extracted from the speech sentence segment that contains a given actor candidate.

The methodology is trained and tested on the Affect Corpus 2.0 (Calix and Knapp 2011) which includes text and speech versions of children's stories as well as actor annotations. All extracted actor candidates were annotated manually to perform the training and testing of the model. Classification results using various supervised learning techniques as well as an analysis of the contribution of speech and text features for actor detection using chi-square feature selection (section 2.7.3) are presented.

7.1 Methodology

In this work, actors are defined as sentient beings such as humans or fairy tale characters capable of human like behavior, in particular, emotions. For detection purposes, actors can be grouped into named entity and unnamed entity categories. Named entities have proper names identified in the story (i.e. Lucinda, Jane, Pickles). Unnamed entities do not have a proper name but are instead referenced typically through a noun phrase (i.e. the cat, the wizard, the big bad wolf).

Tools for named entity recognition (NER) have been addressed in the literature and there are relatively mature implementations available. Unnamed entity recognition methods are less common in the literature. The main approaches to actor detection use named entity recognition followed by disambiguation. Disambiguation refers to the process of finding senses for words. These senses can be obtained using

lexical dictionaries and heuristic rules. For actor detection, noun phrases in a sentence which appear before a verb phrase are good candidates for subjects in the story. Word meanings or knowledge bases such as ConceptNet can be used to obtain senses associated with a word.

Environmental elements consist of things, locations, places and living entities such as animals which are incapable of human like behavior (i.e. thinking). Nouns in noun phrases which are children of verb phrases can be considered as candidates for environmental elements. It is important to note that some environmental elements can be "named entities". Therefore, a filtering implementation is necessary to separate these elements in the text. For this work, filtering is performed by key word spotting only (e.g. checking to see if a word in the NP environment candidate appears in pre-populated lists. Sixteen environment types are used to map to the sixteen cube maps available for the rendering approach (see section 9.2).

7.1.1 Automatic Detection of Sentient Nominal Entities

The objective of the methodology developed in this chapter is to detect sentient actors from text and speech for emotion detection and prediction purposes. The proposed methodology is performed in four phases. First, the speech files are transcribed using an automatic speech recognition (ASR) system. In this dissertation, this step is simulated by using an already annotated corpus. Second, the actor candidates and text features are extracted from the stories (for training and testing purposes, all extracted actor candidates were manually annotated as actor or non-actors). Third, the speech features from the sentence where the actor candidate appears are extracted. Finally, in the fourth phase, the feature vectors are used to train and test the model to classify an actor candidate as actor or non-actor. All non-actor NPs are considered environment candidates. Further selection of these environment candidates is performed by comparison with pre-defined lists. The phases are described in more detail in the following sections.

- **Phase 1: Speech Signal Processing**

This phase focuses on the pre-processing of the files. The goal is to obtain the text and acoustic versions of each file (i.e. story). Acoustic features can be obtained directly from the signal after segmentation. Text

features can be obtained by automatically transcribing the audio signal into text using an Automatic Speech Recognition (ASR) system. See section 2.6.1 for more information on ASR systems.

In this work, the speech-to-text conversion is simulated by using the Affect Corpus 2.0 which has been annotated for this task (Calix and Knapp 2011; Alm 2008). The files in the Affect Corpus 2.0 were manually annotated and are used to extract text and speech features at the sentence level. Speech features are extracted using Praat scripts (Boersma and Weenink 2005). The corpus includes annotations of the actors in each story and their presence in a given sentence.

- **Phase 2: Actor Candidate and Text Feature Extraction**

The methodology developed in this phase for extracting actor candidates is divided into 2 main steps. The first step involves parsing the documents using state of the art taggers, parsers, and other architectures. The second step detects noun phrase candidates and extracts features for each one. The initial pre-processing step to produce the syntactic parses uses the Stanford parser (Klein and Manning 2002). Tokenization and part-of-speech (POS) tagging are also applied.

The syntactic parses produced by the Stanford parser are used for noun phrase detection (chunking). The following short sentence is used to illustrate the methodology for actor detection:

> "In a certain mill lived an old miller who had neither wife nor child, and three apprentices served under him."

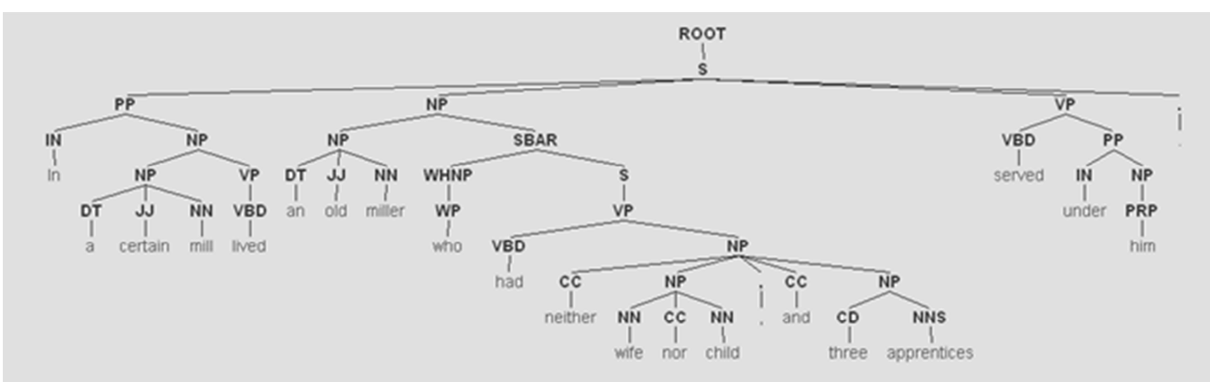


Figure 16: Syntactic parse for "The poor miller's boy and the cat" by the brothers Grimm

Using the previous sentence, an example of a syntactic parse can be seen in Figure 16 where a noun phrase is: [(NP (DT an) (JJ old) (NN miller))]. This Noun Phrase is considered an actor candidate in this work because it is a sub-tree of height 3 where the height is the distance between the head NP and the leaves of the tree (i.e. the words). A sub-tree of height 3 is a portion of the syntactic parse at the bottom of the tree where the acyclic graph ends for that section of the tree and the last nodes (i.e. the words) are children of POS tags that contain nouns and where these POS tags are children of an NP. The noun phrases "an old miller" and "a certain mill" on Figure 16 are both examples of sub-trees of height 3 and NP candidates where "an old miller" is an actor. Special tags like NNP can specifically help to identify named entities in this approach.

After pre-processing, feature extraction occurs by iterating through each sentence in a story and extracting NP candidates and their respective features. The features used in this work are divided into 4 categories which are: syntactic, knowledge-based, relation to pronouns, and general context based (Table 24). These features are collected as feature vectors per actor candidate.

Each story is loaded to memory and the algorithms iterate through it sentence by sentence in sequential order. The procedure is summarized in Algorithm 1 (Figure 17) where ST is the story, "s" is the current

Algorithm 1. Entity extraction algorithm

Input: Story
Output: NP candidates with text-based features

$\forall s \in ST$ **do**
    T ← s.CurrentSentenceParse
    BranchStack ← {}
    ExtractEntities (T, BranchStack)

Figure 17: Sentient nominal entity extraction algorithm

sentence, BranchStack is a stack used to capture the POS tags that make up the branch from the root to an NP candidate, ExtractEntities is a recursive DFS traverse function (see Figure 18 for DFS traverse algorithm) for each syntactic parse, T = (V, E) is the sentence syntactic parse which is a rooted acyclic

connected graph, $E = \{(x, y) \mid x, y \in V\}$ is the set of edges, and V are the vertices. ExtractEntities which is described in the next section is applied to each syntactic parse to find NP candidates and respective features.

- **Sentence Syntactic Parse Traversal and NP Candidate Text Feature Extraction**

A recursive DFS (Depth First Search) algorithm (Figure 18) was implemented to extract NP candidates and relevant features from each sentence. After a sentence is selected, the sentence parse T is extracted and is used by the recursive DFS function ExtractEntities. This function traverses each node V in the tree. The method selects as candidates all NPs of height 3 where the height is the distance between the head NP and the leaves of the tree. For each actor candidate, a branch traversal path (branch stack) is recorded.

BranchStack is a stack used to sequentially collect the tags that make up the path (or branch) from the root node to the leaves of the actor candidate. This stack allows for an actor candidate (NP tree) to be

Algorithm 2. Recursive DFS feature extraction approach for Sentient Nominal Entities

Input: T, BranchStack
Output: NP candidate with text-based features

```
Define Function ExtractEntities (T, BranchStack)
  If node ∈ T do
      If node = 'NP' and T.depth = 3 do
         MakeActorCandidateFeatures(T, BranchStack)
         oldtag ← nothing
         newtag ← nothing
         for child in t do
            if child has node do
               BranchStack.append(child.node)
               oldtag ← newtag
               newtag ← child.node
               EvaluateTagsSequenceSameLevel(oldtag, newtag)
               ExtractEntities (child,BranchStack)
         if BranchStack has value do
               BranchStack.pop()
```

Figure 18: Recursive DFS feature extraction approach for sentient nominal entities

identified as being a child of a PP node, NP node, VP node, and SINV node. Knowing the branch for each NP as a feature helps to identify the subject-object relation of the NP candidate. SINV, for instance, is

used for inverted declarative sentences (e.g. when the subject follows a tensed verb). A breadth first heuristic rule was also used to detect special words that indicate if an NP is a sentient being. For example, for the chunk “said the king”, the algorithm looks in the syntactic parse for sequences of tags and words

Table 24: Actor detection features

Number of features	Names	Type	Category	Heuristic rule for calculation
3	NNP, NNS, CD	Text	Syntactic Part of speech tags	If tag for word in sub-tree == “Tag variable”: Feature (Tag variable) = 1 Note: tag variable stands for NNP, NNS, CD
3	Object, organization, location	Text	Bart tag for Semantic objects	If BART Semantic class for actor == “Sem. variable”: Feature (Sem. variable) = 1 Note: Sem. variable stands for Object, location, org.
1	Said the actor	Text	Word feature	If “said” in sentence: Feature(said) = 1
10	Animal, person, pet, thing, royal, woman, man, good, bad, magical	Text	ConceptNet: IsA relationship	For word get concept IsA: If concept == “Concept Variable”: Feature(Concept Variable) = 1 Note: Concept variable is animal, person, etc.
5	Human capabilities, human parts, human desires, human properties, has fur	Text	ConceptNet: Has relationship or presence in handcrafted lists	If word in List_of_human_capabilities: Feature (Human Capabilities) = 1 For word get concept Has: If concept == “Concept Variable”: Feature(Concept Variable) = 1 Note: Concept variable is fur, human parts
5	Small, medium, large, huge, location	Text	ConceptNet: Location relationship Size	For word get concept At_Location If concept == “Concept Variable”: Feature(Concept Variable) = 1 Note: Concept variable is location, size
5	VP, PP, SBAR, ADJP, SINV,	Text	Branch stack does not contain	If tag in BranchStack: Feature (Tag Branch) = 1
9	SFP, PFP, SSP, PSP, STPM, STPF, STPN, STPG, PTP	Text	Pronouns (PRP) presence in sentence	If word in List (PRP): Feature Binary (PRP) = 1
9	SFP, PFP, SSP, PSP, STPM, STPF, STPN, STPG, PTP	Text	Pronouns (PRP) Word distance metric	Calculate positions for actor candidate and pronouns in sentence Distance = actor_pos – prp_pos Feature distance (PRP) = distance
33	F0 max and avg., intensity max and avg., F1, F2, F3, F4, F5, mean and standard deviation for 12 MFCCs (MFCC1-MFCC12).	Speech		Using Praat scripts
83			**Total**	

(e.g. double quote followed by the word “said”). The ExtractEntities algorithm uses the MakeActorCandidateFeatures function to perform individual analysis for each NP candidate tree and extract features. Here, syntactic, knowledge based, and other general features are extracted. Knowledge

features are extracted using a call to ConceptNet's database (Havasi et al. 2007). The database includes many assertions or concepts about different everyday types of people, things, abstract thoughts, etc. For each entity, ConceptNet can produce many types of relations to concepts such as HasA, IsA, CapableOf, Desires, HasProperty, etc. Each extracted concept is compared to pre-populated lists (see Figure 19) of characteristics to see if the word meets predefined heuristic rules. Special tags for semantic class annotations are also used to identify additional characteristics of the actors such as is an organization, location, or person.

```
listOfLocations = ['field','meadow','forest']
listOfSmallSizes = ['cat','football']
listOfMediumSizes = ['chair']
listOfLargeSizes = ['person','man']
listOfHugeSizes = ['car','elephant']
listOfThings = ['food','tree','rock','stuff']
listOfThings = ['royal','monarchy']
listOfThings = ['castle','palace']
listOfThings = ['royal','royalty','prince','queen','princes','king']
listOfmagical = ['magical','magic','wiccan','pagan','wizard']
listOfThings = ['magical','pagan','cast','spell']
listOfHumanParts = ['hand','eye','glasses','clothes','shoes','body']
listofmale = ['man','male','mr','boy','sir']
listoffemale = ['woman','lady','girl','female','miss','mrs']
listOfHumanActions = ['laugh','walk','feel','love','speak','talk','learn','play','live','think','read']
listOfHumanDesires = ['anger','feel','knowledge','understand','happiness','happy','sad','surprised','afraid','envy','laughter']
listOfHumanProperties = ['person','human','being','sentient','intelligent','social']
```

Figure 19: List of characteristics

Pronoun relation features used heuristics to test for several conditions related to the presence of pronoun types in the current sentence (i.e. singular pronoun, plural pronoun, etc.) and the distance from the actor candidate to that pronoun. The distance from the pronoun to the actor candidate is calculated using a word distance metric. The list of pronouns used can be seen in Figure 20.

```
SingularFirstPerson = ['I', 'me', 'myself', 'mine', 'my']
PluralFirstPerson = ['we','us','ourselves','ourself','ours','our']
SingularSecondPerson = ['thou', 'thee','yourself', 'thyself','thine','thy']
PluralSecondPerson = ['you','yourselves','your','yours']
SingularThirdPersonMale = ['he','him','himself','his',]
SingularThirdPersonFemale = ['she','her','herself','hers','her']
SingularThirdPersonNeuter = ['it','itself','its']
SingularThirdPersonGeneric = ['one','oneself']
PluralThirdPerson = ['they','them','themselves','their','theirs']
```

Figure 20: Pronouns list

- **Phase 3: Speech Feature Extraction**

Acoustic features are incorporated in the sentient actor detection process to capture additional information about the signal. Since emotions occur to sentient beings, it is conceivable to think that emotion related features could be correlated with the presence of actors in a section of an audio file. Many works such as Busso et al. (2009) and Luengo et al. (2010) argue that prosodic and spectral speech features such as pitch average, Mel Frequency Cepstral Coefficients (MFCCs), and formants can help to capture some aspects of emotion in speech. Therefore, this work incorporates these features to see if they contribute to the task sentient actor detection. The speech features used are those from the sentence segment where the actor appears. As such, each sentence where the NP candidate appears serves as a window of additional speech information. Speech features used in this work include pitch (F0), formants (F1-F5), and 12 MFCCs.

- **Phase 4: Classification**

The methodology uses a total of 83 features from the different types described in the previous sections for classification purposes. Classification methodologies were implemented in WEKA and LibSVM and included: Support Vector Machines (SVM) (Burges 1998; Chang and Lin 2001), Naïve Bayes, multilayer perceptron, random forests, and nearest neighbor classifier (Witten and Frank 2005). After classification, a feature analysis was performed to determine which features contributed most significantly to performance. Several feature selection techniques including chi square and information gain ranking were used to select the best features (Witten and Frank 2005).

7.2 Analysis and Results

In this section, the results of the training and testing methodology are presented. The sentient actor detection was performed on a dataset containing 4885 samples of which 2885 samples were labeled as non-actors and 2000 samples were labeled as actors. After performing feature extraction, the data was classified using several machine learning methods including Naïve Bayes, Support Vector Machines, Artificial Neural Networks, Random Forests, and Nearest Neighbor classifier. These different methods are used to determine if the results of adding speech features are consistent across learning methods.

The corpus was divided into two sub-sets; one subset for all entities and one subset where named

Table 25: Actor detection (for all actors) using text and speech features

Multiple Classifier Comparison (F-measure) – Training Set (3908) and Testing Set (977)					
	SVM	**Naïve Bayes**	**Multilayer Perceptron**	**Nearest Neighbor**	**Random Forests**
	Correct	Correct	Correct	Correct	Correct
Actor (Text)	79.1%	74.4%	77.8%	75.6%	77.6%
Actor (Text & Speech)	81.8%	73.9%	77.9%	80.1%	77.4%
Not Actor (Text)	87.4%	83.5%	86.9%	83%	86.3%
Not Actor (Text & Speech)	88.9%	83%	85.8%	87%	87.3%
All (Text)	84.2%	80%	83.4%	80.1%	82.9%
All (Text & Speech)	86.1%	79.5%	82.7%	84.3%	83.5%

entities have been filtered out. The analysis of the F-measure accuracy scores for the first sub-set can be seen in Table 25. The results indicate that SVM performed best overall (86.1%) and that adding speech

Table 26: Results of feature analysis for the entire corpus

Rank	Chi	Feature	Rank	Chi	Feature
1	1569	Is a person? (Text: Knowledge)	16	36	Small (Text: Knowledge)
2	758	NNP (Text)	17	34	Magical being (Text: Knowledge)
3	740	Branch stack has no VP (Text)	18	29	Is a pet? (Text: Knowledge)
4	503	Branch stack has no PP (Text)	19	28	Is a thing? (Text: Knowledge)
5	220	Is female (Text: Knowledge)	20	27	Said (Text)
6	199	SEM Org. (Text: Knowledge)	21	26	Human properties (Text: Knowledge)
7	138	Human capabilities? (Text: Know.)	22	26	Branch has no SBAR (Text)
8	128	Human desires? (Text: Know.)	23	23	MFCC 3 std. (Speech)
9	96	Is an animal? (Text: Knowledge)	24	20	Bad (Text: Knowledge)
10	79	Is a man? (Text: Knowledge)	25	20	SEM obj. (Text: Knowledge)
11	78	NNS (Text)	26	15	Tag CD (Text)
12	70	Royal? (Text: Knowledge)	27	15	Branch has no ADJP (Text)
13	47	Is good? (Text: Knowledge)	28	9	Has human parts? (Text: Knowledge)
14	45	MFCC 7 mean (Speech)	29	9	Large (Text: Knowledge)
15	39	Branch has no SINV (Text)			

features improves actor detection accuracy. The SVM classifier was trained using an RBF kernel with cost of 32 and gamma equal to 0.08.

Results of the feature analysis (Table 26) indicate that semantic features derived from knowledge based approaches have a high contribution to sentient actor detection. Of the speech features, MFCCs performed best overall.

To test the system's ability to detect un-named entities (Nominal Entity Recognition), all named entities were filtered out. Once named entities were removed, the dataset consisted of 4019 samples of which 2735 were labeled as non-actors and 1284 samples were labeled as actors. Table 27 presents the results of the classification using the F-measure metric. The results indicate that the SVM classifier performed best overall (86.7%) and that speech features help to improve classification accuracy. The SVM classifier used an RBF kernel with cost of 32 and gamma of 0.08.

Table 27: Actor detection for un-named entities

Multiple Classifier Comparison (F-measure) – Training Set (3215) and Testing Set (804)					
	SVM	**Naïve Bayes**	**Multilayer Perceptron**	**Nearest Neighbor**	**Random Forests**
	Correct	Correct	Correct	Correct	Correct
Actor (Text)	76.8%	66.9%	72.6%	68.8%	76.1%
Actor (Text & Speech)	**78.6%**	65.4%	75.9%	75%	71.2%
Not Actor (Text)	90.1%	85.3%	88.8%	85.8%	90.2%
Not Actor (Text & Speech)	90.4%	81.6%	88.7%	87.9%	89.1%
All (Text)	85.9%	79.5%	83.7%	80.5%	85.8%
All (Text & Speech)	86.7%	76.5%	84.7%	83.9%	83.5%

The results of the feature analysis (Table 28) indicate that a combination of knowledge-based and syntactic features are useful for actor detection of un-named entities. One important observation from the feature analysis is that the "Is a person" feature was consistently the highest ranked feature for all sets.

This is important because it indicates that more knowledge can improve accuracy scores. For un-named entity detection, it can be seen that speech features had a higher contribution to classifcation accuracy

Table 28: Results of feature analysis for un-named entities

Rank		Feature	Rank		Feature
1	864	Is a person? (Text: Knowledge)	16	59	MFCC 1 std. (Speech)
2	504	Branch has no VP (Text)	17	56	Branch has no SINV (Text)
3	394	Human capabilities (Text: Know.)	18	56	MFCC 8 std. (Speech)
4	379	Female (Text: Knowledge)	19	45	Said word (Text)
5	357	Branch has no PP (Text)	20	45	MFCC 12 mean (Speech)
6	287	Is man (Text: Knowledge)	21	44	MFCC 6 std. (Speech)
7	220	Human desires (Text: Knowledge)	22	36	MFCC 4 mean (Speech)
8	189	Human properties (Text: Know.)	23	35	MFCC 9 std. (Speech)
9	165	Is animal (Text: Knowledge)	24	35	MFCC 10 std. (Speech)
10	162	Royal (Text: Knowledge)	25	33	MFCC 11 mean (Speech)
11	91	Magical (Text: Knowledge)	26	33	Is good (Text: Knowledge)
12	75	MFCC 12 std. (Speech)	27	31	SEM org. (Text: Knowledge)
13	73	MFCC 7 mean (Speech)	28	30	MFCC 7 std. (Speech)
14	61	MFCC 4 std. (Speech)	29	29	MFCC 3 std. (Speech)
15	60	MFCC 5 mean (Speech)	30	28	Is a pet (Text: Knowledge)

when compared to the previous feature analysis presented in Table 26. From Table 28, it seems that most of the Mel Frequency Cepstral Coefficients (MFCCs) used in this work had some influence in the detection model. Finally, these results are consistent with recent results by Pang and Fan (2009a) and Pang and Fan (2009b) for Nominal Entity Recognition.

- **Content Enrichment**

In this section, content enrichment of the story based on automatic content detection is discussed. Once sentient actors are detected in a story, the methodology automatically enriches the content by creating a new XML mark-up file with new metadata about the detected actors. The detected actors and corresponding features are added to the file. A scheme similar to the one used in Mallepudi et al. (2010) is implemented in this work. Figure 21 presents an example of the XML mark-up. This mark-up approach can be used for automatic text-to-scene processing.

```xml
<?xml version="1.0" encoding="ANSI_X3.4-1968"?>
<audio format="MP3">
   <semantic>
      <Rule_inference>
         <type genre="fairy_tales"/>
         <type genre="children"/>
         ...
      </Rule_inference>
      <actors_List>
         <actor name=”Tom” orderid="1" sentenceID="1">
            <characteristics magical="1" is_person="1" >
            <sentence_position w="5"/>
         </actor>
         ...
      </actors_List>
      ...
   </semantic>
</audio>
```

Figure 21: XML format enrichment for story in MP3 format

7.3 Conclusion

In this chapter, a methodology for actor and environment detection for use in automatic text-to-scene processing was developed. An algorithmic and supervised learning approach using text and speech features was used to detect sentient actors. As can be seen from the results, the methodology obtained very good results. Additionally, the results of the feature analysis on un-named entity detection indicate that speech features have an important contribution to the detection of these types of entities. Finally, future semantic analysis work will require incorporating elements related to anaphora resolution to obtain a higher understanding of the semantics of actor interactions in the story. Anaphora resolution is one of NLP’s biggest challenges currently. For fully automated understanding from speech, anaphora resolution is required. However, this aspect is outside of the scope of this study as it is very broad and would require its own PhD dissertation. Instead, several assumptions are made to keep within the scope of this dissertation on emotion understanding. First, since this study focuses on affective dialogue systems, it is assumed that the system can use other mediums besides text to track actors whenever they interact with the system. Mediums that can be used include speech diarization algorithms, voice recognition, and biometrics such as RFID tags. Therefore, with the proposed actor detection algorithms and these mediums it is assumed that the system can detect actors and track them.

[84,Actor 1,0,0,0,0,0,0,Jemima became much alarmed.]
[84,Actor 2,0,0,0,42,66,1,Jemima became much alarmed.]
[84,Actor 3,0,0,0,0,0,0,Jemima became much alarmed.]
[84,Actor 4,0,0,0,0,0,1,Jemima became much alarmed.]

Figure 22: Actor-emotion vectors

Second, after detecting actors and their utterance, the system still needs to learn how to estimate evolving emotional state of actors. To perform this analysis, a simulation is performed (Chapter 8) using an annotated corpus. In this case, the Affect Corpus 2.0 (see chapter 5) has been annotated with evolving emotional states, actors, and all anaphora and other references through the text have been resolved by human annotators. The corpus annotation can be seen in Figure 22 where each line represents an actor emotion vector with emotion magnitudes and classes, sentence and sentence number. This way actors and their evolving emotional state through the text can be tracked and used for emotion magnitude prediction. The model for emotion magnitude prediction is presented in Chapter 8.

CHAPTER 8: ACTOR LEVEL EMOTION MAGNITUDE PREDICTION IN TEXT AND SPEECH

In this chapter, a recurrent (prior-state) feedback methodology is developed to predict emotion magnitudes at the actor level in a corpus of children's stories. Changes in emotional states of actors are assumed to be caused by triggers – the occurrence of actions, words, or new environmental elements that cause emotion changes. The magnitude of the change depends on the specific trigger and its context.

8.1 Methodology

8.1.1 Prior State Feedback Regression Model for Emotion Magnitude Prediction

To predict emotion magnitudes, emotion triggers are mapped to predicted output from a training set to build a regression model. To develop this model, three regression approaches are studied. These approaches include linear regression, multilayer perceptron, and support vector regression.

Models such as linear regression are common for prediction methodologies. However, they suffer from weaknesses related to normality assumptions and linearity assumptions. To address these issues, Artificial Neural Networks (ANNs) have been explored in the past. ANNs can provide good results but can suffer from data over fitting problems. Support Vector Regression (SVR) is more robust to violations of the assumptions related to data normality, over fitting, and non-linearity of the data and may therefore be more suitable for this type of analysis (Wu et al. 2007; Prakasvudhisarn et al. 2003; Bi et al. 2011).

According to Smola and Scholkopf (2004), the SVR methodology tries to obtain a regression function that is as flat as possible (by minimizing the norm of the weight vector) and with low prediction error. The difference between the predicted values and the actual values from the training set (predicted error) should be no more than a certain specified value (E). The "flatness" property of the regression function is achieved by minimizing the norm of the weight vector (w). To deal with non-linear data, a kernel trick is

used to map non-linear data to higher dimensional linear space. Formally, the soft margin optimization problem can be formulated as follows:

Minimize $$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad \textbf{(Eq. 22)}$$

Subject to:

$$Y_i - w\cdot\phi(X_i) - b \le E + \xi_i \qquad \textbf{(Eq. 23)}$$

$$w\cdot\phi(X_i) + b - Y_i \le E + \xi_i^*$$

$$\xi_i, \xi_i^* \ge 0 \qquad \text{for i=1…n}$$

where C is the cost (tradeoff between w and prediction error), “i” is the sample, w are the weights, $\phi(\cdot)$ is the high dimensional feature space function, ξ_i and ξ_i^* are the slack variables that allow for errors to occur during training, with E the tolerated error level. This optimization problem with constraints to find the weight vector (w) and the bias (b) can be solved using Lagrange multipliers. The formulation is quadratic and can provide a single minimum solution using quadratic programming optimization. The fast LibSVM implementation (Chang and Lin 2001) in conjunction with WEKA was used to train and test the SVR model.

- **Steps**

The methodology for predicting the evolving emotional state of actors consisted of three main phases:

- Pre-processing of the input files.
- Extracting text and speech elements which include actors, environmental text elements, emotion triggers (i.e. emotion words, actions, environments), emotion signal features, and speech features in the training corpus.
- Training and testing of a regression model to predict emotion magnitudes per actor.

Changes in emotional states of actors are assumed to be caused by the triggers. The magnitude of the change depends on the current emotion triggers and previous emotional state magnitude. For the purposes

of this work, actors are defined as sentient entities (people and creatures) which demonstrate sentient behavior such as speaking, feeling, and thinking within the context of a conversation or story. In general, such entities can be found in noun phrases (see chapter 7).

The Affect Corpus 2.0 (Alm 2008; Calix and Knapp 2011) has been annotated with actors and their corresponding emotion magnitudes throughout the text to perform the training and testing of the prediction model. A total of 89 texts are used for this analysis.

- **Emotion Triggers for Magnitude Prediction**

A regression approach to predict actor level emotion magnitudes requires combining emotion trigger features in a linear or non-linear model. The following 5 sets of triggers are used as regression factors:

(1) A set of Emotion Tokens (emo) which consists of emotion strong nouns, adjectives, and adverbs (for example, happy, happiness, etc.). These emotion words were collected manually and automatically, and then expanded using WordNet (WordNet 2010) and ConceptNet. These words were manually classified into 5 emotion subsets (Appendix A). A total of 1214 stemmed emotion words were used as Emotion Token features.

(2) A set of Environment tokens (env) which consists of 825 stemmed environment words (Appendix A). These words consist of mainly nouns of places and other settings were an actor might be located. The list of environments was manually collected using noun phrases from the corpus and external sources from the internet. The initial words list was extended using ConceptNet's at location relation property. Examples of environment tokens include road, forest, farm, land, storm, etc.

(3) A set of Action Tokens (act_s, act_o) which consist of 1007 stemmed verbs (Appendix A). For each actor in a sentence, the sentence dependency parse was used to identify if the actor was the subject or object of the verb. Action token examples include words such as kill, attack, stuck, etc.

(4) A set of 36 text based emotion signal features (EmSin). These features consist of counts of the number of emotion words that appear in a sentence per emotion class or polarity. These counts

can be traced through a story to see how emotion intensity varies between positive, negative and between emotion classes. These features use sentiment composition principles to measure the changes and distribution of emotion in sentences. A subset of the 36 features is used to accumulate the emotion intensity from sentence to sentence through a story and to compare the difference in sentiment flow (e.g. the difference in the counts between the current and previous sentence). See Table 9 for further detail.

(5) A set of 33 sentence level speech features (Speech) which include max and average F0, max and average intensity, F1, F2, F3, F4, F5, and the mean and standard deviation for 12 Mel Frequency Cepstral Coefficients (MFCCs).

The text features are extracted using python scripts, the Stanford parser, and NLTK (Bird et al. 2009). All word tokens were stemmed using the Porter Stemmer to reduce dimensionality of the token sets. Speech features are extracted using Praat scripts.

- **Prediction Model**

The dependent variables for the model are the five emotion magnitudes for happy, sad, angry, surprised, and afraid. Each of the 5 emotion magnitudes is allowed to range from 0 (neutral) to 1 (maximum). Once the model is trained, the emotion magnitude for each actor can be calculated as the previous sentence's emotion magnitude plus a weighted sum of the trigger tokens present in the current sentence and associated with the actor. Formally, the model is presented in Eq. 24.

$$E_{d[s],a} = E_{d[s-1],a} + A_{d[s],a} * Q \qquad \textbf{(Eq. 24)}$$

$$Q = [P_{d[s]}^{emo} \cdot \alpha + P_{d[s]}^{env} \cdot \beta + P_{d[s],a}^{act\_s} \cdot \gamma + P_{d[s],a}^{act\_o} \cdot \theta + P_{d[s]}^{sp} \cdot \upsilon + P_{d[s]}^{EmSin} \cdot \psi] \qquad \textbf{(Eq. 25)}$$

where:

$E_{d[s],\ a}$ is a row vector of emotion magnitudes (1 column for each emotion) for actor “a” in sentence s of document d (d[s-1] indicates previous sentence in same document).

$A_{d[s],\ a}$ is a scalar value =1 if actor "a" is physically present in sentence s of document d, and 0 otherwise.

α is a matrix of emotion token weights (5 columns, one for each emotion, by N_{emo} emotion rows where N_{emo} is the total number of emotion tokens in the corpus).

β is a matrix of environmental token weights (5 columns by N_{env} rows).

γ is a matrix of subject action token weights (5 columns by N_{act_s} rows).

θ is a matrix of object action token weights (5 columns by N_{act_o} rows).

υ is a matrix of the speech feature weights.

ψ is a matrix of the emotion signal (EmSin) weights.

$P_{d[s]}^{emo}$ and $P_{d[s]}^{env}$ are 0/1 row vectors whose elements indicate whether the corresponding token (emotion and environment, respectively) is present (1) or not (0) in sentence s of document d.

$P_{d[s],a}^{act\_s}$ and $P_{d[s],a}^{act\_o}$ are 0/1 row vectors whose elements indicate whether the corresponding action token (with the actor "a" as subject or object, respectively) is present or not in sentence s of document d.

$P_{d[s]}^{Sp}$ and $P_{d[s]}^{EmSin}$ are vectors for speech features and text emotion signal measurements (EmSin), respectively.

For the first sentence in a document, $E_{d[s-1]} = 0$, the zero vector (vector of zero values), corresponding to a neutral emotion state is used. The weights correspond to emotion magnitude changes (deltas) to be applied for each emotion if the trigger is present in a sentence.

In this work, a regression / optimization approach is used to learn the weights to assign to each trigger/emotion pairing. Training is performed using linear and non-linear regression (e.g. SVR) approaches. The following section details the methodology for learning these weights.

- **Linear Approach: Linear Regression**

The linear regression is performed over all stories in the corpus with the goal of minimizing the sum of squared error terms between the corpus emotion annotation and the calculated emotions across all actors in all sentences of all corpus stories. The weights are then applied during testing / application to calculate the emotional magnitude of each actor at a given sentence. A linear regression function for the data is formulated formally with an objective function that minimizes the sum of the square errors subject to constraints as follows:

Decision Variables: Token weight matrices α, β, γ, θ, ψ, υ

Objective Function: Min $z = \Sigma_{d[s],a \in c} \varepsilon^2_{d[s],a}$

where $\varepsilon_{d[s],a}$ = error term for actor a, sentence s in document d; and c = corpus

subject to constraints:

$$E_{d[s],a} = E_{d[s-1],a} + A_{d[s],a} * Q \quad \forall \text{ d[s],a} \in \text{c} \qquad \textbf{(Eq. 26)}$$

$$\varepsilon_{d[s],a} = E_{d[s],a} - E^{'}_{d[s],a} \quad \forall \text{ d[s],a} \in \text{c} \qquad \textbf{(Eq. 27)}$$

where $E^{'}_{d[s],a}$ is the annotated emotion level from corpus c for actor "a" in sentence s of document d. Additionally, each weight element of the matrices α, β, γ, θ, υ, ψ is constrained to be between -1 and +1.

- **Non-Linear Approach: Support Vector Regression (SVR)**

The optimization model formulated with equations (22) and (23) produces the weight vector (w) and bias (b) and the regression equation becomes:

$$g(x)_i = \sum_{i=1}^{nv} (\lambda_i - \lambda_i^*) K(x_i, x) + b \qquad \textbf{(Eq. 28)}$$

where K(x_i, x) = $\phi^T(x_i)\cdot\phi(x)$ is the kernel function that maps the input vector to higher dimensional space, λ_i and λ_i^* are the LaGrange multipliers, and nv is the number of support vectors.

8.2 Analysis and Results

The methodology is trained to predict emotion magnitudes per actor at any point in a story using previous emotion magnitudes plus current text and speech features which act on the actor's emotional state. Linear and non-linear regression techniques were used to find the optimal model that can fit the data.

To determine the contribution of semantic information in emotion prediction, the model is trained and analyzed using 2 different approaches. The first approach (Table 29) uses all text and speech features described in the methodology section. The second approach (Table 30) uses only speech features to train the prediction model. Therefore, a total of 3134 features are used for the "all features" approach and 33 features are used for the "speech only" approach. Both approaches include 1 additional feature for the actors' previous emotion magnitude in the story as indicated in the model. Training and testing are performed on each subset using an 80%/20% split. The prediction model is trained and tested using linear regression, multilayer perceptron (MLP), and Support Vector Regression (SVR) techniques.

The SVR is trained using the linear, polynomial, and RBF (Radial Basis Function) kernels. Results of the model accuracy are evaluated and compared using Root Mean Square Error (RMSE) and the correlation coefficient. In total, five prediction equations are trained and tested (one for each of the 5 emotion classes: happy, sad, angry, afraid, and surprised).

Of the three models used, the SVR with an RBF kernel performed the best for both accuracy and speed. This suggests that the data is not linear and that a non-linear approach is required. The model with the worst performance was the multilayer perceptron because of the time required to train the model and high RMSE. The root mean squared errors (RMSE) for the MLP-PCA or MLP-SO (speech only) fell between 21.71 and 36.61. The SVR-RBF model performed the best overall. The SVR correlation coefficients using an RBF kernel fell between a range of 0.69 and 0.80 (Table 29).

Table 29: **Regression modeling results (all features & previous magnitude)**

Regression Modeling – Speech and Text (Train: 80%; Test: 20%)									
	SVR (Linear)		SVR (Polynomial)		SVR (RBF)		Linear Regression		Total Samples
	Corr. Coef.	RMSE	Corr. Coef.	RMSE	Corr. Coef.	RMSE	Corr. Coef.	RMSE	
Happy	0.66	17.74	0.43	23.83	0.76	15.29	0.61	20.13	2640
					g=0.15	c=120			
Sad	0.73	18.33	0.65	27.22	0.77	17.32	0.59	25.10	1590
					g=0.07	c=20			
Angry	0.78	15.79	0.55	25.65	0.80	15.44	0.00	25.72	1192
					g=0.05	c=60			
Surp.	0.64	16.91	0.34	22.09	0.69	16.05	0.52	20.366	3551
					g=0.1	c=90			
Afraid	0.73	18.04	0.60	26.71	0.79	16.34	0.66	21.07	2693
					g=0.1	c=200			

Legend: Corr. Coef. = Correlation Coefficient, g:gamma, c: cost RMSE: Root Mean Squared Error

This result indicates that the model is able to get a good correlation between the features and the predicted magnitudes. The RMSE using the SVR with RBF kernel fell between 15.29 and 17.32. These error results are much lower than the ones obtained with the MLP and the best overall. The linear regression model performed between the SVR and the MLP-PCA.

Feature analysis indicated that the feature with the highest contribution to emotion prediction accuracy was the actors' previous emotion magnitude. This is an important result which suggests that overtime the accuracy of the model can improve and that previous information about the emotional state of the actor is very important. From the comparison between the "all features" approach (Table 29) and the "speech only" approach (Table 30) it can be seen that a multimodal approach using both text and speech features helps to improve prediction accuracy. This result shows the importance of using higher level semantic approaches.

To address the issue of long processing time caused by the high dimensionality of the feature set, the data dimensionality was reduced using Principal Component Analysis (PCA). PCA dimensionality

reduction was performed and evaluated on all 3 models (linear regression, multilayer perceptron, and SVR). The prediction results after PCA for all approaches were worse than the results using all features

Table 30: Regression modeling results (speech features only & previous magnitude)

Regression Modeling – Speech (Train: 80%; Test: 20%)									
	SVR (Linear)		**SVR (Polynomial)**		**SVR (RBF)**		**Linear Regression**		**Total Samples**
	Corr. Coef.	**RMSE**	**Corr. Coef.**	**RMSE**	**Corr. Coef.**	**RMSE**	**Corr. Coef.**	**RMSE**	
Happy	0.46	21.34	0.43	23.69	0.66	19.06	0.48	20.50	2640
					g=0.3	c=12000			
Sad	0.63	21.98	0.62	21.37	0.69	20.11	0.63	21.25	1590
					g=0.07	c=10000			
Angry	0.72	18.4	0.74	17.45	0.78	16.40	0.73	17.42	1192
					g=0.1	c=11000			
Surp.	0.61	18.06	0.65	16.92	0.68	16.34	0.61	17.00	3551
					g=0.1	c=6000			
Afraid	0.67	19.89	0.65	26.37	0.76	17.51	0.68	19.23	2693
					g=0.8	c=240			

Legend: Corr. Coef. = Correlation Coefficient, g:gamma, c: cost RMSE: Root Mean Squared Error

without dimensionality reduction. The dimensionality reduction for each of the 5 emotion classes was as follows: happy (628), sad (531), angry (473), surprised (797), and afraid (670).

Finally, the results of the analysis using all features except previous magnitudes are presented in Table 31. These results help to compare the impact of using a recurrent (prior-state feedback) predictive model vs. not using it. This analysis was performed using the SVR-RBF model. The correlation coefficients of the model without using the previous magnitude as a feature are 0.20 lower than for the models that use previous magnitude as a feature. This shows that the developed methodology can be very useful for this type of applications. Feature ranking in this subset suggest that text based emotion signal features and acoustic features are the most important for emotion prediction.

Table 31: Regression modeling results (All features except for previous magnitude)

Regression Modeling – Speech and Text (Train: 80%; Test: 20%)		
SVR (RBF)		
	Corr. Coef.	RMSE
Happy	0.57	19.86
Sad	0.55	22.81
Angry	0.517	22.2
Surprised	0.32	21.43
Afraid	0.53	22.89

8.3 Conclusions

In this chapter, a model was developed and tested to predict emotion magnitudes for actors in a story. Text and speech features were used and compared. Overall, the model was able to learn and achieved good prediction results. Of the 3 machine learning algorithms used to train the model, the support vector regression technique with an RBF kernel performed the best overall. The SVR correlation coefficients using an RBF kernel fell between a range of 0.69 and 0.80. This result indicates that the model is able to get a good correlation between the features and the predicted magnitudes. The best RMSE results fell in the range of 15.29 to 17.32 and were obtained using the SVR with RBF kernel. These error results are much lower than the ones obtained with the MLP and the best overall.

Feature analysis indicated that the feature with the highest contribution to emotion prediction accuracy was the actors' previous emotion magnitude. This is an important result which suggests that overtime the accuracy of the model can improve and that previous information about the emotional state of the actor is very important. From the comparison between the "all features" approach and the "speech only" approach

it can be seen that a multimodal approach using both text and speech features helps to improve prediction accuracy. This result shows the importance of using higher level semantic approaches.

Table 32: Regression Modeling Comparison to other studies

Study	Method	RMSE	Correlation Coefficient
Dissertation developed model (Chapter 8)	Text and speech features (prosody and spectral)	0.15 – 0.17	0.69 – 0.80
Dissertation developed model (Chapter 8)	Speech features (prosody and spectral)	0.16 – 0.20	0.66 – 0.78
Grimm et al. (2007)	Speech features (prosody and spectral)	0.13 – 0.15	0.46 – 0.82
Wollmer et al. (2008)	Prosody and spectrum speech features	0.10 – 0.20 (MSE)	
Nicolaou et al. (2011)	Audio, facial expressions, shoulder cues	0.15 – 0.24	0.53 – 0.79

Table 32 compares the results of the methodology used in Chapter 8 with results from other recent studies that have used emotion magnitude prediction. Comparing results across different studies is difficult because each study uses different datasets and emotion classes. However, it can be seen from Table 32, that the methodology presented in this chapter appears to perform as well or better than other methodologies.

Finally, once all relevant processing has been done and emotion state magnitudes of actors are estimated, the system can proceed to represent the emotion state of the actors at each position in the story. The final representation uses the extracted actor-emotion vectors to embed XML tags in the story. An XML tag is added for each actor emotion magnitude to the representation in XML format (Figure 21). For example, in Figure 21, the following tag can be added:

<enamex tag="person"><pos tag="nnp"><emotion sent=3 happy=0.6 surprised=0.2/> <word f="Valient"/> </pos> </enamex>

CHAPTER 9: SYSTEM RESPONSE

For a system to provide an automated response to detected emotion content, the following mapping and rendering approaches were developed. These techniques include emotion expression rendering and environment mapping using 3-D graphics.

9.1 Constrained Pose Bone and Heuristic based Rendering Approach

The objective of an automatic 3-D facial expression rendering approach using constrained pose bones is to generate an emotion response on a 3-D puppet based on results from emotion classification which maintains the integrity of the mesh. To achieve this, several architectures must be developed and pipelined to convert the text inputs into the final 3-D outputs. Output parameters are selected by the system based on previously encoded responses to inputs. Inputs to the system consist of testing examples, predicted classes, and inferred parameters.

Once the emotion detection model has been trained (see chapter 6), a system can use this model to try to identify the emotion tone in descriptive sentences. This can be very helpful when rendering 3-D characters from text descriptions because it can be used to automatically render the expression on the face of the 3-D character or puppet. For example, if an emotion class is identified for a descriptive sentence; then, the system can use that class to adjust the face parameters that can render the face expression associated with an emotion class.

Once the parameters are set, a script on any animation software is sufficient to render the facial expression. A rigged 3-D puppet is useful in this application because pose bones with constraints can be adjusted to move the mouth, eyebrows, etc. The puppet's built in constraints simplify the process because the mesh cannot be torn but only deformed within the set boundaries.

The method presented here relies on a frame and heuristic based approach to render expressions in 3-D models. For this initial analysis, the examples of the rendering approach are produced using blender

(Figure 24). However, the approach is similar for other types of 3-D animation software such as Maya, Blender, OpenGL, etc. The set of parameters are conveyed to the 3-D model in Blender by way of the scripting language. Blender uses python scripts and Maya uses the MEL scripting language. The 3-D puppet used was Blender's "Mancandy" rig (Gumster 2009; Blender Foundation 2009). This rig is used because it is freely available and it provides considerable flexibility when rendering facial expressions.

Mancandy's armature for the version implemented in this work consists of 12 controllers. Each controller serves as an imaginary lever or joystick that can be pulled and manipulated to move the muscles, eye track, mouth, eyebrows, jaw, etc. of the face to render any type of expression. Once an emotion state has been determined from the language input, an appropriate combination of facial movements must be selected that can best convey the emotion. The controller's positions are changed by adjusting the x, y, z coordinates of the pose bones via python code. This is very similar to moving an imaginary joystick in a cross fashion: moving the controller up causes smiling, moving it down causes frowning, the center position is neutral. For example, Figure 23 shows the python code that is used to

```
###################################
###   BOTbrows

brows = pose.bones['BOTbrows']
brows.loc[0] = -0.5     ## x
brows.loc[1] = -0.5     ## y
brows.loc[2] = -0.5     ## z

###################################
###   BOTsquint

squint = pose.bones['BOTsquint']
squint.loc[0] = float(0.9)
squint.loc[1] = float(0.9)
squint.loc[2] = 0
```

Figure 23: Python code to adjust pose bones in Blender

adjust the controller's position therefore adjusting the mesh coordinates within the constraints of the "brows" and "squint" pose bones. Heuristic rules, in this case, select the coordinates to use.

It is important to note that the pose controllers in the Mancandy rig follow the framework laid out by Paul Ekman for facial expression in the FACS system (Ekman 1978). Finally, the approach presented here, illustrates mechanisms which can be used in a pipeline-based architecture in existing and emerging

technologies for 3-D graphics generation and animation such as Coyne and Sproat (2001) famous WordsEye (Coyne and Sproat 2010).

Figure 24 includes examples of the 6 emotions used in this study and the possible rendering per correctly classified sentences. Although a good start, the wide breadth of possible facial expression renderings for each emotion class shows that emotion magnitudes are also required to properly render the facial expressions. Section 9.3 presents a different approach where predicted emotion magnitudes are used to determine the level of intensity for each detected emotion class.

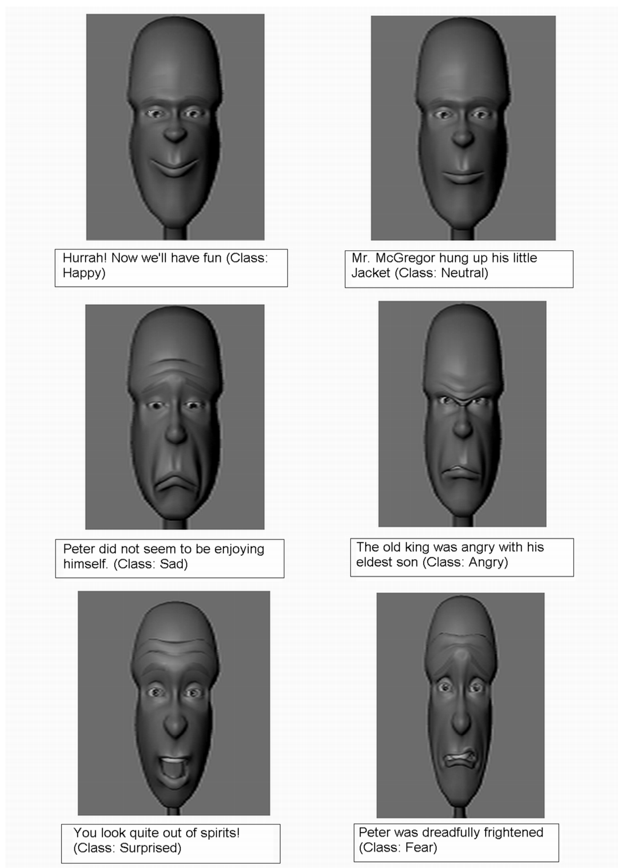


Figure 24: Blender's Mancandy 3D puppet (Transactions on MM © 2010 IEEE)

9.2 Environment or Cube Mapping

Text-to-scene processing is the process of mapping language inputs to visual graphics in 2-D or 3-D. In this section, the detected actors and features from chapter 7 can be used for automatic rendering of story

style characters. Additionally, all NP candidates from chapter 7 which were not classified as actors can instead be considered environmental element candidates. Therefore, the classification task developed in chapter 7 can be used to separate the NPs into these 2 groups. For non-actor NPs, the extracted features can be used to render environments and backgrounds.

As an example of text-to-scene processing, 3-D renderings are provided to illustrate what types of 3-D scenes can be generated from the corpus stories. The colors and type of mesh can be selected based on the actors, environments, features, and emotion classes. The background as can be seen in the image (Figure 25) provides a "dark forest" which can indicate that fear is present in the scene.

The system uses predefined lists of backgrounds to detect environmental information from the inputs. For each sentence, nouns in NP non-actor samples can be extracted and compared to the list of known environments. If an environment is detected, the word is saved for use in the environment rendering

Figure 25: 3-D rendering of a scene with emotion context and actors

approach. For aesthetic and quality purposes, the cube mapping approach (Akenine-Moller 2008) is used to render environment backgrounds in virtual worlds. A total of 16 cube maps are available for 3-D environment mapping. Additionally, the approach uses the detected emotional states as context to adapt and adjust the 3-D graphics based virtual environment and actor (virtual agent). The environment in which a 3-D virtual agent interacts is modified in response to the emotion inputs.

Since most studies focus on faces and gestures, it is important to note that the background environment can be just as critical for affective human computer interaction. Many important aspects such as color,

brightness (Valdez, 1994) and the environment content can reflect and influence human emotions. For instance, certain environments can be more pleasing or even soothing to the user. For instance, for a highly stressed person, a blue background with running water can be relaxing.

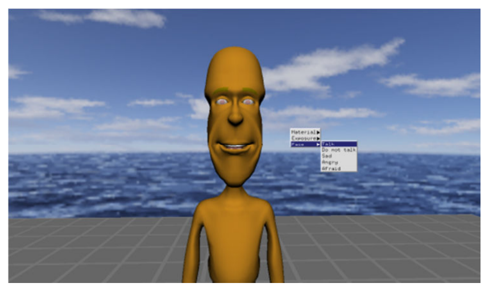
Change in environment for user interaction	
Relaxing sky and water	Emotion gradient background
	
Emotion expression change in an environment	
Frightened look with dark scary sky	Happy talk
	

Figure 26: Emotion sensitive virtual world renderings

Figure 26 presents examples of the rendered virtual worlds using cube mapping and morphing (Akenine-Moller et al. 2008), and environment maps (Cube Maps 2010) using GLSL and OpenGL. Available and used environment maps from (Cube Maps 2010) include: gradient, dark terrain, splatter landscape, grand canyon, hay room, hills, lake, landscape, lava, lobby, mars, night mountains, night sky, red sky, sea, and waterscape.

9.3 Emotion Magnitude Representation and Mapping

In content generation, emotion magnitudes can used for automatic emotional speech synthesis (from text) and emotional facial expression rendering in 3-D graphics (from text and speech). Emotion magnitudes

can be used to adjust pitch, rate or volume parameters in XML based schema such as Speech Synthesis Markup Language (SSML). The intensity adjustments can be used to control the intonation and pitch used to synthesize speech.

In computer graphics, emotion magnitudes can be used as weights to adjust the emotion expression of 3-D character renderings. This approach can be implemented with the morph targets technique (Akenine-Moller et al. 2008) in which vertex level weighted emotion meshes are added to a neutral mesh. In this case, the weight, which determines the level of interpolation between the target and neutral mesh, can be determined by the emotion magnitude predicted by the model.

- **Mapping of Emotion Magnitudes to 3-D Graphics**

In this dissertation, the morph targets technique is used to map the predicted emotion magnitudes to 3-D graphics. The objective of mesh morphing (Akenine-Moller et al. 2008) is to perform an interpolation on a per vertex basis that allows different points to be combined. Therefore, a difference vector is calculated between the neutral mesh point and the target mesh point. These difference vectors are added to the neutral vector and can be adjusted by weights (Equation 29). Using weights to adjust the deformation allows for facial expressions, talking sequences, and gestures to be modified by simply changing the weights of the per vertex difference between a neutral pose and a target pose. These weights are calculated based on the previously described regression models. Formally, the equation is as follows:

$$M_m = M_n + \sum_{q=1}^{t} w_q (M_n - M_t) \qquad \textbf{(Eq. 29)}$$

where M_m is the morphed mesh, M_n is the neutral mesh, t is the number of target meshes, M_t is the mesh for each target t, and w_q is the weight from 0 to 1 assigned to the morphing. This weight is the magnitude that is predicted by the regression model (Equation 24) for each emotion class. The relation can be seen in equation 30.

$$\mathbf{w_q = E_{d[s],a}} \qquad \textbf{(Eq. 30)}$$

Mesh morphing requires the use of different instances of the same mesh in different poses. These meshes can be created in any animation software such as Blender or Maya. Once the meshes are created, they can be read into the system as "obj" files. The important aspect is how to store the data. One approach is to create a data structure to hold the vertex and normal data for the neutral pose plus the data from the target meshes. This new data structure is used instead of the traditional object vertex data structure used for the mesh object. The difference vectors for the vertices and normals can be calculated using vertex shaders. Since the change is performed at the vertex level, it is important to pass each neutral vertex as well as the corresponding pose vertices and normals to the GPU at the same time. To achieve this, the program needs to link each pose's information to the shader via the "attribute" parameters

```
if (morphWeight4 < 0) send_value = 1;
if (morphWeight4 > 1) send_value = -1;
morphWeight4 = morphWeight4 + 0.05*send_value;
```

Figure 27: Talking sequence

defined in GLSL. This way, the GPU will have access to the vertex information for the neutral mesh and all corresponding target meshes. The degree of interpolation can be controlled by passing weights via "uniform" parameters defined in GLSL. An expressive talking sequence can also be implemented by incrementing a mouth-open weight using a step size. The weight grows until a threshold is reached. Once the maximum value is reached, the process is reversed and the step size decreases the weight (Figure 27). In Figure 27, morphWeight4 is the weight that adjusts the mouth interpolation, send_value is the direction in which the counter is adjusted (incrementing or decrementing), and 0.05 is the step size. The speed in which the mouth moves can be controlled by the step size. A larger step size can cause the mouth to open and close more quickly. This approach can also be used for expressive gestures.

The predicted emotion magnitudes can be used as weights for the morph targets rendering approach (Equation 29). The advantage of real valued renderings is that the amount of facial expressions that can be generated is un-limited. Additionally, several emotion expressions can be combined to create new ones. An example of this approach can be visualized in Figure 28. Figure 28 represents a four dimensional

space with axes for happy, sad, angry, and open mouth. Many emotion renderings can be generated by combining the different dimensions in this space. The transformation of the mesh from a neutral expression to other emotions is achieved efficiently, aesthetically, and quickly (Figure 28).

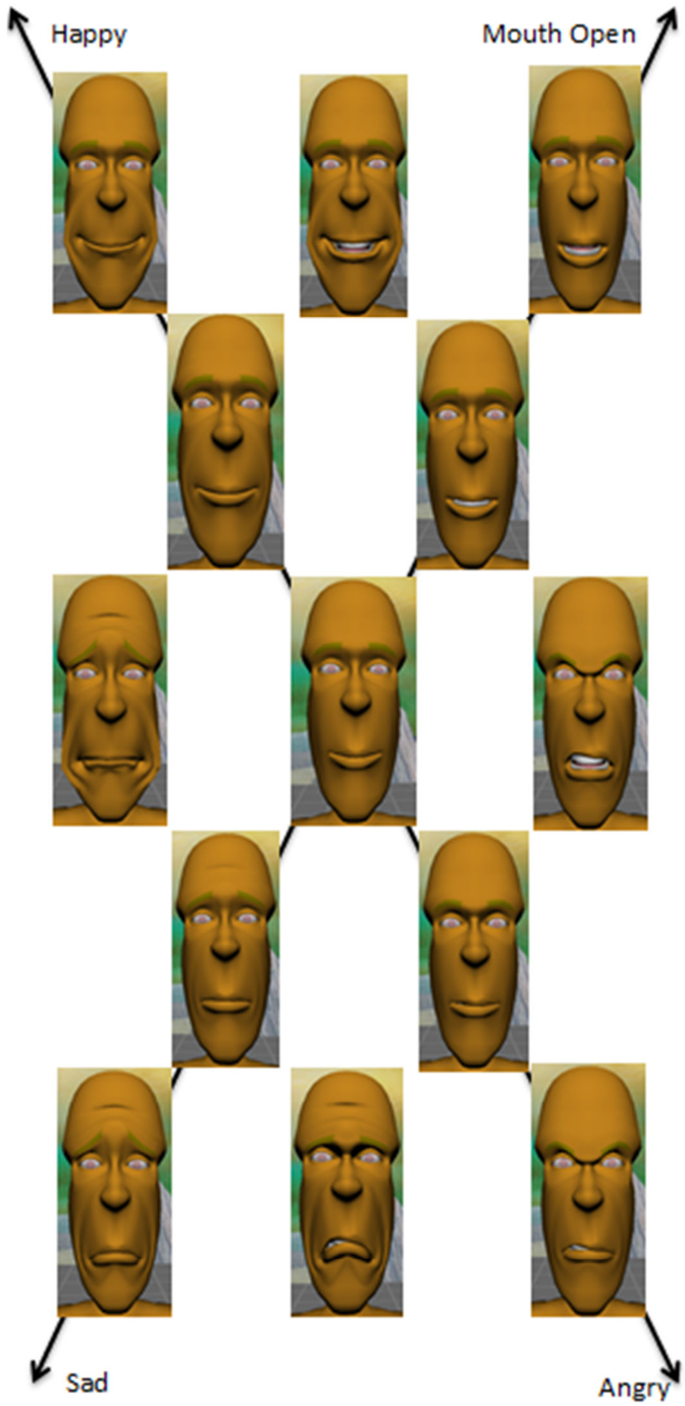


Figure 28: Emotion magnitude renderings

In Figure 28, the neutral expression can be seen in the center of the figure. The lower left hand corner represents the maximum sad expression, the lower right hand corner represents the maximum angry expression, the upper left hand corner represents the maximum happy expression, and the upper right hand corner represents the maximum open mouth expression. Facial expressions in between these points represent combinations of these basic axes.

9.4 Conclusions

The prediction model combined with the morph targets technique (Akenine-Moller et al. 2008) can provide a powerful tool to automatically render emotional facial expressions from emotion detection in text and speech. It is important to note that this technique could be easily extended to gestures and that the application is not limited to emotion expressions alone. Finally, Figure 29 shows the facial expressions as used in an interactive virtual environment.

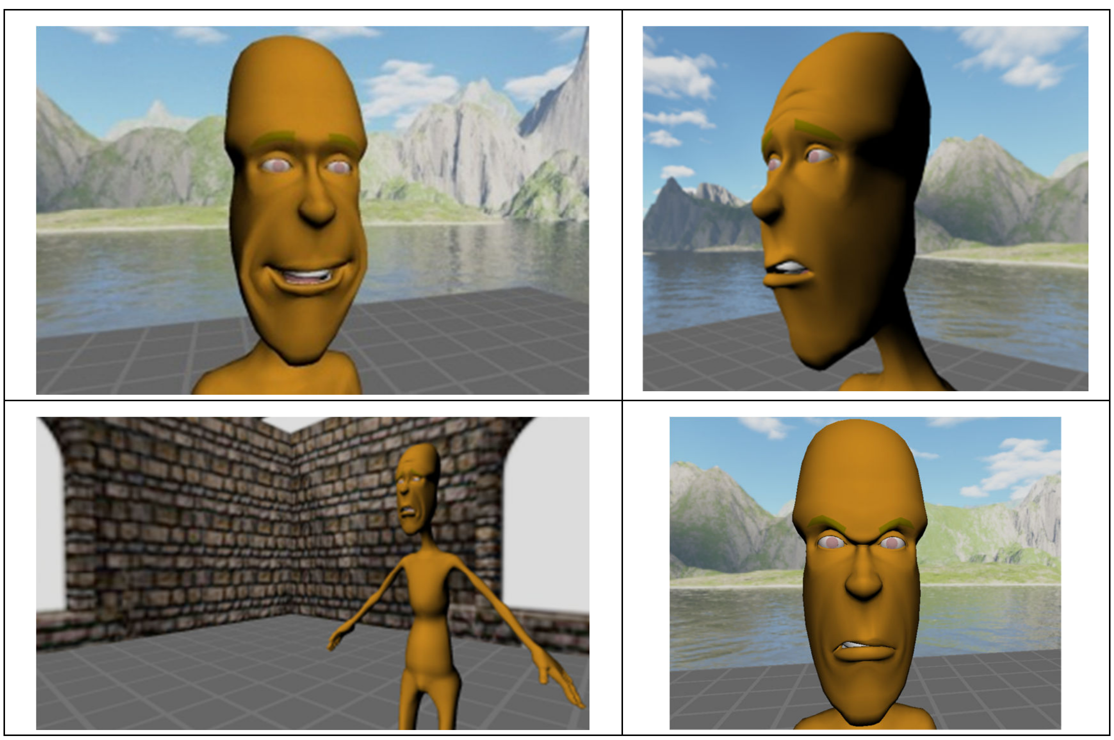


Figure 29: Renderings in virtual world environments

CHAPTER 10: CONCLUSIONS AND FUTURE WORK

This dissertation studied the automated semantic understanding of human emotions in writing and speech towards affective human computer interaction. Several approaches for automated emotion detection and emotion magnitude prediction were developed and evaluated. Additionally, 2 affect corpora were developed to acquire human behavioral information related to emotions. The first corpus was extended with new emotion magnitude annotations and speech information in the domain of children's stories. The second corpus was developed based on medical TV drama conversations.

10.1 Conclusions

Three emotion related methodologies were developed and results were discussed: First, a step-wise classification methodology for affect detection using sentiment composition and sentiment flow features in text and speech was developed and analyzed using two different communication formats (narrative vs. dialogue). Second, a methodology for nominal sentient entity detection was developed which can be used to identify actors. Third, a methodology for actor level emotion magnitude prediction and mapping to 3-D graphics using a prior state feedback regression model and mesh morphing was developed. Additionally, aspects about the challenges of affect detection in NLP such as feature selection, corpus development, unbalanced datasets, and anaphora resolution are also discussed.

The methodologies showed that domain specific applications are required but that there is common ground for emotion detection across applications. Simplifying emotion detection by taking a step wise approach is better that performing one step classification with multiple classes. Additionally, the requirement to render facial expressions from emotion inputs from text and speech showed the importance of predicting an emotion magnitude once a class is detected. For emotion magnitude prediction, non-linear regression models that relax normality assumptions such as Support Vector Regression proved best.

Finally, the methodologies developed in this work can serve as a blueprint for future developments of affective aware systems that use language to understand and respond to humans.

Important contributions of this work include:

- The extension of an affect corpus to include actor level emotion magnitudes, actors, and speech content.
- The development of a methodology to extract emotion word features from annotated corpora using probabilistic emotion content measures
- The development and testing of a recurrent (prior state feedback) method for actor level emotion magnitude prediction
- The development and testing of numeric text-based features that use sentiment composition and sentiment flow principles
- A methodology to map predicted emotion classes and magnitudes to a 3-D virtual agent using the vertex level mesh morphing technique
- The training and testing of a step wise methodology for emotion classification in narrative and dialogue domains
- The development of medical drama corpus based on narrative conversations for emotion analysis
- The development and testing of a methodology for sentient nominal entity detection

10.2 Recommendations for Future Work

Future work will focus on improvements in emotion detection and magnitude estimation, allowing multiple emotion classes to be active concurrently, and studying alternative emotion class models. Further future work will also focus on the speech extension of the Affect Corpus 2.0. More speech, semantic,

syntactic, and lexical features specific to healthcare related data will be added to the models to try to improve the overall performance of the methodology in the healthcare domain.

Additionally, the actor detection chapter (Chapter 7) showed one of the biggest bottlenecks in natural language processing which is anaphora resolution. Therefore, future work will focus on improving the methodology by incorporating better anaphora resolution techniques so that the actors can be tied to all the referring expressions that make reference to them. Finally, additional work will be pursued in improving the text-to-3D model rendering approach and a relevance feedback study will be conducted using human subjects to measure the usefulness of the affective 3-D display and emotion expression renderings.

REFERENCES

Akenine-Moller, T., Haines, E., Hoffman, N. (2008). Real-Time Rendering. A K Peters, Wellesley, Massachusetts.

Alm, C.O., Roth, D., Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 579-586.

Alm, C.O. (2008). Affect in Text and Speech. PhD Dissertation, University of Illinois at Urbana-Champaign.

Artstein, R., Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. Computational Linguistics, Volume 34, Issue 4, pp. 555-596.

Baayen, R., Piepenbrock, R., Gulikers, L. (1995). The CELEX lexical database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Baggia, P., Burkhardt, F., Martin, J., Pelachaud, C., Peter, C., Schuller, B., Wilson, I., and Zovato, E. (2009). Elements of an Emotional ML. [Online]. Available: http://www.w3.org/2005/Incubator/emotion/XGR-emotionml/ [Dec 10, 2009].

Beer, J. M., Rogers, W. A., Fisk, A. D. (2009). Emotion recognition of virtual agent facial expressions: The effects of age and emotion intensity. Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting, San Antonio, TX: Human Factors and Ergonomics Society, pp. 131-135.

Bethard, S., Martin, J., Klingenstein, S. (2007). Timelines from Text: Identification of Syntactic Temporal Relations. In Proceedings International Conference on Semantic Computing (ICSC), pp. 11-18.

Bi, L., Tsimhoni, O., Liu, Y. (2011). Using the Support Vector Regression Approach to Model Human Performance. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 41, No. 3, pp. 410-417.

Bigot, B., Ferrane, I., Pinquier, J., Andre-Obrecht, R. (2010). Speaker role recognition to help spontaneous conversational speech detection. In Proceedings of the 2010 international workshop on Searching spontaneous conversational speech (SSCS '10). ACM, New York, NY, USA, pp. 5-10. DOI=10.1145/1878101.1878104 http://doi.acm.org/10.1145/1878101.1878104.

Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python. 1st ed., O'Reilly Media.

Blender, The software of. (2009). Website: http://www.blender.org/.

Burns, B., Morrison, C. (2003). Temporal Abstraction in Bayesian Networks. In Working Notes of Association for the Advancement of Artificial Intelligence (AAAI), Spring Symposium Workshop: Foundation and Applications of Spatio-Temporal Reasoning, AAAI, Technical Report SS-03-03.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. Proceedings of the 6th international conference on Multimodal interfaces, International Conference on Multimodal Interfaces, pp. 205-211.

Busso, C., Lee, S., Narayanan, S. (2009). Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, No. 4, pp. 582-596.

Boersma, P., Weenink, D. (2005). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/.

Breazeal, C. (2000). Sociable Machines: Expressive Social Exchange between Humans and Robots. Sc. D. Dissertation, Massachusetts Institute of Technology.

Burges, C. J. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, vol. 2, pp. 121-167.

Castellano, G., Kessous, L., Caridakis, G. (2007). Multimodal Emotion Recognition from Expressive faces, body gestures and speech. In Proceedings of the Doctoral Consortium of 2nd International Conference on Affective Computing and Intelligent Interaction, Lisbon.

Calix, R., Mallepudi, S., Chen, B., Knapp, G. (2010). Emotion Recognition in Text for 3-D Facial Expression Rendering. IEEE Transactions on Multimedia, Special Issue on Multimodal Affective Interaction, Volume 12, Issue 6, pp. 544-551.

Calix, R., Knapp, G. (2011). Affect Corpus 2.0: An Extension of a Corpus for Actor Level Emotion Magnitude Detection. In Proceedings of the 2nd ACM Multimedia Systems (MMSys) conference, San Jose, California, U.S.A, pp. 129-132.

Callele, D., Neufeld, E., Schneider, K. (2006). Emotional Requirements in Video Games. 14th IEEE International Requirements Engineering Conference (RE'06).

Callele, D., Neufeld, E., Schneider, K. (2008). Requirements in Conflict: Player vs. Designer vs. Cheater. IEEE 2008 Third International Workshop on Multimedia and Enjoyable Requirements Engineering (MERE'08).

Cassell, J. (2001). Embodied Conversational Agents: Representation and Intelligence in User Interface. AI Magazine, 22(3), pp. 67-83.

Cassimatis, N. (2009). Flexible Inference with Structured Knowledge through Reasoned Unification. IEEE Intelligent Systems, Volume 24, Issue 4, pp. 59-67.

Chang, C.-Y., Tsai, J.-S., Wang, C. J., Chung, P.-C. (2009). Emotion Recognition with Consideration of Facial Expression and Physiological Signals. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 09, pp. 278-283.

Chang, C.-C., Lin, C. (2001). LIBSVM: a library for support vector machines. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16:341-378.

Chen, J. (2008). Automatic Face Animation with Linear Model. NY: Rensselaer Polytechnic Institute Website: http://www.cs.rpi.edu/~cutler/classes/advancedgraphics/S08/final_projects/chen.pdf

Chifu, E., Chifu, V. (2008). A Neural Model for Unsupervised Named Entity Classification. IEEE International Conferences on Computational Intelligence for Modeling, Control and Automation (CIMCA), pp. 1077-1082.

Chuang, Z., Wu, C.H. (2004). Multi-Modal Emotion Recognition from Speech and Text. Computational Linguistics and Chinese Language Processing, Vol. 9, No. 2, pp. 45-62.

Clay, S. R., Wilhelms, J. (1996). Put: Language-Based Interactive Manipulation of Objects. IEEE Computer Graphics and Applications, vol. 16, no. 2, pp. 31-39.

CMU. (1993). The Carnegie Mellon Pronouncing Dictionary. Carnegie Mellon University.

Cortes, C., Vapnik, V. (1995). Support-Vector Networks. Machine Learning, vol. 20, pp. 273-297.

Coyne, B., Sproat, R. (2001). Inferring the environment in a text-to-scene conversion system. In Proceedings of Conference on Knowledge Capture, pp. 147-154.

Coyne B., Sproat, R. (2010). WordsEye: An Automatic Text-to-Scene Conversion System. Website: http://www.wordseye.com/.

Cube maps. (2010). Retrieved from http://www.codemonsters.de/home/content.php?show=cubemaps.

Cunningham, H., Maynard, D., Bontacheva, K., Tablan, V. (2002). GATE: A framework and graphical development Environment for robust NLP tools and applications. Proceedings of the 40th anniversary meeting of the association for computational linguistics (ACL '02), Philadelphia.

Descartes, R. (1649). Passions of the Soul. 1649/1989, Hackett Publishing Co, Inc.

Diesner, J., Carley, K., He says, she says. Pat says, Tricia says. How much reference resolution matters for entity extraction, relation extraction, and social network analysis. Proceedings of the 2009 IEEE symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009), Ottawa, Ontario, Canada, pp. 9-16.

Ekman, P., Friesen, W. (1978). The Facial Action Coding System: A technique for the measurement of facial movement. San Francisco: Consulting Psychologists Press.

Ekman, P., Friesen, W. (1998). Constants across culture in the face and emotion. In: Jenkins, Oatley, & Stein (eds.), Human Emotions: A Reader. Malden, MA: Blackwell, chapter 7.

El-Nasr, M., Loerger, T., Yen, J. (1999). PETEEI: A pet with evolving emotional intelligence. Proceedings of the third annual conference on autonomous agents, Seattle, Washington, USA, pp. 9-15.

Gantz, J., Reinsel, D. (2010). The Digital Universe Decade – Are You Ready? IDC Report, http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm.

Gonzalez, R., Woods, R. (2008). Digital Image Processing. 3d ed., New Jersey: Prentice Hall.

Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., Petajan, E. (2002). Creating Interactive Virtual Humans: Some Assembly Required. IEEE Intelligent Systems, 17(4): 54-6.

Greenberg, S., Hollenback, J., Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In Proceedings of the Fourth International Conference on Spoken Language (ICSLP), Philadelphia, S24-27.

Grimm, M., Kroschel, K., Narayanan, S. (2007a). Support Vector Regression for Automatic Recognition of Spontaneous Emotions in Speech. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), Vol. 4, pp. IV-1085-IV-1088.

Grimm, M., Kroschel, K., Mower, E., Narayanan, S. (2007b). Primitives based Evaluation and Estimation of Emotions in Speech. Elsevier Speech Communication, 49, pp. 787-800.

Gumster, J. (2009). Blender for Dummies. pap/Cdr edition, For Dummies.

Hansen, J.H.L., Bou-Ghazale, S. (1997). Getting Started with SUSAS: A Speech Under Simulated and Actual Stress Database. EUROSPEECH-97: Inter. Conf. On Speech Communication and Technology, vol. 4, pp. 1743-1746, Rhodes, Greece, Sept.

Havasi, C., Speer, R., Alonso, J. (2007). ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge. 22nd Conference on Artificial Intelligence, 2007.

Jenkins, j., Oatley, K., Stein, N. (1998). Human Emotions: A Reader. pp. 63–72. Malden, Massachusetts: Blackwell.

Johansson, R., Williams, D., Berglund, A., Nugues, P. (2004). Carsim: A System to Visualize Written Road Accident Reports as Animated 3D scenes. Workshop on Text Meaning and Interpretation.

Johnson-Laird, P., Oatley, K. (1989). The language of emotions: An analysis of a semantic field. Cognition and Emotion, Vol. 3:2, pp. 81-123.

Jurafsky, D., Martin, J. (2008). Speech and Language Processing, 2nd ed., New Jersey: Prentice Hall.

Kanluan, I., Grimm, M., Kroschel, K. (2008). Audio-Visual Emotion Recognition Using an Emotion Space Concept. 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, pp. 25-29.

Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D. (2005). Semantic annotation, indexing, and retrieval. Journal of Web Semantics, 2(1), pp. 49-79.

Klabbers, E., Mishra, T., van Santen, J. (2007). Analysis of affective speech recordings using the superpositional intonation model. In Proceedings of the 6th ISCA workshop on speech synthesis (SSW6), pp. 339-344, Bonn, Germany.

Klein, D., Manning, C. (2002). Fast Exact Inference with a Factored Model for Natural Language Parsing. In Advances in Neural Information Processing Systems (NIPS 2002), Cambridge, MA: MIT Press, pp. 3-10.

Küçük, D., Yazici, A. (2008). Identification of co-referential chains in video texts for semantic annotation of news videos. In Proceedings of the International Symposium on Computer and Information Sciences.

LDC (1995). COMPLEX English pronunciation dictionary. Linguistic Data Consortium.

Librivox, [Date: Sept, 2010], DOI = http://librivox.org/

Liu, H., Lieberman, H., Selker, T. (2003). A Model of Textual Affect Sensing using Real-World Knowledge. In Proceedings of the 2003 International Conference on Intelligent User Interfaces, IUI 2003, January 12-15, 2003, Miami, FL, USA, ACM 2003, ISBN 1-58113-586-6, pp. 125-132.

Lu, C., Hong, J., Cruz-Lara, S. (2006). Emotion Detection in Textual Information by Semantic Role Labeling and Web Mining Techniques. In Proceedings of the Third Taiwanese-French Conference on Information Technology, Nancy, France.

Loquendo (2002). Loquendo Speech Suite Voice Portal Platform: Integration with Avaya Conversant for ASR. Loquendo.

LSU-NLP, [Date: Sept, 2010], DOI = http://nlp.lsu.edu.

Luengo, I., Navas, E., Hernaez, I. (2010). Feature analysis and evaluation for automatic emotion identification in speech. IEEE Transactions on Multimedia, Vol. 12, No. 6, pp. 490-501.

Mallepudi, S. A., Calix, R. A., Knapp, G. M. (2011). Material Classification and Automatic Content Enrichment of Images Using Supervised Learning and Knowledge Bases. In Proceedings of the IS&T/SPIE International Conference on Multimedia Content Access: Algorithms and Systems V, San Francisco, California.

Mansouri, A., Affendey, L., Mamat, A. (2008). A New Fuzzy Support Vector Machine Method for Named Entity Recognition. IEEE International Conference on Computer Science and Information Technology.

Manning, C., Schutze, H. (1999). Foundations of Statistical Natural Language Processing. 2nd ed., Cambridge, MA: MIT Press.

Mao, Y., Lebanon, G. (2006). Sequential models for sentiment prediction. In Proceedings of the International Conference on Machine Learning (ICML), Workshop on Learning in Structured Output Spaces, Pittsburg, PA.

Mathieu, Y. (2004). Linguistic Knowledge and Automatic Semantic Representation of Emotions and Feelings. In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC), pp. 314-318.

Massaro, D., Cohen, M., Beskow, J., Cole, R. (2001). Developing and evaluating conversational agents. Embodied conversational agents, MIT Press, Cambridge, MA.

McShane, M. (2009). Reference Resolution Challenges for an Intelligent Agent: The Need for Knowledge. IEEE Intelligent Systems, Vol. 24, No. 4, pp. 47-58.

Minato, J., Matsumoto, K., Ren, F., Kuroiwa, S. (2007). Corpus-based Analysis of Japanese-English Emotional Expressions. IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLPKE).

Minsky, M. (2007). The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon & Schuster.

Miller, G. (1995). WordNet: A Lexical Database for English. Communications of the ACM, Vol. 38, No. 11, pp. 39-41.

Moilanen, K., Pulman, S. (2007). Sentiment Composition. In Proceedings of Recent Advances in Natural Language Processing (RANLP 2007), 27-29, Borovets, Bulgaria, pp. 378-382.

Mukerjee, A., Gupta, K., Nautiyal, S., Singh, M.P., Mishra, N. (2000). Conceptual Description of Visual Scenes from Linguistic Models, Image and Vision Computing, 18, 173-187.

Muller, K., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. IEEE Transactions on Neural Networks, Vol. 12, No. 2, pp. 181-201

Neviarouskaya, A., Prendinger, H., Ishizuka, M. (2009). Semantically distinct verb classes involved in sentiment analysis. In Proceedings IADIS international conference on applied computing, AC (1), pp. 27-35.

Nicolaou, M., Gunes, H., Pantic, M. (2011). Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence – Arousal Space. IEEE Transactions on Affective Computing, 16 Mar 2011. DOI: http:/doi.ieeecomputersociety.org/10.1109/T-AFFC.2011.9

NIST. (1990). Timit Speech Corpus, NIST.

Nguyen, H., Cao, T. (2008). Named Entity Disambiguation on an Ontology Enriched by Wikipedia. IEEE International Conference on Research, Innovation and Vision for the Future.

Noh, J., Neumann, U. (1998). A survey of facial modeling and animation techniques. Technical Report, USC, pp. 99-705.

Nuance. (2011). Naturally speaking software. http://www.nuance.com/dragon/index.htm. Accessed 30 March 2011

Opencyc. (2010). Website: http://www.opencyc.org, Date: Feb. 20, 2010

Ortony, A., Clore, G., Foss, M. (1987). The referential structure of the affective lexicon. Cognitive Science, 11, pp. 341-364.

Ortony, A., Clore, G., Collins, A. (1998). The Cognitive Structure of Emotions. Cambridge University Press.

Osgood, B. (2009). The Fourier Transform and its Applications. Stanford University.

Osherenko, A., Andre, E. (2007). Lexical Affect Sensing: Are Affect Dictionaries Necessary to Analyze Affect?. In Proc. of Affective Comp. and Intelligent Interaction, Springer, ACII 2007, 230-241.

Osherenko, A. (2008). Towards Semantic Affect Sensing in Sentences. Proceedings of Artificial Intelligence and Simulation of Behavior (AISB).

Paab, G., Pilz, A. (2009). Named Entity Recognition of Spoken Documents using Subword Units. IEEE International Conference on Semantic Computing, Berkeley, California.

Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol. 2, No. 1-2, pp. 1-135.

Pang, W., Fan, X. (2009a). Chinese Nominal Entity Recognition with Semantic Role Labeling. IEEE International Conference on Wireless Networks and Information Systems.

Pang, W., Fan, X. (2009b). A Novel Framework for Nominal Entity Recognition. Second International Symposium on Computational Intelligence Design.

Pelachaud, C. (2009). Modeling multimodal expression of emotion in a virtual agent. Philosoph. Trans. Roy. Soc. B Biol. Sci., B, vol. 364, pp.3539-3548.

Picard, R. (1997). Affective Computing. MIT press.

Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., Fosler- Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Placeway, P., Chen , S., Eskenazi, M., Jain, U., Parikh, V., Raj , B., Ravishankar, M., Rosenfeld , R., Seymore, K., Siegler, M., Stern, R., Thayer, E. (1996). The 1996 Hub-4 Sphinx-3 System. In Proceedings of DARPA Speech Recognition Workshop.

Plutchik, R. (1980). Emotion: A Psycho Evolutionary Synthesis. New York: Harper & Row.

Porter, M. (1980). An algorithm for suffix stripping. Program. 14 (3), pp. 130-137.

Prakasvudhisarn, C., Trafalis, T., Raman, S. (2003). Support Vector Regression for Determination of Minimum Zone. Journal of Manufacturing Science and Engineering. Vol. 125, pp.736-739.

Qiao, X., Liu, Y. (2009). Adaptive Weighted Learning for Unbalanced Multicategory Classification. Biometrics, Vol. 65, 159-168.

Qiu, L., Kan, M.-Y., Chua, T.-S. (2004). A public reference implementation of the RAP anaphora resolution algorithm. In Proceedings of Language Resources and Evaluation Conference (LREC).

Scherer, K., Wallbott, H.G. (1997). The ISEAR Questionnaire and Codebook.

Schröder, M., Pirker, H., Lamolle, M. (2006). First suggestions for an emotion annotation and representation language. In L. Deviller et al. (Ed.), Proceedings of Language Resources and Evaluation Conference (LREC'06) Workshop on Corpora for Research on Emotion and Affect, Genoa, Italy, pp. 88-92.

Schuller, B., Rigoll, G., Lang, M. (2003). Hidden Markov Model-Based Speech Emotion Recognition. Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, & Signal Processing, April 6-10, Hong Kong.

Sebe, N., Cohen, I., Huang, T. (2005). Chapter on Multimodal Emotion Recognition. Handbook of Pattern Recognition and Computer Vision, World Scientific.

Shafran, I., Mohri, M. (2005). A Comparison of Classifiers for Detecting Emotion from Speech. IEEE International Conference on Acoustics, Speech, and Signal Processing.

Singh, P. (2005). EM-ONE: An Architecture for Reflective Common Sense Thinking. PhD Thesis, Massachusetts Institute of Technology.

Smola, A., Scholkopf, B. (2004). A tutorial on Support Vector Regression. Statistics and Computing, Vol. 14, pp. 199-222.

Song, M., You, M., Li, N., Chen, C. (2008). A robust multimodal approach for emotion recognition. Neurocomputing, 71, pp. 1913-1920.

Soon, W.M., Ng, H., Lim, D. (2001). A Machine learning approach to co-reference resolution of noun phrases. Computational Linguistics, 27(4), pp. 521-544.

Steinberger, J., Poesio, M., Kabadjov, M., Jeek, K. (2007). Two uses of anaphora resolution in summarization. Information Processing and Management: an International Journal, Volume 43, Issue 6, pp. 1663-1680.

Strapparava, C., Mihalcea, R. (2008). Learning to identify emotions in Text. Proceedings of the 2008 ACM symposium on Applied Computing, Fortaleza, Ceara, Brazil, pp. 1556-1560.

Sutton, S., Cole, R., Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M., Yan, Y., Kaiser, E., Rundle, B., Shobaki, K., Hosom, P., Kain, A., Wouters, J., Massaro, D., Cohen, M. (1998). Universal Speech Tools: The CSLU Toolkit. In Proceedings of the International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, pp. 3221-3224.

Tapus, A., Mataric, M., Scassellati, B. (2007). The Grand Challenges in Socially Assistive Robotics. IEEE Robotics and Automation Magazine (RAM), Special Issue on Grand Challenges in Robotics, 14 (1), March 07.

Tapus, A., Tapus, C., Mataric, M. (2008). User-Robot Personality Matching and Assistive Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy. In Intelligent Service Robotics: Multidisciplinary Collaboration for Socially Assistive Robotics, 1, pp. 169-183.

Tawari, A., Trivedi, M. (2010). Speech Emotion Analysis: Exploring the Role of Context. IEEE Transactions on Multimedia, Vol. 12, No. 6, pp. 502-509.

Tawfik, A., Neufeld, E. (1994). Temporal Bayesian Networks. In Proceedings of the First International Workshop on Temporal Representation and Reasoning.

Theune, M., Meijs, K., Heylen, D., Ordelman, R. (2006). Generating Expressive Speech for Storytelling Applications. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, pp. 1137-1144.

Todorovic, B., Rancic, S., Markovic, I., Mulalic, E., Ilic, V. (2008). Named Entity Recognition and Classification using Context Hidden Markov Model, 9th Symposium on Neural Network Applications in Electrical Engineering.

Tokuhisa, R., Inui, K., Matsumoto, Y. (2008). Emotion classification using massive examples extracted from the web. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, pp. 881-888.

Valdez, P., Mehrabian, A. (1994). Effects of Color on Emotions. Journal of Experimental Psychology: General, Vol. 123, No. 4, pp. 394-409.

Valitutti, A., Strapparava, C., Stock, O. (2004a). Developing Affective Lexical Resources. PsychNology Journal, 2(1), pp. 61-83.

Valitutti, R. (2004). WordNet-Affect: An Affective Extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluations.

Versley, Y., Ponzetto, S.P., Poesio, M., Eidelman, V., Jern, A., Smith, J., Yang, X., Moschitti, A. (2008). BART: A Modular Toolkit for Co-reference Resolution. Language Resources and Evaluation Conference (LREC).

Ververidis, D., Kotropoulos, C. (2005). Emotional Speech Classification using Gaussian Mixture Models. In Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), Japan.

Wang, A., Emmi, M., Faloutsos, P. (2007). Assembling an Expressive Facial Animation System. In Proceedings of the 2007 ACM SIGGRAPH symposium on video games, New York, New York.

Wang, W., Wang, W. (2008). A Gray-Scale Face Recognition Approach. Second International Symposium on Intelligent Information Technology Application.

Wang J., Gong, Y. (2009). Recognition of Multiple Drivers' Emotional State. In Proceedings of the 2009 IEEE international conference on Multimedia and Expo.

Wiebe, J., Wilson, T., Cardie, C. (2005). Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation, Volume 39, Issue 2-3, pp. 165-210.

Witten, I., Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. 2d. edition, Morgan Kaufmann Publishers Inc., San Francisco.

Wollmer, M., Eyben, F., Reiter, S., Schuller, B., Cox, C., Douglas-Cowie, E., Cowie, R. (2008). Abandoning emotion classes – towards continuous emotion recognition with modeling of long-range dependencies. In Proceedings of the 9th Interspeech 2008 incorp. 12th Australian International Conference on Speech Science and Technology SST 2008, Brisbane, Australia, pp. 597-600.

WordNet. (2010). Website: http://wordnet.princeton.edu/.

Wu, C., Tzeng, G., Goo, Y., Fang, W. (2007). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Expert Systems with Applications, 32, pp. 397-408.

Yang, Y., Pederson, J. (1997). A Comparative Study on Feature Selection in Text Categorization. In Proceedings of the Fourteenth International Conference on Machine Learning, pp. 412-420.

Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A. (2003). Face Recognition: A Literature Survey. ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458.

APPENDIX A: EMOTION WORD LISTS

Afraid words
accused, adversary, afraid, alarm, alarmed, alarming, mafia, doom, doomsday, day_of_reckoning, end_of_the_world, murdered, shouter, yeller, screams, doomed, agony, damned, Satan, Devil, mist, menacing, somber, duskiness, Death, destruction, killing, deaths, deterrent, danger, distressing, distressful, disturbing, perturbing, worrisome, worrying, predicament, Lucifer, Darkness, Hell, devil, fiend, demon, monster, devilled, devilish, diabolic, diabolical, demonic, fiendish, hellish, infernal, satanic, unholy, cruel, cruelty, hell, devils, despot, grimaces, desperate, danger, catastrophe, disaster, calamity, tragedy, cataclysm, morbid, disfigure, blemish, mar, spoil, scars, defaced, cheeks, vandals, evilness, malefic, malevolent, fearfulness, fright, flee, Fear, timorous, fearfully, fearfully, freak, antsy, frighten, scare, affright, stranger, hangs, frightens, Ghosts, drive, cowered, panicky, panicked, panic-stricken, panic-struck, terrified, desperation, ghost, specter, spectre, haunting, cloak-and-dagger, burning_at_the_stake, victims, ruined, burned-over, burned-out, burnt-out, timid, guardedly, cemetery, graveyard, burial_site, burial_ground, burying_ground, memorial_park, necropolis, tract, burials, demons, bones, irons, shacles, chase, chase_after, frightened, die, decease, perish, pass_away, anxiety, cadaver, corpse, dead, murderer, coward, timidity, Coward, cowardliness, cowardly, faint-hearted, creep, spook, creeping, submission, creepy, creepy-crawly, outcry, howler, riot, specters, disembodied, ghosted, moonlit, haunt, obsess, ghost, haunts, ghoulish, glooming, gloomful, gloomful, darkening, cradle, tomb, burial, tombstone, cemetery, grave, culpt, ghastly, grisly, gruesome, macabre, bellower, screamer, screecher, danger, poverty, obscurity, suffocating, atrociously, abominably, terrify, terrorize, terrorise, terror, terrific, terrifying, terror, threatened, threatens, fainthearted, timidness, timorousness, shyly, timidly, stuttering, sinners, Hell, nightmares, dangerously, intrigues, trembling, shakiness, quivering, shaky, shivering, edginess, inquietude, disquietude, hole-and-corner, mystic, mystical, occult, shiver, chill, quiver, shudder, thrill, tingle, fear, shake, shivering, tremble, convulsively, shuddering, convulsively, slash, whipping, massacre, mass_murder, carnage, butchery, butcher, slaughtered, massacred, apparitional, ghostlike, ghostly, phantasmal, phantom, emanations, tappings, seance, apparition, phantasm, phantasma, fantasm, squeal, screaky, screechy, squeaking, squeaky, squealing, startle, startled, mutilate, homicide, paralysis, palsy, hazard, jeopardy, endangerment, possibility, riskiness, endanger, jeopardize, jeopardise, menace, threaten, imperil, pose, persecuted, phobic, irrational, fears, poisonous, harms, destroys, prisonlike, confinement, confined, dragons, withdraw, pull_away, draw_back, robbers, burglars, robber, thief, steals, robbery, larceny, looting, plundering, riots, wartime, scurry, scamper, skitter, scuttle, scurried, daunt, dash, scare_off, pall, frighten_off, scare_away, frighten_away, dashed, scar, chilling, scarey, scary, shivery, shuddery, shuddery, homicidal, murderous, tendency, thugs, sicken, nauseate, turn_one's_stomach, nauseated, sickening, vile, nausea, nervus, flighty, spooky, unpredictably, nervously, jitteriness, jumpiness, restiveness, jitters, highly, strung, temperament, alarming, anguish, abominable, delirious, frantic, unrestrained, insane, madden, crazy, nerves, maddening, maddened, madman, maniac, lunacy, insaneness, rabies, hydrophobia, phobia, rabid, virus, craziness, rash, blizzard, rabidness, maul, distorted, woeful, ill-being, intolerable, giant, goliath, behemoth, colossus, abnormally, powerful, monster, malformed, mortification, embarrassment, gangrene, mortifying, mortify, necrose, sphacelate, necrosed, gangrenous, mortified, wounded, confine, suspects, imprisoned, trial, incarcerated, prison, captive, confined, jailed, captivity, intimidate, intimidates, threats, intimidated, jittery, jerky, edgy, high-strung, highly_strung, jumpy, nervy, overstrung, restive, uptight, tense, putting_to_death, terminating, death, killed, extinguish, forceful, overwhelm, deprive, lucifer, lugubrious, mournful, gloominess, lugubriousness, sadness, mournfulness, uncheerfulness, recoil, reverberate, unsettle, unsettled, weak-kneed, stalk, stalked, haunted, obsessed, apparitions, heinous, wicked, accusations, horrid, hideous, grossly, stricken, hunt, wild, hunted, hounded, prey, breakneck, rush, hasten, festinate, rush, unnatural, hurriedly, hastily, hasty, hurrying, delirium, frenzy, hysteria, hysterical, neurotic, outbreaks, disturbances, imprison, incarcerate, weariness, deadening, wearisome, capture, weary, aweary, wilderness, uninhabited, untamed, bullet, sword, allegations, burn, burning, burn_mark, burned, burn_down, combust, burns, bite, sting, sharp, stinging, fog, incinerate, Witches, Salem, burn, burn_off, burn_up, war, warfare, fighting, fierce, haunted, firing, beast, deamon, wolf, savage, wildcat, blackness, total_darkness, pitch_blackness, fatal, unfortunate, crashed, pitch-black, pitch-dark, moonless, cellar, blackest, destroyer, cowardice, shed_blood, bleed, hemorrhage, lose, leech, phlebotomize, phlebotomise, bled, extort, bleeding, haemorrhage, hair-raising, nightmarish, darker, darkness, bleak, calamitous, disastrous, monster, ghost, monstrous, shockingly, horrifying, horrible, horror, decapitation, alarming, horrifying, horrible, onslaught, attack, uncontrollable, assault, beat, hostilities, assailed, lash_out, mugger, assaulted, Nightmare, injure, awful, beware, alarmed, suffered, damage, anguished, tormented, tortured, humiliation, anxiousness, nervousness, panic, anxious, dying, eagerly, nervous, queasy, uneasy, unquiet, fraught, anxiously,

apprehensively, aghast, appalled, dismayed, struck, fear, dread, appalling, experience, appalls, merely, dismaying, conditions, apprehensiveness, expectation, taking_into_custody, apprehending, policeman, police, apprehensive, alerted, burglaries, dark, horrific , suspicions, irrationally, raving, hazardous, skydiving, forsaken, desolate, barbarian, uncivilized, uncivilised, civilizing, invaders, evil-looking, witchcraft, witchery, worried, concerned, anxious, darkest, deepest, anxieti, anxieties, Anxiety, anxiou, anxioused, appalling , armi, armies, battlements, battling, brandished, catastrophic , Caution, chickens, chicks, corpses, cowardic, Cowardice, cowardli, cowardly , dangles, dared, darest, deterring, direful , disast, disastrous , Distress, distressed , distrusted, dreadful , fearing, fear-inspiring, fearsome , freak , Fright, frightening , frigid , harmed, horrendous , horribl, horrible , horrifi, horrifying , Horror, horrors, Hysteria, mauled, mortifi, Mortification, murderers, murdering, Nervousness, overapprehensive, overtimid, overtimorous, overwhelmed, paler, Panic, paralysed, plagu, plagued, praying, retreating, robberi, robbers, robbing, scampered, scaring, screamed, shuddered, shuddersome, spectres, squealed, squeals, stabbed, terrible , terrifi, Terror, Timidity, timidli, Torment, trembl, trembled, troubl, troubled , Troublesom, Troublesome, unbear, Uneasiness, villains, Weariness, witches, Worry, darkness, shadow, cryptic, cryptical, inscrutable, mysterious, mystifying, obscure, dim, rituals, depth, glum, glum, superstition, dismal, gloomy, drab, drear, dreary, life-threatening, anxiety, threatening, beg, condemnation, alarming, frightening, anguish, torment, torture, distress, fearful, fearsome, horrendous, horrific, terror, risk, howling, frightful, frightfully, beware, criminal, guilty, punishment, afraid, accused, accuse, impeach, incriminate, accusation, abuse, blame, wrongdoing, misbehavior, opponent, opposer, resister, fear, apprehension, suddenly, snakes, concern, insecurity, aversion, alarm, dismay, consternation, danger, warning, undesirable, alert, warning_signal, alarum, automatic, warning, horrify, alarm, deep, darkest, deepest, dreadful, severe, anguished, anxieti, anxieties, Anxiety, anxiou, anxious, anxioused, anxiously, appalled, appalling, appalling , apprehension, apprehensive, armi, armies, armlet, armour, arms, army, atrocious, attack, attacked, bad, barricaded, battlefield, battlements, battling, beast, beware, black, bled, bleed, bleeding, bloodcurdling, bounded, brandished, burn, burning, burnt, catastrophic , Caution, cautious, cautiously, cemetery, chain, chained, chains, chased, chicken, chickenhearted, chickens, chicks, choke, choked, choking, concern, concerned, corpse, corpses, courage, coward, cowardic, Cowardice, cowardli, cowardly, cowardly , creeps, creepy, cries, criminal, crook, crouched, curse, danger, dangerous, dangles, dare, dared, darest, dark, darker, darkly, darkness, dead, deadly, death, deliverance, deterrent, deterring, devil, devils, diabolical, dire, direful , disast, disaster, disastrous, disastrous , diseased, disfigured, Distress, distressed, distressed , distressing, distrusted, disturbance, disturbed, disturbing, doom, doomed, dread, dreaded, dreadful, dreadful , dreadfully, dreary, eerie, enemies, enemy, evil, Fear, feared, fearful, fearfully, fearing, fear-inspiring, fears, fearsome, fearsome , fled, flee, frantically, freak, freak , fret, fretful, fretted, Fright, frighten, frightened, frightening, frightening , frightful, frightfully, frigid , frilled, frisked, ghost, ghoulish, gloomy, godless, grave, gruesome, hard, harm, harmed, harmless, haunted, heinous, help, horrendous, horrendous , horribl, horrible, horrible , horribly, horrid, horrifi, horrific, horrific , horrified, horrifying, horrifying , Horror, horrors, hunted, hurricane, hurried, hurriedly, hurrying, Hysteria, imprisoned, intimidated, jittery, kill, killed, kills, lucifer, lugubrious, lugubriousness, macabre, mad, maddened, madman, madness, mauled, miserable, misery, misfortune, monster, morbid, mortifi, Mortification, mortified, mown, murder, murderer, murderers, murdering, murderous, nauseate, nauseated, nauseating, nauseous, nerves, nervous, nervously, Nervousness, old, ordeal, overapprehensive, overtimid, overtimorous, overwhelm, overwhelmed, overwhelming, pale, paler, Panic, panicky, pant, panting, paralysed, paralysis, peril, perish, perished, persecuted, phobic, plagu, plague, plagued, pleaded, plight, poison, poisoned, poisonous, praying, prison, prisoner, prisoners, prudently, queasy, quiet, rape, reluctance, Remorse, restlessness, retreating, robbed, robber, robberi, robbers, robbery, robbing, robed, scampered, scared, scaring, scary, scrape, scream, screamed, screamer, screaming, secret, shrieked, shrieking, shudder, shuddered, shuddering, shuddersome, sin, sissy, slain, slashing, slaughter, spectral, spectre, spectres, squealed, squealing, squeals, stab, stabbed, startled, startling, stranded, stress, struggle, struggling, suffocated, suffocating, terrible, terrible , terribly, terrifi, terrified, terrifying, Terror, threaten, threatened, timid, Timidity, timidli, timidly, Torment, tormented, treacherous, treachery, trembl, tremble, trembled, trembling, troubl, troubled, troubled , troubles, Troublesom, Troublesome, troubling, turbulent, unbear, unbearable, Uneasiness, uneasy, unfortunate, unsettled, villain, villains, violent, warn, weak-kneed, Weariness, wearisome, weary, wild, witch, witchcraft, witches, worried, Worry, worrying, wretched

Surprised words
!, aback, abrupt, absurdly, aflutter, aggravated, aggravated , Aggravation, agitated, Agitation, amazed, amazement, ambush, anticipation, apoplectic, appalled, appalling, appalling , arose, arousal, aroused, astonish, astonished, Astonishment, astounding, Awe, majestic, burst, bursts, bustle, captivated, commotion, daunted, daunting, dazzled, dazzling, delirious, enchanted,

enchantingli, enchantingly, enchantment, excit, excitable, excited, excitedli, excitedly, Excitement, excites, exciting, exclaim, exclaimed, exclaims, exclamation, exuberant, exult, exultant, exultation, fascinated, fashionable, havoc, jittery, loud, loudly, mysterious, Mystery, mystified, noise , noises, noisi, noisy, outrageous, perplexed, perturbed, prying, queer, rare, rustling, secret, shining, shiver, shivered, shivering, shivery, Shock, shocked, shocking, shocking , shockingly, spirited, strange, stunned, sudden, suddenly, surpris, surprise, surprised, surprising, suspect, thrill, thrilled, thrilling, tumultous, tumultuous, unknown, unnatural, wild, surprised, aback, surprise, taken, abrupt, sudden, unexpected, precipitous, extremely, surprisingly, unceremoniously, absurd, aflutter, nervous, excited, anticipation, stir, excite, charge, cause, agitated, excited, foamed, bubbled, agitation, extreme, emotional, development, agitating, amaze, astonish, astound, affect, amazes, perplex, mystify, baffle, bewilder, flummox, stupefy, dumbfound, mystery, bewildering, vexing, amazed, astonied, astonished, astounded, stunned, overwhelming, shock, magician, enthralled, vastness, majesty, astonishment, amazement, surprising, ambush, ambuscade, attack, bushwhack, waylay, ambushing, expectancy, expectation, expected, anticipations, seizure, scandalize, scandalise, phenomenon, arousal, rousing, arousing, heightened, awake, awaken, wake, energize, energise, perk, alert, thunder, astounding, dumbfounding, dumfounding, striking, astonishing, staggering, stupefying, impressive, stun, overwhelm, awe, admiration, wonder, suddenly, erupt, trance, catch, captivate, fascinate, bewitch, entrance, enchant, captivated, charmed, dazzle, bedazzle, dazzled, stupefied, dizzied, dazzling, eye-popping, fulgurant, fulgurous, amazingly, impressive, suggestive, flashing, enthrall, spellbound, hex, enchanted, influenced, charms, incantations, bewitchingly, captivatingly, enchantingly, enthrallingly, bewitching, captivation, enchantment, enthrallment, fascination, excitedly, excitation, excitements, exciting, surprise, exclaimed, exclamation, exclaiming, exultant, exulting, curious, transfix, grip, spellbind, fixed, stare, fascinated, hypnotized, hypnotised, mesmerized, mesmerised, spell-bound, transfixed, attention, fixated, baffles, mystified, perplexed, bewilderment, curiosity, Agitation, Astonishment, Awe, bursts, enchantingli, excit, excitedli, Excitement, excites, exclaims, Mystery, Shock, shocking , surpris, tumultous, amazing, glitter, glisten, glint, gleam, glistening, distinguished, shines, shinny, quiver, shudder, thrill, shock, shocks, shocking, dazed, stupefied, unexpectedly, abruptly, quickly, impulse, surprisal, unexplored, alien, fantastic, fanciful, gaga, barbarian, barbaric, savage, wonderment, marvel, wonders

Sad words
abandon, abandoned, abstinence, abuse, abysmal, accidental, accidentally, ached, afflicted, aged, Agony, desolate, lonely, chabby, unfulfilled, meager, meagre, meagerly, stingy, ailing, sickly, sadden, saddened, sadly, unhappily, sadness, deplorably, lamentably, woefully, neglected, unbalanced, sickened, sighing, sigh, sighed, sob, sobbing, convulsive, gasp, inconsolably, humorless, sorrowfulness, irreparable, sorrowful, Proverbs, sorrowfully, dolefully, good-for-nothing, good-for-naught, meritless, spiritless, scrimpy, averageness, mediocrity, ordinariness, consequence, second-rater, second-rate, second-raters, mediocrities, melancholic, melancholiac, melancholia, thoughtful, somber, sombre, solemn, miser, hoarder, wretched, unhappy, prisoners, huddled, stinking, cages, hapless, misfortunate, pathetic, piteous, pitiable, pity, shabby, pathetic, abject, low-down, scurvy, contemptible, rabble, woeful, wretchedness, miserableness, ill-being, affliction, misfortune, slums, intolerable, bad_luck, unnecessary, unforeseen, tough_luck, ill_luck, unfavorable, groan, moan, pain, discomfort, groaned, mourn, mourner, griever, sorrower, lamenter, grieving, died, plaintive, doleful, evoking, eyes, mournfully, bereavement, bereaved, bereft, sorrowing, sorrowful, phrase, `pain_sensation, painful_sensation, discomfort, painful, bites, painfully, distressingly, ignorance, sorely, ail, captured, sulky, depressingly, hopelessly, dispiritedly, hopelessness, mortification, trauma, violence, detriment, negatively, wounded, ailment, complaint, disease, adversity, blows, distressing, poorly, unsatisfactory, ill-fitting, ill-conceived, unfavorably, inconvenience, scarcely, ill-humored, ill-humoured, unwellness, malady, sickness, faineant, indolent, lazy, unemployed, ulcer, leprosy, infectious, sin, destitute, joyless, lamentation, pitiful, laments, mourning, mournful, lamentable, bad, Lamentations, Book_of_Lamentations, Testament, lamenting, desolation, mourned, grieved, wailing, wailful, sinners, mourners, bagpipes, loneliness, solitariness, isolation, forlornness, aloneness, lonesomeness, sole, gloomful, archaic, heartache, heartbreak, brokenheartedness, death, grieve, aggrieve, grieves, dangerous, grievous, life-threatening, threatening, illness, pneumonia, heartbreaking, heartrending, sigh, atrocious, flagitious, unforgiving, unrelenting, placated, grimmest, arduous, backbreaking, grueling, gruelling, laborious, operose, punishing, toilsome, exhaustion, voiceless, bitterness, brokenhearted, heartbroken, heartfelt, condolences, homesick, longing, homesickness, hopeless, bitterly, illnesses, discharged, laid-off, pink-slipped, listless, displease, displeasure, displeased, dolorous, dolourous, lachrymose, tearful, drowned, modest, homeless, embarrass, abash, embarrassed, wipe_out, exhausted, miscarry, miserably, flunk, faintest, fainthearted, timid, faint-hearted, faintness, faintheartedness, dimness, feigning, dissembling, feign, sham, shammed, blaming, infidelity, frail, imperfect, fragile, somberness, sombreness, dispel, gloominess, glumness, gloomily, lugubriousness, mournfulness, uncheerfulness, glooming, gloomful,

uncomfortable, sick_person, diseased_person, sufferer, martyr, sake, woe, troubled, refugees, sulkily, sullen, morose, glum, lowering, darkened, tearfully, tears, lacrimation, lachrymation, tearing, teardrop, tragic, tragical, plight, tragedy, depress, deject, cast_down, embarrassment, Die, disappoints, disturb, disturbed, discompose, discomposed, disturbed, confuse, disconcerting, upsetting, disturbance, inconsolable, disconsolate, unconsolable, dark, dismal, drab, drear, dreary, rainy, dispiriting, desolately, disconsolately, grief-stricken, loneliness, demoralized, demoralised, disheartened, discouraging, depriving, deterring, disgrace, ignominy, dishonor, dishonour, dishonored, degrade, demean, put_down, dishearten, demoralizing, demoralising, disheartening, destructive, morale, dismally, drearily, dreadfully, dreadful, discouragement, disheartenment, dismaying, fired, firing, ignore, dismissed, dismissed, terminated, get_down, dismay, dispirit, demoralize, demoralise, spirits, downhearted, demoralizes, dejected, dejectedly, in_low_spirits, melancholy, depression, Depress, gloomy, grim, dispirited, downcast, down_in_the_mouth, low-spirited, despondency, darkening, mood, pessimistic, despondent, crisis, Depression, severe, impoverish, deprived, disadvantaged, deprivation, unhappy, deprived, charity, despair, desperation, moaned, despairing, desperate, plea, helpless, despairingly, despondently, desperately, heartsick, detachment, isolated, detached, devastated, wept, frustration, doomed, unredeemed, unsaved, danger, poor, damnably, cursedly, damnable, famine, shortage, insufficiency, paucity, insufficient, dying, departure, painful, struggle, Death, plague-bound, killing, deaths, dead_person, dead_soul, deceased_person, deceased, decedent, departed, gone, dishonest, licking, unsuccessful, ending, suffered, difficulties, enemies, shortcoming, failing, deficiency, fault, blemish, outcry, crier, crying, cries, fit, weeping, decease, perish, exit, pass_away, expire, pass, kick_the_bucket, dead, choke, croak, cancer, perished, hate, injustice, impairment, wake, discredited, unjustly, disrepute, cursed, blasted, blame, blamed, blessed, darned, deuced, goddam, goddamn, goddamned, complaints, discontent, displeasure, complains, accusation, complained, complaining, complaintive, dissatisfaction, resentment, complainingly, complaints, disapprobation, condemnation, disapproval, wrong, morally, condemning, evil, disapproved, guilty, criminal, condemn, reprobate, decry, objurgate, excoriate, disapproval, doom, punishable, reprehensible, wrongdoing, confessed, misdeeds, faults, offense, penitent, sinfulness, penance, subdue, conquer, bury, buried, entomb, lay_to_rest, grave, tomb, sink, deeply, unpleasant, memories, calamitous, disastrous, fatal, fateful, unfortunate, dire, crashed, defeat, battle, error, captive, confined, imprisoned, jailed, captivity, unsympathetic, carelessness, neglect, deprive, fraud, uncheerful, inadequacy, childless, rejected, illness, cheated, deceive, deceived, blackness, least, split_up, fall_apart, come_apart, separated, broke, fell, inoperable, ineffective, died, winter, tear, sobs, humbled, humiliated, subdued, violated, upset, disarray, impoverished, injure, bruised, wound, afflict, yelled, insults, dispose, awkward, inconvenience, banished, doom, gloom, bankruptcy, failure, inability, discharge, debts, insolvent, beat, blow, crush, trounce, vanquish, conflict, defeated, beating, thrashing, difficulty, quiver, bewilder, flummox, beaten, trodden, worn, beg, implore, pray, supplication, entreaty, urgently, beggary, begging, mendicancy, penniless, deplore, lament, bewail, bemoan, lamented, besiege, beleaguer, hem_in, besieged, distressed, worried, discouraged, harass, surrounded, betray, betrayed, treachery, Judas, disappoint, failed, crisis, cheat_on, cheat, unfaithful, cheats, denounce, ungainly, inapt, inept, ill-chosen, embarrassing, unenviable, ill_at_ease, uneasy, uncomfortable, unsure, negative, discomfort, restless, regretful, sorry, regret, sorrow, loss, mistakes, unsound, diseased, wickedness, bad, badly, injury, hurts, banish, ban, ostracize, ostracise, shun, cast_out, expel, punishment, relegate, expel, improper, mistreat, maltreat, ill-use, ill-treat, badly, abuses, misused, foul, abusive, misery, stupidity, ache, suffer, pain, hurting, lonesome, dejection, loneliest, solitude, devoid, ill, unwell, trouble, sick, indisposed, alienation, disaffection, estrangement, alienated, alienating, unfriendly, anguish, pains, miserable, anguished, tormented, tortured, humiliation, ashamed, shame, guilt, embarrassment, remorse, lied, torn, unhappiness, betray, intrigues, perfidy, perfidiousness, betrayal, uncomfortably, discontentment, discontent, infelicitous, unlucky, luckless, ill-fated, ill-omened, ill-starred, disturbing, upsets, howl, self-pity, whimper, wailed, drooping, whining, snivel, squeak, screech, creak, screak, skreak, high-pitched, screeching, creaked, whiningly, shrivel, shrivel_up, shrink, wither, shriveled, fade, fading, shrivelled, shrunken, withered, wrinkled, dried-up, woefulness, maimed, Agony, chabby, ailed, alack, Alienation, Anguish, befallen, befell, begged, bemoaned, besieg, bewailed, bewailing, bittersweet , carelessli, chattered, collaps, confesses, conquered, crestfallen , croaked, crying , damag, deceiv, Defeat, dejected , dejecting, Dejection, delicate , depressed , depressing , depressive , desolate , Despair, despairingli, despatched, desperate , despondent , Detachment, disagreeable , disappointed , disappointing , Disappointment, disapproving , disarrang, discomposing, disconsolate , discontented , discouraged , disgrac, disgraceful , disheartened , Dismay, dismayed , dismaying , dispirited , displayed, Displeasure, Disrespect, dissatisfi, distant , distraught , downcast , downhearted , Embarrassment, emot, fainted, fainting, forgiven, forlorn , forsook, Gloom, gloomy , Grief, hearted, heart-rending, Homesickness, Hopelessness, Humiliation, Hurt, hurted, implored, lamentations, livid , Loneliness, longed, Longing, longings, meagr, distress, afflicted, sad, abandon, wantonness, lacking, restraint, reckless, inhibition, worry, forsake, abandoned, Abandon, drowning, victim, dead, vacate, forsake, desolate, desert, lurch, deserted, derelict, forsaken, weed-grown, sadness, abstention, abstaining, maltreatment, ill-treatment, ill-usage, abuse, cruel, inhumane, insult, revilement, contumely, vilification, rude, offend, hurt, death, smite, suffering, plague, stricken, grievously, affected, disease, impaired, elderly, older, depressing, aging, agony, torment, torture, suffering, torments, damned,

excruciation, depopulate, epidemic, depopulated, lay_waste_to, waste, devastate, ravage, scourge, extensive, destruction, ruin, enemy, invasion, bare, barren, bleak, stark, treeless, crushed, grief, depressed, wail, alone, lone, lonely, solitary, ailed, alack, melancholi, Melancholy, melancholy , miserabili, miserable , miseri, misfortun, moaning, Pain, pitied, Pitti, pitty, Pity, pleasureless, Regret, regretful , regretted, Repentance, repented, repents, saddening, Shame, shameful , sighs, sobbed, sombre , sorrowed, sorrowful , sorrowing , Submission, sufferers, Suffering, sulkili, sulky , thumped, unhappi, unhappy , unhealthy , unlucki, weak , weaker, weakhearted, woes, Alienation, Anguish, anguished, ashamed, avail, awkward, bad, banish, banished, bankruptcy, beat, beaten, befall, befallen, befell, begged, begging, bemoan, bemoaned, besieg, besieged, betrayed, bewail, bewailed, bewailing, bittersweet , blackness, blister, broke, broken, bruise, bruised, burden, bury, calamitous, captive, careless, carelessli, carelessly, carelessness, chattered, chatterer, chattering, cheated, cheating, cheerless, childish, childless, collaps, collapsed, complain, complained, complaining, complainingly, condemnation, condemned, confess, confesses, confession, conquered, console, consoled, consoling, continence, crestfallen , cried, cries, croaked, crumbling, cry, crying, crying , damag, damage, damaged, damned, dearth, death, deceased, deceiv, deceive, deceived, Defeat, defect, dejected, dejected , dejectedly, dejecting, Dejection, delicate , delicately, depressed, depressed , depressing, depressing , Depression, depressive , deprived, desolate, desolate , Despair, despairing, despairingli, despairingly, despatched, desperate , desperately, despondent, despondent , Detachment, devastated, die, disagreeable , disappointed , disappointing , Disappointment, disapproving , disarrang, disarranged, discomposed, discomposing, disconcerting, disconsolate, disconsolate , disconsolately, discontented , discouraged, discouraged , discouraging, disgrac, disgrace, disgraceful , disheartened , disheartening, dismal, dismally, Dismay, dismayed, dismayed , dismaying, dismaying , dismissal, dismissed, dispirited, dispirited , dispiriting, displayed, displeased, Displeasure, disquieted, disquieting, disquietude, disreputable, Disrespect, dissatisfi, distant , distraught , dolorous, downcast, downcast , downhearted, downhearted , downtime, dreary, drown, drowned, drowning, dumb, dwell, dwelling, dying, elderly, embarrassing, Embarrassment, emot, emotion, exhaust, exhausted, failed, faint, fainted, fainthearted, fainting, faintness, feigning, fetter, fetters, forgave, forgive, forgiven, forlorn , forsake, forsaken, forsook, frail, funereal, gilt, Gloom, gloomily, gloominess, gloomy, gloomy , Grief, grieved, grieves, grievous, grim, hard, heartbreaking, heartbroken, hearted, heartfelt, heart-rending, heartsick, homesick, Homesickness, hopeless, hopelessly, Hopelessness, Humiliation, Hurt, hurted, hurting, hurts, ill, ill-humored, illness, implored, indolent, innocent, joyless, lament, lamentable, lamentations, lamented, lamenting, livid , Loneliness, lonely, lonesome, longed, Longing, longings, meagr, meagre, mediocrity, melancholi, melancholic, Melancholy, melancholy , miser, miserabili, miserable, miserable , miserably, miseri, misery, misfortun, misfortune, misfortunes, moaned, moaning, mourn, mourned, mourners, mournful, mournfully, mourning, old, Pain, painful, painfully, pains, pale, pessimistic, piteous, piteously, pitiable, pitied, pitiful, pitiless, Pitti, pitty, Pity, pleasureless, poor, poorly, punish, punished, punishment, rape, refuse, Regret, regretful , regrets, regrettable, regretted, rejection, Repentance, repentant, repented, repents, ridiculed, ridiculous, ridiculously, sad, saddening, sadly, sank, scrape, Shame, shameful , sick, sicken, sickened, sickness, sigh, sighed, sighing, sighs, sin, sob, sobbed, sobbing, sobs, solemn, somber, sombre , sore, sorely, sorrow, sorrowed, sorrowful, sorrowful , sorrowfully, sorrowfulness, sorrowing , sorrows, sorry, spiritless, squeamish, Submission, submit, suffer, suffered, sufferers, Suffering, sulkili, sulkily, sulky, sulky , sullen, tearful, tearfully, tearing, tears, terrible, thump, thumped, thumping, tragic, tragical, treacherous, treachery, uncomfortably, unhappi, unhappily, unhappiness, unhappy, unhappy , unhealthy , unlucki, unlucky, unsatisfactory, unwell, upset, upsetting, wail, wailed, wailing, weak , weaker, weakhearted, weep, weeping, wept, whining, wither, withered, woeful, woes, wounded

Happy words

abundance, accomplish, accomplished, achieve, achieved, admirably, admire, admires, admiring, cheerfulness, smiles, innocence, childlike, gaily, harmony, harmoniousness, hurrah, hooray, victory, merit, jollity, jolliness, joviality, jolly, jovial, humor, jolly, kidded, jocund, gleefully, joyously, joyfully, agreeable, conciliate, appease, soothingly, soothing, giggles, giggle, giggled, gladly, gladfulness, gladsomeness, gladsome, hilarity, mirth, mirthfulness, glee, gleefulness, glorious, deserving, magnificent, grandeur, gratify, gratified, gratifying, pleasurable, praise, smile, felicitous, happily, merrily, mirthfully, gayly, blithely, jubilantly, harmony, harmoniously, joy, green, majestic, handsome, noble, admiringly, Adoration, adorned, adornment, agreeable, amiable, amicably, applauded, appreciated, Appreciation, approv, approval, approve, Attraction, babies, baby, bargain, bargained, bashful, beauteous, beauties, beautiful, beautifully, beloved, bemused, beneficial, benefit, benevolence, benign, benignant, bless, blessed, blessing, blest, Bliss, blissful, blissfully, blossomed, blue, bluish, blushed, blushing, boast, boasted, boastfully, brave, bravely, bravery, bravest, bright, brighten, brightened, brighter, brightly, brillianc, brilliancy, bringer, bringing, brink, brisker, business, calm, calmer, calmly, calmness, Care, cared, carefree, cares, caress, caressed,

caressing, Caring, Caring/Care, carnations, celebr, celebrate, celebratory, charm, charmant, charmed, charming, cheer, cheerful, cheerfully, Cheerfulness, cheery, cherish, cherished, children, chipper, chuckl, chuckle, chuckled, chuckling, civil, civilli, civilly, clap, clapping, comedies, comedy, comely, comfortable, comforted, comforting, comical, Compassion, compassionate, content, contented, contentious, Contentment, curtsey, dapper, darling, dear, dearer, dearest, dearli, dearly, delicate, delicious, Delight, delighted, delightful, delightfully, delights, desirable, Desire, desired, doting, eager, eagerly, Eagerness, earnestly, Ecstasy, ecstatic, elated, Elation, elegant, embrace, embraced, embracing, embrasure, enjoy, enjoyed, enjoying, Enjoyment, enjoyments, Enthusiasm, enthusiast, entice, enticing, esteems, Euphoria, euphoric, exalted, excel, excellence, excellent, exhilarated, exhilarating, Exhilaration, extraordinari, fantastic, favorable, favoring, favorite, favourably, favoured, felicita, felicity, festivities, fine, fond, Fondness, fortuitous, fortunate, fortunately, fortune, frolic, frolick, frolicsom, frolicsome, fun, funny, gaily, genial, genteel, gentle, gentleness, giddi, giddy, giggled, gladden, gladdened, gladly, Gladness, gladsome, Glee, gleeful, glorious, good, goodli, goodly, goodness, grateful, gratified, gratifying, gratitude, grinning, happi, happier, happiest, happili, happily, happy, harmonize, harmony, heavenly, hilarious, honest, honorable, hurrah, hurri, innocent, jollili, jollily, Jolliness, jolly, jovial, Joviality, Joy, joyful, joyful , joyfully, joyous, joyously, joys, jubilant, Jubilation, kind, kindest, kindli, kindly, kindness, laugh, laughed, laughing, laughter, lightheart, lighthearted, lively, livest, lovabl, lovable, Love, loveable, loved, lovelier, loveliest, loveliness, lovely, loving, lovingli, lovingly, luck, lucki, luckiest, luckili, luckily, lucky, lucky , lust, magnificent, marvellous, marvellously, mellow, mercy, merri, merrili, merrily, merriment, merry, merrymak, merrymaking, mirth, mirthful, naked, new, nice, nonchalant, overjoy, overjoyed, peppy, perfect, perfection, perky, playful, playing, pleasant, pleasanter, pleased, pleases, pleasing, Pleasure, positive, prettier, pretty, prosperity, prosperous, reconcile, reconciled, redemption, rejoiced, rejoicing, Relief, relieved, respectable, rich, rosy, safe, satiated, Satisfaction, satisfactory, satisfi, satisfied, satisfied , satisfying, smile, smiled, smiles, smiling, smilingli, smilingly, smirking, smite, smitten, splendid , splendidly, splendors, splendour, succeeded, success, successful, sweet, sweetheart, sweeties, sweetly, sweetness, tenderness, terrific, thankful, thankfulness, unharmed, uninjured, upbeat, white, witty, wonderful, wonderfully, young, Zest , cared, cares, Caring, Care, carnations, celebr, charmant, Cheerfulness, chuckl, chuckled, chuckling, civilli, comedies, Compassion, Contentment, dearer, dearli, Delight, Desire, Eagerness, Ecstasy, Elation, Enjoyment, enjoyments, Enthusiasm, esteems, Euphoria, Exhilaration, extraordinari, favoring, favoured, felicita, Fondness, frolick, frolicsom, giddi, Gladness, Glee, goodli, happi, happili, hurri, jollili, jollily, Jolliness, Joviality, Joy, joyful , Jubilation, kindest, kindli, lightheart, livest, lovabl, Love, lovelier, loveliest, lovingli, lucki, luckiest, luckili, lucky , merri, merrili, merrymak, pleasanter, Pleasure, prettier, rejoiced, Relief, Satisfaction, satisfi, satisfied , smilingli, smirking, splendid , splendors, sweeties, zest , happy, smiling, young, support, pleasant, abundance, copiousness, teemingness, accomplish, fulfill, fulfil, achieve, attain, reach, gain, achieved, accomplished, realized, realised, successfully, ambition, overcame, effected, established, settled, securely, unconditionally, admirably, laudably, praiseworthily, commendable, admirable, children, admire, admiration, cheerfulness, cheer, sunniness, sunshine, cheerful, dispelling, spirits, better, smile, grin, grinning, pleasure, amusement, beauteous, poetic, beautiful, sweetheart, beaut, kind, roses, beauties, delighting, exciting, enjoyable, beautifully, attractively, beloved, dearest, honey, endearment, darling, dearly, loved, bemuse, smiled, innocence, ingenuousness, naturalness, innocent, purity, pureness, sinlessness, whiteness, childlike, childly, child, charm, wide-eyed, round-eyed, dewy-eyed, simple, simplicity, wonder, joyousness, joyfulness, happiness, delight, rejoice, gladden, glad, recreational, blue, love, majestic, beauty, serene, composure, gallant, lofty, proud, displaying, dignity, nobility, regal, royal, supreme, golden, splendor, fine-looking, good-looking, better-looking, handsome, well-favored, well-favoured, pleasing, appearance, especially, ideals, pretty, extraordinarily, handsome, bighearted, bounteous, bountiful, loving, imposing, stately, impressive, worship, adoration, admiring, grace, embellish, beautify, attractive, beautiful, flowers, adorned, interest, relieve, plainness, decorating, colorful, interesting, agreeable, amiable, good-humored, good-humoured, disposed, please, affable, cordial, genial, diffusing, warmth, friendliness, amicably, amicable, blessing, sensible, acceptance, satisfactory, indulge, treat, excessive, indulgence, laughed, benevolence, benefaction, kindness, benign, benignant, pleasant, gracious, courtesy, prosperity, blest, favored, fortunate, enjoying, bliss, desirable, enjoy, blessings, peace, blissfulness, cloud_nine, blissful, completely, contented, lovers, ignorance, blissfully, unaware, danger, bloom, blossom, flower, produce, yield, cherry, bloomed, blossom_out, promising, blueness, blueing, whiten, fearless, hardy, intrepid, unfearing, gladness, hopeful, brilliantly, brightness, lively, merry, snappy, zippy, refreshful, calm, excitement, joys, happy-go-lucky, cheerfully, caress, affectionate, lightly, endearing, caressed, fondled, caressing, cuddling, fondling, hugging, kissing, petting, smooching, snuggling, foreplay, lovingness, empathy, friends, carnation, celebrate, holidays, appeal, appealingness, pleases, stimulates, enamour, fascinate, bewitch, entrance, enchant, attract, enamored, cheered, cheer_up, jolly_along, jolly_up, happier, cheers, upbeat, pleasantly, optimistic, cheery, pleasant, cherished, precious, treasured, affection, cheerful, lively, self-confident, chipper, gay, optimist, chuckle, laugh, laugh_softly, polite, clapping, friendly, applause, hand_clapping, comedy, humorous, clowning, funniness, comic, decorous, seemly, fair, comfortable, comfy, comfortable, prosperous, rich, wealthy, alleviate, comforted, comforting, cheering, satisfying,

freedom, amusing, comical, funny, laughable, mirthful, risible, provoking, laughter, stream, pranks, pratfalls, surprise, compassion, compassionateness, sympathy, compassionate, condole_with, sympathize, content, contentedness, relaxed, satisfy, satisfied, satisfaction, contentious, curtsy, curtsey, respectful, Darling, heartfelt, affectionately, exquisitely, fine, butterfly, delightful, delicious, pleasant-tasting, scrumptious, yummy, extremely, pleased, delightfully, positive, ecstasy, elated, ecstatic, enraptured, rapturous, rapt, rhapsodic, elate, lift_up, uplift, pick_up, intoxicate, optimism, Music, exultantly, joyful, winner, excited, gleeful, joyful, jubilant, high-spirited, heart, elation, exhilarating, psychological, optimism, high_spirits, lightness, elegant, refined, tasteful, graceful, suggesting, embrace, embracing, embracement, clasping, affection, hug, fondness, Hug, embraced, hugged, relish, savor, savour, enjoyment, relished, fame, glory, privileges, enjoyed, exuberance, ebullience, overflowing, enthusiast, partisan, partizan, ardent, enthusiastic, euphoria, euphory, euphoric, exaggerated, praise, glorify, honor, thrilled, gladdened, exhilarated, exhilarated, elating, exhilaration, fantastic, fantastical, grotesque, ludicrously, marvelous, marvellous, terrific, tremendous, wonderful, wondrous, great, fantastic!, fanciful, lucky, well-disposed, felicity, felicitousness, festivity, joyous, diversion, hunky-dory, allright, lovesome, affectionateness, tenderness, warmness, warmheartedness, welcome, fortuitous, fortuitously, luckily, frolicsome, frolicky, frolicking, fun, merriment, playfulness, wit, excellently, magnificently, splendidly, famously, gorgeously, magnificence, brilliance, grandness, successful, achiever, succeeder, successes, winners, sweetness, sugariness, angelic, angelical, sweetly, sweetie, truelove, well-liked, pleasingly, pleasantness, tenderheartedness, gratefulness, appreciativeness, unharmed, unhurt, unscathed, uninjured, benevolent, lips, livid, conception, facetious, satirical, eloquent, Young, youthful, youth, vigorous, admires, smiles, Adoration, applauded, Appreciation, approv, Attraction, babies, Bliss, overnice, elegantly, overjoy, overjoyed, peppy, spirited, buoyant, chirpy, perky, liveliness, lightheartedness, pleasance, gratification, positive_degree, confident, positive, positivist, positivistic, positivism, positively_charged, profitable, exuberate, exult, jubilate, fertile, rose-cheeked, rosy-cheeked, fulfilled, fulfillment, fulfilling, hearty, Santa, Claus, smilingly, smiles, smirk, affected, conscience-smitten, awe-struck, infatuated, in_love, exultant, exulting, prideful, rejoicing, triumphal, triumphant, triumph, exultation, jubilance, jubilancy, jubilation, forgiving, benevolent, sympathetic, good-hearted, openhearted, large-hearted, motivated, generosity, warmhearted, laughing, jest, laughed, mirth, laughing, lighthearted, happiest, lovable, loveable, lovely, adorable, comeliness, fairness, loveliness, fondly, lovingly, fondness, luxuria, tall, belief, wonderfully, wondrously, superbly, toppingly, marvellously, terrifically, marvelously, gaiety, merrymaking, jollification

Angry words

abash, abashed, abhorred, abhorrent, abid, abomin, abominable, abominably, aggressive, absurd, aggressiveness, growl, growling, crudeness, sanguineous, slaughterous, ravaging, ravage, pillaging, destructive, raid, wartimes, lay_waste_to, devastate, desolate, scourge, rebellious, resisting, temperamentally, rebelling, rebuff, slight, discourteous, snub, repulse, fight_off, drive_back, attacker, onslaught, blood-red, barbaric, reproached, repugn, dispute, contention, repugnance, repulsion, ill-mannered, bad-mannered, unmannered, unmannerly, ill-bred, uncivil, butcherly, barbarian, uncivilized, brutally, fiercely, harshly, savaged, scolder, chiding, scolding, tongue-lashing, rebuking, shout_out, yell, cry_out, outcry, shouting, yelling, smack, slapped, snapped, snarled, snarling, sneer, leer, scorn, scornful, sneered, bicker, bickering, spat, squabble, cattiness, bitchiness, nastiness, spiteful, offensive, enemy, biliousness, irritability, peevishness, pettishness, snappishness, grumble, rumble, grumbled, rude, growler, complaint, grouch, scold, scolded, malevolent, malignant, meanspirited, mercilessly, pitilessly, unmercifully, remorselessly, merciless, warlike, warfare, militaristic, imbued, militarism, armed_forces, armed_services, military_machine, war_machine, derisive, scoffingly, exasperatingly, blot_out, appal, appall, offense, offence, attacking, violative, violating, offensively, offensively, ravish, assault, quarrelsome, quarreling, arguing, furor, furore, violently, violent, Rage, torment, bedevil, taunt, reproof, reprimand, chide, berate, chew_out, chew_up, have_words, lambaste, grumbles, crankiness, grumpiness, grunts, grunting, grunted, hate, hatred, dislike, indignation, outrage, insult, revilement, contumely, vilification, yelled, insults, affront, diss, rudely, insulted, betrayed, dissed, irate, ireful, protesters, inflame, abhorrence, abomination, detestation, execration, loathing, odium, madden, maddening, lunatic, madman, maniac, lunacy, insaneness, rabies, viral, rabid, craziness, rash, rabidity, rabidness, maliciousness, spitefulness, venom, suffer, malevolence, malevolency, threatening, malicious, badmouth, malign, harmful, detest, detests, ingratitude, mean, malice, hatted, hiss, siss, sizz, hissing, displeasure, ill_will, unfriendly, howl, howling, yawl, impatient, impatiently, impatience, impertinence, insolence, impertinent, indignant, incensed, outraged, umbrageous, unjust, denial, indignantly, protested, indignantly, aghast, anger, angered, clenched_fist, hitting, flare, flame, flared, flames, aflame, inflammatory, attacked, reddening, redden, red, red, reddened, red-faced, fluster, flustered, upset, flustered, hot_and_bothered, perturbed, rattled, thrown, confusion;, forbid, prohibit, forbids, frustration, fume, mad, angry, fuming, furiously, stormy, blowing, screamed, tormentors, hell, frenzy, hysteria, Fury, angered ,

bristle, agitated, bothered, discomposure, spite, aggravate, exacerbate, exasperate, aggravates, infuriate, incense, exasperated, cheesed_off, browned_off, aggravation, exasperation, ferocious, fighting, battle, ferocity, fierceness, furiousness, vehemence, wildness, turbulent, storm, fever, febrility, febricity, pyrexia, feverishness, agitation, feverous, feverishly, demon, ogre, daemon, daimon, evil, violently, fervid, passions, ardent, fiery, impassioned, temper, conflict, combat, fighting, fight_back, fight_down, defend, resist, agitate, exert, vile, despicable, firebrand, instigator, provoker, inciter, instigant, stiff, horrifying, horrible, ugliness, turmoil, upheaval, crabby, cross, grouchy, grumpy, bad-tempered, ill-tempered, crank, cranky, petulant, testy, beatings, slap, kicks, cruelty, cruelness, harshness, tension, blame, blamed, infernal, derision, deriding, despise, despises, scorns, despised, detested, hated, scorned, devour, destroy, Fire, devoured, ardently, fierce, disgust, gross_out, revolt, fed_up, sick_of, tired_of, loathly, loathsome, repellent, repellant, repelling, revolting, offensive, quarrel, argufy, altercate, disagreement, quarreled, enrage, rage, enviable, enviously, jealous, nefariousness, vileness, ungrateful, disagreeable, retribution, Vengeance, revenge, vindictiveness, vengefulness, venomous, vent, erupt, irritating, bewildering, clash, volatiles, rioters, confrontation, warmongering, wrathful, wretchedness, intolerable, yell, abhorred, abid, abomin, angered , angri, angrili, Annoyance, Annoyed, antagonistic , antagonized, antisocial , Aversion, awful , bah, begrudged, bitter , bitterli, Bitterness, choleric , screaming, bristled, Contempt, contemptible , cruelli, cruelti, Cruelty, denying, despis, disgusting , disputing, embittered , enrag, enraged , envi, envied, Envy, exacerbated, Exasperation, exceedingli, fedup, fed-up, fed-up , feroci, Ferocity, fiends, fierc, fiesco, flutterment, forbade, Frustration, furi, furiou, furious , gorged, groaning, groans, Grouchiness, growled, growls, grumbl, grumpi, Grumpiness, Hate, hating, hatr, hissed, Hostility, howled, illtempered, indignant , indignantli, Inflamed , Insult, ireful , Irritation, Jealousy, Loathing, loathsome , malici, merci, mercilessli, mockingli, nasti, Neglect, offended , offenders, Outrage, outraged , outrageous , pissed-off , prowled, quarrelled, quarrelling, rabbity, rancorous , rebelli, reeling, reproaches, repuls, repulsed, Repulsion, Resentment, revenged, revolut, Revulsion, scalding, Scorn, scuffling, snares, Spite, spiteful , tempers, uglier, ugliest, unsatisfi, vengeanc, Vengefulness, vile , vindictive , virulent , warring , wars, Wrath, wrathful , uprise, stand_up, bristles, bristling, react, offended, bristly, spiny, thorny, cruelly, beastly, bestial, brutish, brutal, choleric, irascible, hotheaded, hot-tempered, quick-tempered, short-tempered, outburst, clinch, clench, clasp, clutch, clutches, railing, clenched, clinched, inhuman, insensate, killing, destruction, blow, disputatious, disputative, litigious, abrasive, contempt, disdain, scorn, disrespect, disrespectful, contemptuous, contemptuously, disdainfully, scornfully, thresh, thresh_about, thrash, thrash_about, slash, violently, feverish, thrashed, angri, unruly, irritable, biting, sting, insect_bite, stinger, bitterness, cynicism, dispute, divorce, difficult, acid, acrid, blistering, sulphurous, virulent, vitriolic, corrosive, piercing, piercingly, bitterly, bitingly, bitterness, resentful, disappointed, gall, rancor, rancour, deep, ill-will, acrimony, acerbity, jaundice, tartness, thorniness, rough, bloodthirsty, bloody-minded, violence, bloodshed, tyrants, yells, fury, flaming, damn, all-fired, aggressive, antagonistic, antipathetic, antipathetical, antagonism, antipathy, animosity, brusqueness, antagonize, antagonise, argumentative, arrogance, rapacity, reprehensible, loath, distaste, dislike, rage, wickedness, hurts, gravely, seriously, severe, serious, betray, evilly, wickedly, disobedient, naughty, resentment, severity, severeness, unfriendly, barked, barking, envy, begrudge, envious, battleful, bellicose, combative, combatant, battler, fighter, scrapper, fights, militant, war-ridden, warring, bilious, biliary, bile, angrili, angrily, angriness, wrath, ira, angered, angers, enraged, furious, infuriated, maddened, beating, angrily, raging, tempestuous, violent, inflamed, annoy, irritate, rile, vex, chafe, devil, annoyance, disturb,, irritations, buzzing, irritates, irritation, vexation, botheration, irritated, annoyed, annoying, aggravator, unpleasant, exasperating, pain, infliction, pain_in_the_neck, pain_in_the_ass, peeved, pissed, pissed_off, riled, roiled, steamed, impatience, teasing, snootiness, harassed, harried, pestered, vexed, angry, angry, tantrums, abhor, loathe, abominate, execrate, repugnant, abhors, abhorrent, detestable, obscene, repulsive, offensive, massacre, abominable, execrable, odious, vices, crimes, atrocious, awful, dreadful, unspeakable, bad, displeasing, detestably, repulsively, abominably, odiously, hateful, terribly, atrociously, awfully, abysmally, rottenly, aggressive, belligerent, irrational, preposterous, aggressiveness, aggression, hostility, belligerence, hostile, aghast, appalled, anger, choler, ire, annoy, Annoyance, Annoyed, annoys, antagonistic, antagonistic , antagonized, antisocial , argumentative, arrogance, avarice, averse, Aversion, awful , bad, badly, badness, bah, barking, begrudged, bellicose, belligerent, bilious, bite, bitter, bitter , bitterli, bitterly, Bitterness, blinded, bloodthirsty, bloody, bothered, bristle, bristled, bristling, brute, chauvinist, chauvinistic, choleric, choleric , clench, clenched, cold , colder, coldly, coldness, combative, come down, conceited, condescend, Contempt, contemptible , contemptuously, convulsed, convulsion, convulsively, crabby, cranky, crazy, criticize, cruel, cruelli, cruelly, cruelti, Cruelty, danger, darned, darning, deals, demand, demanded, deny, denying, deplorable, derision, despis, despise, despised, detestable, devour, devoured, devouring, disgust, disgusted, disgusting, disgusting , dismiss, disputed, disputing, embittered , engorged, enrag, enraged, enraged , envi, enviable, envied, envious, Envy, exacerbated, exasperated, Exasperation, exceed, exceedingli, exceedingly, extreme, extremely, fedup, fed-up, fed-up , feroci, ferocious, Ferocity, fever, feverish, feverishly, fiend, fiends, fierc, fierce, fiery, fiesco, fight, fighting, filthy, firebrand, firelight, firestone, firing, fist, flamed, flames, flush, flushed, flustered, fluttered, flutterment, forbade, fought,

Frustration, fumed, fuming, furi, furiou, furious, furious , furiously, Fury, ghastly, gore, gorged, groan, groaned, groaning, groans, Grouchiness, grouchy, growl, growled, growler, growls, grumbl, grumble, grumbled, grumbling, grumpi, Grumpiness, grumpy, grunt, grunted, grunting, Hate, hateful, hating, hatr, hatred, hissed, hostile, Hostility, howl, howled, howling, howls, illtempered, ill-tempered, impatience, impatient, impatiently, impertinence, incensed, indignant, indignant , indignantli, indignantly, indignation, Inflamed , infuriated, insolence, Insult, insulted, irascible, irate, ireful , irritable, irritated, Irritation, Jealousy, Loathing, loathsome , mad, maddened, madman, madness, malice, malici, malicious, malign, mean, merci, mercilessli, mercilessly, militant, militaristic, military, mock, mocked, mocking, mockingli, mockingly, nasti, nasty, Neglect, nonsense, obliterated, obstinate, obstinately, odious, offended, offended , offenders, offensive, offensively, Outrage, outraged, outraged , outrageous , petulant, pissed-off , provoked, provoking, prowled, quarrel, quarreling, quarrelled, quarrelling, quarrelsome, quick-tempered, quiver, rabbity, rabid, rage, raging, rancid, rancorous , ravaging, ravelling, rebelli, rebellious, rebuff, recalled, reckoning, red, reeling, reproached, reproaches, repugn, repugnance, repugnant, repuls, repulse, repulsed, Repulsion, repulsive, resentful, Resentment, restrain, restrained, restraint, revenge, revenged, revolt, revolted, revolting, revolut, revolution, Revulsion, riled, roiled, rude, sanguinary, savage, savagely, scalding, scold, scolded, scolding, Scorn, scorned, scornfully, scoundrel, scuffling, shouted, shouting, shouts, shy, sick, slap, slapped, smacked, snare, snares, snarling, sneer, sneered, spat, Spite, spiteful , strangle, strangled, strikes, striking, temper, temperance, tempers, tempest, tempestuous, terrible, trample, trampled, trampling, ugli, uglier, ugliest, ugliness, ugly, ungrateful, ungratified, unpleasant, unruly, unsatisfi, unsatisfied, vengeanc, vengeance, Vengefulness, venom, venomous, vent, vermin, vex, vexation, vexatious, vexed, vile , villain, vindictive , violence, violent, violently, virulent , volatile, warmongering, warring , warrior, wars, wicked, Wrath, wrathful, wrathful , wrestled, wretch, wretched, wretchedness, yelled, yelling

Phrases afraid
weak knee, ill at ease, total darkness, ran out, ill at ease, ran away oh, heavens, dark forest, dark woods, very scary, dark cemetery, hunt down, track down, horror stricken, horror struck, put behind bars, in haste, very afraid

Phrases angry
very angry, very upset

Phrases happy
walking on air, floating on air, as happy as the day is long, as pleased as punch, as high as a kite, on top of the world, thrilled to bits, as happy as a pig in muck, in high spirits, tickled pink, tickled to death, chuckled, what a pretty, well off, she gave birth, have fun, had a good time, very happy, flying high, too happy, very pleased, felt happy

Phrases sad
very sad, quite alone, all in vain, cried to sleep, very melancholy

Phrases surprised
very surprised

Triggers emotion list
live, old, child, sever, bring, care, death, begrudg, stupid, long, livest, sharp, thought, ill, aros, lie, deep, exclaim, alon, desert, kindli, help, desir, beauti, come, wonder, enchant, merri, happi, music, play, lip, cheek, possibl, light, good, kind, gone, bright, shine, heart, rejoic, new, reach, blind, laugh, fine, torn, asham, forc, creep, hard, pass, delight, poor, magnific, handsom, great, rich, extrem, proud, contemptu, fell, success, fail, succeed, luck, discourag, color, black, white, red, golden, pale, terror, courag, expect, clutch, frighten, taken, hearti, welcom, quiet, anxieti, bite, accomplish, pleas, merrili, fanci, terribl, growl,

fierc, joy, end, careless, gaili, promis, drive, envi, fortun, furi, terrifi, cri, die, beat, kick, better, nice, piti, weep, weari, sorrow, wick, witch, children, bewitch, brightli, pray, dear, wild, beast, tear, wolf, devour, young, wept, rush, tender, content, wilder, hunt, blast, bark, shout, anxiou, beg, fear, rough, pretti, catch, dark, peac, chase, surround, wound, dread, hurt, slight, kill, forsaken, grief, bound, harm, love, happili, cherish, hatr, rose, misfortun, ugli, eye, grumbl, comfort, cold, fire, deadli, suffoc, glad, quickli, cradl, awak, arm, dare, appear, restrain, grace, rosi, evil, guilti, miser, burnt, torment, sole, hang, celebr, bitterli, treat, enviou, lovabl, repuls, winter, blow, green, warm, hastili, thank, arrog, yield, naughti, uglier, angrili, horror, enrag, injuri, compass, stream, violent, suspect, lose, babi, sword, spirit, deserv, wretch, honor, consequ, civil, pain, solitud, trembl, surpris, earnestli, plead, piteous, natur, sweetli, arous, nois, scream, hurri, struggl, rage, trampl, gain, victori, adversari, terrif, despatch, fight, astonish, fate, dead, flame, befal, mysteri, iron, unknown, admir, sank, depth, bluish, excit, curios, exceedingli, delic, fair, preciou, suddenli, deliver, prison, hasten, friendli, kiss, overflow, youth, embrac, lack, friend, stranger, courtesi, agreeabl, faint, caus, awaken, repugn, favor, reveng, mean, punish, pride, disquietud, broken, loudli, discharg, monster, depriv, magician, confin, state, fulfil, marvel, scarc, blue, lucki, support, badli, burden, bless, war, enemi, battl, danger, ruin, conquer, treasur, strang, buri, grave, charm, solemn, splendour, befel, attack, horrifi, corps, royal, snake, unhappi, wondrou, power, stir, troubl, shame, prais, wicked, secret, godless, merci, murder, pierc, sad, sigh, dispos, afraid, warn, prudent, departur, broke, poison, rob, honest, perish, safe, hope, steal, bad, angri, refus, sink, storm, tall, thrown, robber, innoc, passion, hell, devil, favour, cross, entranc, deal, difficult, attent, pure, pri, scold, upset, wither, disturb, curiou, complain, prosper, condit, tremend, eager, sin, creak, frantic, burn, furious, shake, drown, enjoy, heavenli, thief, blister, joyous, anger, sinner, sweet, angel, age, glori, bewar, miseri, dwell, amaz, imprison, fled, abund, confess, obstin, demand, dearli, smile, loss, trial, defend, condemn, repent, extinguish, forgiven, eagerli, anguish, lone, betray, witchcraft, mourn, rustl, abid, smitten, modest, dissatisfi, accus, suffer, treacheri, dumb, vigor, dearest, weaker, gladli, satisfi, mercilessli, howl, beaten, abomin, cellar, struck, alarm, choke, distress, lament, difficulti, unhappili, hiss, fret, alack, tap, depart, wail, lucif, crook, confus, rebelli, pleasantli, impati, temper, bargain, vex, vexat, abus, sake, giant, mere, perfect, horribl, tast, oppos, appeas, snap, stalk, strike, barricad, haunt, interest, serious, lover, affect, belief, dragon, comb, lower, loud, fond, especi, cheer, bloom, brave, tomb, cage, hit, settl, woe, sick, foul, sorri, plagu, hat, gay, flutter, overjoy, least, unawar, dim, worn, fit, curtsi, griev, gasp, nobl, threaten, buzz, flower, shock, forbid, dazzl, busi, belov, lick, unexpect, accident, dash, scald, wrong, chicken, glee, polit, uneasi, denounc, wake, moan, gloom, rabbl, spoil, chafe, bruis, rattl, prey, dearth, groan, bitter, glitter, forsak, reproach, crumbl, entic, malic, mockingli, shrivel, grate, lazi, bone, sadli, whip, smack, troublesom, luckili, butcher, refresh, slaughter, hate, denial, scrape, happiest, befallen, risk, nake, forbad, hasti, brink, poverti, sore, flash, seemli, armi, warrior, battlefield, hurrican, flee, mock, ridicul, armour, stab, queer, festiv, gloomi, fix, prettier, sudden, darl, fieri, meagr, stare, doom, conceit, spite, grisli, avail, unlucki, abash, robe, grand, dreari, sooth, hazard, separ, smite, forlorn, hug, collaps, bleed, harshli, sunshin, homesick, exert, neglect, grunt, blame, cowardli, ail, deepli, deceiv, pitiless, venom, despair, wretched, boast, bewail, reel, comic, funni, fun, chuckl, beauteou, loveli, bemoan, loveliest, sweetheart, bled, snare, fascin, holiday, insol, coward, respect, civilli, jollili, sulkili, reckon, furiou, fright, mourner, compassion, pleasur, dismiss, brighten, entreati, chain, mood, crush, cruel, villain, brute, mad, cruelti, bustl, nasti, filthi, nonsens, calm, dole, foam, lofti, shiver, stormi, thunder, quarrel, experi, esteem, hound, lurch, excel, melancholi, avaric, stiff, clap, bellow, gentl, burst, merriment, pleasant, shabbi, merrymak, despis, sneer, sensibl, seriou, produc, aw, childless, blossom, crash, overwhelm, mist, lightheart, chatter, delici, sob, bare, cheat, grin, wit, charg, remors, concern, croak, profit, vengeanc, simpl, brisker, piteou, plight, amus, influenc, robberi, carelessli, disgust, older, peril, majesti, graciou, deepest, wast, honey, resist, darker, disput, fault, shudder, mistak, darkest, grim, crouch, rare, feign, forgiv, dismal, ghost, scoundrel, savag, yell, vermin, hideou, hurrah, strangl, fiend, destroy, dismay, amic, unharm, distinguish, pant, brighter, scamper, spat, horrid, curs, vent, ambush, suspicion, dangl, deni, attract, braveri, cherri, militari, ravag, subdu, madman, gore, havoc, awkward, warlik, spellbound, scorn, tranc, famin, frolick, fame, helpless, chick, disreput, insult, growler, bloodi, sting, kid, gorg, rumbl, dearer, duski, sulki, plain, woefulli, dishearten, dislik, frolic, stricken, frolicsom, fought, childish, gloriou, sunni, bliss, fade, humbl, afflict, gayli, benefit, gladden, decor, kindest, reliev, posit, chariti, apparit, startl, rous, vice, realiz, error, crime, superstit, spectr, desol, recal, memori, distort, fantast, fever, shadow, spectral, exhaust, flush, gleam, clasp, trodden, childlik, injur, cautious, lightli, uninjur, blush, excitedli, sulphur, deceas, tortur, wanton, abandon, adorn, contempt, feverishli, anxious, joyless, besieg, crimin, unfortun, feroci, violenc, fist, wrestl, awe, elderli, appreci, ignor, complet, proverb, erupt, display, regret, disagre, relief, wearisom, damag, ugliest, sympath, sympathi, develop, harmon, raptur, exclam, bullet, annoy, rude, butterfli, extens, glisten, quiver, ambit, loveabl, highli, uncomfort, fashion, forsook, prowl, favorit, extraordinarili, teas, clench, thrash, bubbl, majest, moonlit, solitari, dapper, offend, treacher, complainingli, cautiou, strand, disobedi, ban, expel, complaint, slain, funer, incens, wrath, splendor, dispel, nobil, convuls, felic, darkli, restless, attain, splendidli, droop, critic, ardent, raini, freedom, consol, jocund, humor, rail, refin, comedi, applaud, worri,

effect, stingi, gall, absurdli, feverish, agoni, disconsol, confid, darest, implor, fondl, plaintiv, heartfelt, eleg, moral, captiv, darn, merit, genteel, paralys, gilt, maul, bah, desper, accept, ordeal, witti, sullen, demon, bristl, giddi, tingl, thump, indign, paler, fondli, calmli, whiten, exquisit, approv, impress, digniti, discont, mortifi, unsatisfi, anticip, ach, brillianc, plunder, squeak, happier, luckiest, extraordinari, revolut, unpleas, laughter, improp, goodli, vast, bewild, policeman, unbear, redden, enthusiast, retreat, triumphant, firebrand, exult, scandal, poetic, banish, puriti, steam, jest, polic, impuls, crazi, establish, tendenc, tragedi, intrigu, reject, absurd, gratifi, screamer, impos, defeat, disfigur, unnatur, miserabili, diseas, whine, rancid, chill, frill, defect, warmth, indignantli, affection, cruelli, wrinkl, calmer, snub, grumpi, burial, harmoni, gratitud, reluct, vandal, disast, fireston, armlet, timidli, debt, unwel, achiev, joyou, intox, gallant, hostil, thrill, satisfact, fetter, tempest, redempt, glorifi, clau, relish, termin, smilingli, hooray, battlement, argu, carnat, huddl, elegantli, frail, clash, slap, jovial, recreat, shriek, deris, unjust, barren, submit, unexpectedli, reconcil, agit, bankruptci, paralysi, lovingli, reverber, enviabl, fiesco, rave, ideal, graveyard, benevol, emot, obliter, grotesqu, disgrac, cemeteri, unquiet, unsettl, cowardic, stress, worship, damn, grandeur, shi, shockingli, slash, fog, wondrous, grimac, outrag, embrasur, malici, noisi, charmant, jolli, enthusiasm, eloqu, despot, mirth, homeless, harsh, ghastli, composur, coldli, fluster, secur, sham, bravest, condescend, colder, persecut, devast, diabol, admiringli, exceed, gorgeous, ordinari, brandish, gloomili, inconveni, harmless, excess, impertin, pitti, scuffl, perplex, sizz, unruli, nerv, disarrang, ravel, provok, triumph, poorli, suggest, forgav, strung, abrupt, squeal, curtsey, peevish, creepi, displeas, bristli, cordial, unsatisfactori, fume, odiou, hurriedli, scuttl, indol, stagger, apoplect, vindict, mystifi, snarl, nervous, scare, giggl, hopelessli, volatil, sweeti, dejectedli, deplor, offens, shyli, distrust, affabl, embarrass, firelight, mown, commot, stun, chirpi, affront, disappoint

Triggers environments list
day, sun, even, next_morn, days_tim, everyth, below, stabl, morn, may_heaven, heaven, next_day, door, bath-room, midnight, to-morrow_night, next_night, water, wed, second_morn, third_morn, winter, ground, wind, splendid_carriag, carriag, year, night, wild_beast, hut, air, fire, hollow_plac, floor, hall, same_day, hous, self-same_day, mighty_empir, life, great_ship, voyag, nightfal, second_night, way, road, tree, little_door, corner, little_cart, little_tre, little_spr, seat, great_forest, forbidden_door, desert, roots_wild_berri, autumn, snow_ic, countri, following_year, hard_ic, straight-way_rain, sunset, thy_work, country_folk, third_day, farmer, seat_thyself, forest, process, room, last_night, daybreak, branch, castl, white_lily_bloom, snow, spring, work, kitchen_window, garden, coach, feet, feast, court, hearth, large_hedg, old_palac, thorns_bush, that_very_day, kitchen, wall, prince_briar_ros, wedding_feast, danc, merry_day, little_mouse_cannot_ventur, upstair, good_even, first_plac, stove, dinner-tim, noon, front_door, moon, brushwood, full_moon, great_fir, hill, week, bridg, tale, villag, wild_man, treasure_gold, world, summer, wild_on, following_day, second_day, stately_k, all_treasur, all_day, setting_sun, gloomy_night, ninth_day, land, east, trade, whole_court, lili, all_court, wedding-feast, train, poor_lili, field_grov, east_wind, west_wind, south_wind, right_shor, heaven_hath, yesterday, leagu, stone, gate, shower, first_day, wild_boar, twilight, long_way, stream, chamber_door, today, window, cottag, outsid, daylight, stair, bed, all_work, good_dinn, brook, compani, handsome_coach, river, war, chamber, third_night, flight, weather, cask, christmas-tim, stables_cow-hous, sea, ship, south, sun_moon_ris, dreadful_storm, space, few_day, good_day, fine_even, castle_g, bank, next_morning_dumml, field, whole_mountain, winter_', all_earth, green_branch, chest, "in_cas, wave, whole_world, entire_villag, thy_lif, ventur, pail, sunday, large_compani, rock, cloud, treasur, "one_night, bench_i, whole_district, wild_dog, honest_farm, coast, all_world, school, days_night, trap-door, lake, fine_illuminated_castl, summer_', whole_town, earliest_dawn, special_dwel, kingdom, in_forest, all_kingdom, next_year, full_year, dinner, path, summer-tim, battl, washing-bowl, heavy_ston, very_way, cart, such_hous, grass-covered_rampart, great_red_hous, build, whole_life_drama, mound, church_steepl, rampart, beautiful_castl, almshouse_window, earth, narrow_lan, low_cellar, spot, spring_day, yard_gat, heavy_sea, journey, evening_bel, deep_wat, northern_region, beech-wood, past_lif, cloud-shadow, grave, many_night, whole_year, shore, bush_tre, field_flow, wild-flow, little_apple-tre, pure_white_curtain, despised_herb, despised_pl, apple-tree_branch, yard, church, same_even, cold_wind, broad_step, dance_mus, scene, place, all_town, citi, public_road, whole_c, casket, large_parti, year_', all_galleri, future_st, everything_beauti, shadow, heavy_show, rain, large_palm-tre, sunlight, little_ship, beautiful_weath, sun_', narrow_street, large_town, town, large_rock, fresh_spring_wat, large_crab-apple_tre, dark_night, great_majestic_sea, natur, woven_velvet_carpet, old_town, island, old_nunneri, deepest_plac, monks_meadow, tower, church_bel, all_other_church_bel, our_scen, great_mound, sand, mountain, convent, crop, green_hil, humble_vehicl, icy_wint, meadow, marsh, storm, altar, ferry-boat, altar_light, angry_wind, wreck, comfortable_room, street, little_window, long_lif, wooden_cottag, long_days_month, great_wav, event, new_moon, great_crowd, military_mus, aerial_ship, in_window-sil, happy_day, youthful_day, depth, violent_thunder-storm, country_peopl, lightn, large_water-drop, employ, churchyard, evergreen, neighboring_grav, public, drawing-room, societi, famili,

day_night, morning_twilight, funer, similar_day, poultry_yard, next_yard, weather-cock, passag, yard_cock, crow, high_road, following_morn, air_light, warm_sunshin, fresh_green_meadow, turf, more_wat, street_boy, white_wal, roof, on_wal, little_wooded_island, wild_swan, in_garden, little_room, to-day, old_hous, basement, interior, clay, ceil, high-road, all_window, arbour, different_lif, withered_oak_leaf, foreign_hothouse_pl, north, tavern, elfin_hil, large_hal, elf_hil, curtain, norway, sea_folk, tingling_frost, water-god, whole_day, in_window, theatr, great_citi, whole_night, ceremoni, long_train, warm_sun, fresh_air, soft_air, trunk, young_lif, many_young_tre, young_tre, winter_summ, dark-green_foliag, christma, great_stov, other_branch, glistening_blaz, folding_door, troop, little_birch-tre, sunshine_fresh_air, star, same_courtyard, tree_', sunshin, rain_descend, genealogical_tre, morning_even, offic, clean_curtain, water-bucket, capital_stori, wedding-day, little_dwel, summit, sacred_mountain, mount_parnassu, neighborhood, many_days_night, mountain-top, wild_thym, solitary_hut, shooting_star, evening_sunlight, eternal_friendship, whole_shop, evening_light, whole_street, greatest_treasur, burning_hous, real_sunshin, hollow_way, clayey_bank, splashing_rain, town_musician, life_limb, battle-field, rainbow, regiment, days_week, last_day, capital_piec, golden_treasur, copenhagen, shape, nearest_corn, accommod, shop, case, publish, first_publish, denmark, english_pir, greatest_wealth, falling_star, atmospher, dawn, obstacl, audienc, future_ev, russian_bath, police_offic, fredericksburg, avenu, every-day_lif, screaming_crow, vaudeville_., water-drop, bending_branch, half-opened_door, whole_affair, small_garden, pine-forest, waysid, delightful_itali, long_white_curtain, esteem, floodgat, bushel, washing_yond, o_kind_heaven, following_sunday, first_year, grave_clos, churchyard_wal, grave_fresh_roses_bloom, bare_floor, lane, little_lan, tan-yard, burdock_forest, whole_famili, clear_pure_air, plant, large_tre, harmoni, wonderful_pl, black_forest_snail, for_plant, old_castl, danish, swedish_flags_wav, danish_arm, bright_morning_light, kronenburg, north_jutland, clear_stream, stick, day_ib, old_alder-tre, water-lili, wild_duck, thicket, wild_wood-path, new_year, following_spr, following_novemb, black_earth, dark_grav, properti, many_summ, rough_wind, wealth, autumn_wind, little_farm, yard_dog, steep_shor, open_seacoast, summer_day, oak, night_drew_nigh, great_tre, topmost_branch, mighty_summit, grand_majestic_oak, bush_herb, birch-tre, lightning_flash, own_way, etern, old_tre, mighty_storm, christmas-day, smallest_hut, bed-chamb, carpet, whole_troop, silent_step, whole_week, shutter, farmer_', farmhous, broad_deep_riv, new_bridg, sunris, good_morn, old_cattle_driv, snowy_hair, cattl, such_large_tre, elder-tree_branch, yonder_corn, afternoon, real_lif, elder_branch, green_lawn, whole_flight, wild_fowl, dark-blue_sea, christmas_tre, valley, whole_even, carnival_rod, to-morrow_morn, next_summ, christmas_light, broad_stream, bush, whole_summ, all_summ, cold_snow, whole_wint, kitchen_cellar, bright_sun, cold_wint, sky, hedg, vine, nest, long_dreary_night, church_clock, mail-coach, clumsy_vehicl, all_plac, happy_new_year, land_cultiv, wood, wagon, little_toy_theatr, railway, sofa, play, little_theatr, stage, domestic_drama, lofty_hal, furnitur, pane, hauschen_street, rain_snow, neighboring_castl, such_even, beech_wood, proud_baronial_castl, whole_apple-tre, stony_mountain, day_molli, apple-tre, pretty_room, good_b, dark_day, old_wartburg_castl, venus_mountain, pathway, histori, snowstorm, bitter_weath, little_corn, snow-storm, own_histori, germani, acacia, cold_autumn_wind, novemb, old_church_bel, chief_town, provinc, great_festiv, military_school, many_heavy_day, dark_days_cold_night, royal_c, bavaria, green_island, foliag, statu, danish_sculptor, large_courtyard, provincial_town, family_circl, long_curtain, balconi, bench, high_step, convent_church, graveyard, paved_street, very_old_hous, at_window, next_sunday_morn, staircas, shop_opposit, snow_drift, step, from_street, fine_new_hous, little_garden, wild_vin, soft_earth, open_sea, old_heathen_day, gabled_window, knightly_castl, ice_crack, golden_cross, buried_town, new_rac, winter_night, old_boundary_post, summer_tim, large_old_carriag, at_church_door, bright_warm_sunshin, snow_man, towards_morn, keen_wind, trees_bush, twilight_hour, window-pan, gentle_sun, last_hous, little_villag, little_way_off, egypt, autumn_arriv, old_day, such_scen, new_court-hous, first_night, canal, church_tow, sentry-hous, barber_', whole_neighborhood, inscription_institut, such_storm, on_grav, toad_rac, dust, elder_bush, in_farm-hous, stork_famili, old_mans, tapestri, cellar, loftiest_tre, mansion, landmark, refug, thick_foliag, garden_tre, chimney, sleepless_night, lordly_lif, frost, jutland, canon_', kind_heaven, such_easter_morn, old_high-road, dark_even, wedding_fest, small_window, tunnel, drain, draught, sunny_spot, snug_retreat, turkey_cock, wild_gees, warm_room, farmyard, hollow_oak-tre, hazel_bush, fourth_day, england, spain, fifth_day, sixth_day, top_oven, shade, all_villag, upon_roof, baking_day, cupboard_door, cat_famili, farmer_potato, bowls_basin, all_night, out-of-door, market_folk, cathedral_clock, westgat, proper_way, sandy_whisk, day_little_luci, great_way, hillsid, rusty_spot, all_way, such_funny_house_!, storeroom, line, float, winter_early_spr, little_stream, stick_hous, way_up, rocks_bush, kitchen_fir, chair, young_famili, kitchen_floor, empty_pail, oven, boiling_wat, all_pig_famili, wide_valley, market, dark_daylight, lawn, next_christma, sitting_room, day_diddl, big_wind, such_lovely_dinn, eve

Triggers verbs list

live, had, serv, said, wa, begrudg, went, came, lay, wait, got, thought, aros, woke, met, took, were, sat, play, put, blew, pull, return, help, tie, brought, wash, dri, stay, saw, gave, did, carri, ask, built, open, stood, rejoic, set, show, reach, laugh, alight, unpack, order, drove,

marri, let, solv, fell, announc, guess, step, heard, turn, won, crack, ate, felt, tri, pleas, like, believ, peep, envi, unscrew, ran, cri, stuck, die, piti, knew, walk, rain, awok, found, touch, wept, unti, led, look, sought, gather, laid, slept, happen, held, rang, beg, agre, sprang, start, knock, jump, rest, chase, limp, crept, told, bound, wore, stretch, answer, rose, left, shut, call, suckl, shook, cover, stroke, spoke, nurs, disappear, dare, watch, receiv, becam, chang, drew, inform, woo, continu, made, wish, enter, began, divid, thank, relat, yield, repli, grew, boil, cut, stop, arriv, bore, pretend, want, seiz, lift, threw, flow, suspect, talk, leap, kept, roll, pronounc, lost, climb, plead, point, need, forc, rush, resound, thrust, caught, resign, seem, requir, push, shot, follow, sound, hesit, obey, sank, resembl, move, rais, recommend, love, embrac, visit, grant, entertain, consent, vouchsaf, remain, declar, pass, dress, rode, pursu, rebound, mutter, depriv, appear, belong, mount, recogn, press, conquer, owe, care, rememb, gaze, hew, place, join, pick, stir, colour, forgot, conceiv, unfasten, sail, fish, row, flew, behav, sent, sigh, growl, warn, fetch, broke, spirt, journey, tore, hunt, hope, understood, inquir, expound, stole, question, refus, offer, float, notic, wrote, read, pri, search, scold, dream, kill, snore, roar, complain, beckon, inspect, observ, fare, promis, accompani, gnaw, torment, beat, perceiv, nod, caus, demand, command, smile, melt, extinguish, descend, fear, unrol, sew, mourn, entreat, strip, smear, accus, use, swept, lack, kiss, chanc, misl, surpris, slip, snif, discov, attack, struck, har, hiss, dug, fret, ceas, greet, expect, welcom, milk, enjoy, invit, prepar, fulfil, lean, gallop, pat, hurt, split, determin, unhar, plough, harrow, hung, snap, summon, ponder, advis, handl, repaid, wonder, hover, cast, stoop, lament, scream, listen, deliv, desir, faint, rear, hit, wrought, plagu, brush, bought, plant, water, danc, drop, stream, squeez, curtsi, fit, reign, griev, wound, rove, bark, buzz, blaze, finish, pack, wrap, besmear, shone, toast, manag, sparkl, lick, clean, stipul, tap, work, tast, scald, bast, occur, sharpen, kick, sold, nudg, intend, confess, meant, present, tail, cudgel, pop, sprinkl, strew, suppos, lock, pour, pray, snatch, toss, groan, glitter, stuf, close, comfort, reproach, share, sang, approach, nibbl, support, jog, smack, spoil, wipe, weigh, drag, scour, last, forbad, pinch, hurri, bid, dip, bent, exchang, flash, depart, echo, snort, oppos, conduct, equip, escap, leapt, shrank, trembl, 'd, end, sorrow, travel, conceit, knight, vow, paid, jeer, chose, miss, nestl, farther, fanci, appeas, fill, collaps, prick, engag, neglect, swore, espi, grunt, ail, clamber, clutch, long, allow, spun, remark, ascend, pierc, wander, roam, wet, soften, fling, unbolt, cost, bewail, bemoan, choke, grasp, bath, resum, roland, dawn, admit, peck, blind, splash, run, labour, spent, propos, lit, circl, reappear, reveal, condemn, scrambl, loosen, aim, hop, bustl, caper, dart, waken, frighten, taught, crawl, tack, wretch, clap, relent, dive, sentenc, rule, inherit, bud, buri, wave, cross, lightheart, sign, seal, cheat, grin, coil, borrow, ought, salt, croak, bargain, unlock, launch, outwit, shave, bind, earn, expati, sneer, butter, wall, tonight, lie, insist, form, father, depend, consum, roast, undertook, succeed, ala, assur, urg, attempt, paus, render, hand, exclaim, dismount, betroth, crouch, heap, curl, settl, stain, unwrap, stoke, "how, comrad, yell, fan, warm, rub, moan, celebr, anger, ventur, breath, howl, wag, counsel, anoint, consecr, spi, ad, march, bray, mew, scamper, dispatch, grope, stab, sow, yoke, hire, swung, prove, learn, bless, trot, skip, land, name, trip, favour, repeat, disclos, increas, embroid, miser, convey, disput, belabour, defend, repent, fled, scorn, lent, blame, threaten, satisfi, consid, save, insult, contain, advanc, bleat, drown, puf, poke, bawl, shall, part, hug, wither, carol, adher, blossom, cherish, bloom, gladden, fade, treat, swell, drift, shudder, glanc, experienc, flit, appli, wail, gleam, express, bask, describ, surround, fold, utter, wrestl, suffer, gush, obtain, absorb, persist, imagin, attribut, underw, lessen, purchas, decid, mould, soil, unfold, strove, lurk, pervad, creat, own, acknowledg, tumbl, clear, glisten, delici, stumbl, beam, titter, beneath, prowl, twist, rippl, produc, rustl, clench, seawe, bow, offend, prevail, mention, rage, career, twitter, rins, cork, outliv, examin, twinkl, extend, display, pleasant, assembl, shout, erect, spread, tree, creak, admir, consol, annoy, applaud, worri, perform, prey, come, persuad, glide, linger, weep, fondl, vanish, awaken, deriv, despis, strut, cluck, chirp, flap, thrive, recov, recognis, flutter, repres, overflow, smoulder, attach, tug, hack, maul, count, spurt, consist, studi, request, recal, pluck, prop, glow, tickl, salut, tip, act, commenc, thump, sprout, explain, hum, resid, prais, appoint, burn, undress, heed, mortifi, regard, dazzl, crackl, chop, cheer, steep, flame, desert, seat, sung, crown, sputter, slope, conceal, shed, mirror, pitch, smoke, remind, skim, tint, fragrant, light, blush, occupi, restor, glimmer, mingl, shiver, illumin, tend, gilt, crumbl, retort, untend, soar, maintain, stare, convinc, refer, contradict, withdrew, fairi, doubt, assert, recur, gain, retain, tar, rig, chatter, reflect, ach, waver, doze, defeat, disfigur, overpow, sting, shade, stammer, overturn, drip, totter, struggl, toil, compli, flock, exist, conclud, test, valu, nail, exhal, expand, improv, carv, chain, unit, grandfath, fire, fought, lash, danish, deserv, toppl, whirl, tower, coo, murmur, gee-up, squeak, measur, flay, trim, clasp, smelt, mark, twine, travers, blow, stamp, interced, cough, jingl, project, huddl, waddl, suck, swallow, exhibit, knelt, renew, clash, array, attend, profess, twirl, merrili, cannot, button, invent, mend, well, entic, unshelt, plunder, uncov, cool, face, treasur, peal, prevent, fed, lower, injur, vote, award, r, dear, starch, contempl, decreas, practis, dispens, chat, throng, crash, prattl, paint, bindwe, writh, wriggl, investig, subtract, neigh, nurtur, pile, dishevel, snow, burnt, fail, splendor, check, replac, procur, gratifi, shoulder, rock, proceed, quack, glare, curv, froze, reviv, endur, await, enabl, devast, destroy, fix, shower, brandish, retir, cram, paddl, scutter, sidl, rummag, wade, scoop, squirm, disarrang, haul, snippet, shuffl, bar, slid, suggest, wink, cuf, perplex, knocked-once-twic, tuck, iron, goffer, scrub, dust, arrang, trickl, punt, tweak, shove, bounc, convers, slap, twitch, brown, flatten, addl, fume, bristl, undid, stagger, rope, endeavor, overbalanc,

```
chew, air, grappl, quarrel, rejoin, accept, implor, sneez, flop, reprov, saunter, dispos, squeal,
grab, unlac, grudg, distrust, snuf, flick, fumbl, race, pelt, apolog, chuckl, 's-been-digging-up,
'tice, thin, streak, crumpl
```

APPENDIX B: PERMISSIONS

LSU

Ricardo Calix <rcalix1@tigers.lsu.edu>

Re: Fw: Use of figures, tables, and text from published IEEE paper

2 messages

j.hansson@ieee.org <j.hansson@ieee.org> Tue, May 24, 2011 at 3:00 PM
To: rcalix1@tigers.lsu.edu

Dear Ricardo Calix :

We are happy to grant you this permission. Our only requirement, in the case of textual material (i.e., using short quotes or referring to the work within these papers), you give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © [Year of original publication] IEEE. In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appears prominently with each reprinted figure and/or table.

Sincerely,

Jacqueline Hansson, Coordinator
IEEE Intellectual Property Rights Office
445 Hoes Lane
Piscataway, NJ 08855-1331 USA
+1 732 562 3828 (phone)
+1 732 562 1746(fax)
e-mail: j.hansson@ieee.org
IEEE Fostering technological innovation
and excellence for the benefit of humanity.

©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©

From:	William Hagen/PUBS/STAFF/US/IEEE
To:	Jacqueline Hansson/PUBS/STAFF/US/IEEE@IEEE
Date:	05/24/2011 03:00 PM
Subject:	Fw: Use of figures, tables, and text from published IEEE paper

Re: Fw: use of figures, tables, and text from published ieee paper
2 messages
J.hansson@ieee.org <j.hansson@ieee.org> tue, may 24, 2011 at 3:00 pm
To: rcalix1@tigers.lsu.edu

Dear Ricardo Calix:

We are happy to grant you this permission. Our only requirement, in the case of textual material (i.e., using short quotes or referring to the work within these papers), you give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © [year of original publication] IEEE. In the case of illustrations or tabular material, we require that the copyright line © [year of original

publication] IEEE appears prominently with each reprinted figure and/or table.

Sincerely,

Jacqueline Hansson, coordinator
IEEE intellectual property rights office
445 hoes lane
Piscataway, nj 08855-1331 USA
+1 732 562 3828 (phone)
+1 732 562 1746(fax)
E-mail: j.hansson@ieee.org
Ieee fostering technological innovation
And excellence for the benefit of humanity.

©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©©

From: William hagen/pubs/staff/us/ieee
To: jacqueline hansson/pubs/staff/us/ieee@ieee
Date: 05/24/2011 03:00 pm
Subject: fw: use of figures, tables, and text from published ieee paper

----- forwarded by william hagen/pubs/staff/us/ieee on 05/24/2011 03:00 pm -----
From: ricardo calix <rcalix1@tigers.lsu.edu>
To: copyrights@ieee.org, w.hagen@ieee.org, ecopyright@ieee.org
Date: 05/24/2011 01:59 pm
Subject: use of figures, tables, and text from published ieee paper

Dear sir/madame,

I am a phd candidate at louisiana state university. I am working of my final dissertation to be submitted to our graduate school on july 8, 2011.

I am the first author of the paper listed below which we published on the ieee transactions on multimedia in 2010. I would like to use figures, tables, and text from different sections of the paper on my dissertation. Therefore, i ask that you please grant me permission to do so.

Also, please let me know if the permission granted via email would be sufficient or if i have to follow further procedures.

###
Paper: emotion recognition in text for 3-d facial expression rendering
Calix, r.a. mallepudi, s.a. bin chen knapp, g.m.
Multimedia, ieee transactions on
Issue date: oct. 2010

Volume: 12 issue: 6
On page(s): 544 - 551
Issn: 1520-9210
Inspec accession number: 11523690
Digital object identifier: 10.1109/tmm.2010.2052026
##

Thanks
Ricardo

J.hansson@ieee.org <j.hansson@ieee.org>   sat, may 28, 2011 at 9:01 am
To: ricardo calix <rcalix1@tigers.lsu.edu>

Comments/response to case id:

Replyto:

from: jacqueline hansson   date: 05/28/2011

subject: re: use of figures, tables, and text from published ieee paper   send to: ricardo calix <rcalix1@tigers.lsu.edu>

cc:

Dear Ricardo calix:

We are happy to grant you permission to reuse the ieee copyrighted figure
Requested in all electronic and printed formats.

Since the figure is for use in your thesis, our only requirements are that
You credit the original source (author, paper, and publication), and that
The ieee copyright line (© 2011ieee) appears prominently with the
Reprinted figure.

Sincerely,

Jacqueline hansson, coordinator

© © © © © © © © © © © © © © © © © ©
ieee intellectual property rights office
445 hoes lane
Piscataway, nj 08855-1331 usa
+1 732 562 3966 (phone)
+1 732 562 1746 (fax)

Ieee-- fostering technological innovation
And excellence for the benefit of humanity.
© © © © © © © © © © © © © © © © © ©

ASSOCIATION FOR COMPUTING MACHINERY, INC. LICENSE TERMS AND CONDITIONS

May 24, 2011

This is a License Agreement between Ricardo A Calix ("You") and Association for Computing Machinery, Inc. ("Association for Computing Machinery, Inc.") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Association for Computing Machinery, Inc., and the payment terms and conditions.

License Number	2675480913601
License date	May 24, 2011
Licensed content publisher	Association for Computing Machinery, Inc.
Licensed content publication	Proceedings of the second annual ACM conference on Multimedia systems
Licensed content title	Affect corpus 2.0: an extension of a corpus for actor level emotion magnitude detection
Licensed content author	Ricardo A. Calix, et al
Licensed content date	Feb 23, 2011
Type of Use	Thesis/Dissertation
Requestor type	Author of this ACM article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Automated Semantic Understanding of Human Emotions in Writing and Speech
Expected completion date	Aug 2011
Estimated size (pages)	130
Terms and Conditions	

Rightslink Terms and Conditions for ACM Material

1. The publisher of this copyrighted material is Association for Computing Machinery, Inc. (ACM). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at).

2. ACM reserves all rights not specifically granted in the combination of (i) the license Rightslink Printable License https://s100.copyright.com/App/PrintableLicenseFrame.jsp?publisherID...
1 of 3 5/24/2011 1:25 PM details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

3. ACM hereby grants to licensee a non-exclusive license to use or republish this ACM-copyrighted material* in secondary works (especially for commercial distribution) with the stipulation that consent of the lead author has been obtained independently. Licenses are for one-time use in a single edition of the work, only with a maximum distribution equal to the number that you identified in the licensing process. Any additional form of republication must be specified according to the terms included at the time of licensing. *Please note that ACM cannot grant republication or distribution licenses for embedded third-party material. You must confirm the ownership of figures, drawings and artwork prior to use.

4. Any form of republication or redistribution must be used within 180 days from the date stated on the license and any electronic posting is limited to a period of six months unless an extended term is selected during the licensing process. Separate subsidiary and subsequent republication licenses must be purchased to redistribute copyrighted

material on an extranet. These licenses may be exercised anywhere in the world.
5. Licensee may not alter or modify the material in any manner (except that you may use, within the scope of the license granted, one or more excerpts from the copyrighted material, provided that the process of excerpting does not alter the meaning of the material or in any way reflect negatively on the publisher or any writer of the material).
6. Licensee must include the following copyright and permission notice in connection with any reproduction of the licensed material: "[Citation] © YEAR Association for Computing Machinery, Inc. Reprinted by permission." Include the article DOI as a link to the definitive version in the ACM Digital Library. Example: Charles, L. "How to Improve Digital Rights Management," Communications of the ACM, Vol. 51:12, © 2008 ACM, Inc.
http://doi.acm.org/10.1145/nnnnnn.nnnnnn (where nnnnnn.nnnnnn is replaced by the actual number).
7. Translation of the material in any language requires an explicit license identified during the licensing process. Due to the error-prone nature of language translations, Licensee must include the following copyright and permission notice and disclaimer in connection with any reproduction of the licensed material in translation: "This translation is a derivative of ACM-copyrighted material. ACM did not prepare this translation and does not guarantee that it is an accurate copy of the originally published work. The original intellectual property contained in this work remains the property of ACM."
8. You may exercise the rights licensed immediately upon issuance of the license at the end of the licensing transaction, provided that you have disclosed complete and accurate details of your proposed use. No license is finally effective unless and until full payment is received from you (either by CCC or ACM) as provided in CCC's Billing and Payment terms and conditions.
9. If full payment is not received within 90 days from the grant of license transaction, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked Rightslink Printable License
https://s100.copyright.com/App/PrintableLicenseFrame.jsp?publisherID...
2 of 3 5/24/2011 1:25 PM and shall be void as if never granted.
10. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
11. ACM makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.
12. You hereby indemnify and agree to hold harmless ACM and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
13. This license is personal to the requestor and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
14. This license may not be amended except in a writing signed by both parties (or, in the case of ACM, by CCC on its behalf).
15. ACM hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and ACM (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
16. This license transaction shall be governed by and construed in accordance with the laws of New York State. You hereby agree to submit to the jurisdiction of the federal and state courts located in New York for purposes of resolving any disputes that may arise in connection with this licensing transaction.
17. There are additional terms and conditions, established by Copyright Clearance Center, Inc. ("CCC") as the administrator of this licensing service that relate to billing and payment for licenses provided through this service. Those terms and conditions apply to each transaction as if they were restated here. As a user of this service, you agreed to those terms and conditions at the time that you established your account, and you may see them again at any time at http://myaccount.copyright.com
18. Thesis/Dissertation: This type of use requires only the minimum administrative fee. It is not a fee for permission. Further reuse of ACM content, by ProQuest/UMI or other document delivery providers, or in republication requires a separate permission license and fee. Commercial resellers of your dissertation containing this article must acquire a separate license.

Special Terms:

APPENDIX C: IRB FORMS

Application for Exemption from Institutional Oversight

Unless qualified as meeting the specific criteria for exemption from Institutional Review Board (IRB) oversight, ALL LSU research/projects using living humans as subjects, or samples, or data obtained from humans, directly or indirectly, with or without their consent, must be approved or exempted in advance by the LSU IRB. This Form helps the PI determine if a project may be exempted, and is used to request an exemption.

LSU
Institutional Review Board
Dr. Robert Mathews, Chair
131 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu
lsu.edu/irb

– Applicant,Please fill out the application in its entirety and include the completed application as well as parts A-E, listed below, when submitting to the IRB. Once the application is completed, please submit two copies of the completed application to the IRB Office or to a member of the Human Subjects Screening Committee. Members of this committee can be found at http://www.lsu.edu/screeningmembers.shtml

– A Complete Application Includes All of the Following:

(A) Two copies of this completed form and two copies of part B thru E.
(B) A brief project description (adequate to evaluate risks to subjects and to explain your responses to Parts 1&2)
(C) Copies of all instruments to be used.
*If this proposal is part of a grant proposal, include a copy of the proposal and all recruitment material.
(D) The consent form that you will use in the study (see part 3 for more information.)
(E) Certificate of Completion of Human Subjects Protection Training for all personnel involved in the project, including students who are involved with testing or handling data, unless already on file with the IRB. Training link: (http://phrp.nihtaining.com/users/login.php.)
(F) IRB Security of Data Agreement: (http://www.lsu.edu/irb/IRB%20Security%20of%20Data.pdf)

1) Principal Investigator: Ricardo A. Calix **Rank:** Graduate Student

Dept: Industrial Engineering **Ph:** 225-315-5655 **E-mail:** rcalix1@lsu.edu

2) Co Investigator(s): please include department, rank, phone and e-mail for each

Dr. Gerald M. Knapp, Industrial Engineering, Designated Associate Professor, 225-578-5374, gknapp@lsu.edu

IRB# E5392 LSU Proposal #

☒ Complete Application

☒ Human Subjects Training

3) Project Title: Automated Semantic Understanding of Human Emotions in Writing and Speech

Study Exempted By:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
203 B-1 David Boyd Hall
225-578-8692 | www.lsu.edu/irb
Exemption Expires: 2-20-2014

4) Proposal? (yes or no) no **If Yes, LSU Proposal Number**

Also, if YES, either
○ **This application <u>completely</u> matches the scope of work in the grant**
OR
○ **More IRB Applications will be filed later**

5) Subject pool (e.g. Psychology students) engineering students

*Circle any **"vulnerable populations"** to be used: (children <18; the mentally impaired, pregnant women, the ages, other). Projects with incarcerated persons cannot be exempted.

6) PI Signature Ricardo Calix **Date** 11/05/10 (no per signatures)

**** I certify my responses are accurate and complete.** If the project scope or design is later changes, I will resubmit for review. I will obtain written approval from the Authorized Representative of all non-LSU institutions in which the study is conducted. I also understand that it is my responsibility to maintain copies of all consent forms at LSU for three years after completion of the study. If I leave LSU before that time the consent forms should be preserved in the Departmental Office.

Screening Committee Action: Exempted ✓ Not Exempted ____ Category/Paragraph ____

Reviewer Mathews **Signature** Robt C Mathews **Date** 2/21/11

Consent Form

Multi-modal Affect research requires large amounts of text and acoustic information of spoken language. The goal of this task is to compile an affect corpus with emotion magnitudes. We are asking that you assign emotion magnitudes to actors in text and that you allow us to record the data. The software to use will be installed in a lab accessible to you or on your personal computer. You many annotate the text at your own pace.

Your participation is voluntary and you may stop at any point. We will make some or all of the data available to the wider research community. No one other than the project staff will have access to any forms you provide to us. Your identity will remain confidential. Some general demographics are typically included in the scientific documentation of corpora and in published findings (e.g. age, gender).

To indicate that you wish to participate as outlined above, please complete the following:
I, (please print name)..
have read this form and agree to take part in the research on these terms.

Signature: .... Date: ....

"This study has been approved by the LSU IRB. For questions or concerns regarding participant rights, please contact the IRB chair, Dr. Robert C. Mathews, 578-8692, or irb@lsu.edu."

Study Exempted By:
Dr. Robert C. Mathews, Chairman
Institutional Review Board
Louisiana State University
203 B-1 David Boyd Hall
225-578-8692 | www.lsu.edu/irb
Exemption Expires: 2-20-2014

Project description

The objective of this project is to annotate a corpus of stories with emotion labels and magnitudes. Human annotators extend an existing corpus manually by extracting the actors in each story and annotating them with their evolving emotional state. A computer program is used to perform this task on pre-selected texts.

Each session, the annotator proceeds to select one of the stories he or she is responsible for. The annotator loads the story into the system by searching for the story's name in the stories directory. When the user clicks on the "begin annotation" button, the first sentence of the story is highlighted. Each sentence in the story is highlighted as the user annotates it. As the user is reading each sentence, he or she will identify actors in the story and determine if they have already been added to the list of actors. If the actor is not on the list, the annotator will click a button to add the actor to the actor's list section. Once the actor is on the list, the annotator can select the actor and adjust emotional state at the given sentence position in the story. This process is repeated for each sentence until the story is completed. The software is installed on the computers in the systems integration lab of Patrick F. Taylor Hall or on the annotator's personal computer. Therefore, there is no constrained time schedule for the annotators to complete the task. Annotation per story takes between 30 minutes to an hour. Audio recordings matching the texts are obtained from the world wide web.

VITA

Ricardo A. Calix was born in La Ceiba, Honduras, on June 11^{th}, 1978. After completing high school at the Instituto Maria Regina, he obtained his bachelor's degree in industrial and systems engineering from the Universidad Tecnologica Centroamericana, Tegucigalpa, Honduras, in 2001. In August of 2004 he started graduate studies at Louisiana State University (LSU) in Baton Rouge, Louisiana, where he obtained a Master of Business Administration in 2006. In August of 2007, he started graduate studies in the college of engineering at LSU as a graduate assistant in the department of industrial engineering. His graduate studies have been in the interdisciplinary program of engineering science with concentration in information technology and engineering. He obtained a Master of Science in Engineering Science degree with concentration in information technology and engineering in 2010. He is a candidate for the Doctor of Philosophy degree in engineering science with concentration in information technology and engineering. The degree will be conferred at the summer commencement 2011.