

Experience Based Adaptive Vocal Interaction System for Domestic Service Robots

Sajila Dimagi Wickramaratne

(168077J)

Thesis submitted in partial fulfillment of the requirements for the degree Master
of Science in Electrical Engineering

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

June 2018

Experience Based Adaptive Vocal Interaction System for Domestic Service Robots

Sajila Dimagi Wickramaratne

(168077J)

Thesis submitted in partial fulfillment of the requirements for the degree Master
of Science in Electrical Engineering

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

June 2018

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

The above candidate has carried out research for the MSc thesis under my supervision.

Signature of the Supervisor(s):

Date:

Dr.A.G.B.P.Jayasekara

Abstract

Domestic service robots have gained an increasing popularity in the recent times due to the lack of personnel to tend to domestic work. Domestic service robots with conversational capabilities can be sought out as an exceptional solution to eliminate the mental distress among the elderly. The conversational skills of the domestic robots are currently not comparable to the human level. Only a few researches have been carried out in order to study how the decisions regarding the conversation are made. Humans use a wide variety of non-verbal entities to determine the state of the conversation and manage the conversation flow to ensure that it is interesting to all the parties involved.

The current conversation systems used in service robots frequently use the linguistic message of the user. Most of these systems consider factual and behavior related conversations. Such systems are unable to comprehend the dynamic situations such as the emotional state of the user when generating a response. Systems which can handle the dynamic states of a conversation are preferred by the users. This research is aimed to close the gap between the use of emotions for conversation systems compared

This research proposes a system can be used to make interaction decisions regarding conversation flow by considering the emotional state of the user. In a human conversation both linguistic and para-linguistic features are used to convey information. The proposed system also use this information from the users' response to identify the emotional state and the actual linguistic message. The system comprises of three segments which are emotion recognition, emotional memory storage and conversation decision unit. These three systems coordinate together to decide whether the decision should be continued or not , if it is going to get continued what should be the response and what are the important elements in the user's response that should be retained in system's long term memory.

Keywords-Conversation, para-linguistic, linguistic, emotions, service robot, human-robot interaction

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. A.G.B.P. Jayasekara who provided continuous support throughout the period to successfully finish this thesis. Further I would like to acknowledge my review panelists Dr. Chandima Pathirana and Dr. Ruwan Gopura whose comments and inputs were valuable for my research.

I would also like to thank the members of the thesis review committee who gave their suggestions to improve this thesis.

This research would not have been successful if not for the many suggestions and valuable inputs from my research colleagues Chapa Sirithunga, Arjuna Srimal and Viraj Muthugala.

Finally I would like to thank my mother and my sister, both whom assisted in numerous way throughout my research providing me the best environment possible for my research.

This work was supported by University of Moratuwa Senate Research Grant SRC/CAP/17/03.

TABLE OF CONTENTS

Declaration	i
Abstract	ii
Acknowledgments	iii
Table of Contents	viii
List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Problem Statement	3
1.2 Thesis Overview	3
2 Literature Review	5
2.1 Assistive Service Robots	5
2.2 Human-Robot Interaction	6
2.3 Importance of Conversation In Human-Robot Interaction	8

2.4	Understanding Emotion	9
2.4.1	Emotion Classification Systems	12
2.5	Verbal Response Generation According to the Emotions	15
2.6	Current status on existing human- robot vocal interaction system	17
2.7	Summary	17
3	System Design	20
3.1	Studies of Human Conversation	20
3.2	An Artificial Agent Conversation System	21
3.3	System Overview	22
3.4	Summary	25
4	Emotion Recognition System	27
4.1	Selection of Emotion Space	28
4.2	Linguistic vs Para-Linguistic Features	28
4.2.1	Feature Extraction for Emotion Recognition	29
4.2.2	Extraction of Prosodic Features	30
4.2.3	Extraction of Spectral Features	31
4.2.4	Identifying Emotion Clusters- Using Unsupervised Learning	33
4.3	Neural Network for Audio based Emotion Recognition	37
4.4	Implementation of Emotion Recognition System	40

4.5	Results	41
4.6	Summary	42
5	Lexical analysis System	44
5.1	Identifying Linguistic Message	44
5.2	Functional Flow of Lexical Analysis	46
5.3	Practical Implementation	48
5.4	Results	50
5.5	Summary	51
6	Emotional Memory Determination	53
6.1	Defining Emotional Memories	53
6.2	Human Emotional Memories Labeling Process	54
6.3	Determining Emotional Significance	55
6.4	Results	57
6.5	Summary	60
7	Conversation Decision Making	61
7.1	Conversation Rule System	61
7.2	Decision Making on Continuation of the Conversation	62
7.2.1	Interaction Decision Determination	62
7.3	System Overview	63

7.4	Implementation of the System	64
7.5	Results	68
7.6	Summary	69
8	Vocal Response Generation	71
8.1	Definition of Empathetic Responses	71
8.2	Factors Affecting Empathetic Behavior	72
8.2.1	User Preferences	72
8.2.2	Effect of the Previous States	73
8.3	Stimulus Events and Expected Behavior	73
8.4	Vocal Response Generation System	74
8.4.1	Implementation	75
8.5	Results	76
8.6	Summary	80
9	Conclusions	82
9.1	Overall Assessment of the System	82
9.2	Limitations of the System	83
9.3	Recommendations for Future Developments	84
	List of Publications	86

LIST OF FIGURES

1.1	Elderly Population Growth Projection until 2050	2
2.1	Asimo produced by Honda	7
2.2	Pepper robot produced by Aldebaran	7
2.3	Romeo produced by Aldebaran	8
2.4	Six basic emotions presented by Ekman expressed through facial expressions	11
2.5	A simplified version of the Circumplex model introduced by Russell	11
3.1	Simplified architecture of a dialog system	22
3.2	Home Screen of the MIVoIS conversation module	24
3.3	The overall conversation decision System	26
4.1	Pitch contours(upper row) and Intensity Contours(lower row) for the four affective states when the user presents the same utterance a) & e) Angry b) & f) Happy c) & g) Sad d) & h) Neutral	29
4.2	Visible voice, Pitch and Intensity Contour for an Angry Utterance	30
4.3	Layered Structure of the System	31
4.4	Spectrogram of the System	32

4.5	Feature Extraction Methodology	33
4.6	Clustering Process To Determine Vocally Distinguishable Emotions	36
4.7	PCA (Principal Component Analysis) graph showing 4 clusters	36
4.8	Flow of the neural network based Emotion Recognition System	39
4.9	Visible voice, Pitch and Intensity Contour for an Angry Utterance	40
4.10	The back-end of the web based emotion recognition module setup	41
4.11	Confusion Matrix for Spectral Feature based Emotion Recognition Neural Network	42
4.12	Confusion Matrix for Prosodic feature based Emotion Recognition Neural Network	43
5.1	Pitch and Intensity Contours of a) & e)happy b)&f)angry c) & g)sad d) & h)calm emotions for the same utterance	45
5.2	Flow Diagram of the lexical analysis system	46
5.3	Parts of speech breakdown of the utterance	47
5.4	Part of Speech Segmentation	47
5.5	Step 1: Statement Input	49
5.6	Step 2 : Analyzing the input	49
5.7	Step 3: System Output	50
5.8	Distribution of Entities of Utterances	51
6.1	Representation of the memory module	56

6.2	Overall System of Emotional Memory Detection	57
7.1	Functional flow of the system implementation	64
7.2	Conversation Initiation of the Robot and the User	65
7.3	Question set used for the topic animals	66
7.4	A participant and MIRob platform during the experiment	66
7.5	User Feedback on Conversation Decision Making without using emotional states	69
7.6	User Feedback on Conversation Decision Making using emotional states	70
8.1	Feeding encoder and decoder data into the Model	75

LIST OF TABLES

2.1	Common Needs and Deficiencies of the Elderly Community	6
2.2	Stimulus events and Expected Behavior of Humans	15
2.3	Summary of the Literature Review	19
4.1	Performance Evaluation of Prosodic and Feature Based Neural Networks	42
5.1	Entity Analysis	48
6.1	Comparison of the Participants Response vs The Robot's Response	59
7.1	Rules System for the conversation Decision System	63
7.2	Results of the Experiment	67
8.1	Possible Stimulus Events and Expected Behaviors	72
8.2	Results of User Assesment of the system generated responses when emotions are not considered	77
8.3	Results of User Assesment of the system generated responses when emotions are considered	78
8.4	User Utterances and Generated Response	79

8.5 The comparison of system generated and human specified utterances 80