Emotion detection from Male Speech in Computer Games

Tarashankar Rudra
Department of Computing
Macquarie University
Sydney, Australia

Manolya Kavakli
Department of Computing
Macquarie University
Sydney, Australia

David Tien
Accounting & Computer Sc.
Charles Sturt University
Bathurst, Australia

Abstract-- **This paper presents the experimental results of our method of classifying emotional states of neutral and anger of male voice from artificial pidgin utterances using Support vector machine (SVM). The objective of the paper is to demonstrate that a new genre of languages called Game pidgin language (GPL) that can not only be used to capture real time speech recognition, but can also generate response from the non-player-character. The emotion which is a floating point value can be used to generate the corresponding emotive response from the non-player-character (NPC).**

I. INTRODUCTION

The development of robust Natural Language Processors is still a distant reality. "CPL is inspired by a frustration at a perceived lack of progress in Spoken Language Research over the last 20-30 years" [1]. It was proposed that a Computer Pidgin Language (CPL) with limited vocabulary and simple set of grammatical rules can be an effective approach for human computer interaction. The objective of this paper is to demonstrate that in speech interactive computer games, a new language called Game pidgin language (GPL) [2] can not only generate real time spoken response from the non-player-character but also capture and express emotions.

Computer games are the most popular means for edutainment today. In his paper, Finn [4] has argued that previously the academic community had ignored the field of Games Technology; it is now that videogames have attained legitimacy in scholarly enquiries. Buoyed by the availability of dedicated software personnel and robust software tools supported by high end game consoles and computers and peripherals at affordable prices, the industry is all set to overtake the film industry in a couple of years. The content of games ranges from kids Information & entertainment (infotainment) to extremely sophisticated virtual reality simulation of real world models thus making them versatile and acceptable to all sections of the society. A CPL [1] is a new spoken language, which is taught to the user and is efficient for dialogues with the computer. In our previous papers [2] [6], we have developed a GPL with limited vocabulary, grammar and syllables for use with speech interactive Computer Games.

Since computer games use 'barks' [2], which is slang's developed for the communications between game agents; we will have two or less syllable words from an artificial language and use them uniquely for expression of emotion. The GPL we have developed will enhance the group bonding especially in multi-player games.

Although Zhang et al [5] cites three reasons for difficulty in voice recognition; with the rapid advancement of hardware technology, speech interactive game is no more a distant reality. All major Game platforms now have speech interfaces and ship with headsets. A game can be made more realistic if the game player is allowed to communicate with the characters in the game. We can have two types of speech communication, Active communication and Passive communication [6].

In our previous study [7], we have shown that we can use support vectors to classify emotions of anger and neutral of uni-word artificial pidgin utterances. In this paper, to investigate the emotion recognition process, we have recorded single and two word pidgin utterances of four GPL sentences with two emotion states of neutral and anger and spoken by male subjects. We have used PRAAT [8] for extracting four statistical prosodic features of pitch [9] - mean, standard deviation, minimum and maximum values. We have then analysed the vocabulary using SVM developed by Gavin Cawley [10]. We have already presented our architecture for modeling emotion in game characters [11].

The paper is organised as follows: in the second section we present our experimental results for emotion classification using support vector machine (SVM), in the third section, we evaluate our results and finally we have the conclusion.

II. EXPERIMENTAL RESULTS ON EMOTION CLASSIFICATION USING SVM

A support vector machine or SVM is an algorithm based on statistical learning theory that constructs a hyper-plane that separates two classes with maximal margin. When the input is not linearly separable, the data is transformed with a non-linear map, producing a higher dimensional feature space in which the data can be separated. The separating hyper-plane in this higher dimensional space forms a non-linear decision surface in the original input space. In our study, we assume that every word in game pidgin language can elicit an emotion during game play. "A classification task usually involves with training and testing data, which consist of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). The goal of SVM is to

1-4244-1272-2/07/$25.00 ©2007 IEEE

produce a model, which predicts target value of data instances in the testing set which are given only the attributes” [12]. In our paper [7], we had shown that we can use SVM to classify emotion from arbitrary utterances. In this study we have trained and classified the data for male voices. In the figures for the datasets, round plots are for neutral voice and square are for anger; we have shown only the plots of mean and standard deviation of pitch in X and Y axis respectively as we cannot plot a graph of four dimensions. The positive values in result column of tables 1(b), 2, 3 and 4 indicate that the utterance has been classified as anger where as negative values signify that the value is neutral, also the emotion code for anger is A, and emotion code for neutral is N and the code for wrongly detected emotion is E.

Dataset 1:

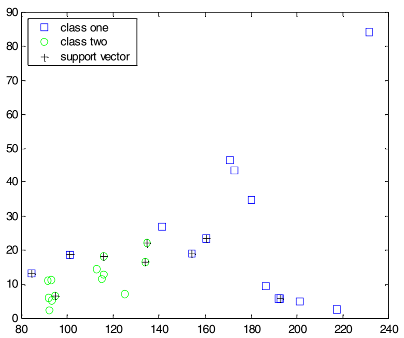

Figure 1 (a)

The figure 1 (a) plots Support vectors of 13 neutral and 14 anger utterances of 7 male voices for the utterance of the word ‘**ab**’, which means go away. The plots are from the first two columns of table 1(a) where X axis plots the mean and Y axis plots the standard deviation of pitch frequency respectively.

In the data given below, we have the four statistical parameters of pitch frequency; the first column is mean, the second is standard deviation, the third column is minimum value and the fourth column is maximum value. The last column in the training data is the classification code; here 2 stands for anger and 1 stand for neutral.

The input to the test data consists of seven consecutive anger data followed by seven consecutive neutral data. The data organized in the result output from SVM consists of positive values for anger and negative values for neutral.

Training data:

Mean	Standard Deviation	Minimum	Maximum	Class code
84.268	13.011	76.270	113.579	2
100.895	18.634	78.592	134.698	2
217.400	2.665	212.656	221.263	2
231.507	84.298	198.144	498.326	2
160.549	23.347	126.627	189.977	2
154.370	18.865	116.840	181.158	2
172.678	43.629	97.265	206.684	2
191.958	5.747	179.144	197.370	2
186.630	9.483	164.671	196.684	2
192.621	5.760	183.256	198.718	2
141.334	26.964	102.764	172.440	2
170.956	46.536	96.974	226.930	2
180.240	34.917	134.584	280.358	2
201.117	4.934	193.419	206.138	2
93.255	5.305	85.803	100.265	1
92.278	2.226	90.140	95.359	1
112.871	14.551	90.194	139.680	1
115.796	18.169	90.973	143.107	1
115.781	12.883	97.236	146.041	1
115.003	11.586	96.740	130.338	1
94.603	6.529	82.766	100.750	1
91.730	6.046	82.236	98.713	1
125.072	7.012	116.829	138.286	1
134.073	16.676	121.577	174.577	1
134.602	22.208	100.151	176.592	1
92.855	11.239	78.931	106.974	1
91.416	10.961	77.338	106.972	1

Table 1 (a)

Test data and result:

In the data below the first seven rows had actual emotion as anger and the next seven rows had the actual emotion as neutral.

Mean	Standard Deviation	Minimum	Maximum	Result	Detected Emotion
129.070	25.257	99.784	163.529	3.5232	A
171.906	38.361	98.776	196.629	31.7336	A
143.690	23.461	105.891	174.909	1.7322	A
169.718	17.771	115.396	180.184	28.6477	A
177.625	30.026	114.454	200.945	22.6693	A
136.586	18.033	104.150	154.464	3.1312	A
192.111	1.137	190.516	193.189	1.5842	A
87.554	2.559	84.031	91.363	-2.4333	N
109.055	15.486	94.470	142.281	-1.0141	N
111.555	11.101	92.994	134.217	-3.9961	N
92.487	6.717	81.812	102.202	-2.5222	N
136.182	7.870	126.363	147.552	-3.8930	N
122.552	19.552	93.451	147.955	0.1482	E
90.749	9.146	72.067	106.781	-2.2998	N

Table 1 (b)

The SVM was tested with the test data for classification of emotions of anger and neutral. We supplied test data for seven neutral and seven anger emotion states, the SVM could classify accurately 13 of the 14 data.

Dataset 2:

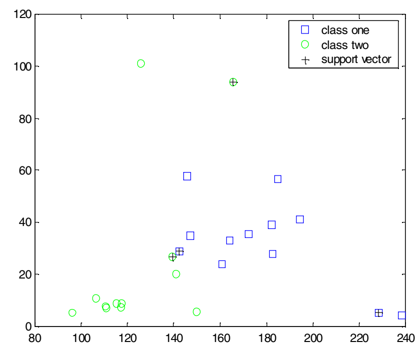

Figure 1 (b)

The figure 1 (b) plots Support vectors of 12 neutral and 12 anger utterances of 7 male voices for the utterance of the word ‘**dem**’, which means go stop.

Training data: It is obtained from PRAAT by providing the word as input. A table similar to table 1(a) is obtained.

Test data and result:

In the data below the first seven rows had actual emotion as anger and the next seven rows had the actual emotion as neutral.

Mean	Standard Deviation	Min	Max	Result	Emotion Code
141.744	14.530	118.684	161.794	-2.2987	E
226.066	19.184	178.983	246.496	9.7798	A
125.239	17.988	100.170	148.404	-3.2083	E
163.752	27.948	105.506	192.750	4.6217	A
188.949	49.394	104.063	262.588	12.7435	A
175.806	41.362	113.898	231.538	6.6265	A
203.633	58.443	82.835	259.809	20.1902	A
97.937	6.717	87.439	106.329	-6.1740	N
124.052	15.522	105.694	155.728	-3.8264	N
116.580	11.165	95.942	131.543	-4.1399	N
101.097	7.663	86.180	114.516	-5.4763	N
111.051	6.314	101.273	120.778	-5.6530	N
136.750	25.683	108.631	175.444	-2.0792	N
95.873	13.940	75.241	119.118	-5.1970	N

Table 2

The SVM was tested with the test data for classification of emotions of anger and neutral. We supplied test data for seven neutral and seven anger emotion states, the SVM could classify accurately 12 of the 14 data.

Dataset 3:

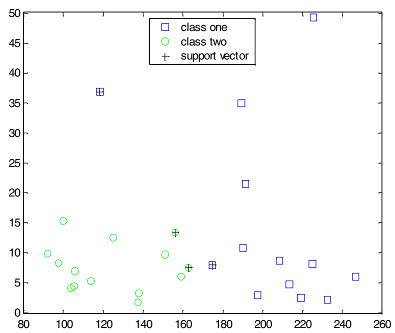

Figure 1 (c)

The figure 1 (c) plots Support vectors of 14 neutral and 13 anger utterances of 7 male voices for the utterance of the word '**fat**', which is an expression of despair.

Training data: It is obtained from PRAAT by providing the word as input. A table similar to table 1(a) is obtained.

Test data and result:

In the data below the first six rows had actual emotion as anger and the next seven rows had the actual emotion as neutral.

Mean	Standard Deviation	Min	Max	Result	Emotion Code
192.139	12.362	171.835	211.724	4.3502	A
178.751	5.746	172.572	188.007	0.8734	A
205.787	13.887	173.012	217.778	7.7389	A
201.473	10.151	184.005	211.596	5.7845	A
192.703	21.308	151.417	220.895	7.3332	A
184.651	6.614	174.671	190.536	2.2522	A
101.552	5.902	93.072	110.149	-8.0300	N
141.965	7.603	130.554	150.593	-3.5171	N
126.630	12.995	109.396	144.767	-4.4211	N
108.989	7.267	96.471	118.295	-7.1332	N
168.845	4.033	164.653	175.921	-0.8554	N
147.922	11.691	131.228	167.384	-2.2512	N
95.478	11.677	75.426	110.489	-7.6533	N

Table 3

The SVM was tested with the test data for classification of emotions of anger and neutral. We supplied test data for six neutral and seven anger emotion states, the SVM could classify accurately 13 of the 13 data.

Dataset 4:

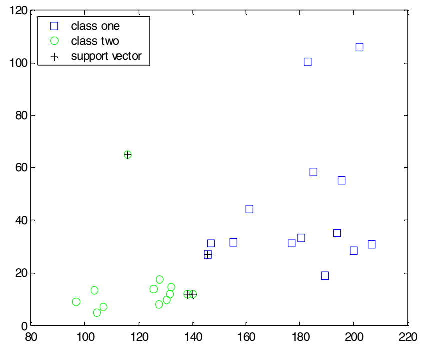

Figure 1 (d)

The figure 1 (d) plots Support vectors of 13 neutral and 14 anger utterances of 7 male voices for the utterance of the word '**Na wok**', which means not work.

Training data: It is obtained from PRAAT by providing the word as input. A table similar to table 1(a) is obtained.

Test data and result:

In the data below the first seven rows had actual emotion as anger and the next seven rows had the actual emotion as neutral.

Mean	Standard Deviation	Min	Max	Result	Emotion Code
218.484	124.432	112.291	487.587	23.1633	A
178.594	18.706	144.710	207.928	2.2679	A
169.839	36.255	129.818	228.441	3.7762	A
142.578	24.732	103.684	176.038	0.5792	A
180.697	21.753	132.779	225.219	3.3599	A
203.530	23.798	150.414	239.023	5.3754	A
158.993	21.835	111.124	189.623	1.6189	A
117.588	76.095	92.791	492.203	-0.2562	N
126.565	9.335	101.124	148.877	-1.7621	N
118.701	13.142	105.911	147.518	-2.1239	N
102.607	5.670	92.001	116.521	-3.3534	N
118.411	7.723	104.880	146.431	-2.4548	N
135.464	11.618	111.992	159.674	-1.2105	N
95.066	6.474	78.907	105.526	-3.4923	N

Table 4

The SVM was tested with the test data for classification of emotions of anger and neutral. We supplied test data for seven neutral and seven anger emotion states, the SVM could classify accurately 14 of the 14 data.

The SVM was tested with the test data for classification of emotions of anger and neutral. We supplied test data for seven neutral and seven anger emotion states, the SVM could classify accurately 14 of the 14 data.

III. EVALUATION

We have already presented our paper [13] on emotion recognition rate for female voice recording of the same words; a comparison of success rate is given in the table below:

Word	Male	Female

Ab	92.86%	87.5%
Dem	85.71%	100%
Fat	100%	75%
Na Wok	100%	100%

Table 5

The average minimum, maximum and mean pitch of male speech for the emotion of anger is given below:

Word	Male		
	Avg. Min	Avg. Max	Avg. Mean
Ab	140.0861	216.0231	170.4659
Dem	124.6123	217.8863	178.7619
Fat	173.9708	217.6782	202.5778
Na Wok	115.2172	268.5176	180.1961

Table 6

The average minimum, maximum and mean pitch of female speech for the emotion of anger is given below:

Word	Female		
	Avg. Min	Avg. Max	Avg. Mean
Ab	168.0545	251.0371	222.0983
Dem	146.1074	275.7466	223.9658
Fat	232.5889	294.3121	267.5283
Na Wok	159.6175	295.9848	236.901

Table 7

The average minimum, maximum and mean pitch of male speech for the emotion of neutral is given below:

Word	Male		
	Avg. Min	Avg. Max	Avg. Mean
Ab	93.1472	127.511	108.4104
Dem	104.3367	198.4864	124.735
Fat	112.7760	137.5926	124.6809
Na Wok	121.5949	103.1893	167.7825

Table 8

The average minimum, maximum and mean pitch of female speech for the emotion of neutral is given below:

Word	Female		
	Avg. Min	Avg. Max	Avg. Mean
Ab	131.1566	204.2600	167.0048
Dem	136.6502	225.0429	157.0575
Fat	157.4576	278.5075	178.6626
Na Wok	137.731	236.8039	169.49

Table 9

We found that emotion classification rate is independent of the number of words from the grammar conforming to our previously established GPL [2][6]. It is also interesting to note the variation of average - minimum, maximum and mean of pitch of the words with respect to gender. Here it is advisable to have the gender to be specified as a precursor to the actual game play. This result is very significant as each computer game will have its own limited set of vocabulary and our aim is to restrict the characteristics of the pidgin utterances to uni-syllable words for quick and easy learning by game players.

IV. CONCLUSION

Our experimental results show that emotion from artificial Pidgin language with arbitrary utterances can be individually classified using SVM for the emotion states of neutral and anger for Human-computer-interaction. This will allow us to program non-player-characters (NPC) to react not only by the meaning of the utterance of the player but also his/her mood or emotional state. We can also make NPC's react to the amount of emotion content in the utterance of the game player. Our concept of pidgin language can be implemented for communicating with speech impaired people, where they speak using a pidgin language with reduced grammar and vocabulary and the computer responds in a natural language. In addition to computer games, this novel approach can be used for disaster recovery, help desk, emergency lines, interactive voice response, creating robots to substitute as pets expressing their emotions in artificial pidgin language, etc.

REFERENCES

[1] Hinde, S., Belrose, G., Computer Pidgin Language: A new Language to talk to your Computer?, (online), URL: http://www.hpl.hp.com/techreports/2001/HPL-2001-182.pdf.

[2] Rudra, T., Kavakli, M., Bossomaier, T., 2003: A Game Pidgin Language for Speech Recognition in Computer Games, proceedings COSIGN 2003, The 3rd International Conference on Computational Semiotics for Games and New Media, pp 90-98, University of Teeside, Middlesbrough, United Kingdom, September 9-12.

[3] Hall, R., A., Pidgin and Creole Languages, 1966, Cornell University Press, London.

[4] Finn, M., 2002, Console Games in the age of convergence, Computer Games and Digital Cultures (CGDC), Conference proceedings, 2002, University of Tampere, pp 45-58, Finland.

[5] Zhang, J., Zhao, J., Bai, S., Huang, Z., 2004, Applying speech interface to Mahjong game, proceedings 10th International Conference on Multimedia Modelling, 2004, pp 86-92.

[6] Rudra, T., Tien, D., Bossomaier, T., a, Spoken Communication with Computer Game Characters, 3rd International Conference of ICITA, IEEE, Sydney, Australia, 2005.

[7] Rudra, T., Kavakli, M., Tien, D., Emotion from Game Pidgin Language Using SVM, International Conference on Computing and Informatics (ICOCI 2006), IEEE, Kuala Lumpur, Malayasia, 2006.

[8] PRAAT, Downloading Praat for windows, (Online), URL: http://www.fon.hum.uva.nl/praat/download_win.html, [Accessed, 8th July 2005].

[9] Dellaert, F., Polzin, T., Waibel, A., Recognizing Emotion in Speech, Proceedings of ICSLP, 1996, Philadelphia, USA.

[10] Cawley, C., Matlab Support Vector Machine toolbox, (Online), URL: http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox/, [Accessed, 20th June 2005].

[11] Rudra, T., Bossomaier, T., Cognitive Emotion in Speech interactive Games, TENCON 05 IEEE Region 10 Conference, Melbourne, Australia, 2005.

[12] Hsu, C., Chang, C., Lin, C., A practical Guide to Support Vector Classification, (online), URL: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, [Accessed, 17th December 2005].

[13] Rudra, T., Kavakli, M., Tien, D., 2007, Emotion detection from Female Speech in Computer Games, Spoken Communication with Computer Game Characters, 4th International Conference of ICITA, IEEE, Harbin, China.