Table III. Verification scenario one- phase one

Classification Method	Training Set (80%)	Testing Set (20%)
SMO	94.85%	96.74%
simple Logistic	94.76%	94.93%
LMT	94.76%	94.93%
Random Subspace	93.04%	93.84%
Random Committee	93.32%	91.67%
Bagging	93.59%	93.12%
Iterative Classifier Optimizer	92.68%	92.03%
LogitBoost	92.68%	92.03%
Random Forest	92.86%	92.03%
MultiClass Classifier Updateable	92.68%	91.30%
Classification Via Regression	93.22%	92.03%
K nearest	91.96%	91.30%
Filtered Classifier	89.61%	90.94%
PART	90.51%	87.68%
Rep Tree	90.24%	89.13%
Attribute Selected	87.90%	86.96%
Jrip	87.44%	85.14%
J48	87.53%	90.22%
MultiClass Classifier	86.09%	82.25%
Decision Table	84.64%	82.25%
Random Tree	82.38%	83.70%
Logistic	79.77%	76.09%
AdaBoostM1	85.00%	81.16%
Decision Stump	79.49%	77.17%
OneR	78.95%	74.64%
BayesNet	74.25%	69.20%
HoeffdingTree	67.84%	65.94%
Naive Bayes	67.48%	64.49%
Naive Bayes Updatable	67.48%	64.49%
Naive Bayes Multinomial	69.29%	61.96%
Randomizable Filtered Classifier	64.68%	69.57%
CV Parameter Selection	53.75%	52.90%
Weighted Instances Handler Wrapper	53.75%	52.90%
ZeroR	53.75%	52.90%
InputMapped Classifier	53.75%	52.90%

Table IV. Verification scenario one- phase two

Classification Method	Training Set (80%)	Testing Set (20%)
SMO	97.97%	94.20%
Simple Logistic	97.77%	92.75%
LMT	97.77%	92.75%
Random Subspace	96.71%	89.86%
Random Committee	96.61%	91.67%
bagging	95.84%	90.58%
Iterative Classifier Optimizer	97.00%	89.49%
LogitBoost	97.00%	89.49%
Random Forest	96.22%	89.49%
MultiClass Classifier Updateable	96.13%	89.86%
Classification Via Regression	95.84%	89.13%
K nearest	94.29%	89.86%
Filtered Classifier	91.58%	87.32%
PART	93.22%	86.96%
Rep Tree	91.67%	88.04%
Attribute Selected	91.87%	87.32%
Jrip	92.35%	86.23%
J48	94.00%	89.49%
MultiClass Classifier	85.19%	82.61%
Decision Table	88.19%	84.78%
Random Tree	86.35%	84.42%
Logistic	84.70%	84.78%
AdaBoostM1	88.00%	78.26%
Decision Stump	84.12%	78.26%
OneR	83.93%	75.72%
BayesNet	80.74%	72.10%
HoeffdingTree	72.02%	64.13%
Naive Bayes	71.93%	64.13%
Naive Bayes Updatable	71.93%	64.13%
Naive Bayes Multinomial	72.12%	63.04%
Randomizable Filtered Classifier	67.47%	67.75%
CVParameter Selection	57.12%	52.90%
Weighted Instances Handler Wrapper	57.12%	52.90%
zeroR	57.12%	52.90%
InputMapped Classifier	57.12%	52.90%

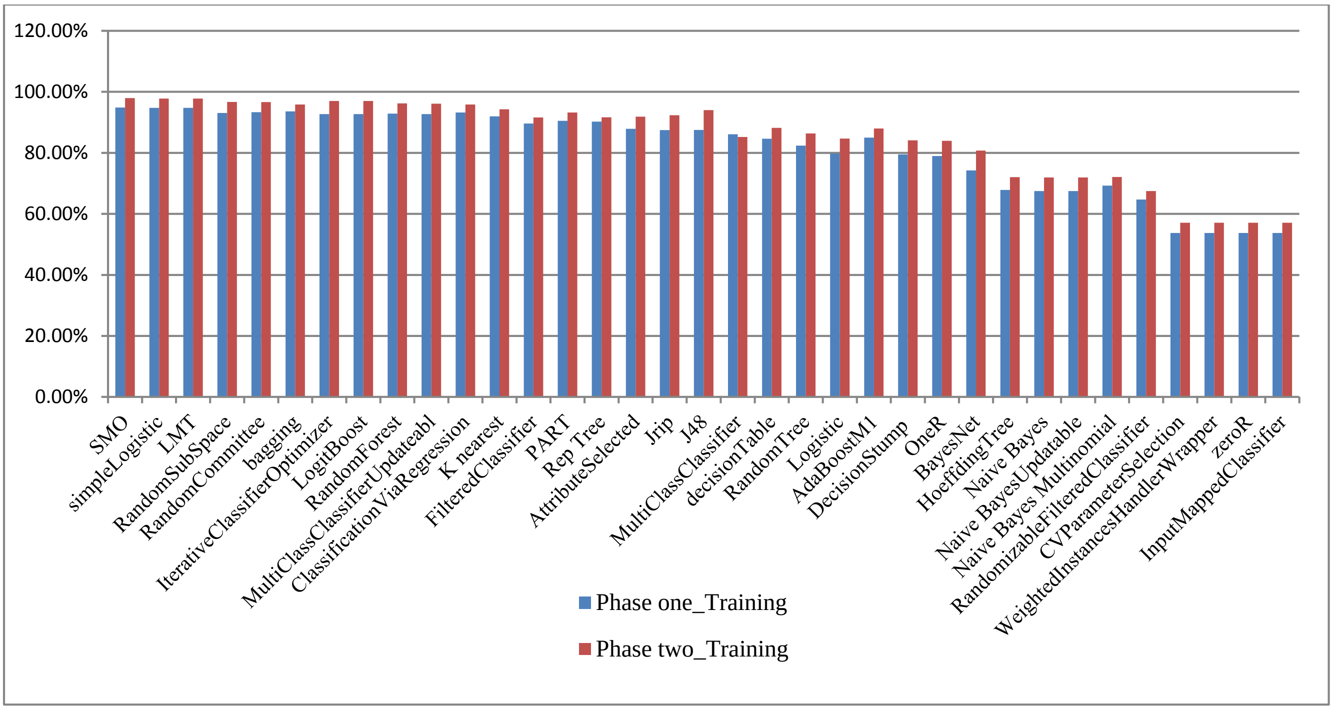

Figure II

In the second verification scenario, the original database into three parts: 60% training dataset (829 records), 20% testing dataset (277 records), and 20% another testing dataset (277 records).The three data sets are mutually exclusive. For phase one, we trained the thirty five classification models with the 60% training dataset and then tested the model with the two testing datasets. Below are the results:

Table V. Verification scenario two- phase one

Classification Method	Training Set	Testing Set 1	Testing Set 2
SMO	93.90%	94.20%	95.95%
simple Logistic	94.35%	93.12%	96.40%
LMT	94.58%	93.12%	96.40%
Random Subspace	92.99%	92.03%	94.59%
Random Committee	92.54%	92.75%	93.24%
bagging	93.67%	91.30%	90.99%
Iterative Classifier Optimizer	93.33%	92.75%	93.24%
LogitBoost	93.33%	92.75%	93.24%
Random Forest	93.33%	90.94%	91.89%
MultiClass Classifier Updateable	92.77%	93.12%	93.24%
Classification Via Regression	91.19%	91.30%	90.54%
K nearest	91.41%	91.30%	92.34%
Filtered Classifier	89.83%	89.49%	89.19%
PART	87.46%	84.78%	87.39%
Rep Tree	89.72%	86.23%	86.49%
Attribute Selected	87.68%	86.96%	87.39%
Jrip	87.01%	85.87%	83.33%
J48	88.36%	89.13%	87.84%
MultiClass Classifier	80.57%	75.36%	80.18%
Decision Table	82.03%	82.97%	81.98%
Random Tree	79.55%	81.16%	81.08%
Logistic	76.05%	75.72%	77.48%
AdaBoostM1	82.60%	80.43%	82.43%
Decision Stump	80.34%	77.54%	76.13%
OneR	78.64%	76.09%	77.03%
BayesNet	73.67%	68.84%	73.42%
HoeffdingTree	65.99%	62.68%	68.02%
Naive Bayes	64.63%	62.68%	69.37%
Naive Bayes Updatable	64.63%	62.68%	69.37%
Naive Bayes Multinomial	69.60%	63.04%	68.02%
Randomizable Filtered Classifier	61.13%	64.86%	69.37%
CVParameter Selection	54.58%	52.90%	50.45%
Weighted Instances Handler Wrapper	54.58%	52.90%	50.45%
zeroR	54.58%	52.90%	50.45%
InputMapped Classifier	54.58%	52.90%	50.45%

Next we applied phase two by removing the videos that were misclassified by more than 16 methods from the training dataset only and the classification models were all trained. The models are then tested twice by the two testing datasets. All classifiers have shown a drawback when applied on the new training dataset except the OneR and the Randomizable Filtered Classifier that have improved 2% and 4% respectively. The PART and Attribute Selected models have shown an improvement of 1% only. For the testing set one, eight methods have shown enhancement between 1% and 3% while the remaining have shown a drawback. The Jrip classifier improved 5% when applied on the second testing set, five classifiers improved 1% up to 2.7% while the remaining have shown a drawback of maximum 4%.

ANALYSIS

In this section we analyze and compare the results of the proposed enhancement model and the two verification scenarios. The two-phase enhancement model aims at building a classification model that learns only correct (well- classified) speech units. Obviously, removing the misclassified videos will improve the accuracy since the dataset contains "good" data. However, in the first verification scenario, we notice an enhancement in the training set but not in the testing set. We believe that this is due to the fact that the training set doesn't contain records of similar feature values as the testing set. Since there is no another ready to use Arabic speech corpora to be used as testing database, we proposed another verification scenario where the training set is 60% of the original database and two testing sets (20% each) are tested. Unfortunately, both training and testing sets didn't show any improvement since obviously the training dataset is considered to be small and not enough for the classifier to learn. A good solution for this problem is to increase the training dataset and add more videos to the corpora so that the classifier will be able to distinguish "bad" records from "good" ones more accurately.

TableVI. Verification scenario two- phase two

Classification Method	Training Set-modified	Testing Set 1	Testing Set 2
SMO	92.61%	94.20%	94.59%
Simple Logistic	93.64%	95.65%	95.50%
LMT	93.64%	93.48%	95.50%
Random Sub Space	91.31%	90.22%	90.54%
Random Committee	91.31%	90.22%	93.69%
bagging	91.44%	90.22%	93.69%
Iterative Classifier Optimizer	91.44%	90.22%	91.44%
Logit Boost	91.44%	90.22%	91.44%
Random Forest	91.57%	91.30%	90.09%
MultiClass Classifier Updateable	91.18%	90.22%	92.79%
ClassificationVia Regression	90.14%	88.77%	87.84%
K nearest	90.14%	92.01%	91.89%
Filtered Classifier	85.73%	87.68%	88.29%
PART	88.46%	86.23%	88.74%
Rep Tree	86.25%	89.13%	89.19%
Attribute Selected	89.11%	87.68%	87.39%
Jrip	85.99%	83.33%	89.19%
J48	86.25%	86.23%	84.23%
MultiClass Classifier	76.13%	75.36%	77.48%
Decision Table	82.62%	82.97%	82.43%
Random Tree	78.99%	81.88%	82.88%
Logistic	75.10%	71.74%	71.62%
AdaBoostM1	82.36%	80.43%	81.53%
Decision Stump	79.51%	77.17%	76.13%
OneR	80.67%	78.26%	77.03%
BayesNet	73.41%	69.93%	75.23%
HoeffdingTree	64.46%	63.04%	68.02%
Naive Bayes	63.42%	63.77%	67.57%
Naive Bayes Updatable	63.42%	63.77%	67.57%
Naive Bayes Multinomial	67.96%	61.59%	68.47%
Randomizable Filtered Classifier	64.72%	67.03%	64.86%
CVParameterSelection	55.12%	52.90%	50.45%
Weighted Instances Handler Wrapper	55.12%	52.90%	50.45%
zeroR	55.12%	52.90%	50.45%
InputMapped Classifier	55.12%	52.90%	50.45%

CONTRIBUTIONS AND FUTURE WORK

In this paper we introduce an enhancement model to improve the recognition of emotions from Arabic speech. This contribution is language independent and may be used by other researchers to improve their results. Increasing the corpus by adding more speech units would be of great benefit to build a successful classification model to recognize happy, angry and surprised emotions from Arabic speech.

REFERENCES

[1] Klaylat, S., Zantout, R., Hamandi, L. and Osman, Z., 2017," Emotion Recognition in Arabic Speech," Proc. IEEE International Conference on Sensors, Networks, Smart and Emerging Technologies, Beirut, Lebanon.

[2] Ekman, P., 1971, "Universals and Cultural Differences in Facial Expressions of Emotion," Proc. Nebraska Symp. Motivation, pp. 207-283.

[3] Ekman, P., 1982, "Emotion in the Human Face," 2nd ed. Cambridge Univ. Press.

[4] Ekman P. and Oster, H., 1979, "Facial Expressions of Emotion," Ann. Rev. Psychology, 30, pp. 527-554.

[5] Clavel, C., Vasilescu, I., Devillers, L. Richard, G., and Ehrette, T., 2008, "Fear-Type Emotion Recognition for Future Audio-Based Surveillance Systems," Speech Communication, 50(6), pp. 487-

503.

[6] Cowie, R. and Cornelius, R., 2003," Describing the Emotional States that are Expressed in Speech," Speech Communication, 40, pp. 5-32.

[7] Kehrein, R., 2002, "The Prosody of Athentic Emotions," Proc. Speech Prosody, Aix-en-Provence, pp. 423–426.

[8] Douglas-Cowie, E., Devillers, L., Martin, J.C., Cowie, R., Savvidou, S., Abrilian, S. and Cox, C., 2005,"Multimodal Databases of Everyday Emotion: Facing Up to Complexity," Proc 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, pp. 813-816.

[9] Devillers, L., Vidrascu, L., and Lamel. L., 2005, "Challenges in Real-Life Emotion Annotation and Machine Learning Based Detection," Neural Networks, 18(4), pp. 407- 422.

[10] Engberg, I., and Hansen, A., 1996, "Documentation of the Danish Emotional Speech Database DES," Center for Person Kommunikation, Institute of Electronic Systems, Alborg University, Denmark.

[11] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. and Weiss, B. 2005, "A Database of German Emotional Speech," Proc. of Interspeech, Lisbon, pp. 1517–1520.

[12] University of Pennsylvania Linguistic Data Consortium, Emotional prosody speech and transcripts, /http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC200228S, July, 2002.

[13] Jovicic, S.T., Kacic, Z., Dordevic, M., and Rajkovic, M., 2004, "Serbian Emotional Speech Database: Design, Processing and Evaluation," Proc. 9th Conference on Speech and Computer, SPECOM'04, St. Petersburg, Russia, pp. 77–81.

[14] Nwe, T.L., 2003," Analysis and Detection of Human Emotion and Stress from Speech Signals," Ph.D. thesis, Department of Electrical and Computer Engineering, National University of Singapore.

[15] Breazeal, C., and Aryananda, L., 2002, "Recognition of Affective Communicative Intent in Robot-Directed Speech," Autonomous Robots 12(1), pp. 83–104.

[16] Zhou, J., Wang, G., Yang, Y., and Chen, P., 2006, "Speech Emotion Recognition Based on Rough Set and SVM," Proc.5th IEEE International Conference on Cognitive Informatics, ICCI, 1, pp. 53–61.

[17] Busso, C., Bulut, M., Lee, C.C. , Kazemzadeh, A., Mower, E., Kim, S., Chang, J.N., Lee, S., and Narayanan, S.S., 2008, " IEMOCAP: Interactive emotional dyadic motion capture database," Journal of Language Resources and Evaluation, 42 (4), pp.335-359.

[18] Schiel, F., Steininger, S., and Turk, U., 2002," The SmartKom Multimodal Corpus at BAS," Proc. 3rd Language Resources and Evaluation Conference, LREC'02, Canary Islands, Spain, pp. 200–206.

[19] Batliner, A., Hacker, C., Steidl, S., Noth, E., D'Arcy, S., Russell, M., and Wong, M. 2004, " 'You stupid tin box' – Children Interacting with the AIBO Robot: A Cross-Linguistic Emotional Speech Corpus." Proc. 4th Language Resources and Evaluation Conference LREC, Lisbon, Portugal, pp.171–174.

[20] Grimm, M., Kroschel, K., and Narayanan, S., 2008,"The Vera am Mittag German Audio–Visual Emotional Speech Database," Proc. IEEE International Conference on Multimedia and Expo (ICME), Hannover, Germany, pp. 865–868.

[21] Morrison, D., Wang, R. and De Silva, L., 2007," Ensemble Methods for Spoken Emotion Recognition in Call-Centers," Speech Communication, 49(2), pp. 98–112.

[22] Wiggins, J. S., 1996 "The Five-Factor Model of Personality". Guilford Press, New York.

[23] Mohammadi, G., Vinciarelli, A., and Mortillaro, M., 2010," The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions," Proc. SSPW, Florence, pp. 17–20.

[24] Schuller, B., Muller, R., Eyben, F., Gast, J., Hornler, B., Wollmer, M., Rigoll, G., Hothker, A., and Konosu, H., 2009, "Being Bored? Recognizing Natural Interest by Extensive Audiovisual Integration for Real-Life Application," Image and Vision Computing, 27, pp. 1760–1774.

[25] Vidrascu, L., and Devillers, L., 2006," Real-Life Emotions in Naturalistic Data Recorded in A Medical Call Center," Proc. 1st International Workshop on Emotion: Corpora for Research on Emotion and Affect (International conference on Language Resources and Evaluation, Genoa, Italy, pp. 20-24.

[26] Koolagudi, Sh., and Sreenivasa Rao, K., 2012, "Emotion Recognition from Speech: a Review," Int J Speech Technol.

[27] Mubarak, O. M., Ambikairajah, E., and Epps, J., 2005, "Analysis of an MFCC-Based Audio Indexing System for Efficient Coding of Multimedia Sources," Proc. 8th International Symposium on Signal Processing and its Applications, Sydney, Australia, pp. 28–31.

[28] Pao, T. L., Chen, Y. T., Yeh, J. H., and Liao, W. Y., 2005, "Combining Acoustic Features for Improved Emotion Recognition in Mandarin Speech," in ACII (LNCS 3784), Springer-Verlag Berlin Heidelberg , pp. 279–285.

[29] Pao, T. L., Chen, Y. T., Yeh, J. H, Cheng, Y. M., and Chien, C. S., "Feature Combination for Better Differentiating Anger from Neutral in Mandarin Emotional Speech,". LNCS 4738, ACII 2007: Springer-Verlag Berlin Heidelberg, 2007.

[30] Davis, S., and Mermelstein, P., 1980," Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences,"IEEE Trans. Audio, Speech Lang Process, 28 pp. 357–366.

[31] Rabiner, L. R., and Juang, B. H., 1993," Fundamentals of Speech Recognition," Englewood

Cliffs, New Jersy, Prentice-Hall.

[32] Chauhan, A., Koolagudi, S.G., Kafley, S., and Rao, K.S., 2010," Emotion recognition using LP residual," Proc.IEEE TechSym, West Bengal, India.

[33] Koolagudi, S. G., Reddy, R., and Rao, K.S., 2010, "Emotion Recognition from Speech Signal Using Epoch Parameters," Proc. International conference on signal processing and communications (SPCOM), IISc, Bangalore, India, pp. 1–5, New York: IEEE Press.

[34] I, A. I., and Scordilis, M. S., 2001," Spoken Emotion Recognition Using Glottal Symmetry," EURASIP Journal on Advances in Signal Processing, , 1(11).

[35] Makhoul, J., 1975, "Linear Prediction: A Tutorial Review," Proc. of the IEEE, 63(4), pp. 561–580.

[36] Murray, I.R., and Arnott, J.L., 2008,"Applying an Analysis of Acted Vocal Emotions to Improve the Simulation of Synthetic Speech," Computer Speech and Language, 22(2), pp. 107–129.

[37] Li, Y., and Zhao, Y., 1999, "Recognizing Emotions in Speech Using Short-Term and Long Term Features," Eurospeech, Budapest.

[38] Wang, Y. Meng, Q., and Li, P., 2009, "Emotional Feature Analysis and Recognition in Multilingual Speech Signal," Proc. Electronic Measurement and Instruments (ICEMI), Beijing.

[39] Vidrascu, L., and Devillers, L., 2007,"Five Emotion Classes Detection in Real-World Call Center Data: The Use of Various Types of Paralinguistic Features", LIMSI-CNRS, France.

[40] Ayadi, M.E., Kamel, M.S., and Karray, F., 2011, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases," Pattern Recognition, pp. 572–587.

[41] Xie, B., Chen, L., Chen, G. C., and Chen, C., 2007, "Feature Selection for Emotion Recognition of Mandarin Speech," Journal of Zhejiang University (Engineering Science), 41(11), pp. 1816–1822.

[42] Chuang, Z.J. and Wu, C.H., 2004, "Emotion Recognition Using Acoustic Features and Textual Content," Proc. IEEE international conference on multimedia and expo, 1, pp. 53–56.

[43] Yu, F., Chang, E., Xu, Y.Q. , and Shum, H.Y., 2001," Emotion Detection from Speech to Enrich Multimedia Content," Proc. 2nd IEEE Pacific-Rim conference on multimedia, Beijing, China.

[44] Hoch, S., Althoff, F., McGlaun, G., and Rigoll, G., 2005, "Bimodal Fusion of Emotional Data in an Automotive Environment," Proc. IEEE international conference on acoustics, speech, and signal processing, 2, pp. 1085–1088.

[45] Hoch, S., Althoff, F., McGlaun, G., and Rigoll, G., 2005 "Bimodal Fusion of Emotional Data in an Automotive Environment," Proc. IEEE international conference on acoustics, speech, and signal processing, 2, pp. 1085–1088.

[46] Lugger, M., and Yang, B., 2007, "The Relevance of Voice Quality Features in Speaker Independent Emotion Recognition," Proc. ICASSP, Honolulu, Hawaii, USA, pp. IV17–IV20. New York: IEEE Press.

[47] Iliou, T., and Anagnostopoulos, C. N., 2009, "Statistical Evaluation of Speech Features for Emotion Recognition," Proc. 4th international conference on digital telecommunications, Colmar, France, pp. 121–126.

[48] Luengo, E., Navas, I., Hernez, I., and Snchez, I., 2005," Automatic Emotion Recognition Using Prosodic Parameters," Proc. INTERSPEECH, Lisbon, Portugal, pp. 493–496.

[49] Lee, C.M., Narayanan, S.S., and Pieraccini, R., 2001, "Recognition of Negative Emotions from the Speech Signal," Proc. ASRU, pp. 240–243.

[50] Litman, D.J. and Forbes-Riley, K., 2006, "Recognizing Student Emotions and Attitudes on the Basis of Utterances in Spoken Tutoring Dialogues With Both Human and Computer Tutors," Speech Comm., 48, pp. 559–590.

[51] http://android-apps.com/apps/skc-interpret/, Android Apps website.

[52] http://appcrawlr.com/android/sprint-mobile-ip, Sprint Mobile IP, App Crawlr website.

[53] Petrushin, V., 2000, "Emotion Recognition in Speech Signal: Experimental Study, Development, and Application," Proc. 6th International Conference on Spoken Language Processing, Beijing, China.

[54] Burkhardt, F., Eckert, M., Johannsen, W., and Stegmann, J., 2010, "A Database of Age and Gender Annotated Telephone Speech," Proc. LREC, Valletta, Malta, pp. 1562–1565.

[55] Schiel, F., and Heinrich, C. , 2009," Laying the Foundation for In-Car Alcohol Detection by Speech," Proc. Interspeech, Brighton, UK, pp. 983–986.

[56] Burkhardt, F., Schuller, B., Weiss, B. and Weninger, F., 2011," 'Would you Buy a Car from Me?' – On The Likability of Telephone Voices," Proc. Interspeech, Florence, pp. 1557–1560.

[57] Mohammadi, G., Vinciarelli, A., and Mortillaro, M., 2010," The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions," Proc. SSPW, Florence, pp. 17–20.

[58] http://www.youtube.com/watch?v=uvhNyAXFTMQ, "Egypt today show", Alfaraiin channel, May 25, 2012.

[59] http://www.youtube.com/watch?v=S1T_EKDpIR8," New cairo show", Al-hayat channel, May 28, 2012

[60] http://www.youtube.com/watch?v=2v6X2VEjb4k , "Laka sumt , AlUraify show", September 12, 2012.

[61] http://www.youtube.com/watch?v=MQv3tKTwm7k, Zain telecommunication, January 22, 2009.

[62] http://www.youtube.com/watch?v=16qNcn03G3s, Prince Sultan bin Fahed call, Alriyadiya sport channel, October 6, 2011.

[63] http://www.youtube.com/watch?v=E4TqhBo1SCk, Althaqafiya channel, January 7, 2012.

[64] http://www.youtube.com/watch?v=Wpf3OxEdJak, "Dairat al dawe", Noon channel, Haifa wehbe call, November 19, 2009.

[65] http://www.youtube.com/watch?v=eBznv9QNU7M, "Musalsalati show", Mona Zaki call, June 12, 2011.

[66] Schuller, B., Steidl, S. and Batliner , A., 2009, "The Interspeech 2009 Emotion Challenge," Proc. Interspeech, Brighton, UK, , pp. 312–315.

[67] Eyben, F., Wöllmer, M. and Schuller, B., 2010, "openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor," ACM.

[68] https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php, "Kruskal-Wallis H Test Using SPSS Statistics", Leard Statistics website.

[69] Platt, J.C., 1998," Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", Technical Report MSR-TR, pp. 98-14.

[70] Vapnik, V., 1995, "The Nature of Statistical Learning Theory," Springer, New York.

[71] Fisher, W. Doddington, G., and Goudie-Marshall, K., 1986, "The DARPA Speech Recognition Research Database: Specifications and Status," Proc. of the DARPA Workshop on Speech Recognition, pp. 93–99.

[72] Vlasenko, B., and Wendemuth, A., 2009, "Processing Affected Speech within Human Machine Interaction," Proc. Annual Conf. of the International Speech Communication Association, Interspeech.

[73] Vidrascu, L., and Devillers, L., 2005," Real-Life Emotion Representation And Detection in Call Centers Data," LNCS, 3784, ACII, Berlin, Springer pp. 739–746.

[74] Zhang, S., 2008, "Emotion Recognition in Chinese Natural Speech by Combining Prosody and Voice Quality Features," Proc. Sun, et al. (Eds.), Advances in neural networks. Lecture notes in computer science, Berlin, Springer, pp. 457–464.