# INTRODUCING UNDERGRADUATES TO PATTERN RECOGNITION AND MACHINE LEARNING THROUGH SPEECH PROCESSING

*Ravi P. Ramachandran, Kevin D. Dahm and Nidhal C. Bouaynaya*

College of Engineering
Rowan University
Glassboro, New Jersey, 08028, USA
ravi@rowan.edu, dahm@rowan.edu, bouaynaya@rowan.edu

## ABSTRACT

This paper describes an educational project experience that achieves a software implementation and performance analysis of a blind signal to noise ratio (SNR) estimation system for noisy speech. The system is based on a pattern recognition paradigm and no clean speech reference signal is available. It is a product of the faculty's research and funded by a government contract. The faculty's research on a real-world issue in speech processing has been converted into an undergraduate project. Assessment results show that the project is viewed very favorably by students. Target versus control group results show that the target group feels better qualified for graduate study and career options in digital signal processing.

***Index Terms***— project based learning, blind SNR estimation, pattern recognition, quantitative assessment

## 1. INTRODUCTION

Project based learning (PBL) [1] has been shown to be an effective method of achieving many undergraduate student learning outcomes. These include but are not limited to improvement in analytical, software, design, communication and critical thinking skills. The projects can also be student driven in that there can be an open-ended component. Projects related to the real world can be introduced to enrich a single course without sacrificing the coverage of required technical content [2]. This can better prepare students for employment and graduate school particularly if there is engagement from industry and/or government [3]. It has also been shown that PBL can be achieved at all levels of the undergraduate curriculum especially through the use of vertical integration [4][5][6] in which concepts and project experiences at a certain level build upon what has been previously learned.

Configuring a project based on a modern topic is essential. The application of digital signal processing for pattern recognition, machine learning and artificial intelligence is a key area that undergraduate students should be exposed to. Artificial intelligence is a major player in today's marketplace. There is great investment of human and financial resources by global giants like Google, Facebook, Microsoft and Baidu [7]. There are widespread applications like speech recognition, natural language processing, biometrics, data mining, computer vision, telemedicine, mobile computing, image understanding, mobile computing and the internet of things. Presently, the coverage of topics in pattern recognition and machine learning is focused at the graduate level. It is imperative to bring these topics to the undergraduate level so that students acquire a basic comprehension of key concepts.

Projects in applying digital signal processing to pattern recognition and machine learning have been configured. Three types of biometric systems (speaker, face and iris recognition) are discussed in [4]. In [7], students learn about different classifiers [8] in the context of a particular application. The $k$ nearest neighbor and support vector machines recognize handwritten numerals. Decision trees identify the type of contact lens. A Naive Bayes classifier filters spam messages. Other examples of real-world signal processing projects include the use of a digital stethoscope to record and extract vital information from a heartbeat signal [9], software defined radio [10], spectrum estimation of electrocardiagram signals [10], brain-computer interface [11] and multispectral signal processing of infra-red and visible images [12].

The project described in this paper teaches the basic concepts of pattern recognition and machine learning to undergraduate students. It can be run either at the junior (Discrete Signal Processing class) or senior level (any course in Signal Processing, Speech, Pattern Recognition or Machine Learning). The objective is to blindly (no clean reference signal) estimate the signal to noise ratio (SNR) of a speech signal corrupted by additive noise and is an education project derived from the faculty's recent research [13]. Estimation of the SNR is important as it can be an important pre-processing step for speaker recognition [14][15], speech recognition [16] and speech enhancement [17].

Section 2 describes how the project evolved and puts the previous work into context. Section 3 gives a detailed descrip-

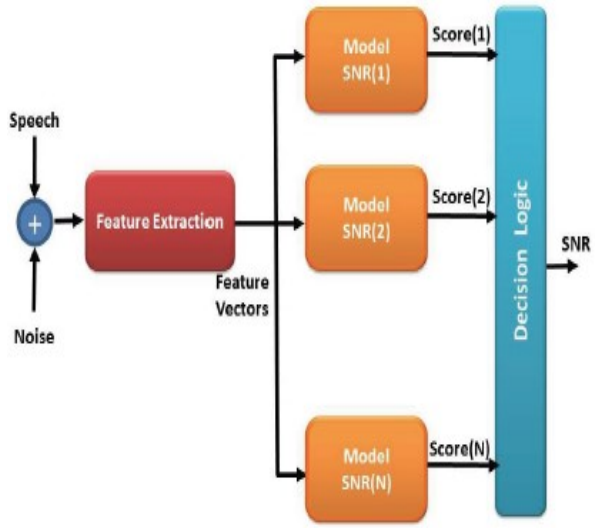


**Fig. 1**. Block diagram of SNR estimation system (taken from [13]

tion of the project. Section 4 gives the assessment results.

## 2. EVOLUTION OF PROJECT

The blind SNR estimation system as shown in Figure 1 is a product of the faculty's research and funded by a government contract [13]. This motivates its use for education as it is a real-world problem. Figure 1 depicts a typical pattern recognition scenario. The TIMIT database is used to get the clean speech which was downsampled from 16 kHz to 8 kHz. Additive white noise from the NOISEX database is used. A speech utterance is corrupted by noise at a particular SNR. Frame by frame processing is done to extract a collection of feature vectors. The feature vector is the linear predictive cepstrum. The feature vectors are passed through a set of classifier models each trained on a particular SNR value. As in [13], the classifier model is a vector quantizer (VQ) codebook designed by the Linde-Buzo-Gray algorithm. A Gaussian Mixture Model classifier [8][18] is found to be similar in performance to a vector quantizer [19]. The value of $Score(i)$ is the output of model $i$ for $i = 1$ to $N$. There are $N = 34$ models trained for SNR values ranging from -1 dB to 32 dB in steps of 1 dB. A hard decision can be taken by declaring the estimate to correspond to the model yielding the optimum score. In [13], a soft decision was found to improve performance. The soft decision takes a weighted linear combination of the SNR estimates corresponding to the best three scores.

The first education task was to have a team of two students (one junior and one senior) achieve a software (MATLAB) implementation of the system in Figure 1, write a laboratory manual for running the project in a course and participate in decisions about distinguishing what can be assigned to a junior and senior course [19]. The two students accomplished this as part of the undergraduate project experience and all required technical knowledge was taught. At the time [19] was published, the writing of the laboratory manual was still a work in progress. This paper gives the details of the laboratory manual and gives assessment results based on running this project in a junior level class.

## 3. DESCRIPTION OF PROJECT

The sections of the laboratory project manual are described. Students need to write a formal report with a title page, table of contents, summary and conclusions and include appropriate references. A minimum of three references is required.

### 3.1. Introduction

Students are asked to discuss the motivation of doing this project and to formulate a list of the objectives.

### 3.2. Background on SNR Estimation

Students read reference [13] and write a synopsis on what they have learned.

### 3.3. Description of TIMIT Database

Students research the TIMIT database and write a description of the database. The following specific questions are to be addressed. What is the significance of the 'sa', 'si' and 'sx' sentences? What other speech processing applications has TIMIT been used for?

In this project, 90 speakers from the TIMIT database will be used. The first 5 sentences of each speaker will be used to train the SNR estimation system. The remaining 5 sentences will be used for performance evaluation.

### 3.4. Adding Noise to the Speech

Students are taught about signal power, noise power and the meaning of signal to noise ratio. The mathematics relating to adding noise to a signal at a specified SNR is discussed.

Students write a MATLAB function to add white noise to a speech signal at a specified SNR. The white noise file from the NOISEX database is used. Students are then asked to take any speech file from the TIMIT database and add noise at SNR values equal to 30 dB, 20 dB, 10 dB and 0 dB. They are to listen to the clean speech and the noisy speech at the various SNRs and record their observations.

### 3.5. Feature Extraction

At the junior level, code to read in a speech signal, do frame by frame processing and extract a collection of 12 dimensional linear predictive cepstrum vectors is provided. At the

senior level, a separate lab on frame by frame processing for feature extraction is assigned prior to the project.

Students randomly select one of the first five sentences for each of the 90 speakers. Using all of the 90 speech files, the feature vectors are computed when the speech is corrupted by noise at an SNR of 10 dB. The distance between each pair of feature vectors is then calculated. These are the intraclass distances for 10 dB. The process is repeated to get the intraclass distances for 15 and 20 dB. A plot of the probability density functions of the intraclass distances for 10, 15 and 20 dB is interpreted.

The distance between each pair of feature vectors for speech corrupted by 10 and 15 dB is calculated. These are the interclass distances between the classes of 10 dB and 15 dB. This is repeated to get the interclass distances between the classes of (1) 15 dB and 20 dB and (2) 10 dB and 15 dB. A plot of the probability density functions for the 3 sets of interclass distances are compared and interpreted. Also, a comparison is done with the results obtained for the intraclass distances.

### 3.6. Training of Vector Quantizer Based SNR Estimation System

As mentioned earlier, the first 5 sentences of each speaker will be used to train the SNR estimation system. A vector quantizer (VQ) codebook serves as the model for each SNR value from -1 dB to 32 dB in steps of 1 dB. Each VQ codebook is designed using 450 training speech utterances. Each of the 34 codebooks has a size of 256. Code to implement the Linde-Buzo-Gray algorithm is supplied.

### 3.7. Performance Evaluation of the SNR Estimation System

As mentioned earlier, 5 sentences from each speaker not used in training will be used for performance evaluation. There are 450 test speech utterances for each SNR value. The tested SNR values range from 0 to 30 dB inclusive. For a speech utterance corrupted with noise at a certain SNR, feature extraction is first performed to get a collection of test feature vectors. A particular test feature vector is quantized by each of the 34 codebooks by finding the closest codevector in each codebook. There are 34 different distances, one for each codebook. This process is repeated for every test feature vector. The distances are accumulated over the entire set of feature vectors. This accumulated distance is the score for each codebook.

There are two methods of estimating the SNR for the speech utterance [13]. For a hard decision, the SNR estimate corresponds to the codebook which gives the smallest score [13]. In the soft decision approach, the 3 lowest scores are used to estimate the SNR [13]. Let the 3 smallest scores be denoted as Score(1), Score(2) and Score(3). The corresponding SNRs are denoted as SNR(1), SNR(2) and SNR(3). The following equations describe the soft decision approach [13].

$$\text{Total} = \sum_{j=1}^{3} \text{Score}(j) \tag{1}$$

The scores are converted to probabilities (denoted by Prob). Lower scores have higher probabilities as given by

$$\text{Prob}(i) = \left[ \frac{\text{Total} - \text{Score}(i)}{2\text{Total}} \right] \tag{2}$$

for $i = 1$ to 3. From the probabilities, the SNR is estimated as

$$\text{SNR} = \sum_{j=1}^{3} \text{Prob}(j)\text{SNR}(j) \tag{3}$$

For each tested SNR value, there will be 450 SNR estimates, one for each test speech file. The absolute error for each test file as —True SNR value - SNR estimate—. The average absolute error (AAE) is the average of these 450 error values. An AAE is calculated for each SNR from 0 to 30 dB.

Students write MATLAB code that accomplishes soft decision. A parallel implementation is not mandatory but if accomplished, will satisfy the open-ended component described below. The code plots the AAE versus SNR. The plot is to be interpreted and observations recorded. The code is written to accommodate clean speech. The basic question is "What is the average SNR estimate for clean speech? Interpret the result."

### 3.8. Open-Ended Component

At the senior level, an open-ended component is compulsory. At the junior level, extra credit is given. Suggestions given include trying another classifier, another feature and other types of noise. Another classifier or feature can be compared to the implemented system and fusion can be attempted. A parallel implementation of the algorithm can be achieved.

### 3.9. Conclusions

Students are asked to clearly state what they learned and how well the objectives they listed in the Introduction were satisfied.

## 4. ASSESSMENT RESULTS

The assessment is based on running the project in a junior level digital signal processing class. This is their first exposure to discrete signals and systems and the prerequisite is Analog Signals and Systems. A target group of 57 students performed the SNR estimation project. A control group of 29 students performed a digital signal processing hardware project involving filter design and implementation. This is the only difference between the two groups as all technical content is the same for both groups.

| 1 - Strongly disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree | |
|---|---|
| Statement | Mean |
| The laboratory project as a whole helped reinforce MATLAB software skills. | 4.33 |
| The laboratory project as a whole helped reinforce written communication skills. | 3.82 |
| The laboratory project provided me with a basic background in an application area of signal processing. | 4.25 |
| The laboratory project helped me gain experience on the performance aspects of a signal processing system. | 4.27 |
| The laboratory project has raised awareness of signal processing as a field. | 4.31 |
| The laboratory project has provided a fundamental background in digital signal processing | 4.35 |

**Table 1**. Project outcome survey results

| 1 - Strongly disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree | | |
|---|---|---|
| Statement | Mean for Target Group | Mean for Control Group |
| I believe that the knowledge set and skills I have obtained in this class make me better qualified for graduate study and/or career options in digital signal processing. | 4.07 | 3.36 |
| I am now more likely to follow popular media news / developments / programs that relate to digital signal processing as compared to the beginning of the semester | 3.29 | 3.23 |

**Table 2**. General survey of target versus control groups

### 4.1. Open-Ended Aspect

For juniors, this was optional. Five students accomplished a parallel implementation of the system and compared the running time to a serial implementation. The reduction in running time was statistically significant.

### 4.2. Target Group Surveys

A survey relating to the learning outcomes of the project was given to the target group only. Table 1 gives the results and shows that the target group viewed the project very favorably.

### 4.3. Target Group Versus Control Group

Table 2 shows that students who experienced the project (target group) also viewed the course overall more favorably than those who did not, in terms of preparedness in the area of digital signal processing. The mean response to the first question about being "better qualified for graduate study and/or career options in digital signal processing" was higher by 0.71 for the target group. This difference was statistically significant based on a one-tailed t-test with unequal variances in that the p-value [20] is less than 0.001. The target group gave a fractionally higher response to the second question about being "more likely to follow popular media relating to digital signal processing", but this difference is not statistically significant.

The target and control groups are asked the question, "Indicate your level of interest in taking each of the following courses electives". A Likert scale [21] of 1-5 is used. The list has 12 possible electives. Three are considered directly related to the project. The mean responses for these three courses are:

1. Biometric Signal Processing, Target: 3.20 , Control: 2.73
2. Advanced Digital Signal Processing, Target: 3.42, Control: 3.27
3. Machine learning / Pattern Recognition, Target: 3.65, Control: 3.77

The goal was to determine whether the project inspired interest in the student to further study the area. The single largest difference is the response on "Biometric Signal Processing". However, this difference is not statistically significant. The control group may have a keen interest in signal processing and pattern recognition despite not doing this project.

## 5. SUMMARY AND CONCLUSIONS

Students are exposed to a real-world scenario of applying signal processing to a pattern recognition task for blind SNR estimation. Many learning outcomes are achieved. A target versus control group comparison shows that the target group feels better qualified for future options in digital signal processing. This quantitative result is obtained with statistical significance.

## 6. ACKNOWLEDGEMENT

This work was supported by the National Science Foundation through the IUSE Grant DUE 1610911.

## 7. REFERENCES


1. A. Guerra, R. Ulseth, and A. Kolmos, Editors, *PBL in Engineering Education: International Perspectives on Curriculum Change*, Sense Publishers, 2017.

2. N. Hosseinzadeh and M. R. Hesamzadeh, "Application of Project-Based Learning (PBL) to the Teaching of Electrical Power Systems Engineering", *IEEE Transactions on Education*, Vol. 55, pp. 495–501, November 2012.

3. M. J. W. Lee, S. Nikolic, P. J. Vial, C. H. Ritz, W. Li and T. Goldfinch,"Enhancing Project-Based Learning Through Student and Industry Engagement in a Video-Augmented 3-D Virtual Trade Fair", *IEEE Transactions on Education*, Vol. 59, pp. 290–298, November 2016.

4. R. P. Ramachandran, K. D. Dahm, R. M. Nickel, R. J. Kozick, S. S. Shetty, L. Hong, S. H. Chin, R. Polikar and Y. Tang, "Vertical Integration of Biometrics Across the Curriculum: Case Study of Speaker, Face and Iris Recognition", *IEEE Circuits and Systems Magazine*, pp. 55–69, September 2014.

5. C. Trullemans, L. De Vroey, S. Sobieski, and F. Labrique, "From KCL to Class D Amplifier", *IEEE Circuits and Systems Magazine*, pp. 63—74, January 2009.

6. L. Whitman, D. Malzahn, V. Madhavan, G. Weheba, and K. Krishnan, "Virtual Reality Case Study Throughout the Curriculum to Address Competency Gaps," *International Journal of Engineering Education*, Vol. 20, pp. 690–702, 2004.

7. W. Sun and X. Gao, "The Construction of Undergraduate Machine Learning Course in the Artificial Intelligence Era", *International Conference on Computer Science and Education*, Colombo, Sri Lanka, pp. 62–65, August 2018.

8. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

9. Y. Alqudah, E. Qaralleh and M. Mace, "Enhancing the Teaching of Digital Signal Processing through Project-Based Learning", *International Journal of Online Engineering*, Vol. 9, 2013.

10. T. Schäck, M. Muma and A. M. Zoubir, "Signal Processing Projects at Technische Universität Darmstadt", *IEEE Signal Processing Magazine*, pp. 16–30, January 2017.

11. J. Katona and A. Kovari, "A Brain–Computer Interface Project Applied in Computer Engineering", *IEEE Transactions on Education*, Vol. 59, pp. 319–326, November 2016.

12. C. H. G. Wright and T. B. Welch, "Image Fusion: An Introduction to Multispectral Signal Processing", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Canada, pp. 7001–7005, April 2018.

13. R. Ondusko, R. P. Ramachandran, M. Marbach and L. M. Head, "Blind Signal to Noise Ratio Estimation of Speech Based on Vector Quantizer Classifiers and Decision Level Fusion", *Journal of Signal Processing Systems*, Vol. 89, No. 2, pp. 335—-345, November 2017.

14. M. Frankle and R. P. Ramachandran, "Robust Speaker Identification Under Noisy Conditions Using Feature Compensation and Signal to Noise Ratio Estimation", *IEEE Midwest Symposium on Circuits and Systems*, Abu Dhabi, UAE, pp. 133—-136, October 2016.

15. J. S. Edwards, R. P. Ramachandran and U. Thayasivam, "Robust Speaker Verification With a Two Classifier Format and Feature Enhancement", *IEEE International Symposium on Circuits and Systems*, Baltimore, Maryland, pp. 1946—-1949, May 2017.

16. H. K. Kim, R. V. Cox and R. C. Rose, "Performance Improvement of a Bitstream-Based Front-End for Wireless Speech Recognition in Adverse Environments", *IEEE Transactions on Speech and Audio Processing*, Vol. 10, pp. 591—-604, November 2002.

17. S.-W. Fu, Y. Tsao and X. Lu, "SNR-Aware Convolutional Neural Network Modeling for Speech Enhancement", *Interspeech*, San Francisco, California, pp. 3768—3772, September 2016.

18. I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Springer, 2002.

19. P. Awolumate, M. Rudy, R. P. Ramachandran, N. C. Bouaynaya, K. D. Dahm, S. Nazari and U. Thayasivam, "A Pattern Recognition Approach to Signal to Noise Ratio Estimation of Speech", *ASEE Annual Conference and Exhibition*, Columbus, Ohio, June 2017.

20. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Brooks/Cole Cengage Learning, 2012.

21. A. Joshi, S. Kale, S. Chandel and D. K. Pal, "Likert Scale: Explored and Explained", *British Journal of Applied Science and Technology*, February 2015.