

A PATTERN RECOGNITION APPROACH TO SIGNAL TO NOISE RATIO ESTIMATION OF SPEECH

Mr. Peter Adeyemi Awolunate P.A
Mr. Mitchell Rudy, Rowan University

Rowan University Electrical and Computer Engineering student.

Dr. Ravi P. Ramachandran, Rowan University

Ravi P. Ramachandran received the B. Eng degree (with great distinction) from Concordia University in 1984, the M. Eng degree from McGill University in 1986 and the Ph.D. degree from McGill University in 1990. From October 1990 to December 1992, he worked at the Speech Research Department at AT&T Bell Laboratories. From January 1993 to August 1997, he was a Research Assistant Professor at Rutgers University. He was also a Senior Speech Scientist at T-Netix from July 1996 to August 1997. Since September 1997, he is with the Department of Electrical and Computer Engineering at Rowan University where he has been a Professor since September 2006. He has served as a consultant to T-Netix, Avenir Inc., Motorola and Focalcool. From September 2002 to September 2005, he was an Associate Editor for the IEEE Transactions on Speech and Audio Processing and was on the Speech Technical Committee for the IEEE Signal Processing society. Since September 2000, he has been on the Editorial Board of the IEEE Circuits and Systems Magazine. Since May 2002, he has been on the Digital Signal Processing Technical Committee for the IEEE Circuits and Systems society. His research interests are in digital signal processing, speech processing, biometrics, pattern recognition and filter design.

Nidhal Carla Bouaynaya
Dr. Kevin D. Dahm, Rowan University

Kevin Dahm is a Professor of Chemical Engineering at Rowan University. He earned his BS from Worcester Polytechnic Institute (92) and his PhD from Massachusetts Institute of Technology (98). He has published two books, "Fundamentals of Chemical Engineering Thermodynamics" and "Interpreting Diffuse Reflectance and Transmittance." He has also published papers on effective use of simulation in engineering, teaching design and engineering economics, and assessment of student learning.

Dr. Rouzbeh Nazari, Rowan University

Dr. Nazari is an assistant professor of Civil and Environmental Engineering at Rowan University. His primary research interests are: application of remote sensing in water technologies and environment, resiliency and water reuse, impact assessment of climate change and extreme weather events on cities. Dr. Nazari is has worked with NASA, NOAA, Consortium on Climate Risk in the Urban Northeast, and New York State Resiliency Institute for Storms & Emergencies teams of active researchers who focus on climate issues affecting the urban corridor encompassing the U.S. Northeast. Dr. Nazari has published several book chapters, journal papers and has presented his work in national and international conferences.

Prof. Umashanger Thayasivam, Rowan University

Umashanger Thayasivam an Associate Professor of Statistics from Rowan University. I have expertise in robust estimation with mixture models. My research interests include mixture models, robust estimation and statistical applications in STEM. My interdisciplinary statistical research has spanned diverse areas including statistical data mining, speaker recognition systems, spoof detection, spectrum sensing and network security. I have more than ten journal publications and several conference presentations.

I have been PI/co-PI for several internal and external grants and collaborations, including the ongoing three year collaboration data mining project with Bristol Myers Squibb pharmaceutical company, Rowan University Seed grant for a study of statistical and data mining techniques in the field of network security and computer forensics, as well as College of Science and Mathematics grant for evaluation of data classification techniques. I also am performing Biomarker research aimed at optimizing and verifying

the utility of autoantibody biomarkers for early diagnosis-Biomarker Discovery Center at Rowan SOM. Where I ensure that all of the data evaluation strategies and methodologies employed in the studies will take full advantage of any recent developments, improvements and alternative analytical approaches.

A PATTERN RECOGNITION APPROACH TO SIGNAL TO NOISE RATIO ESTIMATION OF SPEECH

ABSTRACT

A blind approach for estimating the signal to noise ratio (SNR) of a speech signal corrupted by additive noise has been proposed. The method is based on a pattern recognition paradigm using various linear predictive based features, a vector quantizer classifier and estimation combination. Blind SNR estimation is very useful in biometric speaker identification systems in which a confidence metric is determined along with the speaker identity. It is also highly useful as a pre-processing step in speech and speaker recognition systems so that a proper degree of enhancement can be applied to augment system performance. This paper is a work in progress depicting the investigation conducted by two undergraduate students pertaining to (1) further research in SNR estimation and (2) the preparation of a laboratory manual to be used in an undergraduate class.

INTRODUCTION AND BACKGROUND

Estimating the signal to noise ratio (SNR) of a speech signal has interesting practical applications. Moreover, performing a blind SNR estimate [1] without knowledge of a clean reference signal is more relevant to many practical scenarios especially in the area of voice biometrics [2]. Blind SNR estimation is very useful in biometric speaker identification systems in which a confidence metric is determined along with the speaker identity [3]. The confidence metric is partially based on the mismatch between the training and testing conditions of the speaker identification system and SNR estimation is very important in evaluating the degree of this mismatch. A pattern recognition approach to estimate the SNR of speech based on vocal tract features and a vector quantizer classifier has been proposed [1]. This system has also been used as a pre-processing step in speaker identification and speaker verification systems [4][5]. This pre-processing step has been successful in making biometric speaker recognition systems more robust to noise conditions. The SNR estimation is utterance-based as opposed to segment-based.

The educational impact of this project is two-fold:

1. Undergraduate students are initiated into research/development [6] by working on a team to achieve a software implementation of the SNR estimation system. The students will also evaluate the performance of the system by experimenting with different features and classifiers. Producing a student-authored paper in a refereed technical conference is the objective.
2. The students will also write a laboratory manual for a portion of this project to be run in a digital signal processing and/or a speech processing class. The objective is to have the students produce a paper in a refereed education conference.

The learning outcomes for the students engaged in research and for the students doing the project in a class include:

- Enhanced application of math skills
- Enhanced software implementation skills
- Enhanced interest in signal processing
- Enhanced ability to analyze experimental results
- Enhanced communication skills

This project is a work in progress. The work commenced four months ago with a team of two students performing the research and preparing the laboratory manual. This paper will report on the work completed up to now. Involving undergraduate students in research and educational innovations has been highly successful in motivating them to proceed to graduate school [6][7].

RESEARCH ORIENTED TASKS

The first task is to achieve a software implementation of the SNR estimation system described in [1]. The performance of this system will be used as the baseline and students will comprehend how to achieve a modular design. The block diagram of this system [1] is shown in Figure 1.

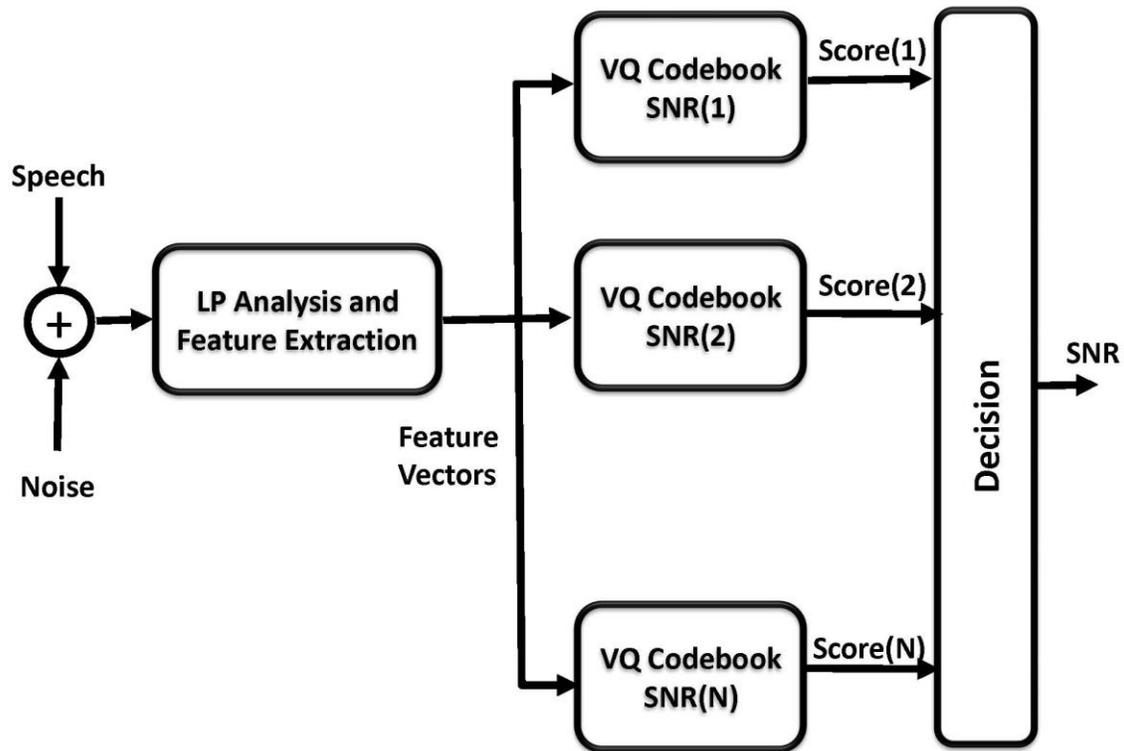


Figure 1 – Block diagram of SNR estimation system (taken from [1])

Due to the time-varying dynamics of the speech signal, the students learn how to perform frame by frame processing as illustrated in Figure 2 [8]. The signal is divided into segments or frames of 30 ms duration. The overlap between adjacent frames is 20 ms. Within a frame, the vocal tract is assumed to be stationary and the calculated features are useful for SNR estimation.

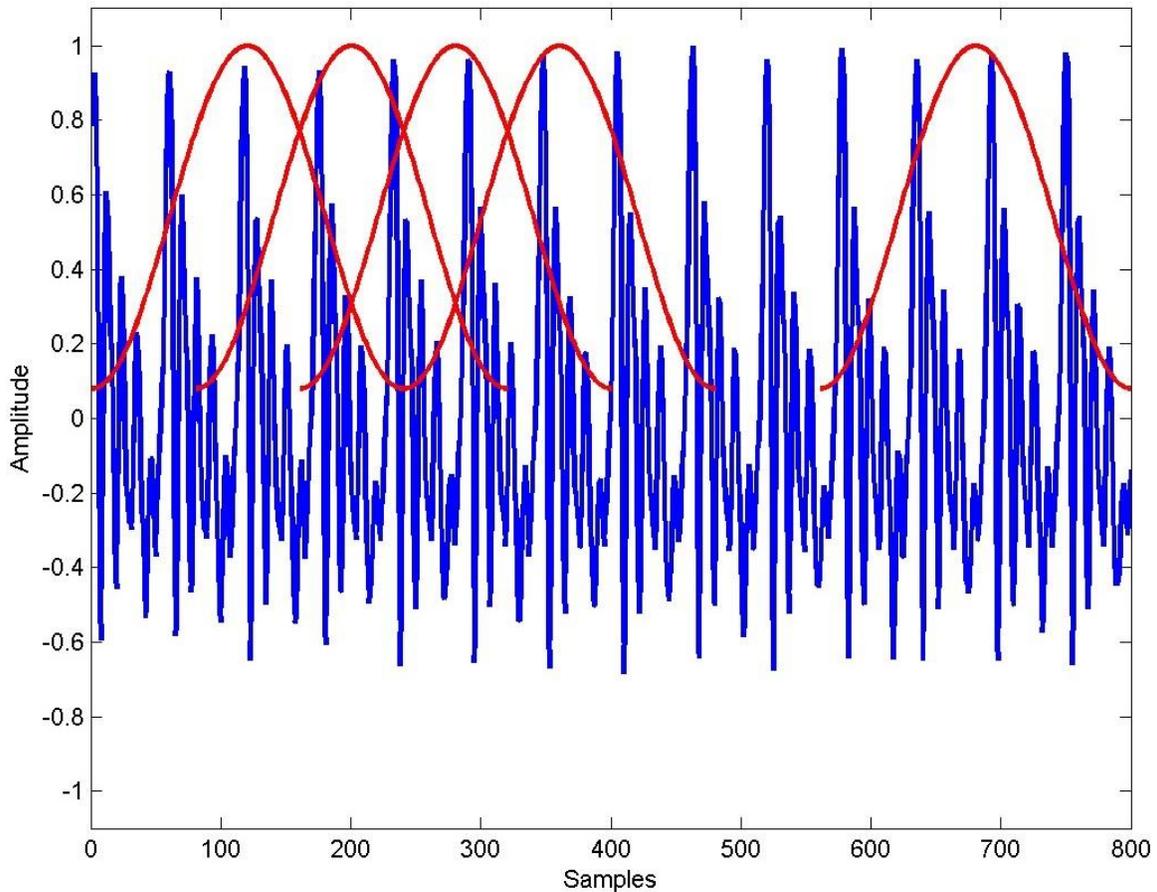


Figure 2 – Illustration of frame by frame processing (taken from [8])

The software modules implemented are as follows:

1. Addition of white noise at a particular utterance based SNR.
2. Linear predictive (LP) analysis on a frame by frame basis to compute the LP coefficients.
3. Conversion of the LP coefficients to the LP cepstrum (CEP) which is the feature used for SNR estimation.
4. Vector quantizer (VQ) codebook training: Use the Linde-Buzo-Gray algorithm to configure codebooks based on the CEP feature computed from speech at various SNRs. In this system, VQ codebooks are designed for SNR values from -1 dB to 32 dB in 1 dB increments.
5. Performance evaluation of the system and interpretation of the results.

Ninety speakers from the TIMIT database were used in the experiments. The utterances in the database were downsampled from 16 kHz to 8 kHz. The first five sentences for

each speaker are used for training the VQ classifier. The remaining five are used for testing. For each SNR value, there are 450 test utterances. For each utterance, the absolute error which equals $|\text{True SNR value} - \text{Estimated SNR value}|$ is determined. For each SNR, there are 450 absolute error values which are averaged to form an average absolute error (AAE) [1]. The testing of the system is performed for SNR values from 0 dB to 30 dB in increments of 1 dB. Therefore, there are a total of 31 AAE values. These 31 values are averaged to form an overall average absolute error (OAAE) as defined in [1].

For a particular test utterance, the CEP features are calculated in each frame and passed through all the VQ codebooks. A soft decision approach to generate the SNR estimate as described in [1] is used. The codebooks with the three best scores are selected. Based on these scores, a weighted linear combination of the SNR estimates corresponding to these three codebooks determines the final SNR estimate. This is known as a soft decision approach [1].

The student research team implemented the VQ based system and a similar system based on a Gaussian mixture model (GMM) classifier. For this case, a GMM model for each SNR value is designed using the Expectation-Maximization (EM) algorithm [9][10]. In achieving this implementation, students gain much insight into the concepts of probability and random variables. The SNR estimate is determined in two ways. The first is using a hard decision in that the SNR specified by the GMM model with the best score is the estimate. The second is a soft decision approach implemented in the same way as for the VQ classifier as described earlier.

The OAAE performance values determined for the VQ and GMM systems is shown in Table 1.

Classifier Approaches	OAAE (dB)
VQ Soft decision	1.80
GMM Hard decision	1.80
GMM Soft decision	1.61

Table 1 – OAAE results

Figure 3 shows the AAE for the three approaches as a function of the SNR that is tested.

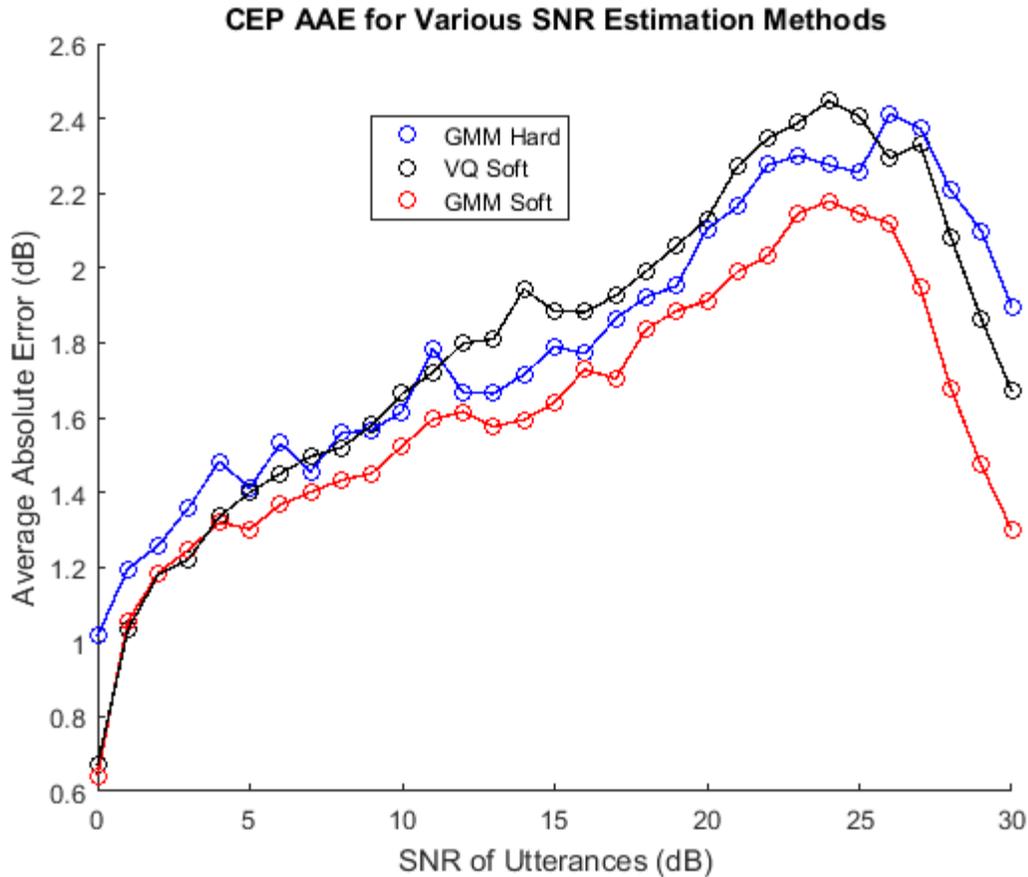


Figure 3 – The AAE plot for the three approaches

Clean speech (no noise added) was also an input to the SNR estimation system. The expected SNR estimate should be greater than 30 dB. Table 2 shows the results.

Classifier Approaches	SNR Estimate (dB)
VQ Soft decision	30.67
GMM Hard decision	31.58
GMM Soft decision	30.82

Table 2 – SNR estimates for clean speech

EDUCATION ORIENTED TASKS

The student team is in the process of writing a laboratory manual for a class project that involves a software implementation of a VQ based SNR estimation system. Each step is

to be clearly explained. Since the students have already implemented this system, they are well versed with the process of a modular design. When this project is run in a class, the deliverables will include the software implementation and a formal lab project report. The report will have a title page, table of contents, introduction and objectives, all the theoretical background, results and discussion, conclusions and references.

ASSESSMENT RESULTS

When the project is run in a class, the assessment instruments will include:

- Student surveys: A target group (students who performed the project) will be compared to a control group (students who did not perform the project).
- Faculty tracking of student achievement of learning outcomes based on evaluation of student work using significant rubrics.
- A concept inventory test given before and after the project is performed.

SUMMARY AND CONCLUSIONS

This paper discusses a work in progress of an undergraduate student team that is gaining research experience by implementing an SNR estimation system, and is also writing a laboratory manual that will facilitate future use of the project in core engineering courses. This project falls within an NSF-funded umbrella project in which big data applications and technology are being vertically integrated throughout the engineering curriculum.

ACKNOWLEDGEMENT

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

REFERENCES

1. R. Ondusko, M. Marbach, R. P. Ramachandran and L. M. Head, "Blind Signal to Noise Ratio Estimation of Speech Based on Vector Quantizer Classifiers and Decision Level Fusion", accepted in *Journal of Signal Processing Systems*.
2. A. K. Jain, A. A. Ross and K. Nandakumar, *Introduction to Biometrics*, Springer, 2011.
3. M. C. Huggins and J. J. Grieco. Speaker identification confidence metrics for heterogeneous model spaces, Proc. of the 8th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, 440–443, 2004.
4. 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 16—19, 2016.
5. J. Edwards, R. P. Ramachandran and U. Thayasivam, "Robust Speaker Verification With a Two Classifier Format and Feature Enhancement", submitted to *IEEE International Symposium on Circuits and Systems*, Baltimore, Maryland, May 28—31, 2017.
 6. Y. Mehta, R. Dusseau and R. P. Ramachandran, "Conducting State-of the-art Research in an Institution with a Strong Undergraduate Education Focus", *ASEE Annual Conference and Exhibition*, Atlanta, Georgia, June 23--26, 2013. (with)
 7. S. Davis, M. Frankle, R. P. Ramachandran, K. D. Dahm, and R. Polikar, "A Freshman Level Module in Biometric Systems", *IEEE Int. Symp. on Circuits and Systems*, Beijing, China, May 19–23, 2013.
 8. R. P. Ramachandran, R. Polikar, K. D. Dahm and S. S. Shetty, "Open-Ended Design and Performance Evaluation of a Biometric Speaker Identification System", *IEEE Int. Symp. On Circuits and Systems*, Seoul, South Korea, May 2012.
 9. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
 10. I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Springer, 2002