

AC 2008-2673: PEDAGOGY OF A COURSE IN SPEECH CODING AND VOICE-OVER-IP

Tokunbo Ogunfunmi, Santa Clara University

TOKUNBO OGUNFUNMI, Ph.D., P.E. is an Associate Professor of Electrical Engineering at Santa Clara University, Santa Clara, California. He earned his BSEE (First Class Honors) from Obafemi Awolowo University, (formerly University of Ife), Nigeria, his MSEE and PhDEE from Stanford University, Stanford, California. His teaching and research interests span the areas of Digital Signal Processing (theory, applications and implementations), Adaptive Systems, VLSI/ASIC Design and Multimedia Signal Processing. He is a Senior Member of the IEEE, Member of Sigma Xi, AAAS and ASEE.

Pedagogy of a Course in Speech Coding and Voice-over-IP

Abstract

The area of Speech Coding and Voice-over-Internet Protocol (Voice-over-IP) has become an important application area of Digital Signal Processing in the Electrical Engineering curriculum. This is due in part to the ubiquity of wireless and wire-line communication devices and appliances that utilize speech coding and voice-over-internet protocol (IP) methods.

We have recently developed a three-course sequence on Speech Coding and Voice-over-IP taken by Senior-level undergraduate and graduate students at Santa Clara University. The first two courses teach Speech Coding while the third course deals with Voice-over-IP. Most of the fundamentals are on Digital Signal Processing but we focus on the applications to speech and voice coding.

In this paper, we first describe the DSP curriculum for both undergraduate and graduate students. We describe our experiences and the challenges encountered in developing these courses. We detail some of the laboratory and teaching materials and the exercises developed, etc.

We discuss as an example the internet low-bit rate speech coder (iLBC) which is used to code speech under packet loss conditions that exists on the internet.

Finally, we present possible future directions in the course development.

Introduction

The area of Digital Signal Processing (DSP) in most universities has grown rapidly over the last 30 years or so. Since one of the most popular original DSP textbooks [1] was written and published in 1975, there has been a tremendous growth in the area measured by the number of textbooks published in the area [2-10] plus other books and recent focus [11-12]. This growth is fueled largely by growth in the semiconductor industry which has enabled more and more transistors to be fitted on smaller and smaller silicon chips. Another reason for the growth in DSP is the growth in application areas where this technology continues to be used such as wireless communications, networking, bio-informatics, consumer electronics, etc. As suggested by industry watchers [13], this growth will not abate anytime soon. In fact, an evidence of this is the fact that DSP education, as a course, is now part of the undergraduate electives in most universities [11-12]. It is likely to be part of the undergraduate core courses in the next few years. As the growth continues, some universities have made a course in Digital Signal Processing (DSP) the first course taken by undergraduates instead of the traditional Analog Electric Circuits course [5].

We also note that DSP education is beginning to trickle down to the high-school level as well as evidenced by the Infinity project [14]. As more of the theory of DSP course percolates down to the junior and sophomore levels, there is a need to include advanced courses in some kind of applications and implementations of DSP into the senior and graduate level curricula. At Santa Clara University, we have established courses in the senior/graduate levels to fulfill this need. These courses complement our existing courses in the area of DSP theory, applications and implementations tailored to the needs of our undergraduates and graduates.

Speech Processing [15-20] is one of the major applications of digital signal processing. In this paper, we focus on three quarter-length courses based on the theory and applications of speech processing. These courses are

- Speech Coding I
- Speech Coding II
- Voice-over-IP

In this paper, we discuss the pedagogy (the principles and methods of instruction) for our follow-up courses on Speech Processing and Voice-over-IP. We focus on the activities of educating or instructing; activities that impart knowledge or skill in these areas.

The paper is divided into six sections. In Section 2, we briefly present the current DSP curriculum for undergraduate and graduate electrical engineering students in our school and we discuss the innovations we have made in that curriculum development. Then in Section 3, we detail the two Speech coding courses we have introduced into the curriculum. In Section 4, we discuss the new speech coder used for internet applications as an example. In Section 5, we discuss the course on Voice-over-IP. We make some final observations and conclusions in section 6 including possible new directions for these courses.

DSP Curriculum

BSEE Undergraduate Curriculum in DSP

At our university, as in most universities, we offer a four-year BS program in Electrical Engineering. In this program, we require that students take basic sciences and mathematics in the first year. In the second year, they take a few engineering courses but still continue to get a strong science and mathematics foundation. In the third year, they complete the set of core electrical engineering courses. In the final year, they take mostly elective courses. During the four year program, the students also have to complete university-required core curriculum courses in areas like English, ethics, political science, religious studies, etc. We also offer flexible Junior Spring term for students to work (as co-op or intern), or study abroad or take more technical electives or graduate classes.

The core of the Electrical Engineering program are the courses which every undergraduate student has to take. It consists of the following 9 courses: ELEN 21 (Logic Design), ELEN 33 (Introduction to Digital Signal Processing Systems), ELEN 50 (Electric Circuits I), ELEN 100 (Electric Circuits II), ELEN 104 (Electromagnetics I), ELEN 105 (Electromagnetics II), ELEN 110 (Linear Systems), ELEN 115 (Electronic Circuits I) and ELEN 151 (Semi-conductor Devices).

There are technical elective courses in the areas Wireless Communications, Analog Electronic Circuits, Digital Signal Processing, Integrated Circuits (IC) Design, Semiconductor Devices and Materials, Control Systems, Robotics, Storage Devices, etc.

The three Speech Processing courses listed above can be taken as part of the electives in the Senior year under the Digital Signal Processing technical electives..

MSEE Graduate Curriculum in DSP

We offer a MS degree in Electrical Engineering with many possible emphasis areas. One of the emphasis areas is Digital Signal Processing. The core of the Masters degree Electrical Engineering program is the following sets of courses: One course in Digital Systems (Logic Design or Synthesis) (2 units) , One course in Electromagnetics (2 units), Four courses in Applied Mathematics (8 units), One course in Applied Ethics (2 units), One course in Electronics (2 units), One course in Electric modern Networks (2 units) and One course in Control Systems (2 units).

Out of a 45 units required for graduation, this MSEE core takes up 20 units. The original intent was to ensure a breadth of knowledge of electrical engineering at the master's level. With the remainder of 25 units, the student can take emphasis areas like Digital Signal Processing.

In the graduate curriculum, we have established a program for the DSP emphasis area consisting of a set of DSP area core courses and advanced courses. See Figure 1 for a diagram of the DSP emphasis area. The arrows indicate required pre-requisites. As shown there are many related areas to DSP (such as Computer Architecture/ASIC, Communication, Speech Processing, Image Processing, etc.) and which can be combined to make a more meaningful program.

The core courses for the Digital Signal Processing emphasis area (which are required for students specializing in this area) are: ELEN 233 (Digital Signal Processing I) (2 units) , ELEN 234 (Digital Signal Processing II) (2 units) , ELEN 334 (Statistical Signal Processing) (2 units) and ELEN 235 (Estimation Theory I) (2 units).

The Advanced Signal Processing courses (second column (tier) of DSP courses) include courses such as hands-on practical implementations on DSP processors or FPGAs, Speech Coding, Multimedia Signal Processing, Estimation, Adaptive Signal Processing, Artificial Neural Networks, etc. Some of these are shown in Figure 1.

The focus of this paper is on the Advanced DSP courses in the area of Speech:

1. ELEN 421 (Speech Coding I) (2 units)
2. ELEN 422 (Speech Coding II) (2 units)
3. ELEN 423 (Voice-over-IP) (2 units)



Figure 1 : Some courses in the Graduate MSEE DSP Emphasis area

The Speech Coding Courses

Many universities, including ours, continually strive to improve their programs by assessing its impact and learning outcomes and modifying, changing or deleting, adding courses based on academic and industrial technology trends. This is actually required by the Accreditation Board for Engineering Technology (ABET) [21] as part of accreditation requirements. In the area of Digital Signal Processing (DSP), many schools offer a single course introducing the theoretical methods used. What is lacking in most cases, is a course on an application area of DSP such as speech coding and processing.

We offer a two-course sequence in the area of Speech Coding. Here we describe the contents of the Speech Coding I and Speech Coding II courses. Each course is a for a 10-week term.

ELEN 421 Speech Coding I Review of sampling and quantization. Introduction to Digital Speech Processing. Elementary principles and applications of speech analysis, synthesis, and coding. Speech signal analysis and modeling. The LPC Model. LPC Parameter quantization using Line Spectrum Pairs (LSPs). Digital coding techniques: Quantization, Waveform coding. Predictive coding, Transform coding, Hybrid coding, and Sub-band coding. Applications of speech coding in various systems. Standards for speech and audio coding. *Prerequisite: ELEN 334 (Statistical Signal Processing) or equivalent.* (2 units)

The textbook for the course includes [10, 15-20]. The aim of the course is to understand the introductory methods of speech coding.

Speech Coding I Lab Contents

The laboratory part of the course consists of five laboratory assignments. Some of the labs require the use of MATLAB. The titles of the labs are :

Lab 1: Review of DSP Methods and Techniques for Speech

Lab 2: Pitch Determination Methods

Lab 3: LPC Model. LPC Parameter quantization using Line Spectrum Pairs (LSPs)

Lab 4: Scalar Quantization: Nonlinear Quantizers in Speech (μ -law and A-law)

Lab 5: Example: G.711 Speech Coder

These labs build on one another and culminates in the student implementing in MATLAB the standardized ITU G.711 voice coder [21] which is commonly used in industrial applications.

ELEN 422 Speech Coding II Advanced aspects of speech analysis and coding. Analysis-by-Synthesis (AbS) coding of speech, Analysis-as-Synthesis (AaS) coding of speech. Code-Excited Linear Speech Coding. Error-control in speech transmission. Application of coders in various systems (such as wireless phones). International Standards for Speech (and Audio) Coding. Real-Time DSP implementation of speech coders. Research project on speech coding. Introduction to speech recognition. *Prerequisites: ELEN 421 (Speech Coding I).* (2 units)

The textbook for the course includes [10, 15-20]. The aim of the course is to understand the advanced methods used speech coding. Concepts such as vector quantization, analysis-by-

synthesis methods used in CELP coders are emphasized. Many examples are given of the concepts as specified in various of the standardized speech coders from ITU and other agencies. In particular, we use the FS 1015 Speech Coder, the G.729A Speech Coder and the internet Low Bit Rate codec (iLBC) as examples.

Speech Coding II Lab Contents

The laboratory part of the course consists of five laboratory assignments. Many of the labs require the use of MATLAB. The titles of the labs are :

Lab 1: Vector Quantization

Lab 2: Analysis by Synthesis Methods: CELP Coders

Lab 3: Example: The FS 1015 Speech Coder

Lab 4: Example: The G.729A Speech Coder

Lab 5: Example: The internet Low Bit Rate codec (iLBC)

Lab 6: Real-Time DSP implementation of speech coders

These labs build on one another and culminates in the student implementing in MATLAB the standardized ITU G.729A voice coder [21] which is and commonly used in industry.

As advances are made in speech coding technologies, we have added new topics to the course contents. One example is the discussion and evaluation of the internet Low Bit Rate codec (iLBC) recently proposed and beginning to be widely-deployed.

Next, we discuss this new speech coder.

internet Low Bit Rate codec (iLBC)

The internet Low Bit Rate codec (iLBC) is a narrowband speech codec developed by a company called Global IP Sound (GPIS) and its structure is defined in [32-35]. It is commonly used for VoIP applications such as Skype, Yahoo Messenger and Google Talk among others.

The continuously increasing demand for carrying voice traffic over the Internet has forced internet-appliance developers to look at different speech coding types to find the one that better fits the characteristics of packet-switched networks. Due to the bursty nature of packet-based traffic and the possibility of packet loss due to network congestion, current CELP codec such as G.729A will not offer the same performance when previous samples are not available and will cause signal distortion at the far end.

Global IP used its expertise and knowledge of the VoIP technology and market requirements to design a new type of codec focused on packet communication that has increasingly become more popular with the deployment of VoIP over Internet. This new codec is called internet Low Bit Rate (iLBC) and was designed to operate a low bit rates and provide good quality speech during low traffic and network congestion situations. iLBC eliminates the dependency on previous frames because each packet is treated independently to achieve a better response to packet loss, delay or packet jitter.

a. Structure

Sampling rate of 8 KHz, 16 bits/sample 2 frame lengths: 30 ms (240 samples) for a bit rate of 13.3 KHz and 20 ms (160 samples) for a bit rate of 15.2 kbit/s. Most of low bit rate codecs limit voice bandwidth to 50-3400 Hz whereas iLBC utilizes the full 4 KHz bandwidth producing higher quality reconstructed voice.

b. Advantages

- As opposed to CELP codec that require previous data to estimate the pitch gain and lag, internet Low Bit rate Codec estimates the pitch of the signal in the same frame eliminating the dependency of previous samples and look ahead delays. This is why iLBC offers a better performance during packet loss conditions.
- Use of LSF/LPC interpolation enhances the performance of the codec during high packet loss rates, compensating for lack of information from previous blocks/frames.
- It also ensures high voice quality during normal network operation with very low packet loss rate.
- Incorporates packet loss concealment techniques.

c. Algorithm

iLBC codec uses adaptive codebook forward and backward in time as follows:

- Determine start state vector inside the speech frame. It contains the highest residual energy especially for voiced signals (dominant pitch).
- Transmit encoded location and waveform of the start state per frame.
- Adaptive codebook is filled with segments of the decoded start state vector.
- Long Term predictive encoding is used from the end of the start state until the end of the speech frame. (Forward in time). Adaptive codebook is updated frequently with the most recent decoded signal.
- Then, the adaptive codebook is filled again with segments of the decoded start state and the first encoded signal segment.
- On the other direction, the long term predictive coding is backward in time starting from the beginning of the start state until the beginning of the speech frame.
- Therefore, each packet will contain start state, gain and lag information that will be used at the far end for proper decoding without depending on previous speech frames. Even with packet loss conditions the adaptive codebook in the encoder and the coder will be the same.

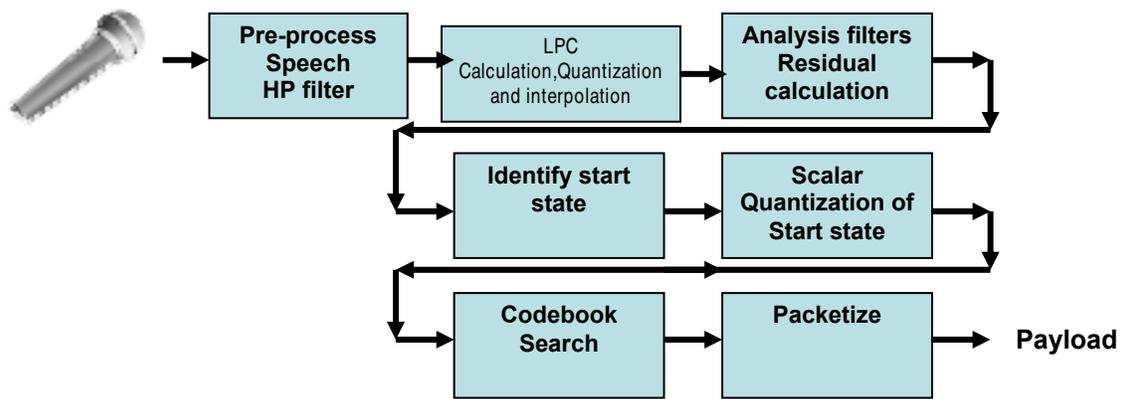


Figure 2 iLBC Encoder block diagram

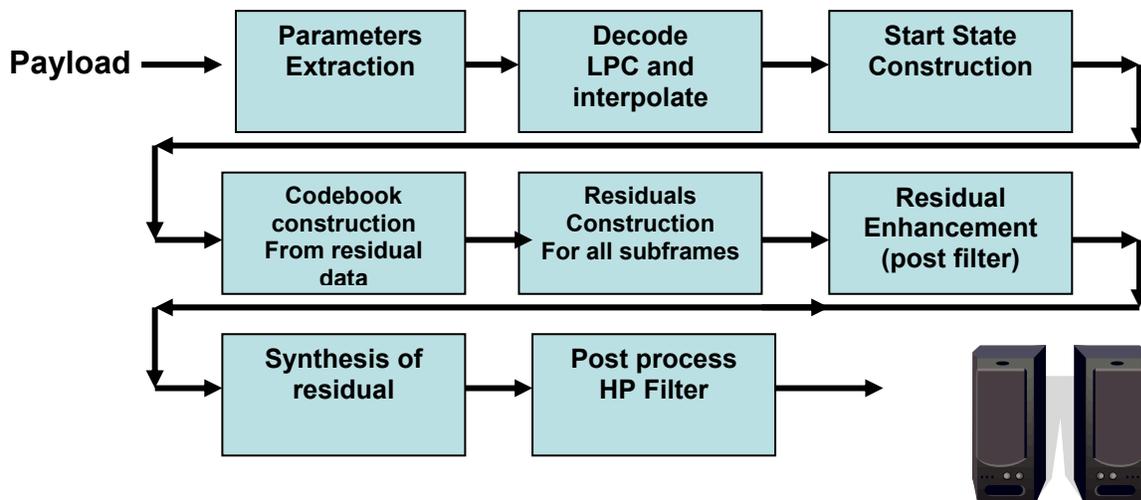


Figure 3 iLBC Decoder block diagram

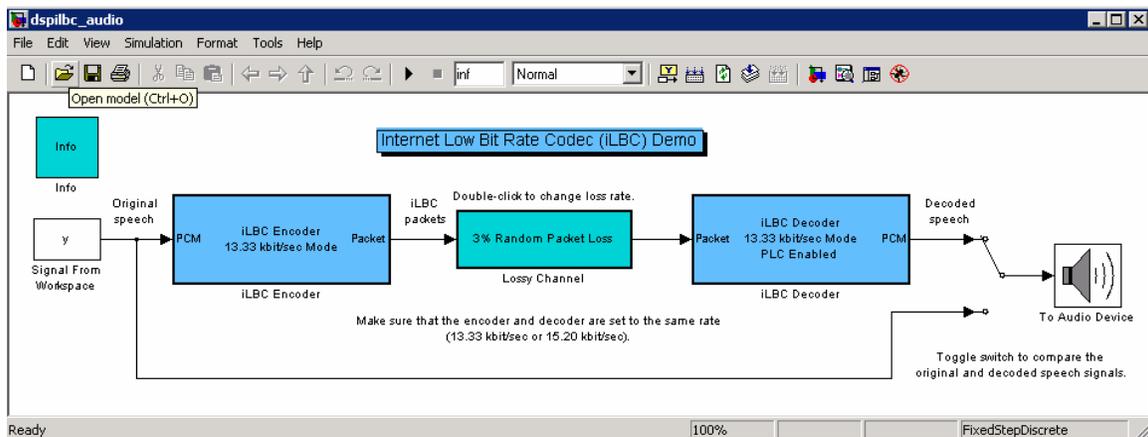


Figure 4 Matlab's iLBC demo model block diagram

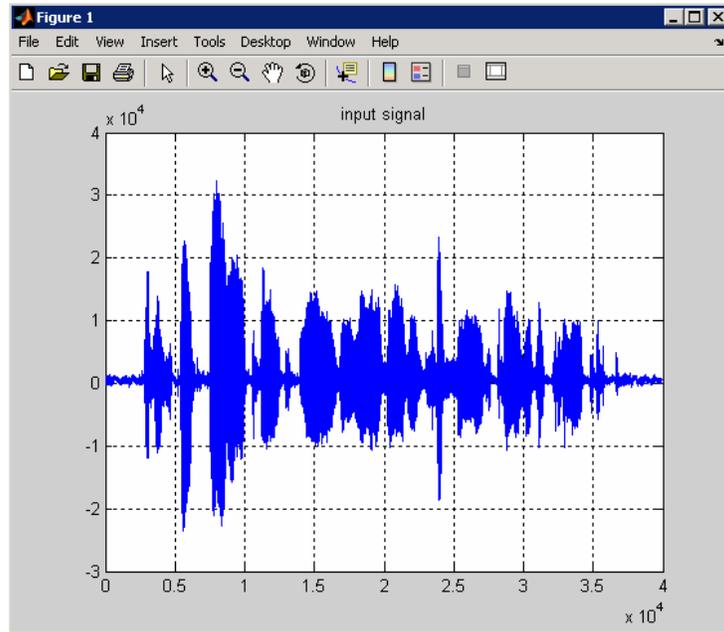


Figure 5 Original speech signal

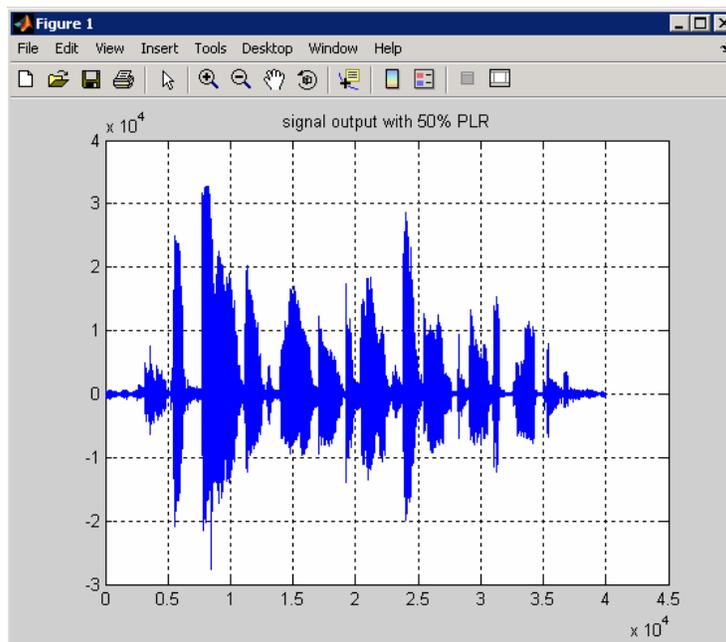


Figure 6 Reconstructed speech signal with 50% packet loss

Figure 2 shows a block diagram of the iLBC Encoder block diagram and Figure 3 shows the block diagram of the iLBC Decoder. Figure 4 shows the iLBC demo program using MATLAB [36]. An example of a speech signal encoded using iLBC is shown in Figure 5. After transmission over the internet, this speech signal is corrupted by a 50% packet loss. It is interesting to see the performance of the iLBC decoder that it is still able to recover the speech signal quite well as shown in Figure 6.

The Voice-over-IP Course

The merging of traditional communications and computer networks means that we have many more choices for communications. It is now very common to transport voice over computer networks. Products from companies such as Vonage, Skype, Google Talk, etc. and other wireless communications providers such as AT&T, Verizon, Sprint, etc. have enabled this possibility.

To satisfy the need for students to understand this rapidly growing DSP application area, we offer a course on Voice-over-IP. The course is a for a 10-week term. Here we describe the contents of the course.

ELEN 423. Introduction to Voice-over-IP Overview of voice encoding standards relevant to VoIP: G.711, G.726, G.723.1, G.729, G.729AB. VoIP packetization and signaling protocols: RTP/RTCP, H.323, MGCP/MEGACO, SIP. VoIP impairments and signal processing algorithms to improve QoS. Echo cancellation, packet loss concealment, adaptive jitter buffer, Decoder clock synchronization. Network convergence: Softswitch architecture, VoIP/PSTN, interworking (Media and Signaling Gateways), signaling translation (SS7, DTMF/MF etc.), fax over IP.
Prerequisite: ELEN 233 or knowledge of basic digital signal processing concepts. (2 units)

The textbooks used for the course are listed in References [22-27]. We are in the process of developing labs for this course and this will be reported in the future.

The course has been quite successful and draws many students each time who are eager to understand how speech and audio can be encoded and transmitted using internet protocols.

In the future, we hope to evaluate the courses' outcomes by using a survey [28-30]. We also hope to report on assessment methods best suited for these new courses. One of the options is to use the Signals and Systems Concept Inventory (SSCI) test on DSP [31] or develop one more suited to the new courses.

Conclusions

In this paper, we presented a pedagogy of the Speech Coding and Voice-over-IP courses developed at Santa Clara University. The area of Speech Coding and Voice-over-Internet Protocol (Voice-over-IP) has become an important application area of Digital Signal Processing in the Electrical Engineering curriculum due in part to the ubiquity of wireless and wire-line communication devices and appliances that utilize speech coding and voice-over-internet protocol (IP) methods.

We also discussed our curriculum development in this area. We have presented details about the coverage of the courses and the use of MATLAB in homework exercises and laboratories. We described our experiences in developing these courses. We detail some of the laboratory and teaching materials and the exercises developed, etc.

We discussed as an example the internet low-bit rate speech coder (iLBC) which is used to code speech under packet loss conditions that exists on the internet.

Finally, we presented possible future directions in the course development such as courses outcomes, assessment tests, etc.

References

- [1] Oppenheim, Alan and Schaffer, Ronald, Digital Signal Processing, Prentice-Hall publishers, First edition, 1975.
- [2] Proakis, John and Manolakis, Dimitris, Digital Signal Processing: Principles, Algorithms and Applications, MacMillan publishers, Second edition, 1992.
- [3] Philips, Parr and Riskin, Signals, Systems and Transforms, Prentice-Hall publishers, Third edition, 2003.
- [4] Lee, Edward and Varaiya, Pravin, Structure and Interpretation of Signals and Systems, Addison-Wesley, NY, 2003.
- [5] McClellan, James, Schaffer, Ronald and Yoder, Mark., DSP First : A Multimedia Approach, Prentice-Hall publishers, 1997.
- [6] Mitra, S.K., Digital Signal Processing: A Computer-Based Approach, (Third edition), McGraw-Hill Publishers, 2006.
- [7] Oppenheim, Alan and Schaffer, Ronald and Buck, John, Discrete-Time Signal Processing, Prentice-Hall publishers, (Second edition), 1999.
- [8] Strum, R.D. and Kirk, D.E., First Principles of Discrete Systems and Digital Signal Processing, Addison-Wesley Publishers, 1989.
- [9] Therrien, Charles, Discrete Random Signals and Statistical Signal Processing, Prentice-Hall Publishers, 1992.
- [10] Rabiner, L. and Schaffer, R., Digital Processing of Speech Signals, Prentice-Hall Publishers, 1978.
- [11] Johnson, Don, "DSP in Rice University's Electrical and Computer Engineering Curriculum", *Proceedings of the Ninth DSP Workshop and First Signal Processing Education Workshop*, Oct. 2000
- [12] Gibson, Jerry, "DSP: Five Degrees of Freedom", *Proceedings of the Ninth DSP Workshop and First Signal Processing Education Workshop*, Oct. 2000
- [13] Strauss, Will, "Digital Signal Processing: The new semiconductor industry driver", *IEEE Transactions on Signal Processing*, vol. 17, pp. 52-56, March 2000
- [14] The Infinity project [Online] <http://www.infinity-project.org/home.html>
- [15] Chu, Wai C., Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, Wiley Publishers, 2003.
- [16] Kondo, A.M., Digital Speech: Coding for low bit rate communication systems, Wiley Publishers, (Second edition), 2004.
- [17] Ogunfunmi, T. and Narasimha, M.J., Class Notes for ELEN 421, 422 and 423. Santa Clara University, Santa Clara, CA, 1997-2007.
- [18] Gibson, Jerry, Speech Coding Methods, Standards and Applications, *IEEE Circuits and Systems Magazine*, pp. 30-49, Fourth Quarter 2005
- [19] Bishnu S. Atal, "The history of linear prediction", *IEEE Signal Processing Magazine*, pp. 154-161, March 2006.

- [20] Robert M. Gray, "The 1974 Origins of Voice-over-IP", *IEEE Signal Processing Magazine*, pp. 87-90, July 2005.
- [21] ITU-T Recommendation G.729 (03/96), Coding of Speech at 8 kbit/s using conjugate structure algebraic-code-excited linear-prediction (CS-ACELP).
- [22] Collins, Daniel, Carrier Grade Voice-over-IP (Second edition), McGraw-Hill, 2002.
- [23] Davidson, Jonathan et. al., Voice-over-IP Fundamentals (Second Edition), Cisco Press, 2006
- [24] Wright, David, Voice-over-Packet Networks (First Edition), Wiley, 2001
- [25] Olivier, Hersent "IP Telephony", Addison-Wesley (1999)
- [26] Minoli, Daniel, "Internet Architectures", Wiley, 1999
- [27] Black, Ulyess, Internet Telephony: Call Processing Protocols, Prentice-Hall, 2001
- [28] American Board for Engineering and Technology (ABET) Engineering Accreditation Commission, *Criteria for Accrediting Engineering Programs*, Baltimore, MD, 2003. (URL: <http://www.abet.org>)
- [29] Felder et. al., Index of Learning Styles, papers,
(URL: <http://www.ncsu.edu/felder-public/ILSpa.html>)
- [30] Felder et. al., Active and Cooperative learning styles, papers
(URL: http://www.ncsu.edu/felder-public/Cooperative_Learning.html)
- [31] Ogunfunmi, Tokunbo "Analysis of Assessment Results in a Linear Systems Course" *Proceedings of ASEE Annual Conference, 2007*.
- [32] IETF organization, RFC3951 Internet Low Bit Rate Codec (iLBC),
<http://tools.ietf.org/html/rfc3951>
- [33] Global IP Sound, "iLBC White Paper",
http://www.gipscorp.com/files/english/white_papers/iLBC.WP.pdf, October 15, 2004
- [34] S.V. Andersen, W.B. Kleijn, R. Hagen, J. Linden, M.N. Murthi, J. Skoglund "iLBC – A linear predictive coder with robustness to packet losses" *Proceedings of IEEE Speech Coding Workshop, 2002*.
- [35] W.B. Kleijn, "Enhancement of Speech by Constrained Optimization" *Proceedings of IEEE Speech Coding Workshop, 2002*.
- [36] Matlab, Signal Processing BlockSet version 6.6 (Matlab's release R2007b)