SYLLABUS
MPATE-GE 2632: Introduction to Audio Coding

Steinhardt School of Culture, Education, and Human Development
Music and Performing Arts
Department of Music Technology

Instructor: Dr. Schuyler Quackenbush
Audio Research Labs
www.audioresearchlabs.com
Email: schuyler.quackenbush@nyu.edu
Office Hours: Immediately after class, meet at classroom

Course Description

This course gives an introduction to the models of the human auditory system: the hearing mechanism and auditory masking, sound stage perception, and sound localization. Aspects of audio perception that can be exploited to achieve audio signal compression will be investigated in detail: the critical band structure of hearing, monophonic frequency masking, monophonic pre- and post-temporal masking, stereo masking, and perceptual correlates to sound localization in the 3-D sound stage. The course will explore in detail now these auditory models are used with signal processing tools such as transforms, filterbanks, predictors, quantizers and entropy coders to build audio coders. These principles will be illustrated by a constructing a simple Matlab-based audio coder during class time. The principles will be reinforced by investigating several MPEG audio coding architectures: MPEG-1 Layer II, MPEG-1 Layer III (MP3), MPEG-4 Advanced Audio Coding (AAC) and MPEG Surround. Students will have a mid-term and a final project.

Learner Objectives

By the end of the course students will:

- Understand human perception of sound and how to exploit the perception mechanisms to achieve audio signal compression,
- Be able to construct a basic audio coder in MATLAB,
- Be able to assess the subjective quality of audio coding algorithms,
- Become familiar with the audio coder that are pervasive in the marketplace.

Prerequisites

The course assumes that the student is familiar with:

- Basic mathematics (e.g. algebra, trigonometry, logarithms).
- Basic concepts of signal processing
- MATLAB programming

Exceptions can be made, and students that have not satisfied the prerequisites should contact the instructor.

Projects

Students will be expected to complete two projects during the term: one due during midterms week and one due during finals week. Each project can be either a *programming* project or a *library research* project.

- *Programming Project* – Students create a portion of an encoder/decoder or some signal-based analysis or analysis/synthesis system, optionally based on MATLAB tools provided by instructor. A list of example projects is shown in Annex 1. Projects should:

    - Include a written overview of what the program does and how it does it; what each module does; and what the student mastered in the project.
    - A written summary of how to run the project and what the code will produce or demonstrate.
    - The MATLAB code for the project. Code should be delivered as a zip archive that, when unpacked, has the correct structure and all support files needed to run from the MALAB command line.

    The grade for the programming project will take into account the quality of the MATLAB code, the written report, the (optional) subjective quality. Since student projects can build upon MATLAB code modules provided by the instructor, it is important that project code should clearly indicate what work was done by the student and show what was learned and mastered by the student.

- *Library Research Project* – Students research and write a paper about a signal processing tool that was reviewed in class or about a tool related to the theme of perceptual coding. There should be sufficient level of detail to demonstrate that the student has mastered the topic presented.

    The grade for the library research project will take into account the scope of the topic presented, how well the paper teaches the topic and what the student has demonstrated to have learned from the project.

Homeworks
Weight: 45% of the final grade

Final Project:
Weight: 55% of the final grade

Readings

Required Text

- M. Bosi and R. Goldberg, *Introduction to Digital Audio Coding and Standards*. Kluwer Academic, Boston, 2003.

Optional Papers (supplied by instructor)

- Quackenbush, S. and Wylie, F., "Digital Audio Compression Technology," Chapter 37, NAB Engineering Handbook, 2007, Academic Press.

- T. Painter and A. Spanias, “Perceptual coding of digital audio,” Proc. IEEE, vol. 88, no. 4, pp. 451–513, Apr. 2000.
- Spanias, A, “Speech Coding: A Tutorial Review,” Proc. IEEE, Vol. 82, No. 10, Oct. 1994.
- Bosi, Marina; Brandenburg, Karlheinz; Quackenbush, Schuyler; Fielder, Louis; Akagiri, Kenzo; Fuchs, Hendrik, “ISO/IEC MPEG-2 Advanced Audio Coding,” JAES Volume 45 Issue 10 pp. 789-814; October 1997
- M. Wolters, et al., “A closer look into MPEG-4 High Efficiency AAC,” 115th AES Convention, October 10–13, 2003, New York, preprint 5871.
- J. Herre, et al., “MPEG Surround-The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding”, JAES v. 56, no. 11, pp 932-55, November, 2008.
- M. Neuendorf, et al., “MPEG Unified Speech and Audio Coding -- The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types,” JAES Volume 61 Issue 12 pp. 956-977; December 2013.

Course Format

Classes will be conducted using lecture instruction and class discussion.

Course Website

This course has a dedicated web site on NYU Classes. The **syllabus**, details about **assignments**, and any other general **course information** will be available on the site.

Course Requirements

1. Reading
It is important that you read assigned materials prior to the class for which the reading is due as the reading material will be the topic for that class. During each lecture, the reading assignments for the next lecture will be indicated.

2. Homeworks
There will be several homeworks that will typically involve some Matlab programming. Each homework will reinforce lecture information and build a foundation for the final project.

3. Projects
Students will be expected to complete a final project that is due during finals week. The expected scope of the projects is discussed above.

Course Outline

1. **Overview of Audio Coding**
   Reading: Chapters 1 and 3

   - Lossless and lossy coders
   - Basic perceptual coder
   - A/D and D/A

- Low-pass filters
- Sampling
- Reconstruction
- Aliasing

2. **Review of Digital Systems and Signals**
   Reading: Chapters 1 and 3

   - Digital representation of numbers
     - Floating point
   - Signal power and dB
   - Impact of sampling rate
   - Complex numbers
     - Magnitude and phase
   - Frequency representation
     - Discrete Fourier transform and fast Fourier transform

3. **Perception of Sound**
   Reading: Chapters 6 and 7

   - Sound pressure level
   - Loudness
   - Organs of hearing
   - Cochlear response
   - Critical bands
   - Absolute threshold of hearing
   - Types of masking
     - Monophonic sounds
       - Simultaneous masking
       - Pre- and Post-masking
     - Stereo sounds
       - BMLD
   - Perception of sound in space
     - Perceptual correlates to localization
       - ITD
       - ILD
     - Perceptual correlates to spaciousness
       - IC

4. **Digital Filtering**
   Reading: None

   - Digital Filters
   - Convolution (filtering)
   - Signal flow graphs

    - FIR
    - IIR
    - Simple example

5. **Transforms and Filterbanks - 1**
   Reading: Chapters 4 and 5

   - Filterbanks
     - 2-band splits
     - Multi-band splits
     - MPEG-1 filterbank
   - Coding with Filterbanks
     - 2-band
     - MPEG-1 filterbank

6. **Transforms and Filterbanks - 2**
   Reading: Chapters 4 and 5

   - Discrete Fourier Transform
     - FFT
     - Redundancy Removal
     - Time/Frequency resolution tradeoff
   - Windowing
     - Frequency smearing
   - Analysis/Synthesis Systems
     - Core of audio coder
     - FFT
       - Windowing, Overlap, Add
     - MDCT
       - Windowing, Overlap, Add

7. **Quantization and Coding**
   Reading: Chapter 2

   - Quantization
     - Mid-Tread, Mid-Riser quantizers
     - Quantizer step size
     - Maximum and RMS error
   - Entropy coding
     - Probability distribution
     - Arithmetic coding
     - Huffman coding

8. **Principles of Audio Coding - I**
   Reading: Chapter 8 and 9

- Basic architecture of a perceptual audio codec
    - MDCT
    - Perceptual model
        - Oddly-stacked FFT
        - Power spectrum on Bark scale
        - Spread spectrum
        - Masking threshold
        - Threshold in quiet
    - Quantizer
        - Mid-tread
    - Entropy coding

9. Principles of Audio Coding - II
Reading: None

- Demo: MATLAB implementation of audio codec

10. Watermarks
Reading: None

- Analysis/Synthesis system
- Perceptual threshold
- Spread Spectrum model for embedding
    - Information sequence
    - Spreading "chip"
- Embedding
- Detection
- Demonstration

11. Subjective and Objective Quality Assessment
Reading: Chapter 10

- Subjective test methods
    - ABX
    - BS.1116
    - MUSHRA
    - Comparative
- Parametric statistics
    - Mean score
    - Individual error variance
    - ANOVA linear model
- Objective test methods
    - PEAQ
- Demo: Conduct subjective test

- Demo: Typical perceptual coder distortions

12. MPEG-1 Layer III (MP3) and MPEG-4 AAC
Reading: Chapter 11, 12 and 13

- MPEG-1 Layer I and II
- MPEG-1 Layer III
    - Hybrid filterbank
    - Perceptual model
    - Huffman coding
- MPEG-2 AAC and MPEG-4 AAC
    - Coding gain of primary coding tools: transform, perceptual model, entropy coding

13. MPEG Surround
Reading: None

- MPEG-4 HE-AAC
    - SBR tool
        - Polyphase filterbank
    - Parametric Stereo tool
- MPEG Surround
    - Inter-aural level differences
    - Inter-aural time differences
    - Inter-aural coherence
    - Concept of 2:1 and 3:1 downmixer

14. The Speech Production Model
Reading: None

- Speech model
    - Acoustic tube model of vocal tract
- Parts of speech model
    - Linear predictive model of vocal tract
    - Model for residual or excitation signal

Required Software

- MATLAB
- Word processor of your choice
- Spreadsheet (Excel recommended)

Statement on Academic Integrity

Students are expected-often required-to build their work on that of other people, just as professional researchers and writers do. Giving credit to someone whose work has helped you is expected; in fact, not to give such credit is a crime. Plagiarism is the severest form of academic fraud. Plagiarism is theft. More specifically, plagiarism is presenting as your own:

- a phrase, sentence, or passage from another writer's work without using quotation marks;
- a paraphrased passage from another writer's work;
- facts, ideas, or written text gathered or downloaded from the Internet;
- another student's work with your name on it;
- a purchased paper or "research" from a term paper mill.

Other forms of academic fraud include:

- "collaborating" between two or more students who then submit the same paper under their individual names.
- submitting the same paper for two or more courses without the knowledge and the expressed permission of all teachers involved.
- giving permission to another student to use your work for a class.

Term paper mills (web sites and businesses set up to sell papers to students) often claim they are merely offering "information" or "research" to students and that this service is acceptable and allowed throughout the university. THIS IS ABSOLUTELY UNTRUE. If you buy and submit "research," drafts, summaries, abstracts, or final versions of a paper, you are committing plagiarism and are subject to stringent disciplinary action. Since plagiarism is a matter of fact and not intention, it is crucial that you acknowledge every source accurately and completely. If you quote anything from a source, use quotation marks and take down the page number of the quotation to use in your footnote.

Consult The Modern Language Association (MLA) Style Guide for accepted forms of documentation, and the course handbook for information on using electronic sources. When in doubt about whether your acknowledgment is proper and adequate, consult your teacher. Show the teacher your sources and a draft of the paper in which you are using them. The obligation to demonstrate that work is your own rests with you, the student. You are responsible for providing sources, copies of your work, or verification of the date work was completed.

Students are responsible for understanding the concept of plagiarism, and knowing and understanding the contents of the University "Statement of Academic Integrity" http://steinhardt.nyu.edu/policies/academic_integrity

Plagiarism will immediately result in a failing grade in the course and the student will be reported to their school's academic Dean.

Students with Disabilities

Academic accommodations are available for students with documented disabilities. Please contact the Moses Center for Students with Disabilities at 212-998-4980 for further information.

Appendix A - Graduate Scale and Rubric

Steinhardt School of Education Grading Scale

There is no A+	
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	65-69
D	60-64
There is no D-	
F	Below 60
IP	Incomplete/Passing
IF	Incomplete/Failing
N	No Grade

Letter Grade Rubic

A—Outstanding Work
An "A" applies to outstanding student work. A grade of "A" features not simply a command of material and excellent presentation (organization, coding, asset management etc...), but importantly, sustained intellectual engagement with the material. This engagement takes such forms as shedding original light on the material, investigating patterns and connections, posing questions, and raising issues.
An "A" assignment is excellent in nearly all respects:

- It is well organized, with a clear focus.
- It is well developed with content that is relevant and interesting.
- It fulfills all the technical and creative requirements of the assignment.
- It demonstrate a clear understanding of the material discussed in class.
- It is engaging

B—Good Work
A "B" is given to work of high quality that reflects a command of the material and a strong presentation but lacks sustained intellectual engagement with the material.
A "B" project shares most characteristics of an "A" project, but

- It may have some minor weaknesses in its implementation, either technical or creative.
- It may have some minor lapses in implementing the one or two required elements.

C—Adequate Work
Work receiving a "C" is of good overall quality but exhibits deficiencies in the student's

command of the material or problems with presentation or implementation.

A "C" project is generally competent; it is the average performance. Compared to a "B" paper:

- It may have serious shortcomings in its implementation or organization.
- It fails to meet two to three requirements outlined in the assignment.
- The functionality of one or more elements has been compromised.

D or F—Unsuccessful Work

The grade of "D" indicates significant problems with the student's work, such as a shallow understanding of the material.

- It is messy in its implementation
- It displays major organizational problems
- It fails to fulfill three of more of the requirements outlined in the assignment
- It is unrelevant to the assignment
- It includes confusing transitions or lacks transitions altogether

An "F" is given when a student fails to demonstrate an adequate understanding of the material, fails to address the exact topic of a question or assignment, or fails to follow the directions in an assignment, or fails to hand in an assignment.

Pluses (e.g., B+) indicate that the assignment is especially strong on some, but not all, of the criteria for that letter grade. Minuses (e.g., C-) indicate that the paper is missing some, but not all, of the criteria for that letter grade.

Appendix B – Project Ideas

Library Research Paper

- 20 hours of work or ~5 pages
- Possible topics
    - Cover some aspect of perception more deeply that we covered in class
        - History of audio perception, e.g. critical bands
        - Spatial audio, sound source localization, sound source separation
    - Don't want a summary of a the contents of some papers. Better if the basic ideas or the progression of ideas over time is documented.

Analysis/Synthesis System

- Using FFT analysis/synthesis (50% overlap)
    - Test Threshold in Quiet model
        - Add noise below threshold. Adjust threshold until it is just audible
    - Test NMR model
        - Use class toolbox
        - Add noise below NMR. Adjust threshold until it is just audible
        - Do above for different classes of signals
        - Harmonic, noisy, etc.

Masking Measurement Tools

- Threshold in Quiet
    - Measure threshold in quiet for 3 subjects and chart results
- Noise band masking tone
    - Measure masking as a function of frequency for 1 subject

Perceptual Coder

- 2048 long-block MDCT
- 2048 block oddly-stacked FFT perceptual model
- Divide spectrum into two regions
    - Low region +/-N level quantizer
    - High region +/-1 level quanitizer
- Quantize spectral values
    - No entropy coding

Embedded data channel

- Read WAV file
- 50% overlap fft
- Determine minimum masking threshold across entire block
    - No threshold in quiet
    - Determine corresponding number of "lsb's" after inverse transform
- Substitute that number of lsb's for that block
    - Consider using actual data message
- Write matching extractor

Subjective Assessment

- Download 3 perceptual coders
  - MP3
  - AAC
  - HE-AAC
- Get 6 signals representing a variety of classes
  - Vocal, instrumental, percussive, continuous, etc.
- Code 6 signals using 3 coders for a range of bitrates
  - Do “pre-listening” to check that results span a range of subjective quality
- Conduct subjective test
  - 3 coders x 3 rates
  - 8 listeners
- Write a test report
  - Interpret results