Speech on the Web: An MIT Lab Course

Janet Slifka†, Stefanie Shattuck-Hufnagel† and Laura Koller‡

† Speech Communication Group, Massachusetts Institute of Technology, Massachusetts, U.S.A.

‡ OpenCourseWare, Massachusetts Institute of Technology, Massachusetts, U.S.A.

E-mail: slifka@speech.mit.edu, stef@speech.mit.edu, lkoller@mit.edu

ABSTRACT

The materials for the Massachusetts Institute of Technology (MIT) course, Laboratory on the Physiology, Acoustics and Perception of Speech have been made available on the Web as part of MIT's OpenCourseWare (MIT OCW) initiative. The course, originally developed by Dennis Klatt and now taught by Ken Stevens, emphasizes the development of analysis skills in the domains of linguistics, articulation, acoustics, and perception. The course offers fifteen labs; a core set of labs aims to familiarize the student with a range of laboratory methods, and a flexible set of labs focuses on current research issues. The site provides downloads of lab instructions and recorded media to support the labs. Where appropriate freeware is not easily available, downloads for software tools are provided. The MIT OCW program is not a distance-learning initiative; instead it provides free and searchable access to MIT course materials to any user. (http://ocw.mit.edu/6/6.542j/f01/index.html)

1. INTRODUCTION

A lab setting provides students of phonetics the opportunity to explore measurement and analysis tools as well as providing an introduction to the relative advantages of various methods for studying aspects of the speech process. The Massachusetts Institute of Technology (MIT) course, Laboratory on the Physiology, Acoustics, and Perception of Speech, has been offered at MIT for 25 years. This course is designed to develop a basic skill set for studying the speech process, provide exposure to emerging techniques, and offer an opportunity to apply these skills in the context of an individual research project. The materials for this course are now available on the Web, free of charge, to any user anywhere in the world with access to the site http://ocw.mit.edu/6/6.542j/f01/index.html. This paper summarizes the background and content of the course, presents the organization of the Web-based materials, and as an example, examines three specific labs in detail.

2. HISTORY OF THE COURSE

This lab course was originally developed and taught by Dennis H. Klatt of the Speech Communication Group at MIT. The first offering was in 1978 and was co-listed under the Electrical Engineering and Computer Science Department and the Linguistics Department, reflecting its multidisciplinary approach. The basic structure of the course still holds today. The course was designed to span a thirteen-week semester. In a typical week, there are two labs, each preceded by a one-hour lecture. The labs are spread across the topics of acoustics, physiology, and perception. The course begins with a focus on tools for acoustic analysis of speech, and these skills re-surface in some of the labs on physiology and perception. Some labs span more than one session, and the last several weeks of the course are devoted to individual research projects. There are no exams.

The specific labs within the course vary somewhat year-to-year. There is a core set of 12 or 13 labs that covers a fundamental skill set. There are 2 or 3 evolving labs that center on a current area of research within the Speech Communication Group during the time that the course is offered. In 1989, Kenneth N. Stevens became the instructor for the course, and in 1992, the course became co-listed under a third academic department, the Health Sciences and Technology (HST) Department. The listing coincided with the start of the Speech and Hearing Sciences graduate degree program within the HST department.

The current version of the course has two guest lecturers: Stefanie Shattuck-Hufnagel conducts two labs focused on prosody, and Joseph Perkell conducts a lab using cineradiographic images to study articulation.

3. MIT OPENCOURSEWARE (OCW)

MIT's OpenCourseWare (MIT OCW) program was launched in September 2002 as a large-scale, Web-based publication of MIT course content. The intent is to provide free, searchable, coherent access to all MIT course materials in an open and permanent format to learners around the world. The MIT OCW project is not a distance-learning initiative, nor it is intended to replace or represent an actual interactive classroom environment. The materials available on the Web, which may include lecture notes, assignments, reading lists, exams, problem sets and solutions, simulations, and video lectures, aim to capture the depth of the educational resources that support each course. The materials on the MIT OCW site may be used, copied, distributed, translated and modified, but only for non-commercial educational purposes that are made freely available to other users. (http://ocw.mit.edu)

ISBN 1-876346-48-5 © 2003 UAB

4. THE MIT OCW VERISON OF THE LAB COURSE

The MIT OCW site for Laboratory on the Physiology, Acoustics, and Perception of Speech offers materials under the categories of: Syllabus, Lab Schedule, Required Readings, Lab Software, Lab Database, and Projects. The Syllabus page gives the expected overview and organizational materials. The Lab Schedule page provides a link to a separate page for every lab in the course. On the page for each lab, there is: a link to a handout that gives instructions for the lab, a list of media for the lab with links to pre-recorded versions of those media, and a list of the tools to be used in the measurement or analysis. The current titles of the 15 labs in the course are listed in Table 1.

Table 1: Labs available online

Lab	Lab Title
1	Recording speech in a sound-treated room, and spectral analysis and waveform editing by computer
2	Spectrographic and spectrum analysis of the English vowel system
3	Broadband spectral analysis of sonorant consonants
4	Spectral analysis of stop and fricative consonants: sound generation from turbulence in the vocal tract
5	Sound generation at the larynx: characteristics of the glottal source
6	Formant frequency calculations from area function data
7	Sound segment reduction
8	Prosody: acoustic evidence for prosodic constituent structure
9	Intonation labeling OR Factors affecting duration
10	Speech intelligibility and confusion matrices
11	Introduction to speech disorders
12	Interpretation of cineradiographic motion pictures of selected utterances
13	Speech synthesis using a formant synthesizer
14	Measuring speech movements using data from x-ray microbeam system
15	Higher-level synthesis with a formant synthesizer, using quasi-articulatory parameters

The Required Readings page displays the entire set of readings for the course as grouped by lab number. When the course is offered to students at MIT, there are a variety of software tools that are used during the semester. These tools are all listed on the Lab Software page and include tools for time-domain and frequency-domain analysis that center on the Klatt Speech Analysis software. However, in general, much of the analysis can be done with freely available software tools, and links are provided to examples of such tools. In some cases, links are provided for download of specialized tools, e.g., scripts to conduct vocal tract area calculations and examine microbeam data.

The Lab Database page lists recorded media that are available on the site as grouped by lab number. These data include (for both a male and a female speaker): a set of isolated words spanning the English vowel and consonant set, a set of related sentences, the Rainbow passage, and examples of spontaneous speech. For the labs that focus on prosody, data can be downloaded for sentences as produced by four speakers with two different prosodic forms, sentences with various intonation patterns, and sentences demonstrating duration differences in relation to sentence prosody. Recordings for listening tests related to speech intelligibility in noise can also be downloaded. A set of recordings focusing on speech disorders can be downloaded to compare speech produced by normal-speaking children and speech produced by children who misarticulate certain sounds as well as speech produced by deaf children. Utterances as produced by normal-speaking adults and adults with dysarthria are also available. Media to study the physiology of speech production include cineradiographic images and x-ray microbeam data.

The final segment of the course is centered on conducting individual research projects. It is suggested that the participants select a topic about midway through the course and propose a limited experiment. Traditionally, a written version (in format of Journal of the Acoustical Society) is submitted at end of term. The Projects page lists examples of research topics.

5. EXAMPLE LABS

One of the primary skills developed in this course is acoustic analysis of the digitized waveform. The materials on the MIT OCW site include seven labs with a primary focus on acoustic analysis tools. One such lab is *Lab 2: Spectrographic and Spectrum Analysis of the English Vowel System.* An outline of the lab is given in Table 2. Part 1 illustrates such concepts as the effects of different analysis techniques on the available frequency-domain information, and time- and frequency-domain techniques for estimating the fundamental frequency. Part 2 encourages the student to explore the construction of a vowel space and to relate it to some distinctive features.

ISBN 1-876346-48-5

Table 2: Lab 2: Spectrographic and spectrum analysis of the English vowel system.
Goal: Investigate a set of spectral analysis tools and use these tools to make acoustic measurements on American English vowels.
Speech materials: A recording of the word "shutter" and recordings of word pairs containing the set of English vowels.
Part 1: Explore the discrete Fourier transform, an LPC version of the spectrum, and the determination of the fundamental frequency (F0) and formant frequencies. Consider various versions of the spectral representation including: the differences between a spectrum calculated with a short time window and a long time window and the information available in an averaged spectrum for both short and long averaging times. Apply these tools to periodic speech sounds (in this case the vowel ʌ) and noise-like speech sounds (in this case, the fricative ʃ).
Part 2: Make a set of measurements for the lowest three resonances of the vocal tract (formants). For simple vowels, make one measurement at the vowel midpoint and for dipthongized vowels, select two points while attempting to avoid the regions of consonantal influence. Plot the results and formulate the simplest possible relationships between the first two formant frequencies and the distinctive features of high, low, back, tense, and rounded lips. Also, make comparisons of the formant data obtained from the spectrogram and from various versions of the spectrum (as explored in Part 1) at the midpoint.
Reference readings: Peterson, G.E., and Barney, H.L. (1952) "Control Methods Used in a Study of the Vowels", *J. Acoust. Soc. Am. 24*, 175-185. Also in R. Kent et al. (1991) *Papers in Speech Communication: Speech Production*, Acoustical Society of America, Woodbury, NY, 585-594. J. Hillenbrand et al. (1995) "Acoustic Characteristics of American English Vowels", *J. Acoust. Soc. Am.* 97, 3099-3111. Stevens, K. N. (1998), *Acoustic Phonetics.* MIT Press, Cambridge, MA, pages 257-322.

In addition to the seven labs on acoustic analysis, there are three labs that focus on the perception of speech. The general topics are prosody (Labs 8 and 9) and speech intelligibility in noise (Lab 10). Lab 10, *Speech Intelligibility and Confusion Matrices,* is outlined in Table 3. This lab encourages students to think about which aspects of the speech signal are robustly perceived in noise and the effect of context on perception.

Table 3: Lab 10: Speech intelligibility and confusion matrices.
Goal: Examine two methods of assessing the intelligibility of speech passed through a communication channel.
Speech materials: Consonant-Vowel (CV) syllables in various levels of white noise. Sentences in various levels of babble noise.
Part 1: Listen to the series of CV syllables in noise and write down the consonant. Organize the group results into a confusion matrix and interpret the results for patterns focusing on distinctive features.
Part 2: Listen to the series of sentences in noise and write down the last word in the sentence. Score the results and interpret on the basis of sentences that provided context clues to the final word and sentences that did not.
Reference readings: G.A. Miller and P.E. Nicely (1955) "Analysis of perceptual confusions among some English consonants", *J. Acoust. Soc. Am. 27*, 338-352. D.N. Kalikow, K.N. Stevens and L.L. Elliot (1977) "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability", *J. Acoust. Soc. Am. 61*, 1337-1351.

Finally, there are five labs that focus on the physiology of speech production. The topics are disordered speech (Lab 11), movement of the articulators (Labs 12 and 14), and speech synthesis (Labs 13 and 15) using a formant-based articulatory synthesizer. An outline of Lab 14, *Measuring Speech Movements Using Data from an X-ray Microbeam System*, is given in Table 4. This lab explores the concepts of invariance and variability in the articulation of speech sounds.

 ISBN 1-876346-48-5 © 2003 UAB

Table 4: Lab 14: Measuring speech movements using data from an x-ray microbeam system
Goal: Make measurements on the movements of the tongue and other articulators during speech production using data from an x-ray microbeam system for tracking the placement of pellets on the articulator surfaces.
Speech materials: Pellet and acoustic data for two speakers reading isolated words and sentences.
Tools: MATLAB (Mathworks, Inc.) and the MAVIS scripts to analyze and display pellet data.
Part 1: Examine the trajectories for the anterior jaw pellet. Determine the times when this pellet is maximally high and propose an explanation.
Part 2: Examine the pellets tracking the motions of the tongue root, upper lip, and tongue body pellets in the middle of a set of 12 vowels. Attempt to explain the differences in the results in the context of vowel features.
Part 3: Compare the x-position of the upper lip pellet during the production of /s/ in "this" and /sh/ in "cash," and explain the differences.
Part 4: Compare the tongue body position in the middle of the vowels in "coat" and "both."
Optional investigations: For the following investigations, make the set of measurements and attempt to interpret them in the context of the speech segments and the articulation required to produce them. Compare the position of the lower lip pellet during the initial consonant closure in the words "flip" and "blend." Compare the x-position of the upper lip pellet during the /s/ in the words "seed" and "sued." Examine the movement of the tongue body as a function of time in the vowels in the words "said", "seed", and "sayed." Examine the movement of the tongue pellets during the production of the /r/ sounds in sentences and compare across subjects.
Reference readings: Westbury, J.R. (1991) "The significance and measurement of head position during speech production experiments using the x-ray microbeam system", *J. Acoust. Soc. Am. 89*, 1782-1791.

A number of the eight labs that focus on perception and physiology also draw on the acoustic analysis skills developed in the first seven labs.

6. SUMMARY

A lab setting offers the opportunity to explore measurement and analysis techniques to reach a better understanding of the types of information provided by the various tools. The MIT course Laboratory on the Physiology, Acoustics, and Perception of Speech spans the topics of speech acoustics, perception, and physiology with an emphasis on relating each area back to the acoustics and to linguistic contrasts. Freely available materials on the Web that support this course as a part of the MIT OpenCourseWare program broaden the exposure of the course to students at many universities and beyond, and offer the opportunity for feedback to improve the course in the future.

ACKNOWLEDGEMENTS

This course would never have come into existence without the vision and broad knowledge of Dennis Klatt. It has been immeasurably enriched by the theoretical insights and experimental grounding of Ken Stevens. The transfer to the Web would not have been possible without the encouragement and support of Helen Hanson and the MIT OCW staff.