UNDERGRADUATE SPEECH PROCESSING AWARENESS

Marc Ressl, Jorge Prendes and Roxana Saint-Nom

BUENOS AIRES INSTITUTE of TECHNOLOGY (ITBA) Argentina,
mressl@itba.edu.ar, jprendes@itba.edu.ar, saintnom@itba.edu.ar

ABSTRACT

Teaching a speech processing course in an undergraduate engineering program is a challenge, especially in a country where research is not a priority. We address the concerns of providing a suitable training to students, of designing a syllabus under many constraints, of keeping students motivated and of finding the appropriate way of grading. Constant innovation has proven to be a key in this course's success, encouraging our students to choose our program among others. The increase in enrollment over the years is both an achievement and a challenge. This initiative creates academic prospect in the future of several students and promotes the development of the SP area in our University

***Index Terms*—** Signal Processing Education, Speech Processing

1. INTRODUCTION

In this paper we shall discuss the teaching methodology and outcomes of a speech processing course. It is delivered in the senior year of a 5-year EE undergraduate engineering program at Buenos Aires Institute of Technology (ITBA) in Argentina. As observed in [1], teaching graduate-level courses in the underground context is challenging. An additional complexity is that most academic programs in Argentina are not science but industry oriented. It was furthermore desired to test new teaching methodologies, and to produce a course taking advantage of the latest tools on the Web.

The first problem is adapting ourselves to this new environment; breaking away from the traditional lecture and evaluation scheme. The professor must also encourage 5th year students to face complicated subjects. Finally, it is important to consider this course as a guide to motivate students to a research career, hopefully leading them towards the development of a scientific vocation.

In order to introduce this course we shall first describe the context in which it is placed.

Electives (Senior): Signal Processing			
First Semester		Second Semester	
Digital Communications	*6*	DSP-FPGA	*6*
Adaptive Filtering	*6*	Image Processing	*3*
Neural Networks	*3*	Speech Processing	*3*

Table 1. Courses in the Signal Processing specialization

Our Speech Processing course is part of a senior year specialization (see Table 1) in Signal Processing. It follows the course *Adaptive Filtering*, which focuses on spectral estimation, LMS and RMS algorithms, and array processing. There are also two other prerequisites: *Analysis of Digital Signals and Systems* and *Random Signals*, which provide the necessary foundation for understanding and developing speech algorithms.

2. SYLLABUS OVERVIEW

2.1. Motivation

Speech processing is a wide body of knowledge. Faculty may face several challenges when designing a course in speech processing at the undergraduate level.

Student motivation is a key factor, but of no less importance is the proper design of the abilities a student should acquire. Furthermore, knowledge of the capabilities of students and professors is also essential.

Finally, we consider that 5th year senior students must get involved in hands-on examples, interactive learning and challenging projects because they represent unfailing incentives.

We are thus driven by the challenge of adapting complex topics to this context [2]. Moreover, we would also like to develop the student's appreciation of speech related problems. Hopefully, fresh blood in some of our R+D projects will be the outcome.

To our advantage is the fact that students have a good deterministic and stochastic signal processing background. On the other hand, we face several constraints:

- we are time-limited by the number of credits to 54 hours
- students are, at the time of this course, engaged in the capstone project
- except for those who decide to go to graduate school, this will be their only speech processing experience

As a result of these particulars, we designed the beginning of the course as an introduction focusing on speech applications. Several state-of-the-art speech synthesis and speech recognition examples display the full potential of the speech processing area.

We then provide an overview of fundamentals, including the speech production system, the auditory system and several topics of psychoacoustics.

An insight of analysis and synthesis tools in the time and frequency domain follows. As a next step we further develop the already familiar spectrogram, and introduce filterbank summation, cepstrum and LPC methods. Interactive materials, from sources like http://cnx.org/ are exposed. It is then that we present a first project where these concepts are to be used and developed.

The last part of the semester is oriented towards major applications: speech coding, speech synthesis and speech recognition. We don't dwell in deep detail because of our restrictions. Basic algorithms are presented, and after providing some tutoring about possible lines of research, we motivate students towards a second project that develops the topics they liked the most.

2.2. Syllabus

In order to support the ideas aforementioned, the full syllabus is described. As state-of-the-art textbook we make use of [3].
There are four blocks:

- *Introduction to Speech Processing*
  Speech signal processing: purpose, language, the speech communication circuit, neuroscience, history, applications, the scope of research, practical demonstrations.
- *Speech Production and Auditory System*
  Anatomy of the speech production system, articulatory phonetics, the IPA alphabet, spectrograms and waveforms, prosody, speech production models, anatomy of the auditory system, psychoacoustics, auditory system models.
- *Time and Frequency Domain Methods*
  Time and frequency domain representations of speech, window characteristics and time/frequency resolution trade-offs, estimation of speech production model parameters: energy and pitch extraction, short time Fourier analysis and synthesis, filter bank summation, the cepstrum, autocorrelation and covariance linear prediction of speech, analysis-synthesis systems, optimality criteria in time and frequency, optimum LPC parametrization.
- *Speech Applications*: Coding, Synthesis and Recognition [4], [5], [6], [7]
  Speech coding: PCM, ADPCM, CELP. Speech synthesis: language processing, prosody, diphone and formant synthesis; time domain pitch and speech modification. Speech recognition: hidden Markov models and associated recognition and training algorithms. Speaker Recognition and Verification. GMM and SVM systems. Speech and speaker corpora.

To get practically oriented MATLAB examples, we use [8] in speech issues. This gives us the opportunity to illustrate concepts discussed in class, and gives students hands-on experience with important techniques.

2.3. The web as a teaching aid

In order to produce a highly interactive course, we utilize the Sakai Content Management System [9]. It is used as a concentrator and archive of information, and allows easy retrieval of documents related to the course, live chat, the ability to receive hand-ins electronically, and also features a grading tool and a calendar that is easily integrated into other applications.

We also make use of teaching material available freely on the web, of which the most valuable to us is video resources. For example, we use video for enhancing the understanding of the anatomy related to the human speech circuit [10], or for visually explaining the differences of several speech codecs. This also serves the purpose of producing a relaxed atmosphere, and helps avoiding mental overload by shifting the means of learning to other channels of communication.

We exploit software for showing concepts interactively: for example, we use an LPC analysis and synthesis tool [11] that produces a 3D visualization of the estimated tube model of the vocal tract in real-time.
This year we've also planned to introduce a highly innovative homework: groups must produce writing assignments on Wikipedia [12], improving pages of certain topics related to speech processing. The main advantage is that they have to deal with real-world situations, hopefully resulting in increased dedication. They also strengthen their ability to think critically and evaluate sources. Finally, their effort remains on-line for reference, instead of being discarded and forgotten.

3. R+D INVOLVEMENT

GEDA, the digital electronics research group of the EE department, features many projects that attract promising students from 3rd and 4th year. Some of these projects are related to signal processing: digital hearing aids, a distributed musical orchestra using dynamic synthesis, and a speaker verification platform.

Students are encouraged toward signal processing through their commitment to these projects, and tend to follow the signal processing specialization.

Therefore, GEDA research projects are advantageous to our Speech Processing course; projects tend to align and produce higher-quality results. The consequent synergy is indeed beneficial to GEDA. We therefore emphasize the value of coordinating real-world research projects with coursework, even in the undergraduate context.

4. SPEECH PROJECTS

As we'd already mentioned, part of the coursework consists of two projects: a mid-term project which focuses on the direct application of synthesis and analysis techniques, and a final project which should integrate the acquired knowledge into a functional application. The midterm assignment also serves as a jump-board towards the final project.

Both projects have a strong emphasis on research, so that students get a hands-on analytical experience and realize they are capable of doing research.

The fact that speech processing has a direct, audible result is fortunate as it keeps motivation high, and shows the potential of learned techniques.

Based on student's capabilities, we encourage the use of MATLAB and its Signal Processing Toolbox. This allows us to put the focus on concepts and not on implementations. We do not emphasize real-time applications and leave that for the correlative *DSP-FPGA* course.

In order to portray achieved results, we shall describe two sample projects that took place in 2010.

4.1. Prosody change based on the PSOLA algorithm

This project is designed for mid-term level. It consists of an analysis and implementation of Pitch Synchronous Overlap-And-Add algorithm (PSOLA) [13]. This algorithm is based on the overlap-and-add method explained in class, and is capable of speeding up/slowing down speech without modifying the formants of the re-synthesized voice.

The student Andrés Totorica based his research on S. Lemmetty's master thesis [14], a review of diverse speech synthesis methodologies. There are different approaches to change prosody using PSOLA, and the student had to choose and justify the most appropriate one. He compared time-domain, frequency-domain and linear prediction PSOLA, and reviewed strengths and weaknesses of each algorithm according to measured and perceived sound quality, computational complexity and applicability. He also tackled the problem of applying PSOLA to unvoiced segments. Finally, he implemented the time-domain variant [15], using autocorrelation as the pitch estimation method. After a thorough analysis of his results, he proposed several improvements, for instance using a more advanced voice-activity detector. A sample result of his work can be seen in Fig. 1:

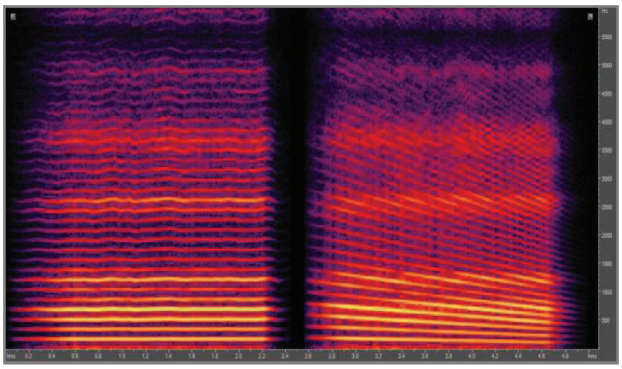

Fig. 1. to the left, a spectrogram of an /a/ phoneme, to the right, application of PSOLA pitch shift.

In this case, the phoneme /a/ was recorded. It can be observed to the left that this phoneme's pitch is not constant over time. The algorithm estimates this pitch, and transforms it into the same phoneme with a linearly varying pitch, where the fundamental frequency ranges from 200Hz to 100Hz.

4.2 Pitch correction based on Spectral Modeling Synthesis

This project is an end-term assignment. It implemented a pitch correcting (auto-tune) algorithm based on Spectral Modeling Synthesis (SMS) [16]. The student Nicolás Baum was quite challenged, as these methods were not lectured in class and are at the Ph.D. level.

SMS detects and tracks the harmonic components of the spectrum, splitting a signal into stochastic and deterministic components. Even though an implementation of SMS is freely available on the web, the student decided to develop it by himself in order to gain a deeper understanding of the theory. This proved to be more of a barrier as the harmonics tracking algorithm was difficult to debug, and unfortunately our student did not achieve good results, affecting his presentation negatively.
Nevertheless, he was still very motivated and insisted on preparing a second presentation, at which he got his harmonics tracking algorithm working almost as well as the one found in [16].

5. ASSESSMENT APPROACH

To us, the balance between practice and theory is very important; hence research and development projects are an significant part of the assessment. Research is a good indicator of how much a student understands theory, and development shows how much practice was acquired. Depending on the number of students, projects are assigned on an individual basis or to small groups.

Projects must include an oral presentation. A written report and the developed software must be handed in. The grading of the project is measured by the quality of the presentation and the written report. The purpose of the presentation is to teach students how to defend the work they have done and prepare them for a competitive environment. Not only may professors ask questions, other students are also encouraged to do so. This contributes to their attitude mark.

In the case of grading courses with several students, assessing individual grades is not convenient. We thus apply four follow-up tests. This practice also intends to keep students up to date with the course. Follow-up tests usually last for an hour, and they are focused on recent topics studied in class. Each individual test does not contribute heavily to the grade, but altogether they account for approximately a third of the total grade.

For our present course, we calculate the assessment with equation (1); PR stands for Projects, AM for Attitude Mark and FT for Follow-up Tests. The attitude mark has the same weight as a follow-up test, and encourages students to participate in class and develop good presentations.

$$AssessmentGrade = 0.6\frac{\sum PR}{N_{PR}} + 0.4\frac{AM + \sum FT}{1 + N_{FT}} \quad (1)$$

In the past, we have experimented as well with role-switching, the exchange of the roles of professor and student. A student is assigned a course subject, and he or she must explore, study and prepare a class. Some of the abilities that are learned are managing time, seeking effectiveness in the way of communicating ideas, giving examples, and being prepared to face complicated questions. The professor must also ask certain questions, so that it is assured that relevant topics are covered.

Unfortunately this method is much more time-consuming and is only appropriate for small courses, but the advantages are considerable: students get a much deeper understanding of a subject, and get exposed to a real-life experience. While grading role-switching classes, professors learn how a student deals with giving the presentation, and that tends to be a much better way of grading a student than just assessing the content of his class.

6. PRELIMINARY RESULTS

Fig. 2 and table 2 represent the evolution of our course, since it was first given in 2009. It is quite remarkable to realize how the speech processing course evolved in the last 3 years: in 2009, 7% of senior-year EE students chose speech processing, while in 2011 this number increased to 27%. This definitely marks the success of our approach. Although the difficulty of the course has been *amplified* due to the improvement of our course material, grades have also increased, which is an indication that *their* motivation *was* improved. It is also interesting to note that the GPA of students taking the Speech Processing course tends to be higher than the average *any given year*. For example, this year the average GPA of students of our course is 8,01 / 10,0, much higher than the average 6,51 / 10,0 of 2011 senior-year students. We are thus attracting the best students.

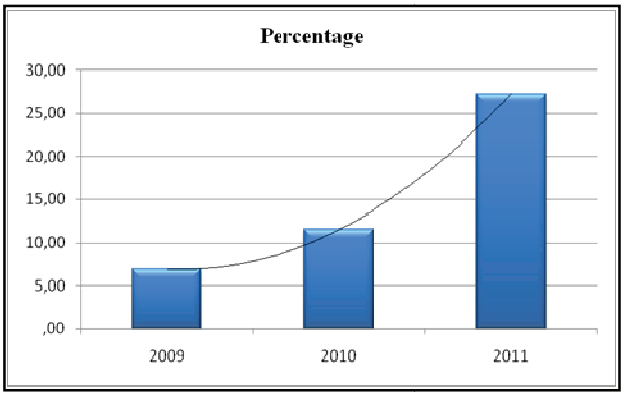

Fig. 2. Students enrolled in the Speech Processing course as a percentage of senior-year EE students

Year	2009	2010	2011
Students in SP / senior-year EE students	2 / 29	3 / 26	9 / 33
Average grade	7 / 10	7,5 / 10	8,2 / 10 *

*: grade was calculated with the available data

Table 2. Progress of the Speech Processing course

In addition, we have conducted interviews to former students, and evaluated their feedback.

Lautaro Carmona was encouraged to study signal processing by way of a language detection project he had produced for the *Analysis of Digital Signals and Systems* course. He is presently working on the capstone project which is related to digital signal processing in the video domain.
http://imxcommunity.org/profile/LautaroCarmona

According to Andrés Totorica, the most interesting part in the Speech Processing course was the implementation of the algorithms, and getting hands-on experience with MATLAB and its toolboxes. He is about to finish an internship at *Institute de Recherche en Informatique de Toulouse*, France where he is applying image processing to ultrasound medical images. His collaboration was applied to the project thesis:
http://www.irit.fr/Sujets-de-theses-et-de-stages,844

7. CONCLUSIONS

A new course in Speech Processing for undergraduate students was designed and implemented. In the context of a new SP specialization of a 5-year EE program, we were able to captivate candidates to an area of limited development in Argentina. Complex algorithms became achievable applications. Students could seize Speech Processing problems that where unknown before. This result is a combination of an appropriate background and a successful course.

Our formula contains a balanced syllabus, challenging projects anchored in our R+D center, modern web tools and forward-looking assessment. Good grades and a growing number of students create a favorable environment to our plans.
The new SP undergraduate specialization is achieving our expected goals [1]. Speech related projects are increasing. There are students who are planning to follow their graduate studies in Speech Processing. There are others already working in SP developments.

We have accomplished the first steps towards Speech Processing awareness among alumnae. It is encouraging to be able to predict more to come.

8. ACKNOWLEDGMENTS

We would like to to thank Lautaro Carmona, Nicolás Baum and Andrés Totorica for their feedback and contributions.

9. REFERENCES

[1] Saint-Nom, R., "Advanced SP topics in an innovative undergraduate EE curriculum," *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, vol., no., pp.2641-2644, April 2008.

[2] Pendse, R.; Johnson, E., "Teaching an undergraduate class vs. graduate class: is there a difference?," *Frontiers in Education Conference, 1996. FIE '96. 26th Annual Conference., Proceedings of*, vol.1, no., pp.59-62 vol.1, 6-9 Nov 1996.

[3] Rabiner and Shafer**,** *Theory and Applications of Digital Signal Processing*, Prentice Hall (2010), ISBN**:** 0136034284

[4] Gersho, A., "Advances in speech and audio compression". *Proceedings of the IEEE [on line]*. Vol. 82, issue 6, pp. 900-918, June 1994.

[5] Picone, J.W., "Signal modeling techniques in speech recognition". *Proceedings of the IEEE [on line]*. Vol. 81, issue 9, pp.1215-1247, September 1993.

[6] Campbell, J.P., Jr., "Speaker recognition: a tutorial", *Proceedings of the IEEE*, vol.85, no.9, pp.1437-1462, Sep 1997

[7] Rabiner and Juang, *Fundamentals of Speech Recognition*, Prentice Hall (1993), ISBN: 0130151572.

[8] McLoughlin, *Applied Speech and Audio Processing*, Cambridge (2009), ISBN: 0521519543.

[9] Sakai Project, http://www.sakaiproject.org/

[10] Brandon Pletsch, *Auditory Transduction*, http://www.youtube.com/watch?v=PeTriGTENoc

[11] Misra A., Wang G., Cook P., real-time LPC visualization, http://soundlab.cs.princeton.edu/software/rt_lpc

[12] Wikipedia, http://en.wikipedia.org/wiki/Wikipedia:School_and_university_projects

[13] Moulines, E.; Charpentier F.; "Pitch Synchronous Waveform Processing Techniques for Text-to-Speech sis Using Diphones", *Speech Communication, 1990, Volume 9, No. 5-6, Page(s): 453-467.*

[14] Lemmetty, S., *Review of Speech Synthesis Technology*, Master Thesis, Helsinki University of Technology (1999)

[15] Chalamandaris A. et al., "An Efficient and Robust Pitch Marking Algorithm on the Speech Waveform for TD-PSOLA", *2009 IEEE International Conference on Signal and Image Processing Applications.*

[16] Serra X., *A System for Sound Analysis / Transformation / Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. Thesis, Stanford University (1989)