THE SAS PROJECT: SPEECH SIGNAL PROCESSING IN HIGH SCHOOL EDUCATION

Dushyant Sharma*, Alankrit Poddar†, Sumanna Manna† and Patrick A. Naylor‡

*Voicemail-To-Text Research, Nuance Communications Inc., California, USA
†International Sahaja Public School, Talnoo, Himachal Pradesh, India
‡Electrical and Electronic Engineering, Imperial College London, UK

ABSTRACT

We describe the Speech And Sound (SAS) outreach project with the aim of introducing high school students to speech signal processing through the real-life example of automatic speech recognition. The syllabus was designed to help students understand how the concepts they learn as part of the physics, mathematics and computing courses relate to real-life applications. The six week project was organized into a mixture of informal lecture and practical sessions and the students were encouraged to engage in informal discussions with the instructors with any questions and ideas. The project was piloted at an international high school in India with 10th, 11th and 12th grade students. By the end of the course, the students had gained a high level understanding of the many technologies that make up such a complex system, as evident by the high overall scores in the final assessment.

Index Terms— Engineering, Outreach, ASR, DSP, STEM

1. INTRODUCTION

We live in a world where engineered technology is ubiquitous and even though the early adoption of new technologies is fascinating to high school children, there continues to be a gap between the learning of the foundation Science, Mathematics and Technology subjects at the high school level and the application of those theories or concepts in the engineering disciplines. A number of outreach activities have been carried out over the past three decades to try and bridge this gap, by providing high school students a hands-on experience of real-life applications in applied sciences and mathematics. It is also well documented that the transition from high school to the first year of engineering programs can be problematic for many students [1, 2], which has inspired many outreach programs that have been designed to familiarize high school students with engineering concepts [2, 3].

The Infinity project [4] [5] aimed at teaching aspects of digital signal processing and digital music to high school students, giving them an opportunity to learn about engineering design principles in the context of information technology. The project was piloted in 10 schools in the state of Texas between the year 2000 and 2001. The course content included waveform synthesis and modeling, a real-time sinusoidal MIDI player and building a low cost loudspeaker. The hands-on laboratory aspect of the course was modeled around a Texas Instruments TMS320C31 evaluation module and a third party visual applications builder software that allows prototyping of real-time signal processing systems.

In the US, Project Lead The Way (PLTW) [6] is a non profit organization that promotes K-12 Science, Technology, Engineering and Mathematics (STEM) programs in a number of areas, including Engineering, Biomedical and Computer Science. PLTW delivers these programs to over 6,500 schools in the USA through a strict program of certification for individual instructors as well as schools. The Ingram School of Engineering at Texas State University partnered with PLTW and the E3 alliance to develop a novel outreach activity that introduced Computer Aided Design (CAD), Micro electromechanical Systems (MEMS) and micro-robotics [7] to high school students. A seven day research camp with 16 high school students was carried out with an objective of designing, prototyping and testing a bridge.

A multifaceted program based on a mechatronics platform was presented in [2]. The University of Ottawa (UO) undertook several outreach activities to help students make the transition between high school and engineering education and also to increase their awareness to the engineering profession. Through the use of an interdisciplinary subject such as mechatronics the program was able to introduce students to aspects of mechanical, electrical and computer engineering. As part of the outreach activities, engineering students worked in teams to solve problems relating to senors, sensor-actuator or control systems and presented their work to high school students. This was found to be very well received by the high schools students, with 97% of students (from a total of 44) who responded to the survey agreed that the activity helped them understand how the math and science they learn at school applies to real life.

Rosen et. al. [8] present an outreach program that used image processing to reach out to artistically inclined K-12 students to stimulate an interest and help them understand the real-life application of math and science concepts. The course content consisted of image color manipulation (RGB channel

analysis), image denoising and edge detection. An outreach project using music technology and DSP to teach STEM topics was proposed in [9], where a one week Summer Music Technology (SMT) program was carried out at Drexel University with the aim of introducing 9th and 10th grade students to modern music technology and how those concepts relate to the STEM subjects. Topics included music recording and production, speaker technology, wave nature of sound, sound synthesis, musical instrument acoustics, digital signal representation and music information retrieval.

This paper describes a novel outreach activity to an international school in Dharamshala, India, where 10th, 11th and 12th grade students (age group from 16 to 18 years) were given an introduction to speech signal processing as part an overall goal of providing the students a high level understanding of the major components of an Automatic Speech Recognition (ASR) system and how each of those components relate to the STEM subjects they learn as part of their syllabus. The example of popular mobile personal assistants and Nuance's Dragon dictation software [10] were used to demonstrate real-life applications of ASR technology. To the best of our knowledge, this is a first outreach activity of its kind in a high school in India. More details of the project aims and methodology are presented in Section 2. The course syllabus is presented in Section 3 and the outcomes of the project are presented in Section 4, with conclusions in Section 5.

2. SPEECH AND SOUND (SAS) PROJECT

The aim of the SAS project was to help senior high school students understand the key components of an automatic speech recognition system, as used in mobile personal assistants and dictation software and thereby help them develop an appreciation of how the STEM subjects they learn as part of their curriculum relates to real-life applications. In order to understand the various components of an ASR system, a number of key topics had to be covered, which included the theory of waves and representation as an analogue signal, sampling and quantization, frequency domain representation and processing as well as machine learning. The main topics covered in the SAS syllabus are presented in Section 3. The SAS project was carried out over a six week period between November and December 2014.

Given the limited duration of this outreach activity and the budget constraints of being at a school in India, the SAS project utilized open source software such as PRAAT [11] and Audacity [12]. The teaching was carried out by the first and second authors of the paper, both having an engineering background. The project was allocated 10 lessons for each year group, including two 45 minute lessons during the physics timetable and two 45 minutes lessons as extra-curricular activity time in the evenings. The teaching was a combination of informal lectures, video presentations and hands-on practical sessions, where the students were encouraged to engage

with the instructors in discussions around the topics being presented. The use of powerpoint presentations helped give a direction to the lecture sessions and bring a change to the typical black-board note taking style of teaching they were used to in other classes.

3. SYLLABUS

Here we present the topics that were taught as part of the SAS project. The syllabus for this outreach activity was constructed such that the underlying physics and mathematics principles would be accessible to the age group from grade 10 to 12. Taking account of insights gained from previous outreach activities published in the literature, in particular [4, 5, 9], a large part of the theory of the acoustics was presented in the context of music technology. For example, the mathematical representation of tones was shown in the context of electronic synthesizers. The five main modules covered in the teaching and practical sessions are presented in the following subsections.

3.1. Physics of waves

This module reviewed the basic theory of waves that was already covered in the physics syllabus. This helped give a gentle start to the teaching by mostly revising some theory the students were already familiar with. The key difference was that this was presented in the context of sound waves, where the idea of synthesizing simple and complex tones using the mathematical formula, $y = A \times \sin(\omega_1 t) + B \times sin(\omega_2 t)$ was presented using the PRAAT [11] software. The measurement of sound power using a logarithmic scale was also presented.

3.2. Digital Signal Processing

In this module, the concept of a signal was first introduced with an example representation of a simple analogue sound signal such as a pure tone. The concepts of sampling and quantization were presented and explained with the help of an audio compression activity in Audacity [12], based on linear Pulse Code Modulation (PCM). The activity involved loading a linear PCM audio file and experimenting with different sample rates and bit depth and assessing the perceptual effects of those processes. The operation of a microphone was explained and also how the acquired signal may need to be amplified appropriately to avoid distortions. The sound recording feature in PRAAT was used to illustrate these effects.

3.3. Speech Processing

With the concept of a digital signal and its proper acquisition having been explained, the physiology of the human speech production system was introduced and the concept of the basic sounds of a language or phonemes was presented [13].
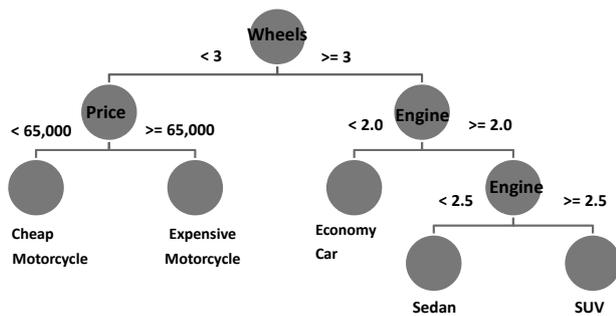
Fig. 1. The CART model constructed using the training set of the car classification toy example.

The concept of frequency analysis was presented as a means of obtaining a more consistent representation of the signal. A practical activity of viewing the frequency spectrum of various speech signals was performed in Audacity. The concepts of bass, treble, reverberation [14] and gain were explained through a hands-on equalization lab activity exploiting a graphical equalizer in Audacity in the context of music production.

3.4. Machine Learning

As most ASR systems are data driven, machine learning plays an important part in the system. In the SAS project, the CART [15] binary tree algorithm was used to illustrate the general concept of machine learning through a car type detection toy problem. The task was to identify the class of vehicle from the predefined set, {cheap Motorcycle, expensive Motorcycle, economy car, sedan, SUV} and the features were the set {ID number, make, color, engine size, price, number of wheels, class}. The students were given a separate training data-set (10 data points) and were assisted with the initial construction of the tree. Table 1 presents the training data for the car classification example and the corresponding CART model in presented in Fig. 1.

Once a CART classification tree had been constructed on the training set, a test set of 10 unseen examples was used to evaluate the models. The sufficiently small size of the data allowed the students to construct the trees by hand. This module proved to be very exciting for the students and gave them a hands-on experience of machine learning. Other applications of such techniques was also presented (fraud detection and speech quality assessment [16]) .

3.5. Automatic Speech Recognition

The last module was an overview of a typical ASR system [17], consisting of audio acquisition, pre-processing (voice activity detection and segmentation into short-time frames), mapping to the frequency domain, conversion to the compressed Mel-Frequency-Cepstral-Coefficient (MFCC)

representation, followed by acoustic and language modeling. In order to make the theory easier to understand, the two key components of an ASR system, the acoustic and language model, were simply presented as black-box modules that map the acoustic features into phonemes and correct for spelling and grammar, respectively. These concepts were demonstrated using the Dragon dictation software [10]. This allowed the students to appreciate the complexity of the problem of ASR, as well as the key concepts behind this technology and the way in which it relates to the theories and concepts taught in the Physics, Mathematics and computing courses at school.

4. OUTCOMES

In this section we present the outcomes of the SAS outreach project, including the assessment criteria and results.

4.1. Assessment

In order to gauge the success of the project, a short assessment was carried out at the end of the course. The assessment was designed to test the student's grasp of the key concepts in the study (rather than checking if they remembered all the technical details) and their ability to link those concepts to the STEM subjects already taught as part of the curriculum. The assessment was organized into a set of eight multiple choice questions covering all five topic ares mentioned in Section 3. In addition, three questions having the format of assigning text labels to images based on a process flow chart were also included. The ASR theory was tested with a flow chart completion question, where the important processes had to be correctly identified as shown in Fig. 2. In addition to the short assessment, the students were asked to complete a short questionnaire about the SAS project. The questionnaire contained six questions for rating the course material, the presentation style (mix of audio, visual and practical demos) and the perceived value of the outreach activity. The final two questions allowed the students to give more detailed feedback on the most interesting aspects of the project and possible improvements for the future.

4.2. Results

The test results are presented in Table 2 and it can be seen that overall, the students seemed to pick up the concepts of different types of waves, machine learning and the basic building blocks of ASR systems very well, as shown by an average score greater than 70% in all modules. The signal processing module, in particular the digital-to-analogue conversion theory was found to the the most challenging as reflected by the lower average score for that module. The ASR and machine learning topics received the highest ratings as the topics of

ID	Make	Price (Rs.)	Engine Size (Liters)	Number of Wheels	Color	Class
1	BMW	30,000	3.0	4	Blue	SUV
2	Mercedes	30,000	2.5	4	Green	Sedan
3	Tata	5,000	1.2	4	Blue	Economy Car
4	Nissan	9,500	1.2	4	White	Economy car
5	Mitsubishi	25,000	3.0	4	Black	SUV
6	Honda	30,000	2.5	4	Green	Sedan
7	Honda	25,000	2.8	4	Blue	SUV
8	BMW	30,000	2.5	4	White	Sedan
9	Royal Enfield	9,500	0.9	2	Black	Expensive Motorcycle
10	Yamaha	3,500	0.6	2	Black	Cheap Motorcycle

Table 1. The training data samples for the car classification toy example. The students were guided in the tree building process and were asked to consider which features are most useful. For example, the ID number or the color are poor choices for features as they would not generalize. The recursive partitioning was explained in terms of considering which feature would help split the data with the most accuracy (analogous to the concept of purity in the CART algorithm).
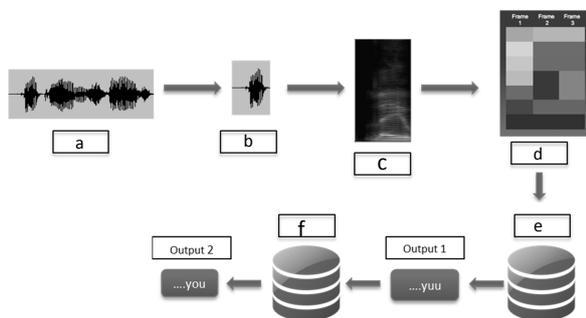

Fig. 2. The flow-chart based ASR process identification question. The correct block descriptions, starting from the top left and moving clockwise are as follows: (a) speech utterance, (b) a small segment of the utterance, (c) convert to frequency representation, (d) convert to MFCCs, (e) Acoustic model, (f) Language model.

Module	Average Score	Minimum Score
Wave Theory	83.3%	66.7%
Signal Processing	70.0%	50.0%
Machine Learning	75.0%	70.0%
ASR	85.7%	37.5%

Table 2. The overall test results of the multiple-choice assessment, split by the topic area.

most interest by the students, followed by the speech processing and wave theory topics. The students found the quality of the teaching material and the practical style of teaching to be extremely good. Some students reported a desire for the course to be more detailed and to run for a longer period. One of the students reported the following, "when there are more practical work, things are memorized much better and the understanding improves" in their explanation for why they prefer a more practical and application based approach to teaching of STEM subjects.

5. CONCLUSIONS

We presented a novel outreach project with the aim of teaching select topics in speech signal processing to 10th, 11th and 12th students at an international high school in Dharamshala, India. The topic of automatic speech recognition was used to illustrate a real-life application of those signal processing techniques and the way in which topics taught as part of their physics, mathematics and computing courses are applied in engineering. Moreover, an innovative machine learning topic was also successfully taught using the CART algorithm. The students were guided with the CART model building process and were able to construct a working model from a small training set and apply it to an unseen test set. The topic of equalization was also very well received as many of the students had a deep interest in music. A final assessment covering the key topics from the course was carried out and the average score in each topic was between 70% and 85.7%. This indicates that the students grasped the main topics very well. Moreover, feedback from the student self-assessment questionnaire was overwhelmingly positive. Students reported that the teaching style, comprising of informal lectures and practical sessions helped stimulate an interest in the subject and was preferred to the traditional note taking style of teaching currently adopted at the school. The SAS project will be re-run in the 2015 academic year after incorporating the feedback and experience from the current study and collaboration is already underway to expand this initiative to other schools.

6. ACKNOWLEDGMENTS

REFERENCES

[1] Dan Bundy, "The freshman seminar: Assisting the freshman engineering studentś transition from high school to college," in *Proc. American Society for Engineering Education Annual Conference & Exposition*, 2001.

[2] R.W.Y. Habash and C. Suurtamm, "Engaging high school and engineering students: A multifaceted outreach program based on a mechatronics platform," *Education, IEEE Transactions on*, vol. 53, pp. 136–143, 2010.

[3] Lance Jackson, "Applying virtual technology: A joint project between the university of queensland and townsville state high school," *Australian Science Teachers Journal*, vol. 46, pp. 19, 2000.

[4] S.C. Douglas, "The infinity project: digital signal processing and digital music in high school engineering education," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001.

[5] S.C. Douglas, M.P. Christensen, and Geoffrey C. Orsak, "Designing pre-college engineering curricula and technology: Lessons learned from the infinity project," *Proceedings of the IEEE*, vol. 96, pp. 1035–1048, 2008.

[6] "Project lead the way," Online, https://www.pltw.org/about-pltw.

[7] W. Stapleton, B. Asiabanpour, H. Stern, and H. Gourgey, "A novel engineering outreach to high school education," in *Frontiers in Education Conference, 2009. FIE '09. 39th IEEE*, 2009.

[8] G. Rosen, J. Silverman, and A. Chauhan, "Connecting artistically-inclined k-12 students to physics and math through image processing examples," in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th*, 2009.

[9] Y.E. Kim, A.M. Batula, R. Migneco, P. Richardson, B. Dolhansky, D. Grunberg, B. Morton, M. Prockup, E.M. Schmidt, and J. Scott, "Teaching stem concepts through music technology and dsp," in *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE*, 2011.

[10] Nuance Communications Inc., "Dragon naturally speaking 12.0 premium [computer software], http://www.nuance.com/dragon/index.htm," 2014.

[11] Paul Boersma & David Weenink, "Praat: doing phonetics by computer [computer program]. version 5.3.51, http://www.praat.org/," 2013.

[12] "Audacity [computer software], version 2.0.6, http://audacity.sourceforge.net," 2014.

[13] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1978.

[14] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[15] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, CRC Press, 1984.

[16] D. Sharma, P. A. Naylor, N. Gaubitch, and M. Brookes, "Short-time objective assessment of speech quality," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Barcelona, Aug. 2011.

[17] G. Saon and Jen-Tzung Chien, "Large-vocabulary continuous speech recognition systems: A look at some recent advances," *Signal Processing Magazine, IEEE*, vol. 29, pp. 18–33, 2012.