Advanced Sinusoidal Modeling of Speech and Applications

Overview

Speech is a fundamental mode of communication. The speech production-perception apparatus duo have evolved in nature to enable effective human-human communication. The speech production takes place under the controlled cognitive guidance. The production apparatus consists of vocal tract system having flexibility due to articulators and at the same time has some inertia. As a result, the system takes some non-zero time to change from one shape to the other. This aspect has been used in practice to justify the short term stationarity assumption employed in speech signal processing. That is, the characteristics of the speech signal are assumed to be stationary typically when viewed in segments of 10-30 ms. Based on short term processing, several methods have been proposed in the literature including short term Fourier transform, cepstral analysis, linear prediction analysis and sinusoidal analysis. The short term Fourier transform is the modified version of Fourier transform using window functions for analyzing non-stationary signals like speech. The cepstral analysis is based on source – system separation by performing a non-linear operation in the frequency domain. The linear prediction analysis involves source-system separation based on prediction process. The sinusoidal modeling is based on estimating amplitude, frequency and phase values of set of sine waves. All these methods are on the assumption of short term stationarity. However, due to the time varying excitation at a faster rate, this assumption is not fully correct as there are variations within short segments of 10-30 ms also. Hence the quest for exploring new methods for speech signal processing.

The advanced sinusoidal modeling is the modified version of conventional sinusoidal modeling based on the motivation to minimize the effect of assumption of short term stationarity. For instance, adapative sinusoidal modeling that refers to adapting the parameters of sinusoidal model to the local characteristics (phase and / or amplitude) of the analyzed speech signal. The conventional sinusoidal modeling will have a set of sine waves whose frequency and/or amplitude are constant. In general, the adaptive sinusoidal is based on the principle of projecting a signal segment onto a set of non-parametric, time-varying, non-stationary set of sinusoidal basis functions inside an analysis window.

The sinusoidal modeling has found widespread applications in the domains of speech and audio processing. The primary application of sinusoidal modeling is in speech and audio analysis. Stressed speech analysis and recognition, speech classification, voice transformation and synthesis are other applications. The course will explain in detail about these applications and also future trends of these applications.

Objectives

- *Overview of speech processing*
- *Conventional sinusoidal modeling*
- *Advanced sinusoidal modeling*
- *Applications of sinusoidal modeling*
- *Possible research directions in sinusoidal modeling*

Course Details	**Dates : 26 December 2016 – 30 December 2016** **Day 1:** *Lecture 1: 9:00 to 10:00 AM* — **Overview of speech processing** *Lecture 2: 10:30 to 11:30 AM* — **Conventional Sinusoidal modeling of speech** *Laboratory 1: 2.00 to 5.00 PM* — **MATLAB exercises on Sinusoidal modeling of speech** **Day 2:** *Lecture 3: 9:00 to 10:00 AM* — **Importance of phase in speech processing** *Lecture 4: 10:30 to 11:30 AM* — **Adaptive sinusoidal modeling of speech** *Laboratory 2: 2.00 to 5.00 PM* — **MATLAB exercises on Phase and Adaptive Sinusoidal modeling of speech** **Day 3:** *Lecture 5: 9:00 to 10:00 AM* — **Stressed speech analysis using sinusoidal modeling** *Lecture 6: 10:30 to 11:30 AM* — **Speech analysis using adaptive sinusoidal modeling** *Laboratory 3: 2.00 to 5.00 PM* — **MATLAB exercises on stressed speech analysis** **Day 4:** *Lecture 7: 9:00 to 10:00 AM* — **Speech classification using adaptive sinusoidal modeling** *Lecture 8: 10:30 to 11:30 AM* — **Speech transformation using adaptive sinusoidal modeling** *Laboratory 4: 2.00 to 5.00 PM* — **MATLAB exercises on stressed speech classification and transformation** **Day 5:** *Lecture 9: 9:00 to 10:00 AM* — **Overview of audio processing** *Lecture 10: 10:30 to 11:30 AM* — **Audio modeling using sinusoidal modeling** *Examination 1: 2.00 to 5.00 PM* — **Course examination**
Who can Attend ?	▪ **Graduate and Post-graduate students at all levels (B.Tech/B.E./M.Tech/M.E./M.Sc./Ph.D) from Computer Science and Electronics (CSE / ECE) background or equivalent.** ▪ **Faculty from reputed academic institutions and technical institutions.** ▪ **Persons from R&D organizations and industries and project staff.**

Fees	*The participation fees for taking the course is as follows:* *Students :* **INR 1000/-** *(Refundable)* *Participants from abroad :* **USD 500 /-** *Industry/ Research Organizations:* **INR 20,000 /-** *Academic Institutions:* **INR 10,000 /-**
Course Co-ordinators	***Prof. S. R. Mahadeva Prasanna*** *: Principal Coordinator* *Department of Electronics and Electrical Engineering, IIT Guwahati , Guwahati – 781039, India* *Tel: +91 361 258 2513 / 2082 (O), 258 4513 (R) , +91 9954008138 (M)* *Email:* ***prasanna@iitg.ernet.in*** ***Prof. S. Dandapat*** *:* *Department of Electronics and Electrical Engineering, IIT Guwahati , Guwahati – 781039, India* *Tel: +91 361 258 2505 (O), 258 4505 (R), +91 9435119011 (M)* *Email:* ***samaren@iitg.ernet.in***
Registration Procedure	**Please follow the following steps for registration :** **1.** Go to GIAN website (**http://www.gian.iitkgp.ac.in/GREGN/index**) and register. You need to pay a one-time fee of **INR 500 /-** for registration. **2.** Select the course you wish to enroll for – **161006002** (*Advanced Sinusoidal Modeling and Applications*). **3.** Once you enroll for the course, the course coordinators will be notified. The course coordinators will **shortlist** the candidates, and **selected** *candidates will be notified by email.* **4.** The **selected** candidates must pay the applicable fees by *online bank transfer / wire transfer / internet banking* to the following bank account. **Please keep the online transfer receipt for proof of transfer.** ***Bank Name*** : **STATE BANK OF INDIA** ***Branch Name*** : **IIT GUWAHATI BRANCH** ***IFSC Code*** : **SBIN0014262** ***MICR code*** : **781002053** ***Account Name*** : **IIT GUWAHATI R&D – MHRD** ***Account No*** : **31151533220** ***Account Type*** : **Savings** **5.** Fill the **course registration form** in GIAN portal. Email the **course registration form** and the **online transfer receipt** to the course coordinators.

The Faculty

Prof. Yannis Stylianou is Professor of Speech Processing at *University of Crete, Department of Computer Science, CSD UOC*, and Group Leader of the Speech Technology Group at *Toshiba Cambridge Research Lab, UK.* Until 2012, he was also Associate Researcher in the Signal Processing Laboratory of the *Institute of Computer Science ICS at FORTH.* During the academic year 2011-2012 was visiting Professor at *AHOLAB, University of the Basque Country, in Bilbao, Spain (2011-2012)*. He received the Diploma of Electrical Engineering from the National Technical University, N.T.U.A., of Athens in 1991 and the M.Sc. and Ph.D. degrees in Signal Processing from the Ecole National Superieure des Telecommunications, ENST, Paris, France in 1992 and 1996, respectively. From 1996 until 2001 he was with *AT&T Labs Research (Murray Hill and Florham Park, NJ, USA)* as a Senior Technical Staff Member. In 2001 he joined *Bell-Labs Lucent Technologies, in Murray Hill, NJ, USA* (now Alcatel-Lucent). Since 2002 he is with the Computer Science Department at the University of Crete while since January 2013, he is also with Toshiba Labs in Cambridge UK. His current research focuses on speech signal processing algorithms for speech analysis, statistical signal processing (detection and estimation), and time-series analysis/modeling. He has co-authored more than 150 scientific publications, and holds 15 UK and US patents, which have received more than 4400 citations (excluding self- citations) with H-index=30. He co-edited the book on *"Progress in Non Linear Speech Processing"*, Springer-Verlag, 2007. He has been the P.I. and scientific director of several European and Greek research programs and has been participating as leader in USA research programs.

Prof. S. R. Mahadeva Prasanna joined the *Department of Electronics and Electrical Engineering of Indian Institute of Technology Guwahati (IITG)* as Assistant Professor in 2004, became Associate Professor in 2007 and Professor in 2012. His areas of teaching and research interest include signal and speech processing. He has supervised 07 PhD Theses, 25 MTech Projects and 30 BTech projects. Currently 12 research scholars are pursuing PhD, 03 MTech and 10 BTech students are pursuing their projects under his supervision. He has over 175 publications in national and international journals and conferences. He has successfully executed many funded projects and currently many projects are ongoing. He has visited many countries to present his group research work and also work on collaborative projects that include USA, UK, Switzerland, Germany, Japan, France, Italy, Belgium, Argentina, and Singapore. He is *Associate Editor for Springer Circuits, Systems and Signal Processing, IETE Technical Review and IETE Education Journals*.

Prof. S. Dandapat joined the *Department of Electronics and Electrical Engineering of Indian Institute of Technology Guwahati (IITG)* in 1997 and became Professor in 2007. His areas of teaching and research interests include signal processing, particularly in the fields of biomedical signal processing, and speech processing. He has supervised many PhD Theses in these areas and published in reputed national and international journals and conferences.