Mining the Meaning

By Pranav Jawale (with Prof. Preeti Rao)
DAPLAB, Deptt. of Electrical Engineering, IIT Bombay

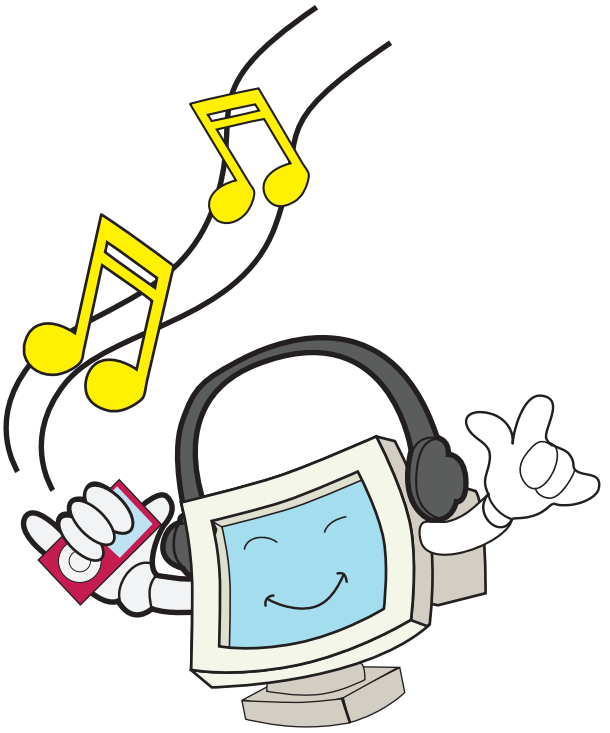
Modern lifestyle has made humans more machine-like, but modern science can also make machines more human-like, so that they can understand each other. Vision, hearing and touch are some of the ways in which humans gather information about their surroundings. We, at Digital Audio Processing Lab (DAPLAB) in the Dept. of Electrical Engineering, work on exploring the sound signal with a view to extract information of interest. It helps to understand how humans listen and make sense of sound. We need to reverse-engineer human perception and cognition of audio and use it to create a system that can match human abilities.

> *"We need to reverse engineer human perception and cognition of audio and use it to create a system that can match human abilities."*

Here we describe/introduce some of the projects which we have been working on.

Music Classification

Inside any music store you will find music categorized according to various descriptors (such as genre, artist, period, region, instrumentation, and maybe even mood or intent). Without these descriptors it will be a horrendous task for us to search for what we like. These descriptors are the musical metadata that is typically prepared via arduous manual tagging by human expert listeners. There are many million titles out there and many more are being added every week. We would like to automate the task of metadata extraction by audio computing. For that, we first need to understand the perceptually important characteristics of the audio signal that enable humans to classify the music. Clearly, musical attributes such as melody, rhythm and timbre play a role in music categorization. An important focus of audio research is to relate these musical dimensions to the low-level signal properties such as the time-varying spectrum, periodicity and intensity of the sound. Finally, pattern recognition methods serve as the bridge to link signal features to the high-level semantic characteristics.



Singing Tutor

Can machines be taught the critical appreciation of music? An immediate application might be a music tutor that delivers critical feedback to a beginner or advanced learner on the fine aspects of a performance. Music is an art form where even its maestros will differ on some things, how can we expect the machines to do the job? Currently we are tackling a small part of this hard problem. Signal processing algorithms, recently developed at DAPLAB, evaluate a singer based on similarity with a model singer on several musical dimensions.

Query by Humming

How often does it happen that you can recall a song but don't remember the exact lyrics? Humans most often remember a song by its tune. The tune or melody of a song is a particular sequence of notes in the song. Notes are told apart based on their pitch and duration. The melody recognition task involves listening to the tune of the song and recognizing the song. Any ordinary person with an ear for music can do that; but it is a challenging problem to write an algorithm for that. Over the past few years DAPLAB has been been developing a tune recognition system. It is a computer program that allows a user to upload a query in the form of humming (eg. la laa la). The audio waveform is then analysed with signal processing techniques to extract the melody. Using pattern matching the extracted melody is compared with the melody patterns stored in the

database. The algorithm ignores absolute pitch and tempo of the humming and makes its decision based on the relationship between successive notes.

Speech Recognition

How does a child learn its mother-tongue? She uses various audio/ visual / haptic cues and gets feedback on imitation. In order to make our interaction with machines more natural, they need to understand (and speak!) our language. This idea has given us many problems to solve, such as speech recognition, speaker recognition, identifying the intent (or emotion) of the speaker, speech synthesis and so on. An application includes automatic transcription of speech in the audio/ video files. Let's say you want to look up the definition of 'Nyquist rate' in the CDEEP EE603 video lectures. For this, you would first need a speech recognition system to transcribe the database. Then the algorithm will intelligently search the transcription, based on the context.

Currently we are part of an ambitious project that attempts to automate agricultural information access over wireless channel using speech recognition. Taking the research to real-world applications such as this involves dealing with variable dialects, accents, user quirks and noise, both acoustic and that from communication.

Overall, our research is focused on extracting meaningful information from audio signals of any origin and developing user friendly technologies. If you are are interested in the exciting field of sound processing, just knock the door of DAPLAB. ■

Gigabit Networking Laboratory

By Ashwin Gumaste
Department of Computer Science Engineering, IIT Bombay

The Gigabit Networking Laboratory (GNL) was set up in June 2006 in the Department of Computer Science and Engineering with a mission to conduct research in the telecommunication domain especially focusing on high-speed networks. The laboratory has grown significantly over the past 5 years by showcasing sustainable research, product development and intellectual property creation. From the research perspective, the laboratory has produced close to 130 publications in top tier journals and conferences. In addition, the principle investigator has been involved in 20 granted US patents with another 20 pending with the USPTO.

At this time, the laboratory runs 11 externally funded projects – exemplifying strong interaction with industry and the government. Projects are broadly classified into three domains – metro core, enterprise networking and access networks – covering the three major network focuses. Metro core research focuses on networks that form the crux of service providers, connecting points-of-presence (POPs) to each other via fiber optic lines. These POPs have equipment that caters to multiple traffic functions such as transport, switching, routing and aggregation. By keeping most of the functionality in the lower layers of the 7-layer Internet hierarchy – and in particular at the optical and data-link layer, we are able to achieve low cost as well as low energy requirements – intrinsic to both CAPEX and OPEX reduction. Principle work in the area of metro core involves network architecture, equipment architecture, protocol design and rapid prototyping. Much of the cellular traffic explosion that we see resides in the fiber and is passed to the central office from where it is connected to other networks or the Internet. The inter-cell-phone tower connectivity is through fiber optic links using optical transport equipment that is also designed by the lab. The lab engages enterprises, from the connectivity as well as computing perspective – designing next generation switches and routers for very high-speed information transport. A critical aspect of end-user research is the work done in collaboration with a European vendor on access networks. Specific high-speed fiber to the home and curb technologies are being investigated under this project.

> **"The OEthernet concept transforms a conventional network to one that supports the entire gamut of communication functions "**

Technology and Research

Specific projects include the creation of a new communication technology paradigm that achieves very low-latency data transport across the Internet using a fundamental technology paradigm called Omnipresent Ethernet (OEthernet). The OEthernet concept transforms a conventional network to one that supports the entire gamut of communication functions (i.e. transport, aggregation, routing, transmission, flow and error control and congestion) to the data-link layer, thus making the best use of the fiber and resulting in very low latency and low cost. This technology has found application