ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Education in Acoustics
Session 3aED: Learning by Listening: Education in Acoustics Based on Listening
(Lecture/Poster Session)

3aED1. Technical ear training: Tools and practical methods

Jason Corey*

*Corresponding author’s address: School of Music, Theatre & Dance, University of Michigan, 1100 Baits Drive, Ann Arbor, Michigan 48109-2085, coreyja@umich.edu

Broadly defined, technical ear training seeks to make associations between aural impressions of sound quality and quantifiable characteristics of audio signal processing and acoustical measurements. Technical ear training typically focuses on attributes of sound such as spectral balance (e.g., filtering and parametric equalization); dynamic range of musical signals (including artifacts produced by dynamics processing); reverberation, delay, and early reflections (from real acoustic spaces or generated artificially); and spatial extent (width and depth). These elements of recorded sound can be broken down into graduated levels of audibility for the development of critical listening skills. With repeated and regular practice of carefully chosen exercises, listeners can gain increased sensitivity to subtle details of sound, as well as efficiency and accuracy in identifying specific parameters of signal processing by ear. With applications primarily in sound recording and production, technical ear training is also highly relevant to the evaluation of acoustical spaces as a complement to objective measurements. This presentation will review a selection of software modules developed by the author to teach critical listening skills to undergraduate students. The author will also discuss some practical methods and exercises used for teaching technical ear training and critical listening.

Published by the Acoustical Society of America through the American Institute of Physics
DEFINITION AND GOALS OF TECHNICAL EAR TRAINING

Technical ear training focuses on relationships between perceptual attributes and physical properties of sound, specifically in the context of applications such as sound recording, loudspeaker development, electronic music composition, and audio signal processing. Letowski (1985) and Miskiewicz (1992) were among the first to write about formal methods of technical or timbral ear training, and they compared the process to musical ear training or solfège. Quesnel (2001) and Olive (2001) have also made significant contributions to technical ear training, having written about methods for developing memory for timbre and increasing a listener’s sensitivity to timbre changes. This paper will describe briefly some techniques and methods for teaching technical ear training and critical listening skills in an undergraduate class for students in sound recording and music technology.

We might describe sound recording as a process through which a sound engineer makes decisions about sound are aesthetic in nature and based primarily on subjective impressions. Technical ear training seeks to introduce a level of objectivity across stages of production work through the development of critical listening skills within the context of controllable parameters found on audio signal processors. Technical ear training aims to help students and practitioners develop:

- associations between perceived attributes of audio and corresponding physical characteristics, especially with signal processing parameters commonly used in sound recording
- memory for perceptual attributes of signal processing parameters

Training is mostly focused on the timbre of sounds, primarily as controlled by audio equalizers, but it also covers time-based and amplitude-based processing.

Audio signal processing parameter settings usually correspond to physical properties of sound, and as such, do not have obvious links with perceptual attributes of audio. The perceived changes in sound qualities produced by a device can usually be described with verbal adjectives, but such adjectives are not as precise as the parameter settings found on the processors themselves. A listener might describe an equalizer’s effect as making a recording sound “hollow,” but in describing the sound it is more accurate to be able to name the precise center frequency that has been boosted and the amount that it has been boosted.

Audio signal processing types can be grouped according to common sound recording practices, in terms of general categories, and are listed here with their respective device names:

- frequency spectrum processing – filters, parametric equalizers
- amplitude-based processing – dynamic range compressors, expanders, gates, limiters
- time-based processing – reverberation, delay, chorus, flanging, phasing

PRACTICE WITH PARAMETRIC EQUALIZATION

In sound recording, a parametric equalizer affords control over the spectral balance of an audio signal, but it can also affect track or mix balances within a recording. Parametric equalizers usually offer independent control over center frequency, gain (symmetrical boost or cut), and bandwidth. Multi-microphone recordings are typically balanced using individual level control for each microphone signal but additional control of balance can be achieved through selective cuts or boosts to frequency bands. Frequency resonances in one musical instrument recording may partially mask other microphone signals in the same recording when the two are mixed together, making the challenge of an optimal balance between signals difficult. When resonances are reduced with equalization, masking is less problematic and balances can be achieved more easily.

To the novice recording engineer, there may be no clear connection between a 12 dB boost at 125 Hz, for example, and the associated perceptual attribute. It may be easy to distinguish a boost at 125 Hz from a boost at 1000 Hz, but consistently matching the perceived sound of an equalized signal with specific physical characteristics (e.g., produced by equalizer settings) is not so obvious without practice. Put another way, there is no innate reference for the perceptual attribute of a boost at 125 Hz, 1000 Hz, or any other frequency. Technical ear training helps students develop mental maps that link perceptual attributes of audio with the respective physical characteristics. By mapping perceptual attributes with objective parameters of sound, students can become more efficient and accurate in the use of signal processing devices. Systematic training can introduce meaningful relationships between perceptual attributes and physical characteristics of audio signals. With knowledge of alterations of specific frequencies by ear, one can use an equalizer with less guesswork and a more systematic approach.
In the technical ear training class that I teach at the University of Michigan, students are encouraged to practice regularly with software modules (a CD-ROM containing software modules for Mac OS X and Windows is included with Corey (2010)). The software practice modules randomly select and apply signal processing to a chosen sound source according to a parameter limitation framework specified by the student. The student’s goal, when working with the software modules, is to identify correctly the processing that was randomly chosen by the computer, within the parameter options specified by the student. Practicing with the equalizer module, a student can choose among the following options:

- frequency resolution (ISO octave or third octave frequencies)
- number of center frequencies affected per question
- gain or gain combination (such as +12 dB only; a combination of +6 or -6 dB)
- sound source to be processed
- specific range of frequencies
- bandwidth or Q (Fc/bandwidth)

It is usually recommended to work within the nine possible ISO octave frequencies (63, 125, 250, 500, 1000, 2000, 4000, 8000, and 16000 Hz) at first, and eventually progress to the 25 third-octave frequencies. Further limiting the range of possible octave frequencies can also be helpful in the beginning stages, so that students can familiarize themselves thoroughly with the sound of each octave frequency. For instance, a student can begin with mid-range frequencies (e.g., 500 Hz to 2000 Hz) the first week, move to low-range frequencies (e.g., 63 Hz to 500 Hz) the following week, and then to upper range frequencies (e.g., 2000 Hz to 16000 Hz) after that.

Students are encouraged to start working initially with pink noise as the sound source. Due to its statistically equal energy per octave, pink noise typically represents frequency content more evenly than most music recordings. Thus when a boost at a given frequency is made in pink noise, we can be assured that a boost at another frequency will produce an effect of comparable strength. Music recordings usually have resonances (more energy) at some frequencies and less energy at other frequencies, thus a boost at an existing prominent resonance could create a stronger effect than a boost at a frequency that has less energy.

Despite the possible inequality of representation across the audible spectrum, music and speech recordings are important sound sources for practice with equalization. The main use of equalization in the practice of sound recording is, of course, with music and speech recordings, and there are many commercial recordings that can be used for practice. No two recordings are identical and each one can offer a chance to learn more deeply about the effects of boosts or cuts at each of the standard frequencies. Research conducted by Olive (1994) indicates that program material can affect listeners’ response accuracy in matching equalization settings. He found that listeners had more correct responses with some recordings than with others.

**PRACTICAL ISSUES IN THE CLASSROOM**

Ideally, technical ear training class meetings are held in an audio control room or similar space with reverberation times around 0.5 s to 1 s across the frequency spectrum. Careful placement of high frequency and low frequency sound absorbers and diffusors in the listening space can help reduce strong reflections and associated comb filtering, and minimize prominent standing waves. Full range, high quality loudspeakers with relatively flat frequency response and low distortion characteristics are crucial for critical listening.

Class meetings focus on listening practice, discussion, paper readings, and quizzes. Listening practice in class generally involves choosing a framework for possible questions (e.g., frequency resolution, gain options, and sound sources for equalization practice) and then listening as a class to practice questions. Switching between flat (unaffected) sound and equalized sound for about 30 seconds (or longer if needed), I invite students to identify the frequency that they think has been altered. After identifying the frequency, they are asked to choose the gain setting, cut or addition. I request comments and discussion about the sound qualities that they hear. We talk about the characteristics of the musical instruments and voices in the recording and the processing that was applied. In my classroom experience, inviting students to describe what they hear not only helps solidify their memory for a particular parameter setting, but it also helps other students identify characteristics that may not have been as salient prior to discussion. For example, a boost at 8 kHz may not only bring out the “sizzle” of cymbals, but may also affect the attack of a bass drum or the sibilance of a vocal; a 16 kHz boost on a Baroque music recording may accentuate the string-plucking sound of a harpsichord more than it affects the sounds of other instruments in the recording.

To help develop memories for the aural impressions of specific frequency alterations, students are encouraged to make notes about the attributes of frequency alterations for each piece of music that we review in class and that they listen to on their own. One of the main learning points is making clear, unambiguous distinctions between memories...
for adjacent octave or third-octave frequencies. For example, the description of the sound of a 4 kHz boost should be unique enough so that it is not confused with the description of a boost at 2 kHz or at 8 kHz. As mentioned above, descriptions are not as specific as stating the specific frequency in Hertz and the amount of boost in decibels. Still, writing down impressions about sound quality can serve as a step towards more precise parameter identification and serve as method to explore the aural impressions of a boost or cut at a single center frequency. Notes about aural impressions of each frequency boost or cut are encouraged but not required for students and descriptions are likely somewhat unique to each student.

**Listening Quizzes**

I assign weekly in-class quizzes that focus on a specific range of parameters of equalization and are limited to 10 questions. At the beginning of the semester, I outline the parameters of each weekly quiz in the course syllabus, stating the specific frequency resolution (octave or third-octave frequencies), amount of boost or cut, frequency range (e.g., 125 Hz to 2000 Hz), and sound recordings or pink noise to be used in each quiz. During an in-class quiz, I play back one of the possible recordings or pink noise, switching between the equalized and effect-bypassed versions. Students write down the frequency that they think has been affected, and they are free to use notes that they may have about each frequency during the quizzes.

During the first class of the semester I give a preliminary equalization listening quiz, which is similar to the regular weekly quizzes, except that does not count towards students’ grades. For each question I choose a third-octave frequency to boost, and apply it to a recording, switching between the original signal and the equalized version. After we finish the preliminary quiz, I list the answers so that students can check their work but I do not collect their answers. This preliminary quiz is for their information only and helps serve as a baseline for their subsequent study, and they can compare it to the final listening exam in terms of accuracy and confidence. By the end of the semester, they will likely have developed the ability to identify the nine octave frequencies with a high degree of accuracy and speed, and they will have some skill in correctly identifying third-octave frequencies. Although students may have the incorrect feeling that they have made little progress by the end of the semester, they can refer back to their preliminary test from the first day before beginning the intensive technical ear training. Because progress is relatively slow through the semester and perhaps also because with increasing awareness students realize more of their shortcomings, they may not be fully aware of the substantial progress they have made by the end of the semester.

As the class progresses, students become more accurate, more confident, and faster in their response times. Working with octave frequencies and 12 dB boosts on pink noise or a known recording, I can expect that students will typically be able to correctly identify alterations by ear within about 5 seconds. I have given this type of test several times during the second half of the course and students consistently make correct judgments most of the time.

**Absolute Identification Versus Matching**

One challenge of in-class quizzes is that students must use absolute identification. That is, they cannot go through the process of making an attempt at a response, listening to their attempt as it compares to the question, and then submitting their response. With quizzes covering third-octave frequencies, absolute identification of correct center frequency is more difficult than with octave frequencies. In “matching” quiz mode, students can audition their response to a question before submitting it. Switching between the student’s attempt and the question, although only possible when a student is taking a practice quiz alone, can present a slightly different learning method than absolute identification. The matching method of working is especially useful for third-octave frequencies with smaller amounts of boost or cut (e.g., 6 dB). When working with octave frequencies and 12 dB boosts, the difference in sound quality of each frequency is significant enough that students can learn to correctly identify them in absolute identification. With third-octave frequencies, especially in the low frequency range (e.g., 63, 80, 100, and 125 Hz) with small amounts of boost, it can be difficult to hear a difference between a boost at one frequency and a boost at an adjacent third-octave frequency, even when switching directly from one to the other. The “matching” learning paradigm allows a student to determine if there is a difference between the question and their response. In certain cases the difference is minute and the challenge of hearing the difference becomes the main goal of practice and testing. Memory for each octave frequency helps get close to the third-octave frequency but switching between the question and the student’s response turns the quiz into a discrimination task instead of solely a memory task.
Compression and Reverberation

Although equalization is the primary focus of the course, we do spend class time listening to dynamic range compression artifacts and sound quality. There are no standard parameter settings on dynamics processors like there are on equalizers so parameter choices are somewhat arbitrary. I have reviewed the typical attack and release time ranges and stepped settings found on commercially available devices, and it appears that most ranges follow a roughly logarithmic scaling, ranging from small steps closest to zero milliseconds to larger steps at longer attack and release times. One challenge particular to dynamics processors is that parameter settings can be dependent on one another. For instance the threshold setting on a compressor will likely determine how audible the attack time is. A lower threshold is likely to make the attack time more prominent than a higher threshold because a lower threshold results in more gain reduction. Class listening sessions for dynamic range compression focus mainly on artifacts produced by various attack and release times. We begin by narrowing down the possible times to only three options for each: 2 ms (fast), 25 ms (medium), and 100 ms (slow) attack times; 50 ms (fast), 320 ms (medium), and 1000 ms (slow) release times. Moving up a level of difficulty gives five possible timing options: 2 ms, 10 ms, 25 ms, 55 ms, and 100 ms attack times; 50 ms, 150 ms, 320 ms, 600 ms, and 1000 ms release times.

Artificial reverberation and delay are also critical to sound recording and some class listening is devoted to comparisons of reverberation decay times and delay times. Usually three to five possible settings for each of these two parameters are compared. For example the starting three decay times are: 0.5 s, 1.5 s, and 2.5 s, and the five initial predelay times are: 0 ms, 40 ms, 80 ms, 120 ms, 160 ms, and 200 ms. Impulsive or transient (and anechoic) recorded material is an ideal starting point for reverberation and delay identification. Sustained or relatively steady state musical material is much more difficult for reverberation decay time identification because the reverberation is masked much more than with transient program material.

CONCLUSIONS

This paper gives a brief overview of some of the techniques and tools related to teaching technical ear training to an undergraduate class. Technical ear training can help increase sensitivity to changes in timbre, develop memory for timbre and its associated physical properties, and increase the speed in which one can make correct identifications of timbral alterations. Recording engineers, loudspeaker developers, electronic music composers, and audio signal processing engineers can benefit from this type of study and practice. Regular practice, classroom listening and discussion, and quizzes are necessary for progression through stages of difficulty and for mastery of the material. Although primarily focused on parametric equalization, critical listening can also encompass other common signal processing devices such as reverberation, delay, and dynamics.

REFERENCES