Using Web Audio for Ear Training of Sound Engineers

Thilo Schaller\textsuperscript{a}, Ian G. Burleigh\textsuperscript{b}

\textsuperscript{a}Digital Audio Arts, University of Lethbridge, Canada
\textsuperscript{b}Quinta Audio, Czech Republic

Web technologies enable the development of interactive multimedia content that can be embedded into web pages or run as web browser applications. The combination of Web Audio and Scalable Vector Graphics is a practical solution for making intuitive, interactive applets that are effective tools for teaching music technology. Interactive applets with visual and auditory feedback help to illustrate various theoretical concepts and their practical outcomes. Several such applets are used for ear training of sound engineering students in the Digital Audio Arts program at the University of Lethbridge. The Noise Spiral applet has been designed to teach frequency band recognition; we describe its use, discuss other similar applets, and present an outlook on future possible integration of Web Audio into technical ear training. The source code is freely available; we hope that it could be a starting point for the development of additional ear training tools that would improve the delivery of this essential part in education of future audio professionals.

Introduction
In the last five years we have been teaching students of Digital Audio Arts (DAA) at the University of Lethbridge.\textsuperscript{1} DAA is a Bachelor of Music program that trains students in audio engineering and production, computer-assisted sound and music, electroacoustic music performance and composition, and other related areas of music technology. All music technology training is done within a traditional music curriculum of music history, theory, and performance. Since the implementation of the DAA curriculum in 2009, it has been altered a number of times and many of the changes have been made in order to meet today’s needs in au dio education. One of the main challenges is the development of the necessary listening skills that students need in order to make artistic decisions and use technology in the most efficient and creative way.

Digital Audio Arts (DAA)
The DAA curriculum is designed for students of music and general fine arts, rather than as a curriculum in engineering, and therefore it attracts and supports students that have talent and experience in arts rather than in sciences. However, for a successful outcome, the students still have to understand concepts of geometry, physics, acoustics and psychoacoustics, and audio signal processing. We have observed that students who enter the DAA program are often not sufficiently prepared by secondary education in terms of abstract thinking and natural science foundations. Therefore, mathematical and scientific concepts are best taught by intuition, demonstration, and experimentation with the help of interactive tools.

In addition to theoretical knowledge and vocational skills, DAA students have to acquire various kinds of listening skills, e.g., the ability to critically assess sonic results of audio signal processing or to recognize flaws in recorded or synthesized sound. Training of these skills is most effective through regular practice and also by demonstration and experimentation with auditory and visual feedback.

Web Technologies Today
The concept of “the web” (World Wide Web) of pages that are served to client devices (desktop computers or mobile laptops, tablets and smartphones) from a network of server computers exists since the 1990s. The original web was primarily a growing collection of hypertext documents with images, written in HTML and interconnected by hyperlinks. The documents were static; each time a change was needed, the server had to compute and serve a new page content. Dynamic multimedia client-side content that would not require these “round-trips” to the server was to be first provided by Java applets: small programs written in the Java language and embedded into web pages as self-contained objects. The intention was that a Java

\textsuperscript{1} Bachelor of Music – Digital Audio Arts, http://digitalaudioarts.ca
program would be written “once” and run “everywhere” on a variety of platforms — different operating systems and web browsers. However, during the infamous “browser wars” among Netscape, Microsoft, and Sun, the compatibility of competing implementations of Java was intentionally broken and its promise to become the main language for dynamic web content fell flat. The other technology that has been widely used for creating dynamic client-side web content, Adobe Flash, is a proprietary software, which is in conflict with the current “share and share alike” culture of today’s “open” web.

Thanks to the open web standards of the World Wide Web Consortium\(^2\), the last several years were a period of consolidation of web technologies. Clearly defined interfaces, the fact that major browser vendors have opened their source code\(^3\) and do strive to adhere and contribute to the standards rather than to break them, and the general culture of open software development created an unprecedented creative environment: an “ecosystem” where teams at major companies collaborate with individuals on peer-reviewed and truly cross-platform, “write once, run everywhere” software.

At the core of current web technologies is HTML5 — the latest specification of HTML designed for web pages and applications with rich multimedia — and a number of associated application programming interfaces (APIs) that are built into all major browsers and that support various services that were formerly available only in desktop environments. The web standards and technologies are evolving rapidly, but apparently in a collaborative and coordinated way. The few differences between how well and how timely the standards are being implemented in web browsers can be resolved by techniques known as polyfills and shims. The web browser has now become the ultimate cross-platform application environment, and the client-side programming language JavaScript the “one language to rule them all”.

JavaScript is evolving into a mature, modern all-purpose dynamic programming language. Although programs written in dynamic languages do necessarily run slower than programs written in compiled languages, modern sophisticated just-in-time compilation techniques greatly improve their performance. There are also several higher-level languages that are “transpiled” (translated/compiled) to JavaScript and that support programming techniques for more efficient and less error-prone development of programs. Namely, the Dart and Typescript languages by Google and Microsoft\(^4\) seem to seek to strengthen the role of JavaScript.

Since most of today’s computing devices frequently connect to the net, they do not only access the latest web content, but also run the latest version of software. Applications are seamlessly updated whenever there are small improvements and application developers and content creators swiftly react to comments and requests of users; application development and content creation on the web is now a community affair.

**Web Audio**

The Web Audio\(^5\) is a JavaScript API for sound synthesis and processing in the web browser. Prior to the introduction of Web Audio, sound in browsers was restricted to playback of sound files and streaming of audio from a server. Web Audio now allows client-side sound synthesis and processing; the API provides a number of common types of audio signal processing modules that can be connected to form a signal-flow graph: gain, mixer, delay, buffer, panner, filter, convolver, compressor, waveshaper, oscillator, etc. User-defined audio signal processing is also possible. The usual modular audio synthesis techniques, popularized in such tools as Max/MSP or Csound, apply similarly in the Web Audio API. Most of the signal-processing code of Web Audio is compiled into the machine code of a browser and therefore runs fast and the slower JavaScript code is used mainly for building the signal-flow graph. The stated purpose of Web Audio is to provide sound for interactive web applications and specifically for in-browser games; as such it has also been designed to support spatial audio.

**Scalable Vector Graphics**

Scalable Vector Graphics\(^6\) (SVG) is a markup language for describing two-dimensional graphics composed from simple elements such as lines, rectangles and polygons, circles and ellipses, various curves, and text. The SVG description of a graphic is embedded as an object into a web page. The elements of an SVG object can be created and modified by a JavaScript program and also when users’ actions activate snippets of JavaScript attached to user interface elements. In turn, the Web Audio signal-flow graph can be altered or the

---

\(^1\) https://www.dartlang.org/; http://www.typescriptlang.org/

\(^2\) W3C, http://www.w3.org/

\(^3\) Mozilla entirely; Google, Apple and Opera partially; Microsoft apparently remains a closed shop.

\(^4\) http://www.w3.org/TR/webaudio/

\(^5\) http://www.w3.org/TR/SVG/
parameters of the audio modules modified, and thus provide auditory feedback to changing graphics.

**Quint.js**
We have written quint.js, a JavaScript library that streamlines and simplifies the creation of interactive SVG/Web Audio applets\(^7\) for audio-visual demonstration and experimentation. More detailed technical description of the library and ideas that it implements can be found in our previous paper [1]. In this text we shall focus mainly on a discussion of several concrete applets and their role in teaching of music technology students.

**The Sensations of Tone**
It seems to be fitting to mention the original motivation for our work: Hermann Helmholtz’s seminal work *On the sensations of tone [Die Lehre von den Tonempfindungen]* [5]. Alexander J. Ellis wrote in his translator’s notice: “Professor Helmholtz’s book … [is] a work which all candidates for musical degrees are expected to study.” Edgard Varèse and Harry Partch, among many others, also credited Helmholtz for fundamentally influencing their appreciation of musical sound. The book is often referred to but very few students of sound arts do read it. It is written in an old-fashioned style and, given when Helmholtz’s work was done, large sections of the text describe the various mechanical devices, such as sirens and resonators, that he had constructed and used for his psychoacoustic experiments. The auditory sensations are also described in many words, because they cannot be otherwise demonstrated in a printed text; to our students the whole work seems, not surprisingly, irrelevant. Our idea and intention were that we could eventually make a “student’s digest”, a web version of Helmholtz’s text with the psychoacoustic experiments recreated in the browser by embedded SVG/Web Audio applets, directly within the text.

**The Pitch Spiral**
Helmholtz’s theory is built on three fundamental observations: that (a) complex tones are formed from series of simple tones (harmonics), (b) the interference of two simple tones is perceived either as auditory beats or a difference tone, but around about 30Hz leads to an unpleasant sensation of a “sensory roughness”, and that (c) tuning a group of complex tones to a pleasant sonority minimizes the aggregate sensory roughness. The system of musical intervals and chords in most music is then a consequence of these phenomena.

The Pitch Spiral applet (Fig. 1) is an applet that has been originally designed to illustrate and confirm Helmholtz’s theory. The spiral curve represents a flattened helical model of musical pitch. Each turn of the spiral corresponds to one octave\(^8\). The filled (colored) discs represent the harmonics of several complex tones and the hollow discs their respective difference tones. Visual coincidence of the discs on the spiral is directly related to the auditory sensation and nicely illustrates many aspects of various tuning systems (Pythagorean, just, tempered), of various divisions of the octave, and of chords and their inversions.

---

\(^7\) The source code is available at [https://github.com/quinta-audio/quint.js](https://github.com/quinta-audio/quint.js) and the applets at [http://quinta.audio](http://quinta.audio).

\(^8\) Although the term ‘applet’ is traditionally intended for Java applets, we consider an interactive visual SVG object that is combined with Web Audio, or any other such small program embedded into a web page to be an applet.

---

\(^9\) The octave, frequency ratio 2:1, is the “equivalence” or “periodicity” relation among musical intervals.
(classical music). Technical listening is the skill to identify “features, characteristics, and sonic artifacts that are produced by various types of signal processing” [2], and technical training is “a type of perceptual learning focused on timbral, dynamic, and spatial attributes of sound as they relate to audio recording and production” [3]. Aesthetic listening is listening in a broader context and in a more artistic way than technical listening. While listening aesthetically, one takes into consideration whether certain attributes (spatial impression, timbre, stereo impression, etc.) work in favor of the particular production.

All these kinds of listening skills have to be developed to complement theoretical knowledge and practical skills surrounding work in a professional recording environment or electronic music studio. The development of all three kinds of listening skills during the undergraduate study is essential for various reasons: Firstly, in the past, listening skills were oftentimes developed during apprenticeships or in entry-level jobs. However, in order to get placement as an intern or to find employment in today’s competitive audio industry and for the creation of a professional portfolio, listening skills have to be developed prior to graduation and entering the audio industry. Secondly, many graduates of the DAA program start freelancing shortly after obtaining their degree and work on their own projects in their own (home) studios without having the chance to observe and learn from industry specialists. Therefore, without a strong foundation in ear training during their formal education, the technical and artistic potential of their music productions and electronic compositions is not realized to its full extent.

Ear Training in DAA

For DAA students, musical ear training is offered during Musicianship courses that have been established prior to the creation of the DAA academic major. Students practice recognition and singing of intervals, scales, and chords, melodic, rhythmic, and harmonic dictation, and sight singing. Specialized courses such as Audio Production Techniques, Advanced Audio Production Techniques, and Audio Spatialization Techniques serve as platform for regular critical listening sessions and discussions which provide students with the chance to develop their artistic listening skills and help forming their sonic vision for their own productions. The method for artistic listening that is currently used is based on the Quality Solfeggio program developed by Cees van der Graft for the degree program “Art of Sound” (formerly “muziekregistratie”) at the Royal Conservatoire in The Hague, Netherlands.

The development of technical listening skills requires regular practice and exercises have to gradually increase in difficulty. Due to the curriculum structure with many electives and the broad range of topics the DAA academic major tries to cover, the integration and development of technical ear training that progresses throughout the four-year study has been difficult. Short semesters (13 weeks), student workload, and the fact that students choose individual courses rather than a coherent course sequence make it difficult to offer sequential technical ear training within the existing course structure. Ideally, students are able to practice outside of class time and as regularly as possible. Therefore, the development of a technical training method that can be extended over several semesters regardless of chosen courses is essential; web-based tools are useful for its creation. Browser-based applets can be accessed on any connected device at any time and enable students to fit in short sessions more often than when they have to visit a dedicated lab or device for their ear training. Online course material eliminates the limitation to a particular classroom, thus making the method available for any course that has access to computers or mobile devices. Furthermore, the technical ear training platform is not affected by times when instruction labs are not accessible. Class time that has to be spent on technical ear training can be kept to a minimum as students are enabled to schedule their own training sessions.

FREQUENCY BAND RECOGNITION

Equalizers have changed from tools for correcting non-linear behavior of recording equipment to creative tools that are used far beyond a mere correction of the recorded material. After an aural assessment of the sonic material, decisions on equalizer settings are made which vary depending on the production context. In order to find the desired frequency bands that need to be adjusted, students are often taught to boost an area of the spectrum with a peaking filter and “swipe it through the spectrum” by changing the center frequency of the filter. While this approach may lead to quick and often good results, it is possible to eliminate this step by ear training and experience [2] in order to accelerate the production process. Some argue that the artistic/sonic vision might change while swiping as one listens to several variations of the sound before finding the correct frequency band [10]. In order to preserve the artistic vision, one has to be able to identify the frequency band that needs adjustment prior to changing parameters of the signal processor. To a certain extent, the development of this skill can be accelerated by using tools for technical ear training.

The Noise Spiral

The Noise Spiral (Fig. 2) — the first applet dedicated to technical ear training in DAA courses — has been developed to facilitate the learning process of frequency band recognition. Similarly to other methods that deal
with this aspect of technical ear training, such as Moulton Laboratories’ Golden Ears or Jason Corey’s Audio Production and Critical Listening: Technical Ear Training, students are asked to identify frequency boosts or cuts within pink noise or musical and other sound material when using the Noise Spiral. Noise is used because any changes to the spectrum are consistent and immediately audible at any given moment [3, 6, 8]. Therefore, students in the first semester of studies are advised to begin their technical ear training with filtered pink noise.

The applet uses a “peaking” Web Audio biquad filter with a user-defined Q to apply a gain change to a certain frequency band. The student is asked to identify the center frequency of the altered band. The user can select the division of the octave (½, ⅓, etc.) which increases the number of possible center frequencies while decreasing the audible difference between adjacent steps. The applet also provides the option to import audio files so that the user can practice with various audio material ranging from recordings of individual instruments to full mixes of all music genres.

![Image](image.png)

Fig. 2: The Noise Spiral

Formant Overlays
The Noise Spiral uses the same graphical representation as the Pitch Spiral. Numerical values of frequencies and formant areas of vowels [7] can be optionally displayed along the spiral. For beginning students, the vowel overlays are a great help. It seems that even people who do not have previous experience working in music or music technology disciplines can still refer to speech quite easily as they are familiar with the timbre of spoken vowels. While the noise spiral currently displays the formant regions for German vowels [7], we plan to include options for other languages such as English or Finnish, male and female voice, etc. [9] in the near future.

The Noise Spiral – while still in development – has been tested in several courses of the DAA curriculum (Acoustics, Audio Production Techniques, Advanced Audio Production Techniques) throughout the academic year 2014–15. The beginning exercises use filtered pink noise, starting with boosts of 12dB and gradually lowering the change to 3dB. The students train to identify the center frequency of octave bands before dividing the octave into more steps. Especially in this context, the formant overlay seems to be a helpful addition. In subsequent exercises, differing audio material is introduced, ranging from fully-produced recordings to recordings of individual instruments. This way, students learn to identify characteristics of frequency bands of ensembles and instruments while aurally verifying tonal characteristics that are presented in their textbook Acoustics and the Performance of Music by Jürgen Meyer. Throughout all courses, frequent online listening quizzes are scheduled to assess students’ progress. While the past academic year served as a test run, closer and more structured integration of the applet use into existing courses is planned for the upcoming years.

OTHER APPLETS
Several other applets have been designed to explain theoretical concepts, with integrated sound that sharpens students’ aural attention. Among those are applets that demonstrate the principles of Fourier analysis, additive audio synthesis, and modular audio processing.

Fig. 3 shows the intuition behind Fourier transform of musical signals that is used to construct sound spectrograms: a complex waveform is multiplied by a sinusoidal probe wave with varied frequency and phase. The positive sum of the wave product indicates a coincidence of the probe wave and some sinusoidal component of the complex wave. When the probe wave frequency and phase approach that of the component wave, one can also clearly notice auditory beats that slow down as the match gets closer.

![Image](image.png)

Fig. 3: Fourier Analysis Applet
Fig. 4: Additive Synthesis Applet

Fig. 4 shows an applet that demonstrates additive synthesis of a complex periodic wave from sinusoidal components. Students can modify the amplitude and phase of the components, see the changing shape of the complex wave, hear the audible result and observe, for example, that a mere change in the phase of the components does not change the perceived sound.

With SVG and Web Audio, one can also create a simple patch for modular sound synthesis. As an example, Fig. 5 shows a simple patch that performs ring modulation of two signals.

Fig. 5: Ring Modulation Patch

FUTURE WORK

The content of the web and the technologies that enable it are evolving rapidly; it seems that for the most part, they move toward consolidation and integration of previously fragmented APIs, improved performance, and advanced development techniques. We are currently redesigning our applets to take advantage of the improvements: we are cleaning up the code, rewriting it in the Google Dart language, and adapting the applets for use on mobile devices with smaller touch-sensitive screens. The new code should be more accessible to others and easier to extend.

The Noise Spiral shall be extended with an option that applies reverberation and only filters the reverberated signal. With this feature, students could practice the identification of timbral differences in reverberation and listen to their effects on the overall sound.

We are designing more applets that will employ other kinds of signal processing and stereo imaging for the delivery of technical ear training. The future version of applets, with their wider selection, could be integrated into interactive web-based lecture notes for courses in music technology. We hope that our work inspires an online community that further extends the possibilities of Web Audio ear training.

REFERENCES