Defining Spectral Surfaces
Robin Bargar†, Bryan Holloway‡, Xavier Rodet*, and Chris Ferrman‡

†National Center for Supercomputing Applications
Beckman Institute, 405 S. Mathews, Urbana, IL 61801, USA
email: rbargar@ncsa.uiuc.edu

*ircam, 31 rue Saint Merri, 75004, Paris, France
email: rod@ircam.fr

Introduction
We propose a sound may be described as a distribution of energy simultaneously in the time domain and the frequency domain, and we refer to a combined time-frequency description as a spectral surface. Using computer graphics a spectral surface may be displayed as a visual surface or "landscape" embedded in a three-dimensional graphical space. Time, frequency and amplitude provide three orthogonal axes. We suggest the spectral surface paradigm can be applied to the representation of the structure of sounds regardless of their source. In order to arrive at a spectral surface description, competing time and frequency representations of a signal need to be reconciled. Graphical display techniques provide a domain for resolving time-frequency surface representation. Visual displays of spectral surfaces may be able to provide a general interface for comparing sounds and combining acoustic characteristics of sounds by modifying sound samples or controlling sound synthesis. We discuss some characteristic principles of spectral surfaces applied to lemur signal analysis, and applied to CHANT sound synthesis.

Graphical display of 2D surfaces in 3D
Display of information in three dimensions is limited to visual devices with special stereo projection capabilities, where a single image is computed for each eye in correct perspective, and off-axis projection is computed for the position of the viewer's head with respect to the 3D location of the visual data. Immersive stereo display is a compelling argument for the effectiveness of graphically-rendered data, and we observe the spectral surface is optimized in a virtual environment workspace. Since these systems are not widely available we will discuss the projection of a 3-D dataset to a 2D representation with the ability to rotate the image around its own axis to obtain 3D visual cues.

Graphics are displayed in terms of points, lines and planes (vertices, edges and polygons). Converting data to graphics usually involves two "imputities:" (1) the loss of some original data points that are not converted into graphical elements, and (2) the generation of visual elements that cannot be traced to individual members of the original data set. These attributes of graphical display are either a liability or an asset depending upon how carefully they are handled in the interface design and the intended use of the display. Data visualizations are not WYSIWYG; they are creations of new information more than translations of existing information. The creative aspect of a display affords us a mechanism for efficient data reduction and data control.

Data resolution and visual resolution
A surface (manifold) is a data set and a visual surface is loosely but conveniently related. Generating visual surfaces involves connecting key graphical points with intermediate graphical points to create the appearance of continuous figures. A visual surface is a function of rules for visual connectivity between structural elements. Our concern is the relation between sound energy data points and the graphical structural elements. Screen resolution (less than 1000 x 1000 pixels on most workstations) does not permit very many sound samples to be displayed, nor every contour in the frequency or time domain. Radical data reduction is required. Visualization of spectral surfaces is a qualitative tool we wish to convert into a quantitative tool for direct manipulation and ultimately resynthesis of sounds. We propose to reverse-engineer the data reduction path and use the graphical display as a method for specifying and altering sounds in the frequency and time domain.

Lemur analyses of sounds
Lemur models sounds as sinusoids with time-varying amplitudes, frequencies, and phases (Fitz). During an analysis, time is divided up into frames - a section of time representing the spectrum of the sound at that point in time. A "peak" in a frame corresponds to a given sinusoid's current amplitude, frequency and
phase. Sinusoids in adjacent frames are connected, constrained by a maximum allowed frequency drift. The connection of these sinusoids between adjacent frames are called tracks (Figure 1).

Lemur is useful in analyzing a broad class of sounds. However, it does poorly with noisy or broadband signals. The result tends to be analyses with an excessive amount of tracks lasting only one or two frames. Harmonic sounds fair much better, due to the definitive presence of harmonics lasting over much of the course of the sound. Lemur is available in the public domain by anonymous ftp via medusa.cs.cmu.edu in the pub/lemur directory.

From Lemur analyses to spectral surfaces

Lemur has no notion of a surface. The tracks are energy peaks but there are no datum for the valleys between tracks (this is distinction to CHANT which defines sound energy as a surface-like structure). Tracks form linear contours in the time domain; from the track contours secondary contours can be triangulated across the frequency domain to form surfaces. Special connectivity rules are needed to accommodate the singularities that arise where tracks begin and end during a sound (at the births and deaths of partials). Track contours are directly related to Lemur track data; the bi-directional surfaces are defined using a more general polynomial description of control points and curvatures.

Conversion from a Lemur analysis to its corresponding spectral surface is done simply by "connecting the dots"; line segments are drawn between each peak and its closest neighbor (Figure 2). This technique is more successful with harmonic sounds or, in general, sounds with noticeable spectral patterns. Noisy signals, modeled by Lemur as a large number of tracks extending only a couple frames, will produce very jagged surfaces providing little insight into the nature of the sound. Harmonic sounds also translate well since they tend to contain energy throughout the entire "3-D space" (time, frequency, amplitude).

In the frequency domain, each track is by definition infinitely thin. Taking in our spectral surface paradigm, we wish to view a Lemur analysis from a macroscopic viewpoint. Therefore, the spectral surface is NOT an exact representation of where energy is present in the signal over the course of its lifetime. Instead, it shows the overall shape of a sound, without making assumptions about whether or not it is harmonic, although it could easily be one or the other. For this reason the notion of a fundamental is also discarded from the spectral surface paradigm. Most importantly, what you SEE with the spectral surface could map onto an infinite number of possible Lemur analyses fitting that same shape.

Manipulating graphical surfaces

Spectral surfaces can be used for creating sounds if we can alter acoustic information by altering its visual representation. The problem is twofold: to efficiently modify a complex graphical structure, and to convey those changes back into sound. In our ideal view the surface might be manipulated by a 3D "paintbrush" or sculpting tool to push around the energy associated with particular time-frequency locations. Implementation of this compelling paradigm is graphically reasonable but complicated by the nature of sound-energy. If a track "moves over" in the frequency domain, what does it leave behind?
To modify a graphical surface we modify the polynomial information stored at various control points. Alternatively, we can perform matrix operations that uniformly translate, rotate and scale segments of graphical structures. Manipulations may be applied regionally with tapering of the operation at the boundaries of the editing region, to smoothly connect with unedited regions. For example, rescaling amplitudes in a surface we use a "light beam" with an adjustable circumference and an inverse square fall-off of the edited values from the center of the region to its boundaries. The region may be shaped independently along the time and frequency axes.

Figure 2.

Spectral surface generated from the Lemur analysis file in Figure 1.

This liability can be converted into an asset by generalizing the notion of a spectral surface and arriving at conversion routines specific to various synthesis engines. In the Lemur case special rules are needed to generate new tracks when energy is placed in a region where no tracks are present. A generalized spectral surface described as a function could originate from another Lemur file or from another source (including CHANT). In this way any spectral surface can modify any existing Lemur analysis by evaluating the function for each peak. The power of a generalized description of a sound, for data storage, synthesis and composition applications is apparent.

Spectral surfaces and CHANT
CHANT implements formant synthesis, where each formant can be uniquely described largely by its center frequency, bandwidth, amplitude relative to other formants, and skirt-width (R001). There is an inverse relationship between a FOF attack in the time domain and the skirt width of the corresponding formant. This width is independent of the formant bandwidth, which is specified at -6 dB.

Our goal was to draw a curve in the amplitude-frequency domain that could be graphically edited, and used to create a sound in CHANT. Formants in CHANT describe a spectral envelope in the frequency domain at each time step, evolving as a spectral surface over time. An editing tool may be used to draw spectral envelopes as curves based upon based on five constraints that can be mapped onto the four formant parameters listed above. The five parameters for each formant corresponding to three points (the peak, and the left and right sides at -6 dB) and two derivatives (at the left and right 6 dB points). We can take these parameters and draw a local curve for that formant. The central point is adjustable in frequency and amplitude, controlling the center frequency and amplitude of the formant. The lower two points are always -6 dB below the peak and symmetrical in frequency on either side. Their location may be specified in the frequency domain, creating the formant bandwidth. The slope below the lower control point is related to the width of the skirt; in CHANT a skirt is determined by the time at which attenuation begins in the FOF corresponding to a formant. We control the skirt slopes from the derivatives at the lower control points.

Multiple formants may be placed on the same time slice. The resulting 2D spectral envelope may be used as a "key frame" to create a spectral surface: key spectral envelopes are placed at intervals along the time axis and the spectral surface is propagated through them in the time domain.
Editing in the frequency domain

The power of CHANT lies in its ability to convert time-domain FOEs into formants, without a fourier-type frequency domain computation. CHANT specifies formant attributes in the frequency domain but computes them in the time domain. For finer control over the frequency domain we prefer to specify sounds directly in this domain, and then specify how the frequency domain evolves over time. To achieve this we need a better understanding of the relation between frequency parameters of formants and the CHANT parameters.

The elusive surface

CHANT bridges the time-frequency discontinuity in signal sound representation. In doing so CHANT approaches the boundaries of the spectral surface paradigm, at least as far as graphical representations are concerned. In CHANT, surfaces are computed only in the time domain though they are specified in our editor as polynomial curves in the frequency domain. The skirt width which a polynomial claims to depict is really not quantifiable in the frequency domain. It is computed in the time-domain and is highly non-linear. The curve along the frequency axis is really a visualization technique connecting discrete control points. However, propagated along the time axis as a spectral surface the imaginary curve gains credibility as a portrait of temporary energy distribution. Editing "imaginary" curves in frequency is an imminently useful tool. By further exploring this paradigm for controlling CHANT, we hope to arrive at more robust relationship between visualizations of sound and numerical models of sound which we wish to control from the visual image.

From sounds to surfaces to CHANT

CHANT synthesis is currently performed using only five formants. For this reason, we wish to find the five most predominant "bumps" within our surface at a given spectral "snapshot" in time. To find the bumps present within a particular snapshot, we use common peak-finding algorithms (used also in Lernur when searching for spectral peaks within an analysis frame) that establish the parameters necessary to generate a single formant within CHANT. Parabolic interpolation is used to determine the absolute center frequency and relative amplitude of a given peak. Thresholds are used to find only the five most predominant peaks. Each bump can be analyzed and mapped onto the four CHANT parameters listed above. Surfaces obtained from other sound sources can be used to generate tones in CHANT. Surfaces can provide a helpful method for arriving at the complex CHANT trajectories needed for interesting sounds.

Directions for further research

The question arises, can we begin from a sound analyzed in a format such as Lernur and converted to a spectral surface, and go another step to convert the surface into a synthesis of the sound using a separate synthesis engine such as CHANT? The pursuit of a reproduction of the original sound makes this an epic quest, for we cannot say an arbitrary sound belongs to the class of sounds available from an arbitrary synthesis engine. Still, the development of a generalization "neutral" encoding such as a spectral surface is a necessary preface to the pursuit of a general sound synthesis notation or language; a notation which furthermore can account for the perceptual attributes of an acoustic event and allow us direct access to their modification. We can observe for example in the computer graphics industry the importance of a general language of perceptual and structural features, the commercial success of computer graphics and the renaissance of graphics hardware and software which followed, can be traced directly to the presence of rigorous general languages for creating and modifying visual objects in three dimensions.

Regardless of our level of interest in a similar economic and commercial boom in sound synthesis, the ability to depict, modify and extract acoustic structures and apply them to the creation of new sounds, crossing from one synthesis engine to another, is a sorely needed acoustic notation and composition "language".

References


376