From Sound Morphing to the Synthesis of Starlight.

Musical experiences with the Phase Vocoder over 25 years

Trevor Wishart

(Received January 11, 2013)

I first came across the phase vocoder in connection with a speculative project to morph the sounds of a speaking voice into sounds from the natural world. I had been working with such morphing procedures in the analogue studio (*Red Bird*, 1973-77) but had become frustrated by the rather hit and miss nature of possibilities in the analogue domain and was aware of the research going on with computer analysis and synthesis in the USA and more recently at IRCAM in Paris. So in 1979 I submitted a project to IRCAM for the morphing of vocal sounds and was subsequently (1981) invited to attend their induction course. Whilst there I discovered that Linear Predictive Coding processes were potential tools to morph vocal sounds into others (these had been used for imposing vocal contours onto noisy environmental sounds, like the sound of the sea). At the conclusion of the course I was invited to realise my project but soon afterwards the hardware system at IRCAM was completely replaced, and I had to wait until 1986 to begin the realisation of what would become *VOX 5*.

I was now advised to look into the newly arrived Phase vocoder program (Mark Dolson et al) but no one at IRCAM had any specialist knowledge about the program apart from the fact that it used the FFT to generate time-windowed spectral data of the input sound. So I broke into the binary output files and discovered tables of data that were clearly amplitude-frequency pairs. My discovery was met with some disbelief by more seasoned programmers as the FFT was known to generate amplitude and phase data (not frequency information) and no-one was aware of the phase vocoder mechanism for deriving the frequency data from the time-changes in the phase information.

Armed with this knowledge however, amplitude and frequency being musician-friendly notions, I was able to develop various programs for transforming the spectrum of my recorded sounds, for example time-stretching through interpolating new windows into the output spectrum. In particular, the amplitude-frequency spectral data of two similar but different sounds could be used to achieve smooth morphing between them.

With the phase vocoder transferred onto a desktop computer with the Composers Desktop Project (founded 1986) I was able to work at home (originally on the Atari ST, later on PC and Mac) and over the next few years developed many processes for modifying sounds in the spectral domain. The frequencies at the amplitude peaks in the spectrum of a singly pitched sound would correspond to the harmonics of the sound and, by comparing peak frequencies, the pitch of a source could be extracted and thence modified. The frequency envelope of the spectrum in a time-window, so long as the envelope window was sufficiently broad, would give a reasonable indication of formants and formant-shifts in the spectrum, so these could be tracked or modified or used to define spectral-domain formant-filters on an input spectrum. Less “rational” processes, but with musical potential, could be explored, such as spectral blurring (losing time detail by retaining only every Nth window and simply interpolating between their values to generate the missing discarded windows) which will make coherent speech sound drunken, or generate spectacular high-frequency sweeps over pitched rhythmic material (in *Globalalia*); spectral tracing (losing frequency detail by discarding the least loud channels in every window) which can be used to clarify a spectrum or generate strange internal partial-melodies from time-stretched inharmonic spectra; spectral tuning, and so on.

These phase vocoder data processes became an essential part of my compositional armoury and I was able to develop a musical aesthetic based on sound transformation, for me a logical extension into the spectral domain of the more traditional practices of transforming pitch and duration data in traditional notated music.

In my most recent project (2011-12) I have used the Phase vocoder data format as a synthesis tool. *Supernova* began as an open-ended research project at the University of Oxford, funded by the Leverhulme Trust, to explore ways in which scientific and musical researchers might cooperate. A joint research project on vocal morphing with the Applied Maths department did not progress as hoped but in the meantime I met or corresponded with many researchers in the physical sciences and began to accumulate various types of data. Bill Chaplin’s team at Birmingham had already mapped seismic waves in stars into musical tones. I discussed the propagation of electron-spin waves in artificial crystals (Alexy Karenowska), the readings from superconducting quantum interference devices (SQUIDS) searching for the magnetic moment of the proton (Hans Kraus), and other data sources. Towards the end of the research period I was put in contact with a Ph.D. researcher who had collected all the available time-series data on the Internet as part of a comparative data-analysis project, which he very kindly donated to me. I devised various interfaces allowing me to convert this data directly into waveforms or to use it as control data for sounds from some other source. But although there are many and diverse ways to sonify such data, many of which might be revealing from a data-display point of view, the problem was to find a musically satisfying reason for animating the data.

I eventually focused my attention on electro-magnetic spectra (rather than time-series data) – the light, infrared, ultraviolet, microwave and radio wave emissions of stars – which I will refer to as light spectra, for brevity. The idea would be to convert these into some kind of sonic-spectrum equivalent. The spectrum of individual stars also tells us something about their material constituents as each chemical element has its own characteristic light spectrum. It seemed it ought to be possible to map the light spectrum into the frequency spectrum of sound, using a suitable scaling of the data.

The first problem with this idea is that the scaled-mapping is essentially arbitrary – there is no particular reason to use one mapping rather than another. As a result, the sonic outcome would be entirely dependent on the mapping chosen. However, if we choose spectral data that varies through time, the *change* in the data will be reflected in the change in the sonic map, whatever mapping we use. As music is a quintessentially time-based art, this seemed the most appropriate route to take. And the most dramatic changing spectra in the heavens are those of supernova explosions.

Pedro Ferreira at the Astrophysics Department had introduced me to various astrophysics researchers, and in particular to Mark Sullivan who kindly provided me with spectral data from both Type Ia and type IIp supernova explosions. These represented readings of the light output on successive days, with explosions lasting 68 and 212 days respectively (!). Not only was the data technically useful. Supernovae are the most spectacular and dramatic events to occur anywhere our universe, and they are responsible for the creation of many of the elements above Iron in the periodic table, elements without which human life could not exist. To paraphrase Carl Sagan, we are all the children of supernovae. So this also provided a strong poetic motive to pursue this particular material.

The supernova data consists of (daily readings of) amplitude-frequency pairs over the frequency range of the electro-magnetic spectrum. These readings can be used to draw a series of smooth curves of the light output across the entire spectrum on each day. Mapping these curves into the frequency domain of sound is simply a matter of changing the frequency scale. We can represent any one of these new sound-spectrum curves as phase vocoder data by assigning a set of discrete frequency values selected from the curve, one for each distinct channel of the phase vocoder. We have moved from the initial frequency-discrete sampling through the continuous graph it represents and back to a new frequency-discrete sampling in the domain of sound.

There are 3 problems concerning this mapping. The first is to decide what frequency to assign in each channel of the phase vocoder, as this will affect exactly what we hear in the sound output. A naïve approach using the channel centre frequencies gives the entire sound output a distinct coloration which is an artefact of the particular channel-count we use for the representation. Alternatively, as each curve typically rises or falls slowly in frequency around the frequency-peaks, then the loudest frequency falling within any channel will lie at the upper or lower frequency limit of the channel, but using these frequency boundary-points in the mapping also gives a distinct coloration to the output. A perceptually more credible approach is to assign a frequency at random within the frequency-limits of the channel, but randomly change this selection as we move from one time-window to the next. These selected frequencies are assigned the amplitude corresponding to their position on the original frequency-amplitude curves. The time-evolving data then more honestly represents the smooth curve input data as, over time, we are representing a large number of points on the spectral envelope (which changes only gradually from time-window to time-window) independently of where the phase vocoder channel boundaries fall.

The second problem is that, given the nature of the input, however we do this assignment we produce a broadband noise output whose variation over time is not aurally spectacular. The clear peaks and troughs in the drawn curve of the visual spectrum do not correspond to particularly striking markers in our aural experience. Looking at the light-spectrum graph what catches our attention are both the spectral peaks and also the various notches in the spectrum, particularly after it reaches its maximum magnitude after the first few days. These notches are physically significant as they represent absorption of energy by the various elements generated in stellar processes. It seems better therefore to map these peak and trough features and ignore the rest so that the aural features are as striking as the observed visual features of the 2-d curve. Mapping peak data to one channel and trough data to another, we can represent in a stereo signal all this key information about the original light curve.

The third problem is perceptual. Our sense mechanisms for perceiving light and sound are quite different. In particular, in the audio spectrum we perceive a clear sonic correspondence between frequencies with simple ratios between them (the intervals of the pure Pythagorean scale), we hear harmonic relationships, degrees of dissonance and, in the case of the octave, even the sense that we are hearing the "same" note, but in a different register. There is no parallel structure in our visual perception of colour - there are no colour "octaves" for example. In general this is not a problem as the spectra produced by our procedure will tend to contain components far from simple frequency ratios - they are strongly inharmonic. Even if we choose a mapping with the specific goal of mapping one time-window to "harmonious" musical intervals, we will not produce persistently harmonious outputs because, as the supernova spectrum shifts constantly through time, our harmonious moment will very quickly drift towards inharmonicity.

On the other hand, the difference between aural and visual perception can be used to our advantage. As harmonic relations have no visual-perceptual significance, we can use the harmonics of our selected frequencies to make them aurally brighter or duller in line with some property of the visual spectrum. It seemed appropriate to map the narrowness of the visual peaks or troughs into the count of aural-harmonics, narrower means brighter sound as the energy is focused more acutely in a narrow peak or trough. This is the mapping I used, and as a consequence the light-spectrum is mapped initially into the lower audible-frequency-range, so that several harmonics are available above this range to colour the resultant spectral frequencies.

Having found a solution to the mapping problem there are still some aesthetic issues to resolve. First of all the overall evolution of the supernova spectrum takes place over months. Mapping the time-scale of the supernova on a one-to-one basis would be possible but hardly likely to attract an audience, or only an audience which dropped by for a snapshot of the evolving piece and would therefore miss the overall evolution of the event. In order to present the entire phenomenon we need to compress the time to a manageable scale. In my realisation one day becomes approximately 9 seconds, so that 68 days is compressed into a 10 minute experience.

Secondly, as supernova explosions are both hyper-energetic and very far away (otherwise they would destroy us) they are generally only noticed by observers with powerful telescopes, against a background of existing prominent star constellations, planets, and the moon and sun. Do we treat this as a mysterious distant event, which might suggest a subtle disturbance in a piece of quiet contemplative music, or is this an overpoweringly energetic event which might suggest powerfully loud music? I (arbitrarily) selected the latter, and in order to convey something of the energy of the event, the changing spectra pass through the 8-channel listening space, as if travelling towards us from the source in the far distance in front of the listener, and are rhythmically animated by a semi-random juddering of the material. Doubling in lower octaves adds to the weight of the loudest sounds.

Thirdly, the overall evolution of the supernova spectrum is a relatively quick rise to maximum intensity followed by a long decline to zero. Although a musical event with a long decaying loudness tail would thus best represent the shape of the supernova event, this is not usually an ideal shape for a piece of music (though it has been adopted in e. g. Penderecki's *Threnody for the Victims of Hiroshima*). As much of the interesting information in the spectrum (the absorption lines of the spectra of the created elements) emerges at it fades away to nothing I decided to elaborate this aspect of the event by displaying the spectra of various chemical elements over the decaying tail of the supernova sound. I mapped the spectra of the elements into sound in a similar way to the supernova mapping, but here the mapping is more straightforward as the element spectra consist of individual narrow spectral lines corresponding to distinct frequencies as electrons jump between quantum states in the atomic shells.

However, this created a different set of musical problems. The chemical spectra are time-static phenomena. A spectrum is a spectrum is a spectrum is a spectrum. And although the (sound-equivalent) spectra of different elements sound different to each other, this does not necessarily make us want to continue to listen to these events once our initial curiosity has been satisfied. Therefore, after synthesizing blocks of raw, unvarying spectra corresponding to the different elements, I created the "Varibox" meta-program (a combination of existing sound-shaping tools) to generate more lively events over these bare spectra, events which pulse, or swell and decay, or flutter or are stacked into attack-resonances – or combinations of these – with a certain randomisation of the parameters of the sound-morphology to produce varied sets of interesting events. These chemical signatures are then distributed over the 8-channel surround-sound space as the sounds of the supernova decay towards zero.

The piece *Supernova* is the resulting 8-channel sound-surround piece.