Title	Adapting and Parameterising Auditory Icons for use in a Synthetic Musical Instrument
Authors(s)	Phelan, Cormac, Bleakley, Chris J., Cummins, Fred
Publication date	2009-06-10
Publication information	Phelan, Cormac, Chris J. Bleakley, and Fred Cummins. “Adapting and Parameterising Auditory Icons for Use in a Synthetic Musical Instrument.” The Institution of Engineering and Technology., June 10, 2009. https://doi.org/10.1049/cp.2009.1695.
Conference details	20th IET Irish Signals and Systems Conference (ISSC), Dublin, Ireland, June, 2009
Publisher	The Institution of Engineering and Technology.
Item record/more information	http://hdl.handle.net/10197/3859
Publisher's version (DOI)	10.1049/cp.2009.1695

Adapting and Parameterising Auditory Icons for use in a Synthetic Musical Instrument

Cormac Phelan, Chris J. Bleakley, and Fred Cummins

School of Computer Science and Informatics
University College Dublin,Ireland

E-mail: `cormac.phelan@ucdconnect.ie`

Abstract — **In this paper we describe the adaptation and parameterization of environmental auditory event structures for use in a real-time musical synthesizer. In doing so, we have developed a new software musical instrument based on the parametric representation of ecological sound structures, and which facilitates the application of typically non-musical auditory events in a musical context. Since this approach is to be realised within a low-latency, performance oriented synthesizer, our aim is not the development of computationally expensive physical models, but rather to effectively convey complex auditory events using arrays of simple sinusoidal components, while still retaining the key perceptual features of those events. By offering a performer ecologically-derived parameter control, we aim to encourage an *Everyday Listening* approach to electronic music performance, while also allowing the user to combine and develop typically unrelated elements of environmental auditory events.**

I Introduction

The *Ecological* approach to audition is firmly rooted in an attempt to understand environmental sound in terms of its real-world causal root [1]. By parsing the wealth of complex sonic data present in our environment into *Auditory Events* [4], we define our perceived sonic surroundings in terms of the physical properties giving rise to those events. As 'Everyday Listening', this approach allows us to perceive a great deal of information about a sound's source, both in terms of environmental cause and dimensional properties [2]. In contrast, the 'Musical Listening' approach essentially disregards any notion of environmental cause, instead perceiving a sound purely in terms of its timbre and sonic properties. With occasional exception, music performances have traditionally been perceived within the framework of 'Musical Listening'. For example, an expressive violin part is more often perceived as a richly harmonic and dynamic sustained tone, rather than the 'Everyday' approach of perceiving a rapid, friction-based excitation of a taught string by a coarse bow.

Today's electronic musical synthesizers concentrate on either a purely 'Musical Listening' approach, without any linked concept of ecological cause, or on an 'Everyday Listening' approach to simulating *existing* musical instruments. While there is certainly creative and technological merit to both approaches, there is still scope to explore the application of an ecological approach in areas other than instrument simulation. In addition, parameter controls in modern synthesizers are typically expressions of abstract terms such as filtering frequency or rate of oscillation. While this is appropriate for a strictly 'Musical Listening' approach, it does little to suggest any concept of environmental cause, offering the performer no choice but to interact with their instrument in non-ecological terms.

In contrast, we aim to adapt typically non-musical auditory events for use in a musical context, and allow the performer to bring together otherwise distant ecological event structures. Rather than exhaustively mimicing the sonic characteristics of existing musical instruments, we instead aim to attribute elements of ecological familiarity and perception to tonal synthesis, an area that does not usually display such traits. In addition, we present the performer with ecologically-defined interactive elements, expressing the parameter space of our synthesizer in terms of the ecological phenomena giving rise to its sonic characteristics (e.g. an 'Object Hardness' paremeter affecting bounce period and spectral content) and further emphasizing our focus on applying an 'Everyday Listening' approach to musical synthesis.

Essentially, our aspiration is to provide a novel approach to both pitched tone synthesis and parameterization of interactive control, both rooted in *Everyday* listening rather than *Musical* listening.

To ensure low-latency operation, the broadest possible platform support and stability in a range of host environments, our synthesizer was developed in C++ using Steinberg's Virtual Studio Technology (VST) SDK [15].

The following sections of this paper introduce related research, describe example techniques for the musical adaptation of environmental auditory events, describe an ecological approach to parameterization, and present experimental results from a listener survey comparing real-world recordings and synthesized events.

II Related Work

Previous work by Gaver [1] [2] [3] laid the foundation for current work in the field of ecological event perception, popularising the concept of 'Everyday Listening' and Auditory Icons, while Warren and Verbrugge's [4] analysis of breaking and bouncing auditory events paved the way for over two decades of study in ecological Acoustics.

The recovery of material properties in contact sounds was examined Klatzky, Pai, and Krotkov [5], and highlighted the crucial role played by decay properties in the accurate perception of cues in virtual environments.

Bilbao, Arcas, and Chaigne [8] developed a computationally expensive but highly realistic model of plate reverberation based on a direct numerical simulation of the equations of motion of a thin linear plate. In contrast, a real-time approach is taken by Aramaki and Kronland-Martinet [10] to reproduce impact sounds by using a subtractive model based on the dynamic filtering of a noisy input signal.

Peltola, Erkut, Cook, and Vlimki [11] developed a real-time Pure Data library to synthesize hand clapping sounds, and offered ecologically-derived parameter control such as 'number of people' and 'synchronization'

None of these studies, however, approach the topic of synthesizing and adapting typically non-musical auditory events for use in a musical context. Furthermore, they do not examine the scope for combining otherwise contrasting ecological events within a single simulation, and implementing this simulation in a real-time musical instrument.

III Adapting Simple Environmental Auditory Events for use in a Musical Context

Simple interactions between real-world objects and materials can be convincingly synthesized using a set of simple sinusoids, each corresponding to the Eigenmodes of our objects, and each with associated attack and decay coefficients [5]. In this section, we detail the adaptation of an example auditory model for use in a musical context while still retaining its perceptual meaning.

a) Weighting Harmonic Partials

Within the 'struck object' class of environmental auditory events, the recovery of ecological characteristics is largely determined by the relative length of decay times for each frequency partial [5]. These decay times are a product of material damping, which is in turn determined by the internal coefficient of friction and varies greatly from one material to the next. With this in mind, the amplitude of each sinusoid within a sounding struck body model is proportional to

$$exp(-b(f)t)sin(2ft\pi) \tag{1}$$

where f is the frequency of the mode, $b(f)$ is a frequency-dependent damping component and t is time. The nature of b depends on the material model [5].

However, many real-world materials exhibit a dissonant relationship between these frequency partials, rendering them less usable in a note-based musical context. In order to apply such ecological auditory events in a musical context, we must adapt their harmonic structure, while still retaining the key sonic features which give rise to the successful recovery of the original material properties. A good way to achieve this delicate balance is by allowing our synthesized tone to retain a similar set of frequencies to that of its real-world counterpart, but applying a set of weighted coefficients to these partials in the hope of emphasizing a certain dominant frequency. The extent to which we need apply these weights is determined by our Environmental Event's position in a spectrum ranging from strongly dissonant to strongly consonant.

Tending toward the less dissonant end of the spectrum, an impact interaction with a *glass object* yields a relatively widely spaced, isolated set of clearly defined harmonic partials, many of which are strongly harmonic (Figure 1) [9]. While it is true that dissonant partials are often present, the sparsity and sharp definition of these partials allows us to easily 'quantize' their frequencies to those of consonances during synthesis, without compromising their implied ecological root (Figure 1). In listener tests, we observed equally

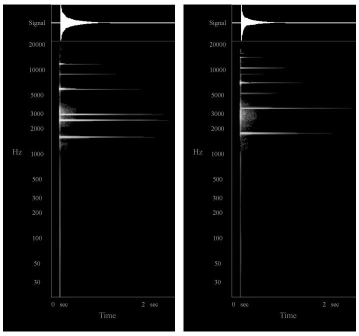

Fig. 1: Non-Adjusted and Adjusted Glass object interactions

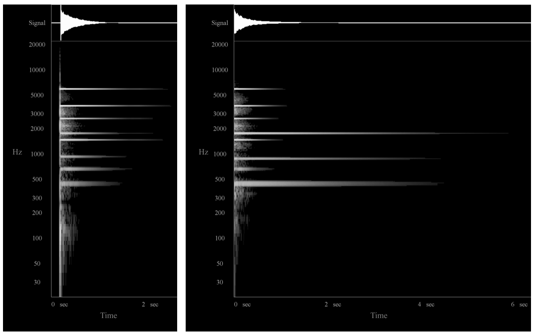

Fig. 2: Non-Adjusted and Adjusted metal bar interactions

satisfactory levels of recognition between a non-adjusted synthesized glass object (69.81%) and a harmonically quantized glass object (66.03%). We examine the results of these listener tests later in the 'Listener Evaluation Experiment' section. Temporally speaking, decay times shorten considerably toward higher frequencies and it is this, combined with a relatively sparse set of harmonic partials, that contributes to our perception of 'a glass object being struck'. It follows that with such a strong presence of a fundamental frequency, there is little need to adapt this Environmental Auditory Event for use in a musical context, beyond a possible quantizing of strong dissonances.

In contrast, an impact interaction with a *metal bar* is sonically characterised by a dense clustering of dissonant slow-decay harmonic partials, with the distribution of energy across these partials tending towards higher frequencies both in terms of amplitude and decay times (Figure 2). Perceptually speaking, the strongly dissonant relationship between partials is the key defining property of this ecological event, making its adaptation for use in a musical context much more challenging. However, by recognising that the attack portion of an ecological event is hugely significant in defining its perceived environmental cause, we can apply carefully chosen weights to the decay times of each harmonic partial, with the intention of allowing a dominant set of strongly harmonic frequencies to emerge over time (Figure 2). In this way, the attack phase of our synthesized Auditory Event retains a distinctly 'metallic' sonic property, since the set of characteristically clustered, dissonant harmonics are still present. With this perceptual link established, our weighted set of strongly harmonic partials are allowed to decay much more slowly than the range of dissonant partials, resulting in a musically usable tone with a striking, metallic attack phase. This is achieved by assigning smaller values for $b(f)$, our frequency dependent dampening component, when f is a strong harmonic of our fundamental. In listener tests comparing the subject's recognition of a real-world recording of a struck metal object (75.47%) to that of a synthesized, adjusted metal object (73.58%), we observed similarly high levels of identification. Furthermore, this adjusted 'struck metal bar' event proved easily recognised as a musical tone, with 98.11% of listeners correctly identifying a short piece of popular music (the 'Happy Birthday' tune) based on this model.

b) Modeling Varied Object Interactions

While the harmonic relationship between the partials present in an Auditory Event strongly suggests the material and nature of a struck body, the distribution of energy across its range of partials is closely associated with the nature of the original environmental interaction with that struck body [2].

Broadly speaking, we can model the variable hardness of a mallet (or, indeed, *any* object of excitation) striking a body by modifying the relative amplitudes of high and low frequency partials (Figure 3), while also adapting the global attack time of the Auditory Event [6]. With the interaction of a *hard* object, there is a tendency to observe a concentration of energy in the higher frequency partials, coupled with a very fast, immediate attack time.

In contrast, we can model a *soft* mallet interaction with precisely the same metal object by shifting the concentration of energy to the lower partials, coupled with a slightly slower attack time (Figure 3). While the set of harmonic partials present remains completely unchanged, our differing set of weighted amplitudes and decay times gives rise to a distinctly different timbre, and one that is clearly distinguishable as the result of a 'soft' excitation interaction. In listener tests comparing synthesized 'metal bar' interactions with varied mallet hardness, 58.49% of subjects were capable of accurately judging *both* perceived hardnesses from the auditory cues, while 73.58% and 77.35% correctly identified either 'hard' or 'soft' interactions respectively. We may apply these principles of harmonic distribution and attack style to

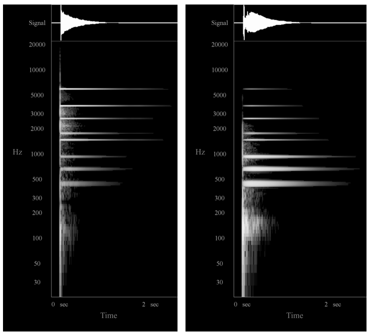

Fig. 3: Hard and soft mallet interactions

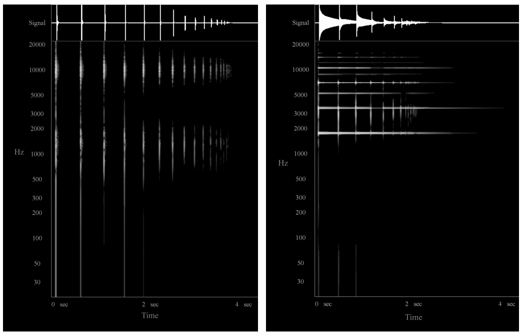

Fig. 4: Ping-Pong Bouncing Event (left) Combined 'Glass' and 'Bouncing' Events (right)

any of our adapted 'Struck Object' interactions, offering the user additional scope to combine ecological characteristics.

IV Adapting Compound Environmental Auditory Events

More often than not, ecological Auditory Events are perceived as composite groupings of related simple events, rather than isolated single-object interactions [14]. While *simple* ecological auditory events are significantly defined by the harmonic relationship between their partials, the temporal patterning of *composite* events plays an important role in listener perception. The richer temporal structure of compound events can reveal a great deal of ecological information to the listener [1], coupled with the spectral content of each separate simple component. Just as it has proved possible to adapt *simple* auditory events for use in a musical context, so too may *compound* events be similarly adapted.

While this raises possible questions of ecological *plausibility*, these issues are not necessarily detrimental to our study, since we are not aiming to strictly simulate existing real-world auditory events. In fact, the potential to combine ecologically implausible event structures may provide the performer with unique timbres, with any retained perceptual associations serving to enrich the experience of the listener.

a) Synthesizing A Real-World Compound Event

An immediately recogniseable (88.67% perceptual accuracy in listenter tests) yet simple compound auditory event is the bouncing pattern of a ping-pong ball on a hard surface, which follows the broad structure of a series of sharply defined, short lived impact events of changing period and amplitude [4]. By modeling this structure as a series of band-filtered, very fast decay noise bursts with decreasing interval and amplitude, we can synthesize a convincing simulation of this auditory event (Figure 4). The high filtering bands chosen follow Klatzky, Pai and Krotkov's principle that 'frequency increases with stiffness' [5]. As listeners, we perceive this compound event very clearly as a bouncing object, largely due to the distinct temporal pattern adhered to by each individual impact event. In listener tests, a real-world recording of a ping-pong ball yielded an 88.68% rate of recognition, followed closely by an 84.9% rate of recognition for a synthesized event.

While there is certainly a distinct and significant rhythmic element to this auditory event, it lacks a steady and dominant pitched component, making it of limited musical use in a tonal sense. In fact, due to the rarity of harmonic purity and stability in real-world auditory events, this is a common obstacle in applying environmental models in a musical context, but one that may be overcome in a similar manner to our previous tonal adaptations of simple events.

b) Tonal Adaptation of Compound Event Structures

If our 'Bouncing Ping-Pong' compound structure is significantly perceptually defined by its temporal pattern, then it follows that we should be able to apply a similar pattern to a series of tonally-adapted simple Auditory Events (such as our previous 'metal' or 'glass' examples), yet still retain the ecological association of a bouncing object.

By mapping a harmonically weighted instance of our simple 'Glass Object' event onto the temporal pattern of a compound bouncing event (Figure 4), a number of key issues are addressed. Most critically, our previously non-tonal auditory event is adapted for a musical context with a clear and sustained set of harmonic partials. These sustained, widely spaced and strongly harmonic partials serve to tonally anchor the auditory event to a root frequency, while also attributing a perceived 'glassiness' to the bouncing event (60.37% of listeners correctly identifying both 'bouncing' and 'glass' elements to the sound). As a result of the event's new temporal complexity, ecological parameters such as surface hardness, initial velocity and object rigidity are now detectable by the listener, in keeping with Gaver's assertion that the richer structure of compound events allows a listener to perceive more detailed information about the event and its environmental cause [1]. Later, we will explore how these environmental parameters can be represented and manipulated in real-time by a performer, resulting in a rich, ecologically-defined synthetic musical instrument.

In combining a temporally complex compound auditory event with a musically-adjusted single impact event, we have struck a balance between the approaches of *Musical Listening* and *Everyday Listening*. On one hand, the presence of a dominant set of closely-related harmonics lends this auditory event to a *Musical* listening approach, where a sound event's environmental cause is disregarded in favour of its timbral properties (or Helmholtz's 'Tone Colour' [7]). On the other hand, the application of an ecologically-based temporal model lends itself to an *Everyday* listening approach, where an auditory event is perceived principally by its physical, environmental cause.

V Ecologically-Based Parameterization and Implementation

The ecological characteristics of our synthesized events are the product of many variable weighting coefficients, affecting both spectral content and temporal structure. However, in presenting a user with interactive parameters, such mathematically abstract terms as a 'Decay Time' or 'Harmonic Weighting' are problematic and potentially counter-intuitive to the performer. Instead, our synthesizer aims to express its interactive elements in terms of their ecological meaning, and consequently offer a more approachable interface to the performer.

By offering real-time control over the ecological properties of a synthesized compound auditory event, the user is given the capability to tailor the auditory event to their need, while also creating a more versatile, dynamic musical tone. This is in contrast to the traditional approach of parameterising sonic features of our synthesized tone according to abstract mathematical terms. For example, in the case of our adapted 'Bouncing Object' event, an interactive parameter termed 'Object Size' determines the weighting coefficients of our harmonic partials, emphasizing lower frequencies for larger values. Altering the temporal structure of our synthesized environmental sound, an 'Object Rigidity' parameter has a threefold effect on the coefficients of bounce interval [4], resonance decay time, and harmonic structure [5].

Even in this very simple example, we can see how the coefficients affected by each parameter may overlap and interact with each other, as is the case with harmonic structure in the previous example. Not only do the two specified parameters of 'Rigidity' and 'Size' have a shared effect on the spectral content of our auditory event, but other parameters such as 'Mallet Hardness' or 'Material' also contribute significantly to the set of weighting coefficients affecting that spectral content. This interaction between user-accessible parameters leads to a nuanced sonic palette, with a convincing real-world interaction between those parameters. In addition to these two example linked parameters, many others emerge within our framework. 'Object Regularity' affects the uniformity of bounce interval, and 'Initial Velocity' affects both amplitude and harmonic distribution, to name but two more. When we begin to consider even this relatively small set of control parameters, it is immediately apparent that by using an ecologically-derived parameter set, the result is an interactive instrument whose parameters do not seem to be isolated, abstract components. Rather, they are a set of environmental criteria, the effects of which closely interact and inform each other.

From the examples already explored, it is clear that many environmental auditory events can be convincingly synthesized using small, weighted sets of simple sinusoids corresponding to the dominant Eigenmodes of the synthesized object, and modulated according to ecological properties. It follows that an *additive synthesis* approach is most suitable for the development of a real-time, performance oriented musical synthesizer, capable of simulating and evolving a range of ecological auditory events.

Depending on the nature of the user-selected ecological event, a single root note will generate a bank of both harmonic and dissonant oscillations, each corresponding to the spectral properties of our synthesized materials, with the balance between harmonic and dissonant dictated by material type. For example, the amplitudes of a 'glass' tone's partials are weighted heavily towards harmonic frequencies, while a 'metal bar' tone has far more dissonant content. In addition, features such as 'mallet hardness' are evoked using a broad weight balance between high and low frequency partials, while global decay times are also weighted according to material, as explored earlier. The many timbral possibilities offered by this configuration are expanded further by applying environmentally-modeled amplitude envelopes to each synthesized note (known as a 'voice') separately, allowing us to model a hypothetical group of bouncing, harmonically resonating objects, for instance.

VI Listener Evaluation Experiment

To evaluate the effectiveness of our range of synthesized auditory event structures, we constructed an online listening survey, allowing participants to listen to audio clips then specify their perceptual associations from a list of options. Choices ranged from four to nine options, but participants were also permitted to enter their own perceptual association, should none of these options be satisfactory. Since we used both synthesized and real-world recordings in this test, we obtained data concerning the ecological validity of our synthesized auditory events, expressing accurately identified perceptual features as a percentage of 53 total participants.

Data for the example recordings of musically-adapted auditory events suggests that perceptual meaning is largely retained, despite changes made to the event structure. For example, a strongly dissonant recording of a real-world struck metal object yielded a 75.47% accuracy in recognition, compared to a 73.58% figure for a synthesized, harmonically adjusted ‘metal rod’ event.

Perceptual accuracy in compound auditory events was, for the most part, significantly higher than that of simple events, supporting Gaver’s claim that the richer structure of compound events makes it possible for a listener to obtain more detailed information about the event and its environmental cause. Bouncing objects such as a real-world ping-pong ball yielded an 88.67% rate of recognition, with a similarly high figure of 84.9% for our synthesized version of the event. Most critically, these levels of ecological validity are maintained through our combined and musically-adapted examples, with a ‘bouncing metal’ example yielding a 90.56% accuracy in identification. This supports the suggestion that listeners are not only capable of recognising and extracting ecological features from a complex synthetic tone, but also that listeners are comfortably capable of linking typically unrelated ecological events within a single musical tone.

VII Conclusion

In this paper we have outlined techniques for the musical adaptation of simple and compound ecological events, by both adapting their temporal structure and by adjusting their harmonic partials, yet still retaining their ecological meaning. We have employed these techniques in a real-time software synthesizer, and also outlined the significance of providing ecologically-based control parameters to the user. We have tested the ecological validity of a range of resultant synthesized tones in a series of listener tests, and obtained satisfactory recognition results.

Our approach for ecologically-based parameterisation also comes with its own set of advantages and disadvantages. Most significantly, it encourages the performer to adopt an ‘Everyday Listening’ approach to musical performance, while also adding a desirable degree of interaction between parameter controls.

The technique of modeling environmental auditory events using sets of simple, weighted sinusoidal components is significantly less nuanced than that of large-scale physical modeling [8]. However, our focus is not on exhaustive computational models but instead on distilling a key, core set of perceptual features in a range of auditory events, then allowing the performer to combine those features. A significant benefit of this simplified technique is the computational efficiency of our implementation, which allows extensive polyphonic performance in real time without exceeding realistic demands on even modest system resources.

This ecologically-based synthesis technique is an alternative to the approaches of traditional synthesis and contemporary physical modeling, and is intended for performance purposes. Depending on the requirements of the user, a more detailed and computationally expensive approach may be demanded, but for many musical performers a real-time instrument offering a novel, ‘Everyday Listening’-based approach to synthesis and parameterization may indeed be more desirable.

References

[1] WW Gaver, *What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception* Ecological Psychology, 1993

[2] WW Gaver, *How Do We Hear in the World? Explorations in Ecological Acoustics* Ecological Psychology, 1993

[3] WW Gaver, *Everyday listening and auditory icons* 1988 - portal.acm.org

[4] WH Warren, RR Verbrugge, *Auditory Perception of Breaking and Bouncing Events: A Case Study in Ecological Acoustics* Journal of Experimental Psychology, 1984

[5] RL Klatzky, DK Pai, EP Krotkov, *Perception of Material from Contact Sounds* Presence: Teleoperators and Virtual Environments, 2000 - MIT Press

[6] DJ Freed, *Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events* The Journal of the Acoustical Society of America, 1990

[7] H Helmholtz, *[Book] On the Sensations of Tone* - 1954

[8] S Bilbao, K Arcas, A Chaigne, *A Physical Model for Plate Reverberation* Acoustics, Speech and Signal Processing, 2006

[9] B Giordano, *Material categorization and hardness scaling in real and synthetic impact sounds* The Sounding Object, 2003, 73-93

[10] M Aramaki, R Kronland-Martinet, *Analysis-synthesis of impact sounds by real-time dynamic filtering* IEEE transactions on audio, speech and language processing, 2006

[11] L Peltola, C Erkut, PR Cook, V Vlimki, *Synthesis of Hand Clapping Sounds* IEEE Transactions on Audio, Speech and Language Processing, 2007

[12] B McClimens, J Nevitt, C Zhao, D Brock, JA Ballas, *The Effect Of Pitch Shifts On The Identification Of Environmental Sounds: Design Considerations For The Modification Of Sounds In Auditory Displays* International Conference on Auditory Display, 2005

[13] C Stevens, D Brennan, S Parker, *Simultaneous manipulation of parameters of auditory icons to convey direction, size, and distance: Effects on Recognition and Interpretation* ICAD 2004

[14] F Cummins, *Auditory event structure and speech.* In Proceedings of the 4th International Conference on Speech Prosody (2008), pages 631-634, Campinas, Brazil.

[15] Steinberg GMBH, *Steinberg VST 2.4 Software Development Kit (SDK)* URL: http://www.steinberg.net/en/ps/support/3rdparty/

This work has been funded in part by a Principal Investigator grant, 04/IN3/I568, to Dr Fred Cummins by Science Foundation Ireland.