A HYBRID APPROACH TO TIMBRAL CONSISTENCY IN A VIRTUAL INSTRUMENT

Sean O'Leary

Department of Computer Science
University of Limerick, Ireland
Sean.OLeary@ul.ie

Niall J.L. Griffith

Department of Computer Science
University of Limerick, Ireland
Niall.Griffith@ul.ie

ABSTRACT

The aim of this work is to make an instrument that is timbrally consistent over pitch and loudness. This particular work is not attempting to reproduce an existing instrument's timbre, but to produce a timbrally dynamic virtual instrument that can be designed by the user.
In this paper there is a brief introduction to timbre and synthesis methods, followed by a proposal on how to make timbrally consistent virtual instruments out of given timbres.

1. INTRODUCTION

Traditionally music has been written for certain instruments with specific timbres. One significant contribution of computers to music is the ability to create new timbres, from scratch or by transforming existing sounds. Sound synthesis, as well as creating new timbres, has been used to mimic these existing instrument timbres. An early method of sound synthesis was sampling, which recorded an instance of a tone. However, when transposed in pitch and loudness this was unconvincing because the acoustic properties of an instrument's tone changes over pitch and loudness; it is not just a case of multiplying the amplitude and frequency. The consistency of the timbre of a traditional instrument over pitch and loudness arises because it originates from a physical source with consistent properties and this in turn forms our expectations of what the instrument should sound like. A partial solution to this 'consistency' problem was to sample over several instances of pitch and loudness, and to interpolate instances that hadn't been sampled, but this was not completely successful. Physical modelling is a current approach that attempts to solve the problem by modelling the mechanics of the instrument, which then can consistently recreate the variations of the spectra over pitch and loudness.

Spectral modelling synthesis [Serra] is a powerful tool for sound transformation and timbre modification. The problem remains as to how to use these novel timbres, or virtual timbres, to make instruments that are consistent over pitch and amplitude.

While physical modelling allows for the design of virtual instruments that can't exist in reality, for example resonating the vibrations from a plucked string with the body of a drum, it doesn't allow for explicit modification of the spectral content of a tone.

So the problem remains: once we have a timbre designed, how do we make an instrument out of it?

2. TIMBRE

Timbre is defined by the American National Standards Institute as "that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different".

Perceptual research on timbre has identified that the main physical attributes that determine our perception of sound are the spectral distribution of energy, and its variation over time. Timbre is a time dependent attribute that evolves. It cannot have an instantaneous value. It describes an attribute of a continuous sound wave, of which an instantaneous value is just a measurement of an single relative air pressure level of the sound wave. This is all but meaningless in the context of timbre without its history or destiny. Timbre has to be analysed over time.

Harmonic sounds are made up of partials which have frequencies that are integer multiples of a fundamental. The summation of the amplitudes of the partials gives the amplitude of the sound wave at any point in time. Musical timbres however, are not purely harmonic, they usually have stochastic elements to them, and the partials are usually not at all well behaved. Also, there are limits to the analysis we can currently perform on real world timbres. Trying to formulate the physical behaviour, and recognise and categorise the perceptually salient aspects of timbre are important ongoing topics in the field of computer music research.

The notion of timbre is also blurred by the fact that acoustical properties, e.g. the spectral and amplitude envelopes, vary over loudness and the pitch range of a given instrument. For instance, many timbre parameters of a high-pitched piano sound are closer to the parameters of a high-pitched flute sound than to those of a low-pitched piano sound, but it is still easily distinguishable as a piano sound [1]. Also, if the spectral components are kept constant over its entire pitch range the instrument is unrecognisable.

2.1. Similarity Tests

In order to arrive at a strategy that will facilitate the design of a sound from a model of its spectral content and temporal evolution it is necessary to correlate physical properties of the sound with perceptual attributes.

Similarity tests are a common method used in the search for the perceptually salient aspects of musical timbre. A number of musical tones of similar pitch and duration are played in pairs to listeners, and the listeners rate the similarity of the presented tones. A multidimensional scaling algorithm is applied to the test results and the first two or three dimensions are examined to find a physical correlation for each dimension.

Grey (1977) [2] performed a classic example of such an experiment. The stimuli Grey used, consisted of sixteen different tones from a range of musical instruments, which were digitally synthesized and perceptually equalised for pitch, loudness, and duration. The results indicated that a three dimensional solution accounted for most of the differential in the judgments. The first dimension represented the spectral energy distribution of the tone. The second reflected the synchronicity of transients in the higher harmonics and the spectral fluctuation of the tone with respect to time. The last represented the level of high frequency, low amplitude and usually inharmonic energy during the attack segment. Grey observed that the second dimension seemed to group the instruments into their respective families.

More recently experiments by McAdams et al. (1995) [3] and Lakatos (2000) [4] again confirmed that attack and spectral distribution, or brightness (centroid), accounted for most of the differential judgments. These studies also used a mix of musicians and nonmusicians to see if they would rate timbre differently. No obvious correlation between musical expertise and timbre judgment was found.

While attack and brightness are recognised as the two most important dimensions of a timbre space for source identification; as timbre is a time dependant attribute there remain other aspects of a sounds structure that are important to the aesthetic quality of a timbre, these include its spectral shape and how this evolves over time.

3. SYNTHESIS

Physical modelling [5] and spectral modelling [6] are the two main synthesis techniques in modern sound synthesis.

Physical modelling attempts to recreate sounds by describing the mechanics of its origin. For example, a string with a resonator in which the string may be plucked or bowed. It has a density and tension, it excites a resonator that has certain dimensions, and is made from a material that has certain distributions of density and flexibility. A mathematical model of such a system can be used to generate a sound. Generally, a musical instrument can be modelled to comprise two elements: a source (or exciter) and a resonator. The source initiates the vibrations, and the resonator amplifies these to produce most of the sound that we hear. A designer can define the various elements of such an instrument with flexibility over many parameters - correlating to physical properties, and put them together in various permutations to create virtual instruments that would be impossible to create in the real world. A strong point in favour of physical modelling is that it produces sounds that adhere to our expectancies of musical sound; for example transients in the attack, or micro changes in the envelopes of the spectrum or amplitude that make the sound more interesting. Also, the intuitive, or expected change in spectrum with pitch and loudness that we associate with acoustic timbres is preserved.

Spectral modelling, on the other hand, is based on the spectral content of sounds. Instead of defining the physical properties of the origin of a sound, a sound is analysed and described in terms of its spectral components. The Fourier transform is used over successive periods in the temporal evolution of the sound to take snapshots of the spectral content of a particular portion of the sound. For pitched sounds the partials are tracked and temporal envelopes mapped for each one. The sound can be re-synthesised through additive synthesis for the harmonic part of the sound, and subtractive synthesis for the stochastic (noisy, residual) part. With this information various transformations can be performed on the sound. This allows for intuitive alterations to the sound as the FT model approximates aspects of the spectral analysis that takes place on the Basilar membrane of the ear. For example following this approach we can alter the brightness of the sound, stretch it over time; or we can add voice like elements to the sound by changing its spectral envelope with respect to time.

Physical modelling and spectral modelling are the dominant techniques in modern synthesis. They both have desirable attributes. For example physical modelling is true to our notion of timbre because the spectral characteristics of the sound produced will change over the models pitch range in a way that is consistent with our recognition of the timbre an acoustic instrument. Also it offers intuitive mappings for control. Spectral modelling allows the intuitive alteration of salient attributes of a given timbre, such as attack time and spectral centroid, and enables various transformations and effects to be performed on the sound.

However, a significant problem with spectral modelling is that to provide all the information required to make a musically interesting sound is largely impractical, even more so the task of defining it over a pitch and amplitude range, which is probably impossible. Thus, there are two problems inherent in current synthesis techniques: The problem with physical modelling is that the spectral shape produced cannot be influenced in an intuitive fashion. The problem with spectral modelling is the degree of detail that is required to achieve intuitive aesthetics and instrumental consistency.

4. CONCEPT

One solution to these problems is to use a physical model as the source and a spectral model as a resonator. The source may be a physical model, or an actual sound that is consistent over a dynamic and pitch range. For example we might use an electric guitar as an input to a spectral model that is derived from another instrument. The intention is to create a playable instrument that is timbrally consistent across its pitch and loudness range, yet is modified according to a spectral resonator that is modelled on the spectral differences between one instance of two sounds (input and object). The input sound can be looked at as the exciter and the spectral model as a spectral resonator. Thus the aim is not to recreate an existing timbre, but to create a novel timbre while providing it with variation over pitch and loudness which is consistent with the sound being used as input.

The elements of timbre that need to be controlled, disregarding the noise elements, are the frequency value of the partials and the overall temporal evolution of the spectral components. There are an infinite number of possibilities here, and for meaningful control we need to understand the perceptual correlates of the structure of the sound. For example, brightness and attack have been recognised as key dimensions in a musical timbre space, however these dimensions do not define the evolution of the partials. Brightness and attack are not unique values with respect to spectral content. The partial evolution is dependent on both the source sound and the resonator.

The resonator for this model can be designed using an object sound. For one instance of pitch and amplitude the spectral model can be constructed so as to make the input sound's partials follow trajectories that are consistent with those of the object sound. By recording the transformation, and making some alterations to generalize its use, we can then perform the same transformation on the input sound, varying its pitch and loudness, and the resulting transformation will be relative to the input sound, its spectral content varying over pitch and loudness relative to the input.

One of the important factors attributed to achieving consistency over pitch and amplitude for instruments with a static resonator, such as the violin or guitar, is that the resonances of the body tell us something about its structure, and that the consistency in recognition is partly realised by the source being resonated by (effectively) a static filter [7]. If the object sound is a product of a source resonated by a body then LPC can be used to decompose the object into a source and filter. The resulting filter can be used independent of the frequency content of the input sound. The remaining transformation can now be modelled on the difference between the source of the object sound and a sample of the input sound, and this transformation will be relative to the fundamental frequency of the input.

Temporal and pitch control will be achieved through the input sound and spectral transformations correlating to timbral attributes performed by the spectral model.

5. TIMBRAL CONTROL

Using the object sound to create the spectral resonator is useful as potentially it allows us to alter the resulting sound further during the design of the instrument. The specifications involved in designing a sound over its pitch and loudness range are still problematic, hence the proposal of an input sound. Likewise, although it would be ideal to be able to create any imaginable sound it would be difficult to specify it's the control of such an open space of spectral evolution intuitively, hence the desirability of using an object sound. To give more control over timbre there are certain meaningful attributes we can vary.

Brightness: The spectral centroid of a sound corresponds to its brightness. The brightness of a sound can be altered by multiplying it by a frequency dependent function, or by adding or removing partials.

Attack: The attack can be altered either in the frequency domain or the time domain by adjusting the amplitude of the spectrum/waveform.

Traditional source / filter decomposition: Using LPC, if it is relevant, the object sound can be decomposed in to a source/resonator model, for example a guitar tone could be split into the energy from the string, and the frequency response imposed on it by the body. The resulting frequency response could be treated in the usual way, the response being independent of the fundamental frequency of the input sound, while the transformation modelled on the differences between the decomposed object source and the input sound, is performed relative to the fundamental of the input.

Frequency values of spectrum: the degree to which the frequency values of the output sound match those of the object sound relative to the fundamental of the input.

Spectral shape of the sound: the degree to which the spectral shape of the object sound influences that of the output.

Harmonicity: this will affect the distance between the partials relative to the distance between the partials of the input sound. It is a multiplying factor for frequency values; a value of one will leave the frequency values of the spectrum unchanged.

Feedback to input: to strengthen the relationship between the input and the resonator, part of the output could be fed back to the source, this could be proportionate to the filter being used, or again could be a time varying attribute. This mimics the sympathetic vibrations present between a string and the instrument body of a traditional string instrument [9]. It doesn't have to model the response of an actual system, it could just be a feature of the instrument design.

It would be desirable for computational reasons for as many of these attributes as possible to be dealt with in the sound design stage of model, i.e. designing the object sound. Brightness and harmonicity should all be dealt with at this stage unless they are to be interactive during performance. Frequency values of the spectrum, and spectral shape are only necessary if the degree to which the object influences the input sound is to be variable, again, if they are not to be interactive they can be included in the one transformation. The LPC decomposition does not have to be done, all the difference in the sound can be modelled in the frequency dependent transformation, but this would remove the resonances that are uniformly present in the sound, and may affect its consistency.

6. ISSUES

When designing the spectral resonator it may be desirable to have the two sounds of the same duration for the processing to be done on a frame-by-frame basis. This may require time-stretching of the object sound [9]. Of course, we will not always know the length of the object sound prior to processing it, so perhaps the object length should be stretched to a reasonable maximum, and if the input exceeds this then the last frame should be held onto for the remainder of the duration.

Detecting the start of a new note could also be a problem, depending on the input sound attack detection could be feasible, or a midi input to a physical model would solve the problem. Also polyphony is not tackled in this model.

7. SUMMARY

The proposed model allows the spectral variation present in a physical system, e.g. a vibrating string, over pitch and loudness to be applied to an abstract timbre with the intent of achieving a consistent timbre over a pitch and loudness range.

This model is based on partial tracking, and as such is not generalised to inharmonic sounds. In order for it to be more general differences in spectral shape could be modelled instead of the difference of partials. Further work will also include research in to Higher Level Attributes [10], and their possible role in this model.

8. REFERENCES

[1] Jensen, K, "Timbre models of Musical sounds", PhD Dissertation, DIKU Report 99/7, 1999.

[2] Grey, J.M., 1977. "Multidimensional perceptual scaling of musical timbre", *Journal of Acoustical Society of America*, vol. 61, pp. 1270-1277.

[3] Mcadams, S., Winsberg, S, Donnadieu, S., De Soete, G, Krimphoff, J, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes", *Psychol Res*, vol. 58, pp. 177 – 192, 1995.

[4] Lakatos, S, Eds., "A common perceptual space for harmonic and percussive timbres", *Perception and Psychophysics* , vol. 62, no. 7, pp 1426-1439, 2000.

[5] Cook P.R., "*Real Sound Synthesis*", A K Peters, 2002.

[6] Serra, X., "A System for Sound Analysis / Transformation / Synthesis based on Deterministic plus Stochastic Decomposition", PhD Thesis, Stanford University 1989.

[7] McAdams, S, "Recognition of auditory sound sources and events" *Thinking in sound: The Cognitive Psychology of Human Audition*, Oxford University Press, 1993.

[8] Bonada, J., "Automatic technique in Frequency Domain for Near Lossless Time-Scale Modification of Audio" in *Proc. International Computer Music Conference*, San Francisco, 2000.

[9] Karjalainen, M,, Valimaki, V., Janosy, J, "Towards High-Quality Sound Synthesis of the Guitar and String Instruments" in *Proc. International Computer Music Conference*, Tokyo, 1993

[10] Serra, X, Bonada, J, Sound "Transformations Based on the SMS High Level Attributes", in *Proc. Workshop on Digital Audio Effects (DAFx-98)*, Barcelona, Spain, pp. 188-191, 1998.