ACOUSTIC-AGGREGATE-SYNTHESIS

Paul Clift
Department of Music
Columbia University, New York, NY

ABSTRACT

Acoustic Aggregate Synthesis describes a real-time performance tool which attempts to fuse synthetic and acoustic sound sources in order to achieve a *semi-acoustic re-synthesis* of a pre-defined acoustic model. At the heart of this project is the desire to maintain the diffusion patterns, attack/sustain/release characteristics, acoustic amplification etc. of a given instrument whilst 'overriding' its timbral characteristics in favour of the creation of a contrasting, readily identifiable, secondary timbre.

1. BACKGROUND

1.1 Additive analysis/re-synthesis

The process described in this essay is essentially one of additive re-synthesis; such processes, even those which make use of the most up-to-date analysis/synthesis tools are well-known and require little elaboration here. Nonetheless, while such procedures have long proved effective as a means of producing/manipulating complex timbres (or, more pertinently here, *hybrid* timbres) in deferred-time, or, when used in real-time, as the basis of orchestrated synthetic accompaniment to instrumental performance, the notion of unifying acoustic & synthetic sound sources in the creation of a perceived single, hybrid timbre with historically-defined semantic characteristics (*e.g.* recognisable as a flute) remains a fecund territory for exploration. Worthy of note is the ambitious research in the hybridisation of timbre undertaken by Jonathan Harvey *et al*, which culminated in the premiere of *Speakings* (2007-8) for orchestra.[1]

1.2 Augmented instruments: the integration of transducers into acoustic instruments

AAS also follows upon a long tradition of 'augmentation' of acoustic instruments through the integration of diffusion transducers. Indeed, a potent factor in the success of timbral homogeneity (in particular when a combination of acoustic & synthetic sounds is used) is perceived directional singularity. Although relatively little has been written on this subject, countless attempts have been made, with varying degrees of success, at establishing fusion between acoustic & electronic sounds through maximal spatial proximity of multiple sound sources. This phenomenon may be enhanced when, as is practicable in certain cases, electronic sounds are subject to a given instrument's acoustic diffusion patterns (*i.e.* the electronic sound source is diffused not directly, but rather after passing through one or more instrumental mechanism(s)).[2] This notion is discussed further in section 2.5 of this essay.

2. ACOUSTIC AGGREGATE SYNTHESIS (AAS)

This process, using the *MaxMSP* interface, makes a real-time comparison between an *instrumental-template* (in actuality a list of the 64 most prominent sinusoidal components of a given timbre, created in deferred-time) and an incoming signal; in response to each component detected in the incoming sound-source, three possibilities exist:

- where the incoming signal contains a component which is present (or is within relative-proximity to a component, as determined in the 'margin-of-frequency-deviation' variable) in the instrumental-template but of lower intensity, the difference in intensity is calculated and applied to the electronic diffusion of that frequency;
- where the incoming signal contains a component which is absent in the instrumental-template, or vice-versa, nothing is diffused;
- where the incoming signal contains a component which is present (or is within relative-proximity to a component) in the instrumental-template but of greater intensity, nothing is diffused;

Thus, in the re-synthesis of the sound source used in the creation of the instrumental-template, a variable proportion is generated acoustically and is therefore subject to that source's aforementioned characteristic nuances of acoustic diffusion, attack, sustain, decay etc.

2.1 The creation of instrumental-templates

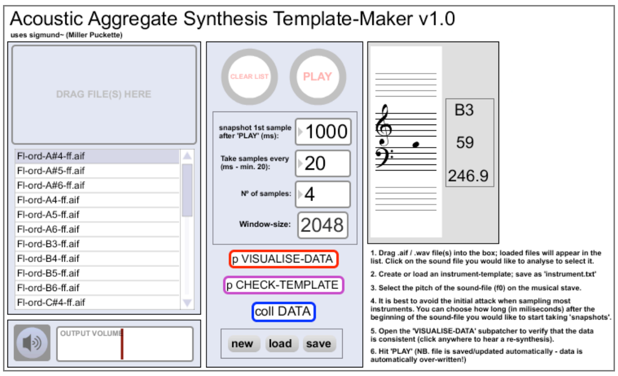

Figure 1. screenshot of AAS instrumental-template maker; using *sigmund~*, multiple 'snapshots' are taken at pre-defined intervals and compiled into a text file;

[1] *Speakings* utilises a system of real-time additive-synthesis (*i.e.* acoustic signal + synthetic components) in the creation of intelligible sequences of vocal formants. The techniques used in the realisation of this work are described in "Marking an Orchestra Speak" (2009), Harvey, Jonathan *et al.*

[2] The most commonly-found example is, for obvious reasons, the use of loud-speakers inside a piano; one such example is Alvin Curran's *Twentieth Century* (1994) for disk-klavier & electronics.

Instrumental templates are simply lists of frequencies & intensities organised according to MIDI pitch-values. I have already developed a number of instrumental-templates[1] which span the entire pitch *and* intensity range of common orchestral instruments. They may be implemented *statically*, *i.e.* taking just one pitch/dynamic from the template data and transposing it to match the *F*0 of the incoming signal, or *dynamically*, by approximating the *F*0 of the incoming signal to the nearest match in the template, then performing minor corrective transpositions (read: multiplications) of template frequency data. Also, multiple instances of a given pitch (2-32) may be created in order to avoid the sensation of timbral 'stasis'. An incoming signal with a progressive change in timbre, such as a note struck on a piano, or a bell (or indeed any other sound with a natural decay) may be used to great effect. It goes without saying that the comparative model will generate no output once the comparative model reaches zero-intensity. Sounds which change timbre sporadically (such as polyphonic textures or combinations of percussion instruments) have generally proven to be ineffective with 64-component analysis/re-synthesis.

2.2 User-interface

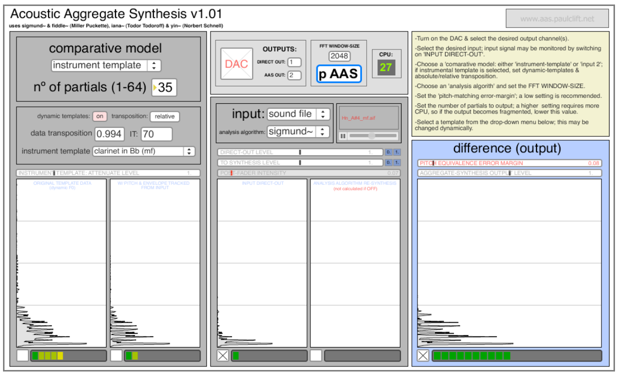

Figure 2. Screenshot of AAS interface; in this example, a template for **clarinet in Bb** is dynamically matched to an incoming signal of a **horn** playing Bb4.

2.2.1 Pitch/register matching

If the intention is to achieve a faithful reproduction of the timbre used to generate the instrumental-template, it is of course essential match it with an incoming signal whose fundamental is relatively close; obviously, attempting to reproduce a flute by means of a contrabass' low-E would generate a timbre which is denatured beyond recognition. Assuming the incoming signal has as its fundamental a frequency corresponding to a pitch which is present (or, as is most often the case, one within relative proximity to a pitch which is present) in the template, the process of dynamic-matching will ensure maximal preservation of the timbral characteristics of the acoustic model.

It may, of course, also be desirable to match a template with a sound source of contrasting register. In this case, a given template pitch may be selected and transposed dynamically to match the *F*0 of the incoming signal; the generated timbre will, of course, become progressively more denatured the further it deviates from its original pitch, which can be used to great effect in certain cases.

2.2.2 Intensity matching

Dynamic template changes based upon the intensity of the incoming signal are also possible. Thus, a flute *crescendo* may be modelled upon a corresponding *crescendo* performed by a clarinet, for example, with both dynamic *and* timbral evolution present in the resynthesised sound.

2.3 Summary of the process of synthesis

Below is a series of examples which illustrates the multi-step process in the generation of output:

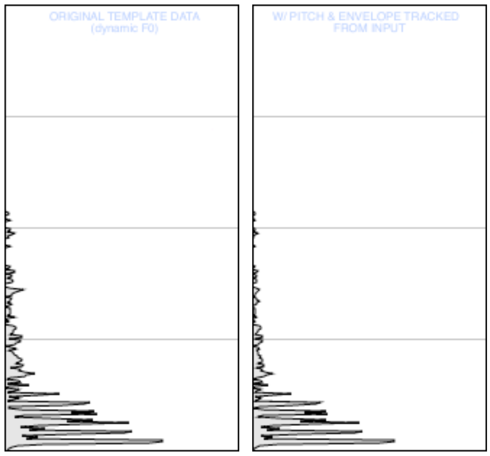

Figure 3. *spectrogram~* visualisations of: (left) a flute template (B^{b}4, *mf*), selected dynamically to best match pitch & intensity of the incoming signal; (right) the same template 'fine-tuned' (transposed & attenuated/ amplified) to match pitch/intensity of the incoming signal precisely.

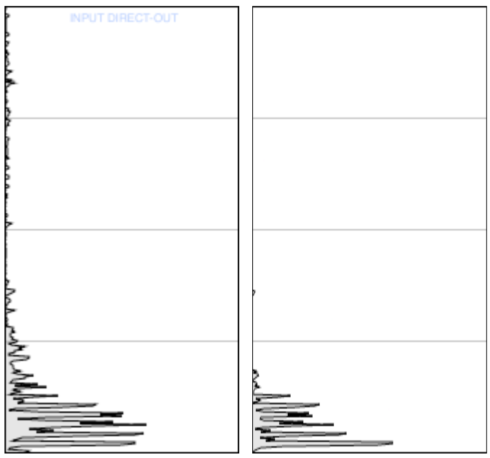

Figure 4. *spectrogram~* visualisations of: (left) analysis of incoming signal, an oboe (B^{b}4, *mf*); (right)

[1] flute, oboe, clarinet in B^{b}, bassoon, trumpet (*ordinario*, w/ harmon mute), horn (*ordinario*, stopped, brassy) & trombone (*ordinario*, w/ harmon mute) etc.

AAS output: the dynamically generated ‘difference’ between incoming signal & template data.

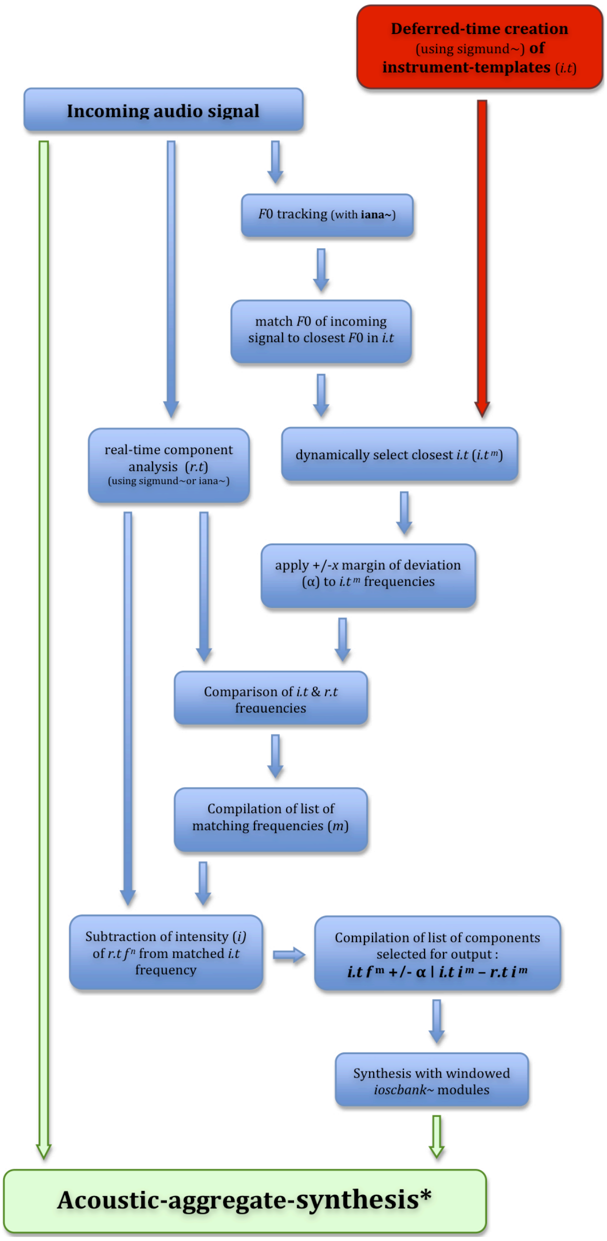

Figure 5. a simplified flow-chart summary of the process

2.4. Empirical results in controlled tests

The asterisk in the above example denotes an ambiguity with regards to the output: of course, an informed choice must be made with regards to the timbres of the incoming signal, the selected instrumental-template, and the registers of both. As such, I have compiled a comprehensive table of ‘efficacy’, describing combinations of instrumental-templates and incoming-signal sources. The results obtained did not always correspond to what one might expect.[1] A scale of increasing efficacy is used, in which combinations are rated from 1 to 5:

1. either *nothing* is diffused (*i.e.* all detected components from template are present in the incoming signal at a *greater* intensity) OR little to no fusion occurs (*i.e.* synthesis is not achieved: two distinct sound sources remain perceptible);
2. some fusion occurs but the result is not in any way evocative of the instrumental-template;
3. a high level of fusion occurs but the result is only minimally evocative of the instrumental-template;
4. complete fusion between sound-source & synthesis with a moderate to high degree of evocation of the instrumental-template;
5. a convincing reproduction of the secondary, acoustically-modelled timbre is achieved;

↓ INSTRUMENT TEMPLATE		INCOMING SIGNAL →	piccolo	flute (in C)	oboe	clarinet (Eb)	clarinet (Bb)	clarinet (bass)	bassoon	trumpet (in C)	horn
flute (in C)	*pp*			-	58: 4 70: 2 82: 3		50: 2 62: 3 74: 3		34: 1 46: 1 58: 2 70: 2	56: 1 68: 2 80: 3	43: 1 55: 1 67: 3
	mf			-	58: 5 70: 3 82: 2		50: 2 62: 4 74: 4		34: 1 46: 1 58: 2 70: 3	56: 2 68: 3 80: 3	43: 1 55: 2 67: 3
	ff			-	58: 4 70: 5 82: 4		50: 2 62: 4 74: 4		34: 1 46: 3 58: 3 70: 3	56: 3 68: 3 80: 3	43: 2 55: 3 67: 5
flute (in C) harmonics	*p*			59: 3 71: 4 83: 4	58: 1 70: 4 82: 5		74: 3		70: 1	68: 1 80: 3	67: 3
	f			59: 3 71: 3 83: 4	58: 1 70: 2 82: 3		74: 3		70: 3	68: 1 80: 3	67: 3
oboe	*pp*			59: 2 71: 2 83: 4	-		50: 2 62: 3 74: 2		34: 2 46: 2 58: 3 70: 4	56: 3 68: 2 80: 2	43: 3 55: 3 67: 3
	mf			59: 2 71: 3 83: 4	-		50: 2 62: 4 74: 3		34: 2 46: 2 58: 3 70: 3	56: 3 68: 3 80: 3	43: 3 55: 3 67: 3
	ff			59: 2 71: 2 83: 2	-		50: 3 62: 5 74: 3		34: 3 46: 3 58: 4 70: 3	56: 3 68: 4 80: 4	43: 3 55: 4 67: 4
clarinet (Bb)	*pp*			59: 2 71: 3 83: 4	58: 2 70: 2 82: 3		-		34: 1 46: 1 58: 1 70: 1	56: 2 68: 1 80: 2	31: 1 43: 1 55: 1 67: 2
	mf			59: 3 71: 3 83: 2	58: 2 70: 2 82: 3		-		34: 2 46: 2 58: 2 70: 3	56: 4 68: 3 80: 3	31: 1 43: 2 55: 3 67: 3
	ff			59: 3 71: 2 83: 1	58: 2 70: 3 82: 3		-		34: 4 46: 2 58: 3 70: 3	56: 3 68: 3 80: 3	31: 1 43: 3 55: 2 67: 2

Figure 6. An extract of a table of observed ‘efficacy’; *y* axis lists instrumental-templates, *x* lists instruments used as incoming signal; efficacy is measured at varying registers (given here in MIDI pitch values) & dynamics.

[1] A comprehensive discussion of this is beyond the scope of this short article, but will be included in future writing on this procedure.

2.4.1 *Choice of timbres*

It should be obvious at this point that it is desirable to emulate a sound which is of ‘greater’ timbral-complexity than that of the incoming signal. For example, generally speaking a flute template functions effectively with a clarinet as sound-source, but less so with an oboe. Nonetheless, we find that anomalies and exceptions are commonplace. For example, the template for a flute playing *fortissimo* in the ‘third’ register (D6-A7) is effective when used with an oboe *mezzo-piano* in the same register; the flute template contains multiple components which are stronger than the corresponding components in the incoming oboe. One may complement any timbre with any other provided that there is some degree of timbral inequity. As mentioned, there will be further discussion of this principle in subsequent, more comprehensive articles.

2.5. Diffusion

As mentioned earlier, in order that the resulting aggregate timbre be perceived as such (and not a mere juxtaposition of acoustic sound-source and electronic synthesis), it is beneficial to place the loud-speaker diffusing the aggregate-synthesis in very close proximity to the acoustic sound source.

2.5.1 *Incorporation of loud-speaker into the instrument*

Notably with the bass clarinet and all members of the saxophone family (with the exception of the *soprano*), it is possible to incorporate the loud-speaker & microphone *into* the instrument itself.

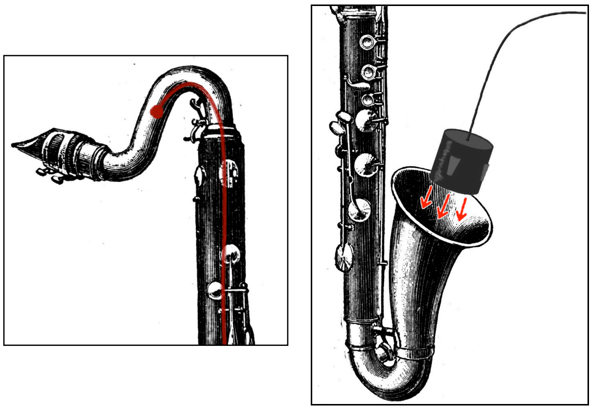

Figure 7. (left): A capsule microphone & small cable (represented here in red), may operate *inside* the bass clarinet without affecting tone or passage of air; the cable runs the entire length of the instrument and comes out of the bell; (right): a loudspeaker may, with minimal preparation, be inserted into the bell, as one would insert a mute into the bell of a brass instrument. This only compromises the *lowest* fingered-pitch perceptually; effect upon other pitches is negligible.

With a loudspeaker directed *into* the instrument but at maximal distance from the microphone, problems of feedback are minimised (the great majority of output from the loud-speaker is diffused through holes in the instrument’s tube and does not carry through to the embouchure). The advantages of this approach are considerable: homogeneity (through the sharing of acoustic mechanisms of amplification & diffusion associated with the instrument) is maximised, thus contributing significantly to an effective aggregate-synthesis.

Similar possibilities exist with string instruments (in particular, cello & contrabass) whereby a transducer is attached to the rear of the instrument and a bridge-microphone is used.

2.5.2 *Use of loud-speaker in proximity*

In the case of instruments for which the integration of a loud-speaker is impractical, such as flute or violin, a speaker placed directly in front of the performer at the vertical acme of the instrument’s average-maximum-acoustic-diffusion has proven effective in achieving a high level of fusion.

3. ACKNOWLEDGEMENTS

I would like to thank Miller Puckette, René Caussé, Jean-Baptiste Barrière, Gérard Bertrand & Alain Terrier for their ongoing assistance in the realisation of this project.

4. SELECTED BIBLIOGRAPHY

Bresson, Jean, & Carlos Agon. “SDIF Sound Description Data Representation and Manipulation in Computer Assisted Composition.” In Proceedings of the International Computer Music Conference. Miami, FL: ICMC, 2004, 520–527.

Fitz, Kelly, & Lippopld Haken. “On the Use of Time-Frequency Reassignment in Additive Sound Modeling.” Journal of the Audio Engineering Society 50, 11: (2002) 879–892.

Grey, John M. “Multidimensional perceptual scaling of musical timbres.” Journal of the Acoustical Society of America 61, 5: (1977) 1270–1277.

Klingbeil, M. K. “Spectral Analysis, Editing, and Resynthesis: Methods and Applications”. Ph.D. dissertation, Columbia University, 2009.

Lerdahl, Fred. “Timbral Hierarchies.” Contemporary Music Review 1, 2: (1987) 135–160.

Lyon, Eric. “Spectral Tuning.” In Proceedings of the International Computer Music Conference. Miami, FL: ICMC, 2004, 375–377.

Nouno, Gilbert, Arshia Cont, Grégoire Carpentier & Jonathan Harvey. “Marking an Orchestra Speak” (2009); retrieved May 1, 2012: http://articles.ircam.fr/textes/Nouno09a/index.pdf

Pampin, Juan. “ATS: A System for Sound Analysis Transformation and Synthesis Based on a Sinusoidal plus Crtitical-Band Noise Model and Psychoacoustics.” In Proceedings of the International Computer Music Conference. Miami, FL: ICMC, 2004, 402–405.

Smith, Julius O. “Virtual Acoustic Musical Instruments: Review and Update.” Journal of New Music Research 33, 3: (2004) 283–304.