PROJECT MUSE®

Nick Collins
Department of Informatics,
University of Sussex
Falmer, Brighton, BN1 9QJ, UK
N.Collins@sussex.ac.uk

Automatic Composition of Electroacoustic Art Music Utilizing Machine Listening

Abstract: This article presents Autocousmatic, an algorithmic system that creates electroacoustic art music using machine-listening processes within the design cycle. After surveying previous projects in automated mixing and algorithmic composition, the design and implementation of the current system is outlined. An iterative, automatic effects processing system is coupled to machine-listening components, including the assessment of the "worthiness" of intermediate files to continue to a final mixing stage. Generation of the formal structure of output pieces utilizes models derived from a small corpus of exemplar electroacoustic music, and a dynamic time-warping similarity-measure technique drawn from music information retrieval is employed to decide between candidate final mixes. Evaluation of Autocousmatic has involved three main components: the entry of its output works into composition competitions, the public release of the software with an associated questionnaire and sound examples on SoundCloud, and direct feedback from three highly experienced electroacoustic composers. The article concludes with a discussion of the current status of the system, with regards to ideas from the computational creativity literature, among other sources, and suggestions for future work that may advance the compositional ability of the system beyond its current level and towards human-like expertise.

A great challenge to the automated production of musical works is the critical role of the human auditory system within the design cycle. Human compositional activity over a musical form provides for continual feedback at multiple timescales, from selecting and refining momentary material, to the control of flow between sections and across the whole work (Roads 1985; Eaglestone et al. 2008). Algorithmic composition has not extensively engaged with problems in the audition of the generated music, including the psychology of musical form (Collins 2009). Previous algorithmic composition "critics" (software modules that evaluate the output of compositional algorithms) have operated mainly in a pre-established, symbolic domain—for example, as fitness functions in a genetic algorithm search (Miranda and Biles 2007). As machine-listening technology advances, however, there are great opportunities to build artificial listening capabilities into algorithmic works that assess their own intermediate and final outputs in a manner analogous to a human composer.

Although full equivalence of human and machine-listening capabilities remains out of reach at present, much research advancement has occurred in this domain in recent years (Klapuri and Davy 2006), and continues to be pushed particularly through music information retrieval (MIR) research (Casey et al. 2008). Indeed, machine listening applied directly to audio signals has become a strong feature of work in live interactive systems (Rowe 2001; Hsu 2005), and it is somewhat surprising that it has not taken place to the same degree within algorithmic production of fixed works.

This article outlines Autocousmatic, a project that seeks to incorporate computational audition into the algorithmic design cycle. The musical sphere selected for this work is that of electroacoustic art music intended for acousmatic presentation; that of fixed, intensive, spectromorphological tape pieces for spatialized diffusion in concert. Timbral transformation tends to be a primary aspect of such works, and although there are no hard and fast rules on the weighting of rhythmic and pitched attributes relative to timbral ones, electroacoustic music tends to operate in a context of experimental art music where less "popular" musical parameters such as timbre or space receive greater investigation. Like all musical categories, exact definitions are controversial and mutable in any still-evolving cultural discourse (Landy 2007), but we shall assume that the readership of this journal is familiar with the landscape of such music, not least through concerts at such gatherings as the International Computer Music Conference (ICMC) or the Society for Electro-Acoustic Music in the United States (SEAMUS). Because intensive listening is so important to the practitioners and audiences for this music (Barrett 2007; Norman 2010), it provides a clear challenge

Computer Music Journal, 36:3, pp. 8–23, Fall 2012

for an algorithmic study incorporating machine listening.

Electroacoustic composers, inevitably in the current age, are computer-savvy, and there have been a number of attempts to render aspects of electroacoustic works automatically. Interesting precedents include Horacio Vaggione's exploration of automated mixing strategies for the work *Octuor* (Vaggione 1984). Larry Austin has built a new, computer version of John Cage's *Williams Mix* (1953), called *Williams [re]Mix[ed]* (1997–2001), that automatically generates new realizations according to Cage's score of chance operations and suggested source-sound families (Austin 2001).

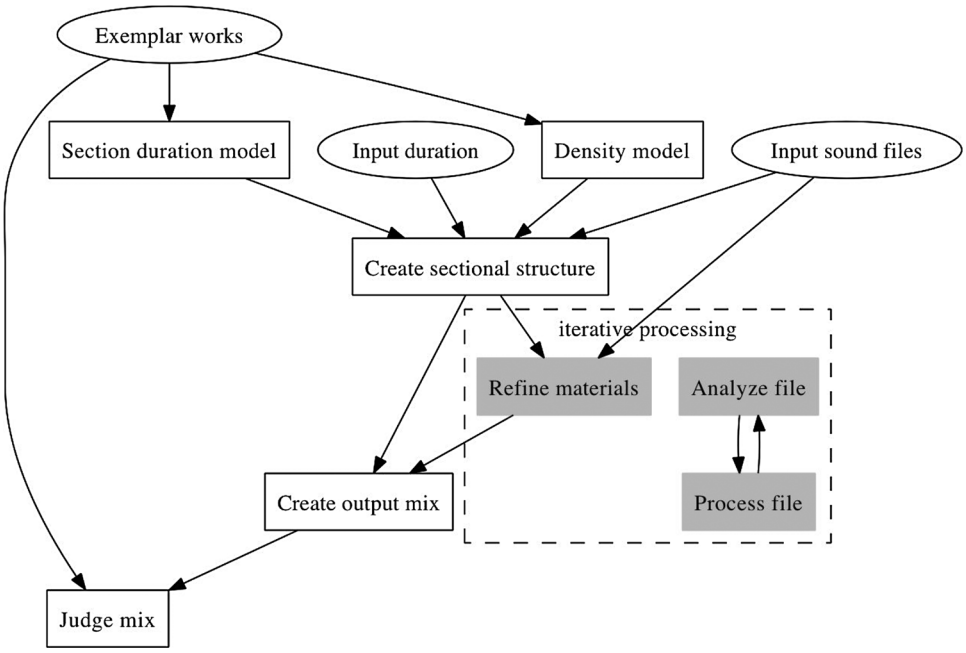
Automated sound-file manipulation has appeared in many computer music pieces and digital artworks. Csound composers are familiar with the algorithmic composition of event scores of rendering instructions, and similar investigations are possible in systems such as CMix, SuperCollider, or Max/MSP. As standalones built from Max/MSP, Karlheinz Essl's *REplay PLAYer* (2000–2007) is a granular-synthesis based texture generator operating on any starting sound file; and Leafcutter John's *Forester* series of programs (available at leafcutterjohn.com/?page_id = 14) work with any sound files fed to them. As a standalone program, *ThonK* (available for Mac Power PC computers before Mac OS X, www.audioease.com/Pages/Legacy) provided various preset texture-generation methods on input audio files. Net artists have also been interested in creating collages using automatic harvesting programs that draw from the vast quantities of audio materials available online. Peter Traub's *bits & pieces: a sonic installation for the World Wide Web* (1999) collects new materials each day from any online sound files it can gather, generating novel mixes every two hours.

The act of collage, often magnified to exhausting degrees, has been a fundamental part of certain fixed works. John Oswald's *Plexure* (1993) is a "plunderphonics" work formed of hundreds of copyright-breaching fragments, assembled by hard, manual labor with a sampler (Landy 2007). R. Luke Dubois's *Timelapse* (2006), meanwhile, automates rendering through spectral averages of 857 Billboard No. 1 chart hits. David Kirby's *maximalism* (2005) claims tens of thousands of snippets; and Johannes Kreidler's *product placements* (2008) deliberately places some 70,000 samples within a time span of about 30 seconds as a protest against German sample-registration policy. Most relevant, perhaps, to the feature-led decisions of the system described herein is concatenative synthesis, which empowers large-scale creation of new works on the model of old works, driven by feature data (Sturm 2006).

Nonetheless, these projects do not have a machine-listening component that analyzes the suitability of audio materials, before and after effects processing transformation, or that could judge a final mix in its entirety, independent of human intervention. Although automated mixing and mastering assistants are being developed (Pope and Kouznetsov 2004; Perez Gonzalez and Reiss 2007), they treat problems such as panning, spectral profile, and inter-track balance, rather than the full compositional process.

One close precedent to the work presented in this article, however, grew out of John ffitch's curation of the Door Project for ICMC 2001. A recording of a bathroom door from the ICMC 2000, in Berlin, was made available as the starting point, and composers submitted pieces that remixed the door recording. One sentence in the call for works read: "Anyone may enter, even if they did not hear the door in question" (ffitch 2001). An anonymous entry was submitted, titled *Even if they did not hear the Door,* devised by creating a software tool chain that generated the work from the original source recording without the composer's ever hearing the original door (an alternative door recording was used for some testing before the final run, and the software included filtering effects alongside basic automatic mixing, including root-mean-square [RMS] level checking). As well as the performance in Cuba and the project CD, this piece was subsequently played at a 2003 placard headphone festival in London (www.leplacard.org), and has probably had around 200 concert listeners. After a decade, the original composer (hereby outed as this article's author) has still not heard a single second of the finished piece, or the original door sample.

Figure 1. Main breakdown of processing flow in Autocousmatic.



We proceed by describing the design of the Autocousmatic program in detail. We then consider its evaluation through a number of means, including through entering created works into electroacoustic music competitions, the release of the software to composers, expert evaluation of outputs, and assessment in relation to the developing field of computational creativity. We conclude with some recommendations for future work in this field. If spoiling the denouement, it may relax some readers to know that electroacoustic composers are not likely to be immediately replaced by machines, but that this project provides an illuminating alternative perspective on the effectiveness of the repertoire.

System Design

The Autocousmatic program has gone through a number of iterations of development from 2009 to 2011. The version described here is the final, publicly released version from the summer of 2011. The public software missed out on one further component, however: a final critic that can choose among a number of candidate output mixes to find the "most suitable" mix. This component is described in a subsequent subsection.

The main principle of the software is to start from a directory of sound files that correspond to the basic materials of the electroacoustic composer. The composer specifies a desired total duration and number of channels for the output work, and goes for a cup of tea or coffee while the system churns away. Various iterations of processing are carried out on the materials, each time running a suite of machine-listening operations to find "useful" or "effective" parts for further processing. Awkward moments, particularly overloads and other digital nastiness, as well as silence and low activity, are filtered out. The form of the output piece is built up by mixing these processed parts, following models of sectional form and activity level derived from exemplars of existing electroacoustic music. Multiple output mixes are created. A further optional component is available to judge the output mixes for their overall flow, in comparison to some existing "inspiring" work.

Figure 1 shows a processing diagram for the main flow of calculation in Autocousmatic.

The exemplar works were fixed for this project, and the database consisted of around half an hour of electroacoustic material, namely, works by Trevor Wishart (*Vox 5* [1986; 6 min 6 sec]), Stéphane Roy (*Trois petites histoires concrètes: Micro-confidences* [1998; 3 min 50 sec]), Bernard Parmegiani (movements 4 and 10 from *De Natura Sonorum* [1975; total duration 4 min 54 sec]), and Denis Smalley (*Tides-Sea Flight* [1984; 12 min 47 sec]). Although a larger database could have been gathered, the human annotation of formal sections in these works demanded a manageable size (extensions to larger databases will be discussed later in the article). The database of works was used to create models of section duration and within-section density to guide algorithmic composition. An individual work from the database (Wishart's) was also used as a "gold standard" for automatic judging of the best from a set of candidate algorithmic mixes, with respect to a similarity measure. These processes are discussed subsequently in more detail.

In Autocousmatic, sections are created with different guiding parameters, including such aspects as the overall density of events, and the abruptness of transitions into and out of the section. Each section is assigned one to four source files as its primary materials. When two or more source files are available in a given section, there is the potential for cross-synthesis of materials in the processing stage.

The processing stage generates a certain number of processed sound files per section, based on the overall density. This module has a bottom–up iterative construction: Source files are analyzed, then "interesting" parts are processed (see the subsequent discussion), then the processed files become the starting point for the next iteration. Each output file is the result of taking a portion of an input file (or more than one segment from one or more files, in the case of cross-synthesis effects) and applying effects from a suite of available processes, for up to five iterations. After each process creates an intermediate file, this file is analyzed by a critic process to determine its eligibility to continue processing, or, after the final iteration, to add to the list of processed files for that section. This framework is reminiscent of "generate and test" techniques, as classically applied in pioneering 1950s algorithmic composition (Hiller and Isaacson 1959).

The audio-analysis components regulating the selection of appropriate segments of sound files, and the rejection of unsatisfactory distorted results, are based on using percussive onset detection (Stowell and Plumbley 2007) and perceptual loudness, as well as on RMS and absolute-peak amplitude. Features are extracted in windows of size 1,024 samples, with a 512-sample hop size, at a 44,100-Hz sampling rate for operations. Post-processing after feature extraction attempts to find viable segments of a given sound file, and to exclude sustained areas of silence or distortion. Tests on sound files include a check for clipping: No more than N floating point values in a row can occur, where all values are greater than 1.0, or alternatively all less than -1.0 ($N = 10$ is the default). Any given sound file is broken down into viable and silent regions. Silences are marked by 10 or more feature frames in a row with low perceptual loudness and low RMS. Non-silent regions are turned into segments between detected onsets; so segments either have a duration of an inter-onset interval, or last from an onset to the position of the next detected silent region. These derived segments are also excluded if their peak amplitude is too low, or if they are too short (less than 100 msec long), but they can otherwise be the basis of further processing.

There are one to five iterations of processing, each weighted by a probability that the processing will not stop before that iteration: 71 percent for iteration 1, 16 percent for iteration 2, 11 percent for iteration 3, and 1 percent each for iteration 4 or 5. These weightings were chosen so that dense iterations of processing would be rare, because two or three different chained effects quickly leads away from the source sound. Available effects processing touches on many classic sound transformations associated with electroacoustic composition work, including granulation, delays, filters, time stretching, and various forms of cross-synthesis (via ring modulation, linear-predictive coding, vocoding by an amplitude-envelope tracking filter bank, and phase-vocoder spectral operations, like phase substitution and magnitude multiplication). SuperCollider's

ability to dynamically specify unit generator networks (via SynthDef construction) is used, so that whole effects units and the control parameters of their constituent UGens are created algorithmically, afresh for every new sound-processing occasion. All synthesis processes can cope with multichannel input files, and can produce as many channels of output as are currently available; spatial movements of various forms are built in, from static position to active movement of all channels independently, conditional on available channels. A new sound file is written for each iteration; floating-point-bit resolution is used to retain dynamic range. Intermediate and final processed files are not normalized; this would contradict the checks for healthy levels and, instead, amplitude compensation is applied at the final mix stage.

Where the iterative effects processing loop on smaller scale materials could be seen as bottom–up, the formal design is a more top–down construction. The imposed formal structure is a kind of "material interaction form"; sound materials, assigned to groups (denoted here by A, B, . . .), are then selected for each section; their co-occurrence can be in the mix, or also can be instantiated in cross-synthesis in the iterated processing. So, a form derived in this manner might consist of sections that are monothematic (e.g., assigned to only A), or have two or more subjects (e.g., AB, ABC, BCDF). A density envelope (described next) controls the activity level through the section, with materials having a greater probability of assignment when higher activity indicates thicker layers. As already indicated, each section may have a tightly synchronized start and/or end, to give greater drama to section transitions; sound files can also be laid down end to end, or with different probabilistic gap models. Checks are in place to watch for situations where a sound file may be too long for the available space (in which case a subsegment of the file is taken, if at all possible).

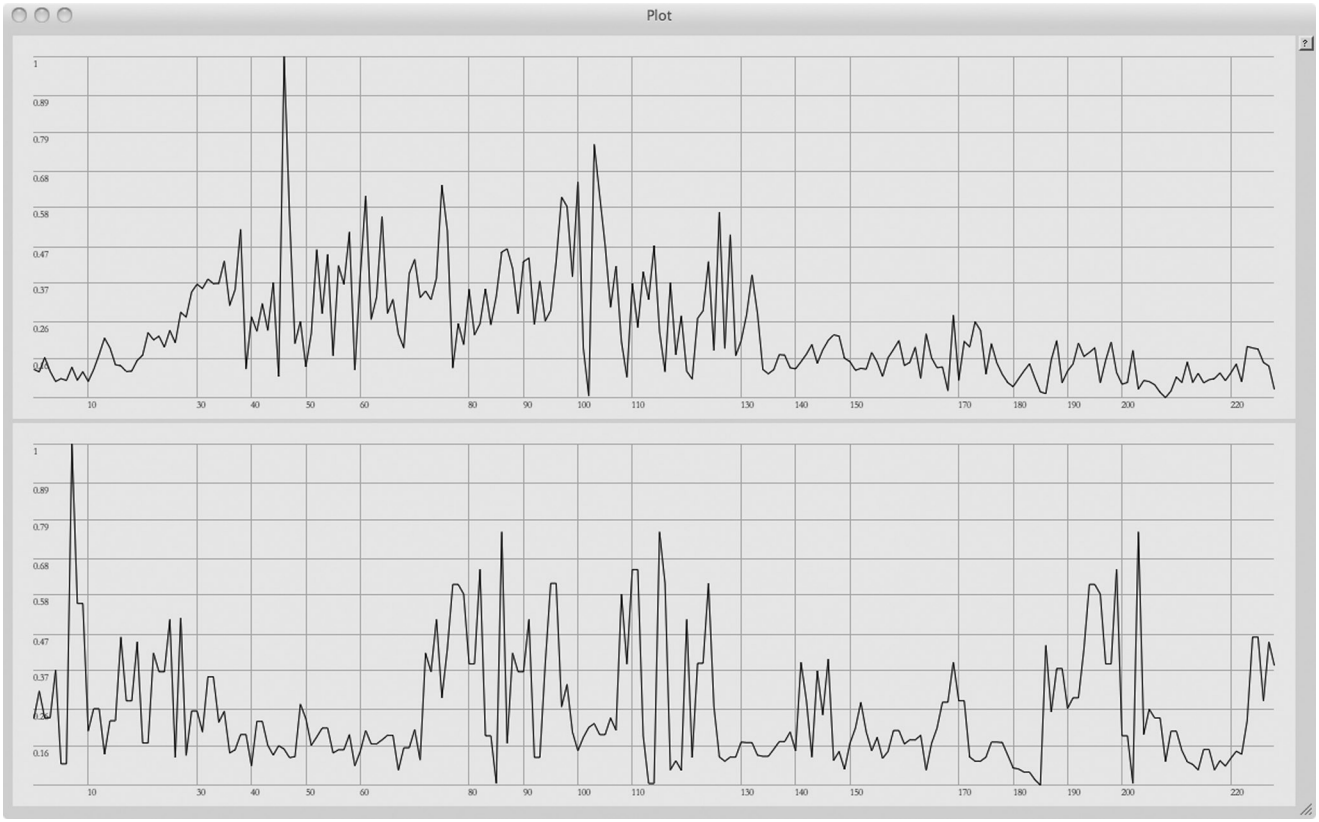
The density model was derived from the exemplar works by extracting four features assumed to be highly correlated with overall activity level, and combining them to make one derived feature. The four intermediate features were obtained from a model of perceptual loudness, a sensory dissonance measure (posited as revealing thicker textures and increased tension, following the spectral-peak pairing comparison method in Sethares [1998]), and two onset-detection functions (pre-peak picking signals, from the complex-domain onset detector and modified Kullback–Leibler divergence described in Bello et al. [2005]).

These four values were extracted with a windowed-analysis process (window 2,048 samples; hop 1,024; sampling rate 44,100 Hz), averaged over time within one-sec segments, and summed up with equal weighting, to obtain single values per second of source data. Although averages could have been taken within whole sections (knowing the annotated section boundaries for the exemplar works), taking values every second gave a much richer time-varying data set.

The model was made generative by treating it as a database of trigrams, that is, working over three consecutive density values in a row corresponding to the last three seconds, a reasonable time scale of the perceptual present. New density curves were created one triple at a time, outputting the first two values from each triple. To get the next triple, a nearest-neighbor method compares the last value of the current triple to the first of all those in the database, selecting the closest in value (to avoid loops, there is a 30 percent probability of a free choice from the database); the process iterates. Figure 2 plots two example density curves. The top curve is from Stéphane Roy's *Trois petites histoires concrètes: Micro-confidences,* with one density value per second. The lower plot was generated with the model. The local nature of the matching model is evident, and larger-scale structures of the original are not emulated closely; nevertheless, there is some relation in pattern, as seen from the use of common *y*-axis points, and some alternation of low bubbling and more tower-like, choppy signal values. A more sophisticated time-series modeling technique could be introduced in the future to tighten the relation, though some use of chance operations in the generating process is meant to avoid over-literal replication of the training set.

The model for duration was created in a similar way. Section durations were manually annotated rather than automatically extracted, to avoid problems with section boundaries which did not

Figure 2. Original Roy density curve (top), and generated density curve (bottom).



correlate well with the sort of large-scale energy change that an MIR section-boundary detector might look for (this lack of correlation was typically due to some overlap of layers). The final model again used trigrams, where the closest match was searched out from the last output value of the previous trigram to the first of any trigram in the database. A randomly generated run of 10 section lengths looks like the following:

21.966, 7.844, 59.558, 63.834, 39.692, 73.013,
80.235, 15.714, 24.483, 25.791

Similar-length sections often occur next to each other, due to the construction method. A more developed model of section duration transitions would follow from a much wider annotation project over existing works; on the other hand, the formal sectional lengths annotated in Wishart's *Vox 5* were

77.093, 16.886, 24.483, 15.53, 25.419,
40.361, 32.663, 25.125

which themselves involve some closely related section lengths. A relation between the *Vox 5* durations and those generated here is apparent, in particular in the 77.093, 16.886, 24.483 which starts the Wishart sequence, and in the 80.235, 15.714, 24.483 later on in the generated sequence (to avoid literal repetition of originating data, some small variation, of up to a few seconds in value, is a part of the duration model).

Choosing the "Best" Mix Out of Several Candidates

For full automation, a mechanism was developed to choose the "best" mix from multiple candidate

mixes for a work. Candidate mixes differ in their particular arrangement of processed sound files, following distinct density curves. The choice procedure acts with respect to an "exemplar" work whose transitions in feature space are taken as worthy of emulation. We typically used Trevor Wishart's *Vox 5* for this, not least for the variety of successful textures within it, but also for the esteemed transition behavior over time.

The machine-listening critic operates by considering the mix to be rated in terms of non-overlapping, ten-sec segments. Each segment is compared to all ten-sec segments in the guide piece, and the best match sought by the dynamic time warping (DTW) distance on the derived feature vectors (Jehan 2005; Casey et al. 2008). The DTW calculation is restricted to a leeway of up to six feature-vector points either side of the $x = y$ diagonal path, that is, a central diagonal strip through the similarity matrix of maximum extent, at most 0.14 seconds off the diagonal in time displacement. Typical features consisted of psychoacoustic loudness, sensory dissonance, onset-detection functions, various spectral descriptors like the spectral centroid and spectral falloff, and timbral information from Mel-frequency cepstral coefficients. Pseudocode for the procedure is provided in Figure 3. The rationale is that proximity of timbral-textural-activity behavior to *Vox 5* is a worthy aim. Even if this final listening model is based on one piece rather than a larger corpus, it is a fine exemplar work, and, in principle, the technique discussed here could be extended to a larger-scale model, as discussed later in the paper. The search is across all segments to avoid imposing the time structure of *Vox 5* on the generated works, but to respect its timbral character and local flow as a guide to effective electroacoustic composition. Although this procedure could be critiqued as inadequate to match human listening, it does at least attempt to respect local feature time series by using the dynamic time-warping comparison rather than summary features.

Figure 3. Pseudocode for finding the best mix from a set of candidates, with respect to an exemplar work.

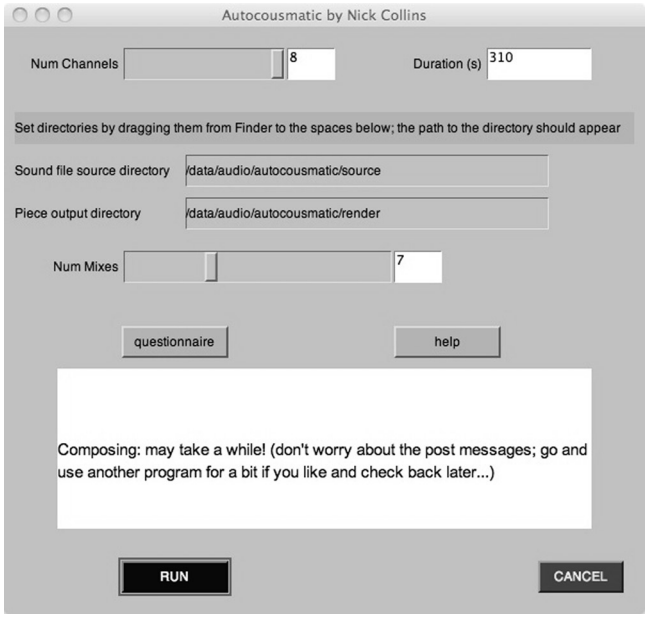
```
findbestmix(exemplarwork, mixes)
     segmentsize =  integer(10 seconds * feature vectors per \
     second)
     e = extractfeatures(exemplar work)
     bestmixscore = infinity
     for each candidate mix M:
          c = extractfeatures(M)
          mixscore = 0
          for each segment s1 of length segmentsize in c
               bestmatchscore = infinity
               for each segment s2 of length segmentsize in e
                    distance = dynamic time warp distance in \
       feature similarity matrix from s1 to s2
               if (distance < bestmatchscore)
                    bestmatchscore = distance
               mixscore = mixscore + bestmatchscore

          if (mixscore < bestmixscore)
               bestmixscore = mixscore
               bestmix = M
     return bestmix
```

Implementation

Implementation of Autocousmatic was achieved with the SuperCollider (SC) 3 audio programming language (McCartney 2002; Wilson, Cottle, and Collins 2011) making extensive use of the scsynth synthesis engine's non-real-time rendering mode. Indeed, scsynth is called upon thousands of times in a single run of Autocousmatic, to run machine-listening UGens to examine sound files, as well as to run effects processing and for the larger-scale construction of final mixes. The SuperCollider Music Information Retrieval (SCMIR) library (Collins 2011) was also used for additional machine-listening capabilities, such as for deriving the feature-based model data, and for comparison of each candidate mix to the template work.

A few SC-specific technical problems are worth briefly noting. Machine analysis is carried out using the synthesis-server application scsynth, but returning the analysis data to the separate SuperCollider language in non-real-time mode can be awkward. Special UGens are used for this purpose, such as Dan Stowell's Logger (available as part of the MCLDUGens from

Figure 4. The user interface of the Autocousmatic standalone application.



http://sourceforge.net/projects/sc3-plugins/). Logger operates via an auxiliary buffer, saving data (even though not audio) to disk as a WAV file at the end of a run; this can then be loaded from the SuperCollider language via the SoundFile class, and the floating point values interpreted as the data. SCMIR provides a separate FeatureSave UGen which works by directly saving ASCII files. The language's main thread has to wait, blocked on the running of the separate scsynth application as a Unix process. The number of buffers on the server had to be increased to cope with long multichannel mixes, in which there were easily more than 1,024 separate sound files in operation. Buffer needs could not be easily managed through dynamically freeing buffers, given the sheer complexity of the usage scenario.

Rendering times vary with density parameters and with the number of processing iterations allowed, but are generally slower than real time for default settings. For example, on a MacBook Pro (2.8 GHz Intel Core 2 Duo, 4 GB RAM, SuperCollider 3.4.3) a two-channel, 52-sec work took 6 min 44 sec to render; an eight-channel, 2-min 30-sec work took 32 min 6 sec. Because of the probabilistic nature of the section densities in particular, there is not a fully predictable linear relationship, though something like two to four times as slow as real time would be expected.

Using Autocousmatic directly from the SuperCollider language depends on setting up directory paths for source materials, intermediate processed files, and final rendered output (default locations, such as the /tmp directory for the intermediate files, can be used). The two main calls, explicitly naming the arguments in SuperCollider code, are:

```
a = Autocousmatic(numChannels:8,
    duration:150.0);
a.go(densityfactor:20, nummixes: 10);
```

This client code demonstrates the simplicity of putting Autocousmatic to work, and the high degree of automation embedded in it. The full source code for Autocousmatic is available under the GNU General Public License (GPL) 3 at www.sussex.ac.uk/Users/nc81/autocousmatic.html.

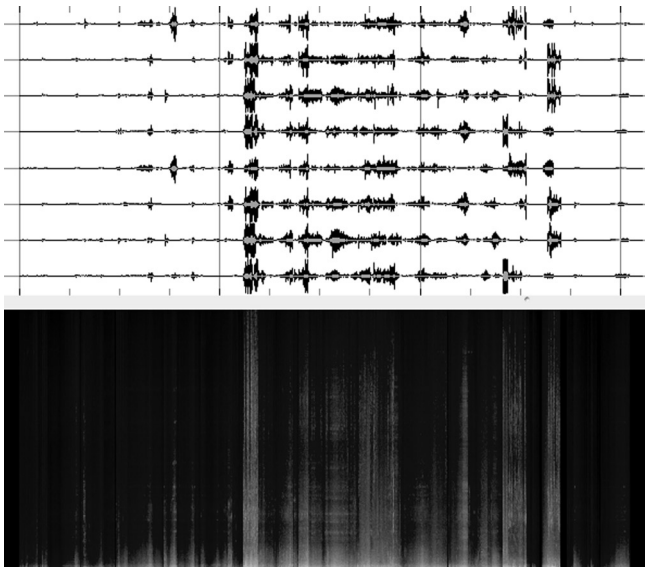
For potential users who may be put off even by a few lines of SC code, the software is also available at the same Web address as a Mac OS X standalone, and its release was used to obtain some of the user feedback which is discussed subsequently. Figure 4 shows the relatively minimal user interface of the released standalone, which was built with SuperCollider but requires no specialist knowledge of SC code to operate.

Figure 5 shows a 3-min, eight-channel work, *Galeic wit*, created with the software, taking as its source material clattering sailboats tied up on a beach in high winds. Different section densities and spatial differentiation of parts are clearly apparent in the time-domain plots.

Evaluation

Evaluation of algorithmic composition programs is critical, in order to avoid the sort of malaise described by Pearce, Meredith, and Wiggins (2002); we certainly do not want to claim without rigor that "the outputs of the system sound good," but instead, to investigate more thoroughly, through expert

Figure 5. Eight channels of time-domain waveforms (top) plus the spectrogram (with 0–22,050 Hz plotted) of all channels summed (bottom). (Figure created in SonicVisualiser.)



opinion other than the designer's, the effectiveness of the generated compositions. Most assessment is carried out with respect to system outputs, rather than the whole system itself; however, the release of the standalone version of the Autocousmatic program in the summer of 2011 was an attempt to get feedback on the system as a generative tool itself, beyond the creator's own assumptions, and we will further discuss the examination of both process and product in the context of computational creativity research. Thus, we seek critical feedback on system design to inform future revisions of the project.

Three main evaluation strands were explored: (1) the submission of Autocousmatic outputs to conference and festival calls; (2) the release of the software as code and standalone, and a set of example outputs on SoundCloud, to a general audience of electronic musicians; and (3) the direct solicitation of constructive critiques on the example outputs from three highly experienced, established electroacoustic composers. We discuss the status of the project further after treating these three strands in turn.

Submission of Outputs

Though clandestine schemes to get Autocousmatic's generated works into competitions and festivals might seem to be attractive publicity stunts, they proved less than informative. It is not just that the Autocousmatic project could not realistically hope to be immediately performing at graduate-composer level or beyond (much as a few comments to be discussed below are encouraging), but that even accepted compositions often receive little feedback for their composers at these sorts of events.

Nonetheless, to test the waters for this sort of evaluation method, as it were, Autocousmatic outputs were entered in response to a number of peer-reviewed calls (one unique piece per competition). To anticipate any worries of an electroacoustic composer: No Autocousmatic-generated piece has yet been successful in such a call. When pieces were submitted, they were never "cherry picked"; that is, the author was not allowed to listen to the pieces in any way before submitting them, but had to trust a rendering run (only pieces of this nature are in the SoundCloud public examples to be discussed subsequently). A pseudonym was used to avoid any identity bias (though an unfamiliar name may lead to a different sort of bias).

The one form of feedback provided by the festivals was the rejection count (sometimes a little approximate; for example, the NoiseFloor festival stated that there were 100+ entries, and ICMC 2010 that there were around 1,700, and NYCEMF only provided the acceptance ratio and no other information). Multiplying together rejection ratios allows an overall probability of rejection for a system to be calculated, assuming that juries use random selection independent of any perceived quality factor or other bias. This is not the most reasonable assumption given human judges, but it does give a baseline for consideration.

Table 1 groups data from seven competitions. For the first five competitions listed, a prototype version of the system, earlier than the system described herein, was used. The prototype had a less developed model of form—it lacked underlying section length, section transition, or activity level models, and files were packed into each section of the mix using a simple probabilistic spacing decision. Its overall probability of random rejection over the five competition entries was 0.37. Version 1.0 of Autocousmatic, the main version described

Table 1. Rejection Data from Seven Competition Works Generated by an Autocousmatic Prototype and Version 1.0

Version used to generate entry	*Festival*	*Number of submissions*	*Number of acceptances*	*Rejection ratio*
Prototype	Miniaturas Confluencias 2010	211	13	0.94
Prototype	NYCEMF 2010	685	137	0.80
Prototype	Noise Floor 2010	100	40	0.60
Prototype	ICMC 2010	1700	200	0.88
Prototype	Seoul International Computer Music Festival 2010	200	15	0.93
Version 1.0	ICMC 2011	854	138	0.84
Version 1.0	60 × 60 2011 (main competition, no resubmission for local versions)	800	60	0.93

herein, was also entered into two competitions in 2011; its (unbiased) rejection probability was 0.78. The combined random rejection probability across both system versions over all submissions was 0.29, much larger than the 0.05 normally required for significance, if we could only trust juries to behave in a random manner.

Unfortunately, across all competitions entered, none gave feedback on the judges' reasons for rejection. Other evaluation methods were critical to obtaining more qualitative insight into the system and its outputs beyond bare rejection data. Despite the probability of random rejection, it is clear that assuming that at least some festival juries would pick up on higher-quality pieces, the system is not performing at a level competitive with human experts. A postgraduate or professional composer of the kind who would normally submit to such events with success would be creating more convincing electroacoustic music than does Autocousmatic.

Users of the Autocousmatic Program

The Autocousmatic program was released in August 2011, with announcements on a number of mailing lists, namely, the SuperCollider users, Canadian Electroacoustic Community, and UK Sonic Arts Network lists, as well as social posts on Facebook and Google+. Alongside the standalone application and the source code, eight example pieces in stereo or stereo mixdowns were released on a SoundCloud page for Autocousmatic (soundcloud.com/autocousmatic). It was made clear that feedback would be welcomed, and an optional questionnaire was built into the standalone which could be emailed back to the author if a user was kindly disposed to do so.

At the time of writing, some SoundCloud tracks have received well over 100 plays, which is not a huge amount, but equivalent or better than much electroacoustic-tagged music on SoundCloud (as revealed by searching for the tag) and much larger than the unpublicized two tracks of electroacoustic music released on a separate page for Autocousmatic's competition-entering pseudonym (where the number of listens over the two tracks are five and seven, respectively). Tellingly, none of Autocousmatic's tracks have been "liked" or publicly commented on (perhaps due to concerns about being the first to do so), though 13 individuals are intrigued enough to have registered to follow Autocousmatic. The SoundCloud tracks are further examined in terms of expert assessment subsequently.

More specific feedback arose from users of the publicly released software. One immediate "gotcha" was that a few users instantly pushed the software to its limits, pointing it to a directory of hundreds of megabytes of sample data as the source, and asking for a 20-min eight-channel piece. This somewhat overloaded the software; one brave soul let the software try to render for two days before being advised against this. A few other users

were not familiar with the (SuperCollider) textbox convention of pressing enter once a number was typed, and found that they could not change the desired output duration. These warnings show that even creating a standalone application can include residual conventions of the originating environment, and that electroacoustic composers will immediately push on large-scale tests of form.

A short qualitative questionnaire was enclosed with the software; it probed a respondent's background in electroacoustic music, how much they had used the software already, their feelings on the quality of output, and its relation to any other projects. Although only a small number of respondents (five) directly filled in the questionnaire, others responded with emails with their own observations (some pertaining to technical problems). Respondents tempered some enthusiasm for the software and its principles by noting perceived shortcomings of the results.

"I love the idea of it," one questionnaire from an experienced composer was kind enough to say, yet qualified this with practical problems observed over three renderings: "its box of tricks just needs to be larger, or more generously spaced out, it tries too many different ideas in a short space of time." They noted that a run on a unified long source (a piano recording) had come out much more effectively than when acting in other runs on a folder of short sounds. Although "it's effective in that it uses many of the processing tropes associated with this music, and often comes up with interesting textures and combinations," the program had problems with structure, enveloping, and some overly literal transformations: "It lacks a sense of agency as the gestures often seem arbitrarily placed."

Other comments highlighted the system as an assistant more than an autonomous composer in its own right: "The program could be a valuable tool in an exploratory phase of composition," where the same kinds of processing choices and densities could even be manually reconstructed, if inspiring. If "the elements are effective, fulfill expectations," the same questionnaire revealed that "the results lack the buff and shine I would expect a human musician to give his or her work." The users were not shy in pointing out differences in quality level with respect to their experience of polished human compositional work, with enveloping, overall structural decisions, a variety of processing, and "buff and shine" being the main criticisms. The machine listening in Autocousmatic failed to convince some respondents. For instance, one saw the software as primarily "selecting individual sounds and applying effects to them."

Intriguingly, further evidence of the use of Autocousmatic has begun to crop up independently on SoundCloud. These provide some insight into behavior of the software in the wild; for instance, Serge Stinckwich's *Hanoi Soundscape Autocousmatic* (soundcloud.com/serge-stinckwich/hanoi-soundscape-autocousmatic) definitely highlights problems with abrupt enveloping when operating on sparse materials.

Expert Evaluation of Outputs

Three expert judges, all established composers with extensive experience of electroacoustic music, including having had their own pieces performed at peer-reviewed festivals, were kind enough to give critiques of the SoundCloud-published output pieces from the Autocousmatic program. They were instructed to comment as if they were providing feedback to a developing composer, and responses are here anonymized (one composer asked for their responses to be "autononymous"). The composers were highly critical, but when they discussed shortcomings, they presented strong arguments as to the perceived deficiencies of Autocousmatic's compositions. Although there might be a possible bias against machine composition as opposed to direct human studio craft (all composers were aware of the provenance of Autocousmatic as a software program), none of the composers seemed overtly disposed against the idea of an algorithmic composer per se. Further, exactingly high professional compositional standards remain an essential benchmark in evaluating Autocousmatic.

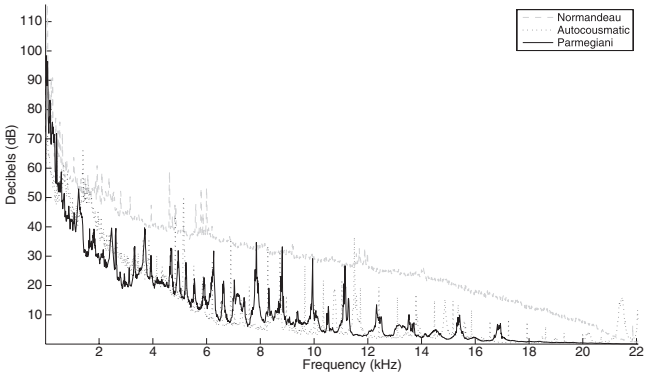
One fierce yet fair correspondent made explicit the need to have the highest standards of listening at every stage of operation, including in selecting viable source materials at the outset: "a critical listening to

source materials, and selection of those which will provide the most promise for development… The source materials lack 'character,' and the processing is rather ordinary." They attacked the abrupt and frequent changes as overly statistical, and that "changes / transitions appear to neither develop or adhere to an internal logic that I am able to make out." Form was a common response: "overall the issues are to do with larger forms, transitions, and questions of taste, e.g., dropouts are considered distracting in most EA [electroacoustic] music. I suppose what's needed is more finesse in terms of how material is assembled (both locally and on a larger scale), which might necessitate further processing."

Some comments pertained to the assumed educational level of the supposedly human composer whom the respondents had been asked to imagine as the music's creator. The music was noted as being at early undergraduate level, or of a level which would fail to achieve entry onto a specialized program (though not rated as the worst applicant!). While damning other pieces, two composers picked out *Accumulator* from the SoundCloud page as a better attempt; curiously, this piece was the longest made available, at 5 min 6 sec: "poor in the timbral variety within each piece, with the possible exception of Accumulator which is the one that could maybe possibly go into a programme", "*Accumulator*: Manages to get through the issues of a longer piece, okay. A bit meandering." More harshly, on the control of tension and release in examples: "little or no drama, meaning that I felt I don't have any expectations set up and thus none quashed. This of course means that the pieces lack (from the reception point of view) directionality."

A very critical concern in general was the spectral allure of the material, for example, "The resulting mix, paramount for tape music, is very flat." Denis Smalley and Robert Normandeau were given as examples of composers whose work has a sparkling and lively sheen of a type to be emulated. For comparison, Figure 6 demonstrates the difference in averaged spectral profile between three works: Robert Normandeau's *Ouverture* to *Clair De Terre* (1999; duration 6 min 12 sec), *Accumulator* by Autocousmatic (2011; duration 5 min 6 sec), and

Figure 6. Comparison of the average spectral profile for three electroacoustic works, one each by Autocousmatic, Bernard Parmegiani, and Robert Normandeau.



Bernard Parmegiani's *Rêveries* (2007; duration 13 min 47 sec). The profiles were calculated by averaging binwise across all windows of a 4,096-point Fourier transform (with a hop size also of 4,096) over each work, using log-power spectrum values calculated from $10 \times \log(\text{magnitude}^2 + 1)$ for each bin. It is evident that the Normandeau work has the smoothest dropoff and greatest overall spectral power. Both the Normandeau and Parmegiani have enhanced power in the lower frequencies compared with Autocousmatic. Although the Parmegiani has some resonant spikes at particular frequencies, it is still smoother than Autocousmatic, which has introduced tight resonances through some processing selections. The Normandeau's spectral sheen, in particular, indicates very careful balancing in mastering, evident in listening to the whole of *Clair De Terre*.

The expert composers undertook less practical work with the system, as they only treated end products, but their highly informed feedback is very valuable as a gauge of research progress, and as an indicator of directions for the path ahead.

Discussion

A problem with evaluating any algorithmic composer is the potential bias of humans against computational automation. This has appeared in a number of studies, anecdotally in the writings of David Cope concerning the reaction of some musicians, critics, and audiences to Experiments in

Musical Intelligence (Cope 2005), and more formally in the work of Moffat and Kelly (2006), who found, in a controlled study, that people, especially musicians rather than non-musicians, had a significant bias against computer-composed music.

It would have to be admitted, however, that the current project did not directly encounter such bias, at least not in any way explicitly declared by participants. Whether due to electroacoustic composers' familiarity with musical algorithms, the increasing profile of generative art in culture (driven especially by the mobile-app boom), or a clear deficiency in software outputs, Autocousmatic was not viewed as a threat. Indeed, when it was presented at a seminar for the 2010 SuperCollider symposium in Berlin, and again in 2011 at the University of Birmingham for Birmingham ElectroAcoustic Sound Theatre (BEAST) composers, people actively engaged with the ideas of increased automation in the compositional process.

Another direction from which to approach Autocousmatic is the developing field of computational creativity (Cardoso, Veale, and Wiggins 2009; Colton, Lopez de Mantaras, and Stock 2009). Here, artificially intelligent systems for creative action are the object of study, algorithmic musical composition being one aspect of this, and computer musicians have much to draw from a growing literature here. A number of evaluative tools have been proposed to study artificially creative systems. We will apply Colton's Creative Tripod (Colton 2008) and the FACE model (Colton, Charnley, and Pease 2011), and avoid any parallels with the Turing Test, which has been much critiqued as a tool in these contexts (Ariza 2009; Pease and Colton 2011).

The Creative Tripod model (Colton 2008) assesses systems with respect to the three supporting struts of skill, appreciation, and imagination. It acknowledges contributions to any creativity assessment from the original programmer, the program itself, and the observer of the program's outputs, and tries to reflect both the end products (compositions in our case) and the process of creating them. One weakness in the trichotomy is the potential circularity, where imagination is a near-equivalent to creativity. Colton's article examines two of his own projects: HR, a mathematical theory-proving system, and The Painting Fool, an automatic visual artist. The Painting Fool's artistic aspiration is closer to Autocousmatic's, and its assessment has also involved attempts to engage with judges (though not with artists themselves as potential users of the program). In parallel with Autocousmatic's machine listening, the Painting Fool incorporates a machine vision subsystem; Colton further claims that the system has a degree of self-appreciation of its own work. This extends to a link between vision and emotional context, to a high-level degree beyond the similarity function of the output critic in Autocousmatic; our system would need to be able to "grow" as an artist as it worked, to reflect a richer human experience, and to have a more human-like appreciation. The imagination of Autocousmatic is tightly locked to the bounds of its exploration of a space of permissible procedures, (Boden's [2003] "exploratory" rather than "transformative" creativity), even if it conducts itself with respect to analyses of the source material and intermediate constructions. As the evaluations above have shown, its skill level can be questioned (especially for mastering quality and enveloping/transitioning). So all three levels of the tripod provide challenges to any claim that Autocousmatic is a computational, creative system of human-like creative degree.

A further alternative perspective is provided by the FACE model (Colton, Charnley, and Pease 2011). This tries to establish the profundity of generation in a system, split between: (F) framing information—the system's own explanations of its work; (A) aesthetic measure—the model it uses to make its creative choices; (C) concept—the procedure of creating works; and (E) expressions of concepts—output works themselves.

The FACE model makes a further distinction between "g" and "p" creative acts for each of the four, where a system may be able to just *generate* in a particular mode, or further be able to *produce* new ways of generating in that mode. Autocousmatic cannot create entirely new aesthetic measures as it works (which would require modeling of processes of personal development and growing experience), but arguably has in its final mix critic an implicit aesthetic measure. Any facility with framing information is absent; Autocousmatic has

no natural-language facility to write its own program notes! With respect to the FACE model's notations, Autocousmatic could at most subscribe to the tuple $\{A^g, C^g, E^g\}$, which is similar to the state-of-the-art computationally creative systems discussed by Colton, Charnley, and Pease (2011).

Moving to other sources of inspiration, electroacoustic composers' writings on the compositional challenges of their work provide a healthy source. A heightening of a composer's analytical mindset is conducive to improving their compositional mindset (Young 2004), and this should go for machine composers as well. John Young also notes the importance of morphological understanding of sounds, suggesting a need for temporal listening models much advanced from the feature-based systems underlying Autocousmatic: "effectively controlling the dynamic relationship between the nature of electroacoustic transformation processes and the projection of perceptually workable generative relationships amongst source sounds and the network of transformations developed with and around them" (Young 2004, p. 10). Knowledge beyond music itself can be critical; the sorts of rich, individual, cultural, memory-dependent decisions evoked by Katharine Norman (Norman 2010) are exactly the most challenging for a disembodied, "music-only" computer to engage with, and they echo comments that the present article has made about the overall human framing of creative work.

Barry Eaglestone and colleagues' series of studies on the cognitive preferences of electroacoustic composers also provide a rich basis for extension (Nuhn et al. 2002; Eaglestone et al. 2008). Their documentation of composer work habits includes revelations such as composers' simultaneous work on multiple pieces, switching between audio applications and thus processing paradigms as a productive tactic mid-piece, moving away from the computer for reflective breaks, and the role of different cognitive styles in composition. In particular, the distinction between "refinement" and "synthesis" approaches (Eaglestone et al. 2008) highlights that a top-down approach is not the only way into composition. The "synthesis" approach of bottom–up discovery, with larger-scale form highly motivated by exploration and categorization of sound objects and their morphology, again illustrates an area where Autocousmatic needs more work. The complex processes of introspection and revisionism that go into composition (with composers sometimes abandoning pieces entirely after much effort), are not really reflected in Autocousmatic's own generate-and-test filtering. More may be learned from documentation projects, such as Barry Truax's intensive consideration of his own compositional process, with extensive multimedia documentation of works through their development (Simoni and Fosdick 2010).

To summarize this discussion, we select four important recommendations arising from the project, to inform future work in this field: (1) full mastering-quality processing of the most exacting standards; (2) a richer model of effective structures in electroacoustic music, with machine analysis resting on a much larger database of prior art; (3) an auditorily acute and temporally expectant listening model, deployed for and adaptable to decisions at all timescales; and (4) a concern with multiple models of compositional process and aesthetic decision making, within which a system might itself develop over time.

As audio-analysis research directed at electroacoustic music continues (Park, Li, and Wu 2009; Couprie 2010; Collins 2011), we can expect to unravel further layers of sonic detail, which can have immediate consequences for generative systems founded on larger databases of prior art. Machine-learning techniques will be necessary to accelerate the musical training of computer programs: A machine equivalent of three years of practice in Schaeffarian reduced listening to sound objects (Landy 2007) certainly would not hurt the prospects of a future algorithmic electroacoustic composer!

Conclusions

As electroacoustic composition is a challenging domain where listening is paramount, it provides an ideal test case for the integration of machine listening into algorithmic composition systems. Autocousmatic, although revealed through the evaluations and discussion herein as only a

stepping stone, provides a contribution to analysis-by-synthesis of electroacoustic work. Formalization of the processes of electroacoustic composition is stimulating to human composers and a reflection on compositional design, providing insights into the creation of works and compositional strategies.

If we were to undertake this project afresh, many aspects of the system would be upgraded. The "inspiring set" would be much extended, and the level of machine listening pushed towards models of expectation that reflect the human time scale of auditory engagement and memory. Four main recommendations are provided at the close of the previous discussion, which indicate the scale of the challenge, but also the richness of work that may result. Commensurate with the prospects for AI in general, the role of wider human experience and social life, in what otherwise seem domain-specific tasks, proves a difficult challenge to overcome, necessitating further modeling of human lifetime experience. Temperamental systems may need compositional biases in their influences, and may relish a poor review as a jumping-off point for renewed struggle; they may vary over time, analogous to a composer's ongoing, developing preferences and discoveries in prior art. Although this sort of emotional attribution may seem far fetched from Autocousmatic in its current state, it points to the adventures ahead.

Acknowledgments

With much gratitude to the three anonymous reviewers, the *CMJ* editors, and all who gave feedback in seminars, questionnaire responses, and consultation.

References

Ariza, C. 2009. "The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems." *Computer Music Journal* 33(2):48–70.

Austin, L. 2001. Liner notes from *Octo Mixes*. Albany, New York: Electronic Music Foundation EMF 039, compact disc.

Barrett, N. 2007. "Trends in Electroacoustic Composition." In N. Collins and J. d'Escrivan, eds. *Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press, pp. 232–255.

Bello, J. P., et al. 2005. "A Tutorial on Onset Detection in Music Signals." *Institute of Electrical and Electronics Engineers Transactions on Speech and Audio Processing* 13(5):1035–1047.

Boden, M. 2003. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

Cardoso, A., T. Veale, and G. A. Wiggins. 2009. "Converging on the Divergent: The History (and Future) of the International Joint Workshops in Computational Creativity." *AI Magazine* 30(3):15–22.

Casey, M., et al. 2008. "Content-based Music Information Retrieval: Current Directions and Future Challenges." *Proceedings of the Institute of Electrical and Electronics Engineers* 96(4):668–696.

Collins, N. 2009. "Form and Algorithmic Composition." *Contemporary Music Review* 28(1):103–114.

Collins, N. 2011. "SCMIR: A SuperCollider Music Information Retrieval Library." In *Proceedings of the International Computer Music Conference*. Available online at www.sussex.ac.uk/Users/nc81/research/scmir.pdf. Accessed May 2012.

Colton, S. 2008. "Creativity Versus the Perception of Creativity in Computational Systems." In *Proceedings of the AAAI Spring Symposium on Creative Systems*. Available online at www.aaai.org/Papers/Symposia/Spring/2008/SS-08-03/SS08-03-003.pdf. Accessed May 2012.

Colton, S., J. Charnley, and A. Pease. 2011. "Computational Creativity Theory: The FACE and IDEA Models." In *Proceedings of the International Conference on Computational Creativity*, pp. 72–77.

Colton, S., R. Lopez de Mantaras, and O. Stock. 2009. "Computational Creativity: Coming of Age." *AI Magazine* 30(3):11–14.

Cope, D. 2005. *Computer Models of Musical Creativity*. Cambridge, Massachusetts: MIT Press.

Couprie, P. 2010. "Utilisations Avancées du Logiciel iAnalyse pour l'Analyse Musicale." In *Proceedings of Journées d'Informatique Musicale*. Available online at jim10.afim-asso.org/actes/43couprie.pdf. Accessed May 2012.

Eaglestone, B., et al. 2008. "Are Cognitive Styles an Important Factor in Design of Electroacoustic Music Software?" *Journal of New Music Research* 37(1):77–85.

ffitch, J. 2001. "The Door Project." Available online at people.bath.ac.uk/masjpf/door.html. Accessed 24 April 2012.

Hiller, L., and L. Isaacson. 1959. *Experimental Music: Composition with an Electronic Computer*. New York: Greenwood Press.
Hsu, W. 2005. "Using Timbre in a Computer-Based Improvisation System." In *Proceedings of the International Computer Music Conference*, pp. 777–780.
Jehan, T. 2005. "Creating Music by Listening." PhD thesis, Media Lab, Massachusetts Institute of Technology.
Klapuri, A., and M. Davy, eds. 2006. *Signal Processing Methods for Music Transcription*. New York: Springer.
Landy, L. 2007. *The Art of Sound Organisation*. Cambridge, Massachusetts: MIT Press.
McCartney, J. 2002. "Rethinking the Computer Music Language: SuperCollider." *Computer Music Journal* 26(4):61–68.
Miranda, E. R., and J. A. Biles, eds. 2007. *Evolutionary Computer Music*. London: Springer-Verlag.
Moffat, D. C., and M. Kelly. 2006. "An Investigation into People's Bias Against Computational Creativity in Music Composition" In *Proceedings of the 3rd International Joint Workshop on Computational Creativity* (pages unnumbered).
Norman, K. 2010. "Conkers (Listening out for Organised Experience)." *Organised Sound* 15(2):116–124.
Nuhn, R., et al. 2002. "A Qualitative Analysis of Composers at Work." In *Proceedings of the International Computer Music Conference*, pp. 572–580.
Park, T. H., Z. Li, and W. Wu. 2009. "EASY Does It: The Electro-Acoustic Music Analysis Toolbox." In *Proceedings of the International Symposium on Music Information Retrieval*, pp. 693–698.
Pearce, M., D. Meredith, and G. Wiggins. 2002. "Motivations and Methodologies for Automation of the Compositional Process." *Musicae Scientiae* 6(2):119–147.
Pease, A., and S. Colton. 2011. "On Impact and Evaluation in Computational Creativity: A Discussion of the Turing Test and an Alternative Proposal." In *Proceedings of the AISB symposium on AI and Philosophy 2011.* Available online at www.doc.ic.ac.uk/~sgc/papers/pease_aisb11.pdf. Accessed May 2012.
Perez Gonzalez, E., and J. Reiss. 2007. "Automatic Mixing: Live Downmixing Stereo Panner." In *Proceedings of DAFx-07*, pp. 63–68.
Pope, S. T., and A. Kouznetsov. 2004. *Expert Mastering Assistant (EMA) Version 2.0: Technical Documentation*. FASTLab Inc. Product Documentation. Center for Research in Electronic Art Technology, University of California Santa Barbara.
Roads, C. 1985. *Composers and the Computer.* Los Altos, California: William Kaufmann.
Rowe, R. 2001. *Machine Musicianship*. Cambridge, Massachusetts: MIT Press.
Sethares, W. A. 1998. "Consonance Based Spectral Mappings." *Computer Music Journal* 22(1):56–72.
Simoni, M., and K. Fosdick. 2010. "Barry Truax: Acoustic Communication and Compositional Techniques." *Computer Music Journal* 34(2):96–98.
Stowell, D., and M. D. Plumbley. 2007. "Adaptive Whitening for Improved Real-Time Audio Onset Detection." In *Proceedings of the International Computer Music Conference*, pp. 312–319.
Sturm, B. L. 2006. "Adaptive Concatenative Sound Synthesis and its Application to Micromontage Composition" *Computer Music Journal* 30(4):44–66.
Vaggione, H. 1984. "The Making of Octuor." *Computer Music Journal* 8(2):48–54.
Wilson, S., D. Cottle, and N. Collins, eds. 2011. *The SuperCollider Book*. Cambridge, Massachusetts: MIT Press.
Young, J. 2004. "Sound Morphology and the Articulation of Structure in Electroacoustic Music." *Organised Sound* 9(1):7–14.