Crowdsourcing the Corpus: Using Collective Intelligence as a Method for Composition

David Plans Casal

In the best of all possible worlds, art would be unnecessary... The professional specialization involved in its making would be presumption... The audience would be the artist and their life would be art.

—Glenn Gould [1]

In the work discussed in this article, I used component-wise subspace decomposition, a spectral analysis technique first described by Casey and Westner [2], to derive note-like components from an original recorded piece and then built these into a simple web application offering each component as a Human Intelligence Task (HIT), modeled after Amazon’s Mechanical Turk [3], in which contributors were each asked to record their own interpretations of components. Each HIT was thus gathered as a reinterpretation of an original, in order to re-synthesize the original corpus [4]. This was then used to build an improvisation system, which I used in performance. Today’s hyper-attentive audience, who have at their disposal amateur technologies that offer previously professional capabilities, have to some extent erased the distinction between professional and amateur. The audience often co-creates, not just consumes, media. Thus, they realize Gould and McLuhan’s vision that the audience would progressively become the composer [5] and embody the post-capitalistic future Jacques Attali foresees in Noise; their activity “involves a radical rejection of the specialized roles (composer, performer, audience) that dominated all previous music” [6]. In this paper, I use Bernard Stiegler’s ecology of attention paradigm [7] to situate crowdsourcing as an emerging form of public participation in music-making and Glenn Gould’s ideas on performance and public access to position this participation as an act of composition. The work is offered as an illustration of the author’s individual process as a composer for finding new notational pathways for collaborative practice.

I designed several scoring systems for Frank, some based on textual representation (using LilyPond) and some graphic, which included algorithmic variables such as the frequency and speed of evolution over generations and the Euclidean distance between major timbre clusters. An example score can be seen in Fig. 1, in which “BreedHZ” refers to how often new generations are bred in the genetic algorithm, and values at the top starting with D refer to the Euclidean distance between pitch centroid clusters.

Increasingly, my algorithmic design centered around timbre clusters, where the algorithm found the strongest fitness (in terms of the pitch centroid). The fitness functions of such algorithms rewarded innovative timbre, which was defined as timbre clusters most distant from the original material (that which was heard at the beginning of each run, from both human and computer). I built improvisational structures out of transitions to and from these clusters, giving myself potential roadmaps for improvisation by pointing at naturally salient points of fitness, as discovered in practice and rehearsal. After some years of pursuing this practice, I wanted to move away from the static use of sonic corpora (harvested from recorded sources) and into using the audience itself as a partner in the creative process. The question then became how to notate parts for an audience whom I wanted to participate through on-line environments. Below I discuss results from my first attempt at this process.
DECOMPOSING AUTUMN

In a recent article, I outlined the use of component-wise subspace decomposition (and audio source separation technique) as a compositional device [15]. Here I address its results in practice. A component-wise decomposition of Autumn in Warsaw, the sixth in György Ligeti’s second book of piano études, was used as the corpus. Each separated audio component from this corpus, as derived by Probabilistic Latent Component Analysis (PLCA) [16,17], was offered as a HIT to contributors, who were given the chance to record one or many components. Whilst the initial system used actual pentagram notation in conjunction with a recording of the component, it was immediately obvious that many of the contributors, not familiar with traditional notation, found it an impediment. I then redesigned the system to offer a waveform-only display of the components, which contributors could re-record using their own voices or anything to hand to reinterpret the sonic qualities of the original as they wished. When gathered and sequenced in the right order, these components formed a re-synthesis of Ligeti’s original étude, which I then used as a corpus to build a music system for live performance.

Navigating the crowdsourced corpus using live matching algorithms such as AudioDB [18] and corpus-based concatenative synthesis [19], I improvised on melodic and rhythmic structures inherent in Ligeti’s original étude. Figure 2 illustrates this process.

In previous experiments in musical crowdsourcing [20,21], notably Eric Whitacre’s Lux Aeterna [22], sheet music was sent out to auditioned participants, who then contributed YouTube performances—250 in all—of their parts. Since the improvisational systems I build are based on pitch clusters and timbre analysis alone, the project that inspired me most was Aaron Koblin and Daniel Massey’s Bicycle Built for Two Thousand, which decomposed the song “Daisy Bell” from a MIDI file into 2,008 components contributors could re-record using Mechanical Turk. Following their example, I wrote a custom audio-recording tool that I could embed in a web application. HIT contributors in such a system are invested individuals, who inevitably want not just to imitate but also to give their own takes on the sonic material. The overall process becomes one of analysis and reinterpretation by collective intelligence. Contributors offered their own analytical performances, thus allowing me to listen to the original étude through a group mind analysis. In his analysis of the temporal object through melody après Husserl, Stiegler writes: “To apprehend isolated tones in a melody or phonemes in a language, as an artifact, is just as useless as trying to understand the crackling sound of a fire through the study of wood” [23]. Isolated spectral components are indeed de-contextualized, but when the HIT is presented within the context of the original piece, a notion of the potential context of each component emerges from the flux. In the perception of a unit, there inherently exists “a temporal, pre-individual ground on which the unitary figure of tone has always already emerged” [24]. My on-line collaborators contributed, without fail, interpretations of each component that were imbued with a clear sense of context, which in this case (Autumn in Warsaw, the Ligeti étude) has an ethereal, otherworldly, but progressively rich timbre and character. The great majority of submitted interpretations subsumed the character of the original piece. Many went further, offering not only the pitch and general timbre of the original component but small micro-improvisations on its general sonic qualities. It would be most appropriate to see these contributions as strictly performative and my later corpus-building and concatenations as compositional. I would like to argue that each contributor’s investment can be seen as not only performative but re-compositional.

In this sense, this web application functioned as a primitive score system (which could be vastly improved), catalyzing a collaborative process into action. It is far from ideal: In future incarnations, contributors should be given more control over the final quality of their interpretation. Filtering and envelope manipulation should be made available, for example. In my upcoming work, a similar application will run as a server-side core, with multiple clients requesting components and submitting interpretations. One of the clients will be web driven and function as a Mechanical Turk application; the other two will run as iPhone and Android applications, allowing each contributor to add minimal effects and edits and to contribute spontaneous recordings of acousmatic material. The server-side application will assemble several nodes (clusters of potential notes/components) to offer the contributor a chance to do more than contribute a single element. They will thus begin to contribute form.

CONCLUSION

In exploring my interest in a phenomenological view of the audience as composer, my ambition was to respond to Bernard Stiegler’s exposition of technologies such as MPEG7 as a crucial effort that nonetheless sees “deep disinterest from the artistic and intellectual communities” [25]. Technologies such as MPEG7 allow media to become “navigable and searchable according to finite lists of criteria” [26], allowing a new kind of listener-

![Score for Back and Forth](https://example.com/score.png)

**Fig. 1.** Score for *Back and Forth*, a piece for piano and algorithm, 2006. (© David Plans Casal)
as-producer to emerge, as attested by the presence of collaborative production sites such as Indaba [27] and Ohm Studio [28], and music recomposition technologies such as mHashup [29] and Concat [30]. By using this technology to involve the audience during the composition process, I also wanted to investigate Stiegler’s “new objectification of sound,” in which he argues (as did Gould and McLuhan, and Bartók before them) that recording and digitization technologies have realized a de-instrumentalization of the ear [31], in that audiences have increasing control over the consumption of their media and need no longer understand it critically or technically in order to play it or recompose it. In Stiegler’s theory, the de-instrumentalization of our ears will be the price we pay for the implementation of another organology of listening: that of analysis [32].

This “age of listening,” which is that of the listener as consumer that Bartók predicted in “La Musique Mecanique” [33], is now converging with previously academic processes such as synthesis and analysis to bring about not only the de-instrumentalization of the ear but the deserialization of consumption. I attempt to build systems that implement, catalyze and celebrate this phenomenon; I cannot lament the loss of individuation that Stiegler proposes global mnemotechnological industries are causing (which is undoubtedly true nonetheless), as I see the inherent potential to give the amateur more control.

Jonathan Impett recently argued that in the design of systems that need to derive knowledge of their performers, “the knowledge cannot come into being independently of its technical implementation,” as “each path necessitates a new technical system” [34]. Impett refers to Stiegler’s exposition of invention and innovation’s relationship (that of tekhné and epistešme). Stiegler also speaks of the increasing speed of transfer between invention and innovation itself, between scientific discovery and technical invention. Technologies such as Probabilistic Latent Component Analysis (PLCA) and MPEG7 have been so quickly put in the hands of artists such as myself that they are still “new” by the time they reach the audience, with their framing (Gestell) made transparent and invisible, always already available to them, as though they had been there forever. In the audience’s use of these technologies, and in their interpretation of their product, they are creating the media they consume within Gérard Granel’s “large now” [35], proving in their wake Stiegler’s argument that temporality can no longer be conceived in terms of the simple Heideggerian “now.” The audience-as-composer, when using notation that itself is recompositional in nature (an on-line recording tool), is a perfect example of the already-there as inherited, not lived.

My interest in exploring Stiegler’s theories by making work of this kind has a second, necessary reason: In the current ecology of hyper-attention (where no one has a linear attention span anymore), it is impossible to ask entire crowds to lend their time to cooperatively imagine and create a future music, if their participation demands study of the past and time-consuming, protracted investment. This is because, as Stiegler points out, “the source of stimulus for hyperattention as it ‘surfs’ . . . is what does not last” [36]. The multiplicity of available data streams on the Web alone engenders a multiplicity of tasks, all competing for attention, which is, after all, merchandise. Because this means that the potential participant audience, the composers in potentia, are hypersolicited, a improviser/programmer must compete with these other “programming industries” for each transindividual, for each construction of the object of attention.

One must build an organon that can invite the surfer to compose, to inscribe, to improvise by combining with the already ongoing attentional processes taking place in what Impett (via William James) lucidly calls the specious present.

To do this, one must consider one’s network’s scattered potential as a series of non-linear time and computation investments, or as Impett aptly describes, potential “moments of inscription” [37]. But they are more than that. When taken up, these moments of inscription are instantly acts of composition; they are acts both of deliberate creation and interpretation. It may appear that, in their non-linearity, these inscriptions are devoid of the linear attentiveness and temporal sensibility that traditionally constitutes music-making. To accept their nature as intrinsically musical, one must recall that each contribution to an exercise in distributed compositional affordance is rooted in time-based references that the individual invests independently of others. As such, the state of all contributions to a potential, distributed act of composition is non-linear. However, each contribution itself is not only linear but also traditionally time-referential.

In the end, in my escape from the theater of performance into the design of distributed improvisation systems, I am simply trying to fulfill Glenn Gould’s vision:

---

**Fig. 2. The Autumn in Warsaw system, outlined, 2009. (© David Plans Casal)**

---

Gérard Granel’s “large now” [35]
As limited as it is, the manipulation of dials and buttons is an act of interpretation. Forty years ago, a listener could only turn his turntable on or off, and maybe, if it was an advanced model, adjust the volume somewhat. The variety of controls available nowadays requires a capacity for analytical judgment. Yet these controls are still very primitive in comparison with the means of participation that will be offered to the listener when the current very sophisticated techniques still in the labs are integrated into domestic appliances [38].

References and Notes


4. For the purposes of feature analysis and extraction, a corpus can be understood as a large set of audio files on which queries are made.


Discography/Videography

Arkanoïd Paraphrase, recorded for Prix Ars Electronica 2011, <www.youtube.com/watch?v=apNar2F0XsQ&feature=youtu.be>.


Manuscript received 1 January 2011.