AUTOMATIC FOR THE PEOPLE: CROWD-DRIVEN GENERATIVE SCORES USING MACHINE VISION AND MANHATTAN

Chris Nash
Department of Computer Science and Creative Technology, University of the West of England,
Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK
chris.nash@uwe.ac.uk

ABSTRACT

This paper details a workshop and optional public installation based on the development of situational scores that combine music notation, AI, and code to create dynamic interactive art driven by the realtime movements of objects and people in a live scene, such as crowds on a public concourse. The approach presented here uses machine vision to process a video feed from a scene, from which detected objects and people are input to the Manhattan digital music notation [1], which integrates music editing and programming practices to support the creation of sophisticated musical scores that combine static, algorithmic, or reactive musical parts.

This half- or full-day workshop begins with a short description and demonstration of the approach, showcasing previous public art installations, before moving on to practical explorations and applications by participants. Following a primer in the basics of the tools and concepts, attendees will work with Manhattan and a selection of pre-captured scenes to develop and explore techniques for dynamically mapping patterns, events, structure, or activity from different situations and environments to music. For the workshop, scenes are pre-processed so as to support any Windows or Mac machine. Practical activities will support discussions on technical, aesthetic, and ontological issues arising from the identification and mapping of structure and meaning in non-musical domains to analogous concepts in musical expression.

The workshop could additionally supplement or support a performance or installation based on the technology, either showcasing work developed by participants, or presenting a more sophisticated, semi-permanent live exhibit for visitors to the conference or Elbphilharmonie, developing on previous installations.

1. INTRODUCTION

In *The Open Work* [2], Umberto Eco presents a vision of Interactive Art as "works in movement", based on "openness", in which audiences become agents in the completion of open-ended artworks, left unfinished by their author and completed during the performance (which may never reach a 'close') through a partnership of performers and audience. These works give rise to a multiplicity of meaning, and democratisation of the creative process, but are also realised within the possibilities and constraints afforded by the artist and their tools.

This paper describes a technology-based platform for such works, in the form of a crowd-driven music system, where a notated musical work can make use of detailed visual scene analysis of a physical space, to map dynamic motions of objects and people in a scene (e.g. a crowd, audience, or public space) to the music in realtime, in order to create, manipulate a live performance.

The proposed workshop will practically explore musical opportunities and challenges presented to artists and audiences, and issues of agency, aesthetic, attribution, and adaptability. Participants will be invited to develop crowd-driven musical works, using a platform coupling object-detection (via machine vision) with Manhattan [1] (Section 2.5) – an accessible, yet scalable digital music notation for composition and programming, which flexibly enables users to edit musical phrases and patterns manually, but also insert formulas and code fragments that dynamically manipulate a piece during performance. The notation supports a wide gamut of generative applications, from simple expressions for individual notes to sophisticated algorithmic processes that generate entire works, and event handling that enables interactive works, responding to live input of musical or non-musical data.

This paper begins with a definition and brief discussion of crowd-driven music, followed by a description of the technology and its previous use in public installations, before outlining the particulars of the proposed workshop, installation, and performance.

2. CROWD-DRIVEN MUSIC SYSTEMS

Crowd-driven music is defined here as a situational score in which the live performance or playback of a notated work in some way responds to the realtime dynamic processes or changing structures in an external chaotic system, such as a crowd of people.

The specific technology and techniques featured in this workshop support a closed-loop system (Figure 1) wherein a live video feed of a public space (e.g. a train platform) is analysed in realtime using machine vision (machine learning used to detect and classify objects and people), with the results streamed to a generative music environment (Manhattan, Section 3) that automatically composes (and renders) a score, using a mixture of static (fixed) and dynamic (processed) elements, the acoustic result of which is played back live into the space.

This system of interaction, and specifically the use of machine learning to conduct detailed and comprehensive visual scene analysis, extends previous artworks that use manual input from the audience or public (e.g. [3], [4]), to affect an autonomous, persistent live performance.

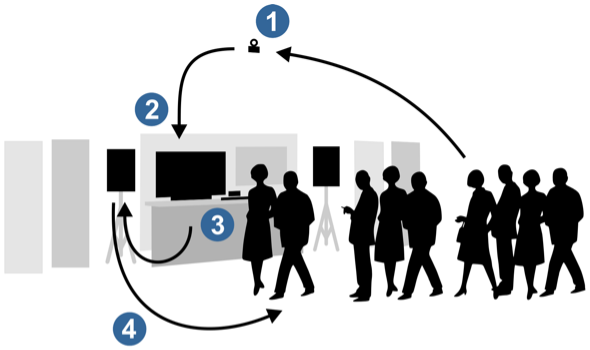

Figure 1 Crowd-Driven Music System

The feedback loop in the system makes it possible for human subjects to take both active and passive roles in a piece; that is, any performance is the product of natural patterns and processes in the crowd, but can also influence crowd behaviour, and be itself subject to influence from individuals and groups within the crowd. This creates both challenges and opportunities for artists to explore new aesthetics and experiences; the agency afforded the crowd versus the integrity of the composer's voice, dictated by the processes and functions employed (and available) to map events and structures between the physical (visual) and musical domains.

2.1 Technical Overview

The system used for the workshop uses machine learning to analyse a scene and detect objects in realtime, using the Darknet *convolution neural network* (CNN) framework [5] and an OpenCL port of the YOLO ("You Only Look Once") v2 realtime object detection system [6]. In common with other machine learning practices, objects are detected using a classifier previously trained on a known set of labelled images, which is then be applied to new images (or frames of a video), detecting objects based their similarity to previously seen examples. YOLO is a performance-optimised detection system that enables the process to run in realtime on modest hardware by reducing the number of convolution processes required per image. In realtime use, using as the graphics processing unit (GPU) of a laptop (MacBook Pro with AMD Radeon 455), the system is able to analyse footage at roughly 7 fps, easily enough to track changes in a scene for the purposes of controlling music.

2.2 Mapping Strategies

The adopted model for mapping scene data to the musical performance is to maintain the state of objects in the scene and allow the playing music to read (pull) the data on-demand. This ensures that changes in the scene are introduced to the performance at musically coherent points or intervals, as decided by the composer – the start of a note, bar, phrase, or section – supporting both fast reactions (events and triggers) and slower gradual processes and context shifts.

While crowd activity can be characterised by noise-like (seemingly random) behaviour, any environment contains hidden patterns and order – structures and processes that can be exposed and exploited in music. Such crowd patterns are not intrinsically musical, placing the onus on composer and code to manipulate them to a given aesthetic. Pre-processing detected objects using techniques such as clustering (grouping similar objects based on proximity) can be useful to extract gestalts that facilitate the mapping process. For example, a naïve 1:1 mapping of individuals to separate musical pitches will overwhelm most systems of harmony or counterpoint, creating a cacophony for anything other than sparse scenes.

Carefully calculated clustering techniques, however, can reduce large crowds to a more manageable finite sets of groups, while also preserving an individual's agency within the music through their influence over the make-up of a group or movement between groups. Such grouping mechanisms allow pieces to respond to higher-level structure in a crowd, which accordingly can be mapped to more abstract musical processes (e.g. tempo, form, structure, harmonic progression). Other notable techniques include linear mapping (e.g. number of people ~ tempo), parameter smoothing (preserving gradual changes without responding to noise), constraints (e.g. quantising pitch to scales or harmonic pitch sets), and condition (i.e. if-the-else; to detect events or categorise data into ranges).

2.3 Performance Platforms

The presented system was originally developed to support a public installation for *BBC Music Day 2018* [7,8], in collaboration with the BBC and Great Western Railways, whereby two crowd-driven pieces were commissioned and installed on the main platform of Bristol Temple Meads station for the day. BBC Music Day is an annual event (covered by BBC TV and Radio) to get the public more involved in music through a series of performances and activities across the country. In Bristol, the crowd-driven music installation was developed to transform travellers passing through the main railway station into composers.

The following objectives were developed to guide the development of the technology and experience:

- ***Agency*** – Members of the public should be able to appreciate their influence on the music.
- ***Accessible*** – The music should have broad appeal to an audience composed of the general public.
- ***Aesthetic*** – The technology should support a coherent musical voice, idiomatic of the artist.
- ***Adaptive*** – The piece should vary over time, responding to context and avoiding repetitiveness.

Each piece tracks platform activity with respect to four objects: people, trains, luggage, and bicycles, using their positions, numbers, or grouping to create and control musical patterns or processes, such as melody, harmony, dynamics, tempo, or instrumentation. Objects were chosen to provide elements that support constantly changing contexts over time, at different rates (constant, regular, occasional, rare). Figure 2 shows an example scene with object detections, as displayed to onlookers on the platform; also showing detected clusters (groups of people) and their relative sizes as blocks along the lower edge.

Installations have since featured in events such as the *Sofia Science Festival*, with new public art commissions for UWE's new Engineering building and local museums.

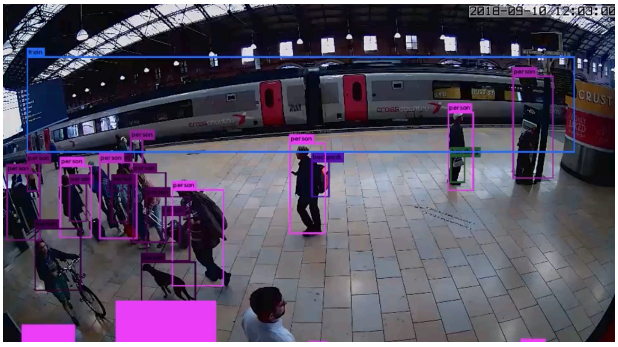

Figure 2 Realtime Object Detection on the Platform

2.3.1 Not So Different Trains

The first piece (Supporting Video 1) used scene data in the live composition of a minimalist piece for piano and virtual choir (singing the suburbs of Bristol using concatenative synthesis), where: the distribution of the crowd determined harmonic pitches and chord voicing; its size and density (quiet vs. busy) determined tempo, instrumentation, and dynamics; and specific events (such as bicycles) varied chord quality. An algorithmic process was used to calculate pitch sets and control semi-randomised progression between tonic, sub-dominant and dominant functional chord families, to provide a tonal palette for crowd-mapped processes, based on an extended *Tintinnabuli* technique, inspired by Arvo Pärt.

2.3.2 Massive Railtrack

The second piece (Supporting Video 2) applies similar processes in development of a "trip-hop" musical style, in homage of Bristol-based band, *Massive Attack*. The harmonic language follows a simplified I-IV-V process derived from the previous piece, synthesised using a multi-sampled local community choir (*Rising Voices*, Bristol Drugs Project's recovery choir) with a synth bass part tracks the largest group of people within the crowd and appropriately selects from the tonal palette. Further drum programming uses untuned samples recorded by the choir (as well as loop), layered according to the relative busyness of the platform (number of people, trains, etc.).

2.4 Supporting Videos

The following videos are available online [8] from:

http://nash.audio/manhattan/tenor2020

Performance Video 1 *(Classical style; ~12mins, HD/MPEG4)*
Screen capture of live music generation with inset platform footage, classical style using piano and synthetic choir (programmed to sing the neighbourhoods of Bristol).

Performance Video 2 *(Trip Hop style; ~12mins, HD/MPEG4)*
Screen capture of live music generation with inset platform footage, trip-hop style using synthesisers and sampled *Rising Voices* choir (for both voice and drum sounds).

Early Technical Demo *(annotated initial experiments; 2:34)*
Annotated screen capture of early experiments with basic music mappings of people locations to notes (pitch and rhythm) and percussion density to crowd density. Annotations are provided to explain the system and process.

2.5 The Manhattan Software

Manhattan [1] (pictured in Figure 3) is a digital music platform developed for learning and creativity in both music and programming – designed to extend traditional computer music practices (such as sequencing) through code fragments situated in the music notation, to support algorithmic, reactive, and dynamic pieces.

Manhattan offers an accessible, yet scalable introduction to music programming for computer musicians, and is used in teaching to develop computational thinking skills in non-coders, part of a wider initiative to support digital literacy and widen participation in coding. The environment exploits the grid-based pattern sequencer style of *soundtrackers*, made of cells specifying notes or other musical events, and applies a spreadsheet metaphor to introduce formulas to musical playback, inheriting many of benefits that have made spreadsheets one of the more successful models of end-user programming. [9]

Unlike other programming languages, the visibility of the data (music), rather than code (formulas), is prioritized, enabling a traditional sequencing/editing workflow, but where the effect of code on the music is apparent. [10] Users can balance manually edited music sequences (static patterns) with varying levels of engagement with programming abstractions (dynamic processes), from simple expressions for isolated dynamic behaviour at specific moments (e.g. conditional repeats, random elements) to formulas for generating entire pieces (e.g. algorithmic music, minimalism, aleatoric music). The environment similarly supports event handling of realtime user and data input to support reactive and interactive applications for live performance, live composition, improvisation, interactive installations, or other sonic art. As such, Manhattan supports continuum of musical practices and aesthetics, from conventional popular and classical styles, to more experimental art and contemporary forms.

Manhattan is being developed as part of a research project involving artists, universities, and schools that is looking at tools to support and extend creative and pedagogical practices in both music and programming. The software is free to download[1] for MacOS and Window, and includes everything required to start writing and programming music (including built-in sounds and synthesisers, plus extensive interactive tutorials).

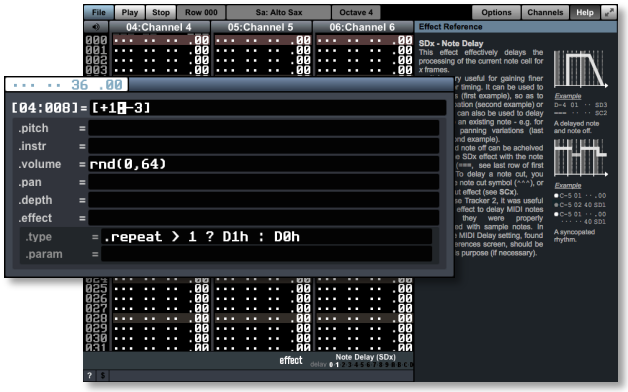

Figure 3 Manhattan music software

[1] http://nash.audio/manhattan

3. WORKSHOP

The workshop is designed with the following objectives:

- Discuss and explore practical applications and expressive opportunities in mapping complex or chaotic systems (e.g. public crowds) to both live and prepared musical works.
- Experimentally develop works of crowd-driven music, using AI (machine learning) and generative techniques, supported by the Manhattan software.
- Identify effective musical analogies and ontologies for structures and processes in non-musical systems, and the aesthetics they engender.
- Establish directions and collaborations for future research or artistic projects.

3.1 Proposed Schedule

A representative draft itinerary for a half-day (3-4hr) workshop is presented below, adaptable to the conference programme. A full-day (5-6hr) alternative is also outlined (changes marked *), specifically allocating more time for practical development and discussion. The extended full-day programme is recommended if conference organisers select the option to exhibit participants' works.

00:00 – Welcome and Introductions *(15m, all)*
Host and delegates introduce themselves, briefly describing their background and respective areas of interest or expertise.

00:15 – Opening Presentation *(30m, organisers)*
Audio/visual presentation introducing relevant concepts and technologies (i.e. *Manhattan* and machine learning concepts), with demonstrations of previous artistic works, basic syntax and expressive mapping techniques.

00:45 – Manhattan Primer *(30m or 60m*, all with support)*
Simple practical exercises using prepared materials, designed to introduce delegates to the fundamentals of the Manhattan tool and syntax, demonstrating core coding concepts (e.g. variables, data, iteration, functions, conditionals) using musical examples. This group exercise will encourage open discussion of techniques and participant's interests and backgrounds, used to help frame and guide subsequent sessions.

[15m BREAK or LUNCH*]

01:30 – Music for One *(30m, all with support)*
Initially, mapping concepts are explored through event handling of simple live inputs (using provided MIDI controllers), applied to affect change in repeating musical patterns, in preparation for more complex scenes and scenarios.

02:00 – Music for Many *(60m or longer*, all with support)*
Participants choose from a selection of pre-captured footage featuring various public settings (e.g. foyers, stations, streets), applying what they have learnt, to develop musical mappings and generative processes that respond to live stream of data about detected objects, such as their position, number, groupings, etc. Within each example setting, a diverse range of objects are tracked (e.g. people, vehicles, animals, luggage).

03:00 – Closing Discussion *(15mins, all)*
Review of issues and findings (or research goals) that have emerged, and call for interest in further research / collaboration. Participants are invited to continue developing their pieces for future exhibition as an installation or performance, possibly as part of the TENOR conference programme.

3.2 Intended Audience

This workshop is suitable for all TENOR delegates, especially those with an interest in interactive or generative music. The technology is designed to be accessible to any musician or computer user, and requires no specific expertise in programming. However, the workshop would particularly suit those with backgrounds, research interests or experience in: notation, composition (modern or common-practice), sequencing, programming (usage and semantics), live installations or performance, and artistic applications of AI/machine learning.

3.3 Required Resources

The workshop has no special requirements. Depending on attendance, it will require a single room with a capacity of 20-30, with table space for each delegate (boardroom, u-shaped, cabaret or classroom layout). There should be a projector/screen with VGA/HDMI connection and a high-quality stereo system for computer audio.

The workshop can be hosted using delegates' own machines or in a onsite computer room, running either Mac (preferred) or Windows, so long as third-party software is supported. Footage is pre-computed, obviating the need to apply processor-intensive machine learning in realtime, such that participants should be able to use any Windows/Mac machine.

Manhattan is freely available and has no specific (e.g. hardware or admin) requirements. Participants can download the software in advance, and retain it after the event. Workshop registration should indicate their preferred OS (Mac or Windows, including version and language). Tablets (e.g. iPads) are not supported. Access to WiFi (e.g. eduroam) is desirable.

Where possible, tea and coffee facilities are desirable. No fee is required for participation in the workshop, unless deemed appropriate by the conference organizers.

3.4 Exhibition of Works

For exhibition of workshop outputs, in the form of a video installation, demo, or presentation; a large TV or projected screen with USB input and loudspeakers would be required, to host a playlist of rendered videos showcasing the participants' works for their chosen scene. The video will be prepared at the conclusion of the workshop.

4. INSTALLATION / PERFORMANCE

Separate from the workshop, a public installation based on the live crowd-driven system is also proposed, showcasing an original interactive work designed for a specific open space, such as a public area of the Elbphilharmonie.

The installation would be adapted from the form used for *BBC Music Day* [7,8] (discussed in Section 2.3/4), with relatively modest technical requirements (TV or projected screen, loudspeakers – see Figure 4), but will require some consultation to identify an effective space and live scene to support an expressive performance. For example, the original installation, a train platform, offered a constant symphony of motion, including proximal and lateral movement of people, groups, and objects, at

varying paces, sometimes stationary, while fluctuating in density over time (from very quiet to very busy). The system provides considerable flexibility with respect to mapping options, where different composition strategies can accommodate a wide range of settings, but expressive possibilities are ultimately tied to the entropy of the scene itself, and will suffer from too quiet or uniform a scene.

The work(s) presented can adapt previous material (e.g. adapting Bristol themes for Hamburg) or be newly commissioned with respect to a defined musical brief (with respect to the hosting organisation or intended audience) or a specific interactive role (with respect to the space or activity therein). An experimental aesthetic, for example, might suit an audience of TENOR delegates more than the general public – and, in a public or civic space, works must also be considerate of the space's function; music cannot negatively interfere with normal affairs or inhabitants, being either intrusive, distracting, or annoying. On the train platform, for example, the priority was the smooth and safe running of the station.

The installation should be able to run for several days (or longer) unattended (using a dedicated machine designed to support a permanent installation, currently in development) and can connect to its live video feed remotely (through an ad-hoc wireless network), enabling flexible and discreet location of the camera module. Options for online or local (smartphone) streaming are also possible.

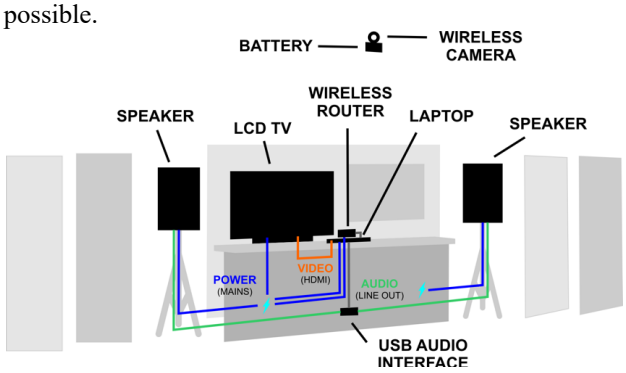

Figure 4 System Overview for Public Installation

4.1 A Note on Privacy

For live or recorded works, this installation is compliant with EU laws concerning privacy and the processing of personal data, such as the GPDR. In live use, captured video footage and data detections are not stored, but processed exclusively in realtime before being discarded. The footage and all data derived from it (i.e. object detections) contain no personally-identifiable information (individuals are identified as "person" only), which is transparently displayed at time of performance. Unless otherwise arranged, only public spaces and scenes of crowds where individuals have no reasonable expectation of privacy are captured. Previous public installations and performances have been ethically reviewed and approved by the BBC, National Rail, and UWE Bristol. Provided previously-captured footage (used in closed workshops or academic settings) has similarly been collected and curated to ensure ethical use of data and protection of privacy.

5. REFERENCES

[1] C. Nash, "Manhattan: End-User Programming for Music," Proceedings of New Interfaces for Musical Expression (NIME) 2014, 2014, pp. 28-33.

[2] U. Eco. *The Open Work.* Cambridge, MA: Harvard University Press, 1989.

[3] Y. Wu, L. Zhang, and N. Bryan-Kinns. "Open Symphony: Creative Participation for Audiences of Live Music Performances" in *IEEE Multimedia*, January, 2017.

[4] S. Bhagwati. *Making Music*. Ulm: April, 2000.

[5] J. Redmon. *Darknet: Open Source Neural Networks in C*. Available: http://pjreddie.com/darknet, 2013.

[6] J. Redmon and A. Farhadi. *YOLO9000: Bigger, Better, Faster.* arXiv:1612.08242 [cs.CV] Cornell University, 2016.

[7] C. Nash. *Track by track: Generative music installation for BBC Music Day*. Technical Report: http://uwe-repository.worktribe.com/output/860147, Bristol: UWE Bristol, 2018.

[8] C. Nash. *Manhattan software, videos and materials.* Available: http://nash.audio/manhattan/shared, 2018.

[9] J.F. Pane and B.A. Myers. *Usability Issues in the Design of Novice Programming Systems*. Carnegie Mellon University, Technical Report CMU-CS-96-132, 1995.

[10] C. Nash, "The Cognitive Dimensions of Music Notations," in *Proceedings of the Second Internation Conference on Technologies for Notation and Representation of Music (TENOR)*, Paris, 2015.

About the Organizers / Speakers

Dr Chris Nash (principal organizer) is a professional programmer and composer – currently Senior Lecturer in Music Technology at UWE Bristol, teaching software development for audio, sound, and music (DSP, C/C++, Max/MSP). His research focuses on digital notations, HCI in music, virtuosity, end-user computing, systematic musicology, and pedagogies for music and programming.

Corey Ford (technical support, recorder) is a post-graduate researcher at UWE Bristol, completing an MRes Data Science studying technology for learning music notation and programming, supervised by Dr. Nash.

Hamburg and London-based music software developers, *Steinberg Media Technologies GmbH*, creators of the *Cubase*, *Nuendo,* and *Dorico* music editors, have also been approached to participate and support the workshop, based on a previous working relationship with the organisers. The extent of the company's involvement will be confirmed closer to the event.