Crowd-driven Music: Interactive and Generative Approaches using Machine Vision and Manhattan

Chris Nash
Creative Technologies Laboratory
Dept. of Computer Science and Creative Technologies, UWE Bristol
Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK
chris.nash@uwe.ac.uk

ABSTRACT
This paper details technologies and artistic approaches to crowd-driven music, discussed in the context of a live public installation in which activity in a public space (e.g. a busy railway platform) is used to drive the automated composition and performance of music. The approach presented uses realtime machine vision applied to a live video feed of a scene, from which detected objects and people are fed into Manhattan (Nash, 2014), a digital music notation that integrates sequencing and programming to support the live creation of complex musical works that combine static, algorithmic, and interactive elements. The paper discusses the technical details of the system and artistic development of specific musical works, introducing novel techniques to mapping chaotic systems to musical expression and exploring issues of agency, aesthetic, accessibility, and adaptability relating to composing interactive music for crowds and public spaces. In particular, performances as part of an installation for BBC Music Day 2018 are described.

The paper subsequently details a practical workshop in crowd-driven music, delivered digitally, exploring the development of interactive performances in which the audience or general public actively or passively control live generation of a musical piece. Exercises support discussions on technical, aesthetic, and ontological issues arising from the identification and mapping of structure, order, and meaning in non-musical domains to analogous concepts in musical expression. Materials for the workshop are available freely within the Manhattan software.

Author Keywords
crowd-driven music, audience participation, open works, interactive art, generative scores, machine learning, machine vision

CCS Concepts
• **Applied computing → Sound and music computing**; *Performing arts*; • **Human-centered computing →** *Mixed / augmented reality* • **Information systems →** *Music retrieval*;

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.

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

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, mapping dynamic motion of objects and people in a scene (e.g. a crowd, audience, or public space) to music in realtime, to create and manipulate a live performance.

An accompanying workshop practically explores the musical opportunities and challenges presented to artists and audiences, and issues of agency, aesthetic, attribution, and adaptability. Participants explore 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.

The 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 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 realtime dynamic processes or changing structures in an external chaotic system, including but not limited to crowds of people, in a manner defined by the user (artist).

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, see Section 2.2) that automatically composes (and renders to audio) a score, using a mixture of static (fixed) and dynamic (processed) elements, the acoustic result of which is played live, back 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. [13], [1]), to affect an autonomous, persistent live performance. By default, such open works have indefinite duration, but can simultaneously exhibit both structure and infinite variation, depending on the ontological map crafted by the artist, as driven by the emergent patterns or seemingly random noise in crowd behaviour and busy public spaces.

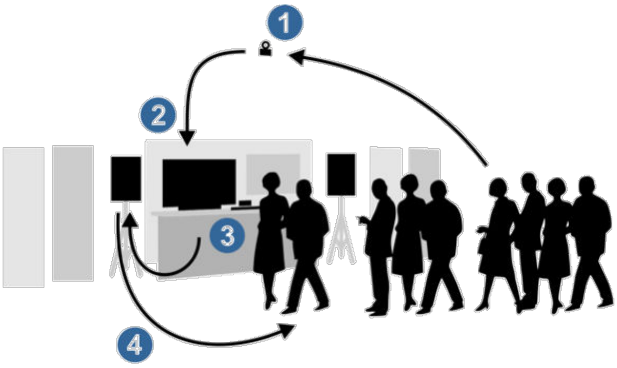


Figure 1. Crowd-Driven Music System

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 their behaviour, and be itself subject to influence from individuals or groups within the crowd. This creates both artistic challenges and opportunities to explore new aesthetics and experiences; the agency afforded the crowd vs. the integrity of the artist'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 performances uses machine learning to analyse a scene and detect objects in realtime, using the Darknet *convolution neural network* (CNN) framework [10] and an OpenCL port of the YOLO ("You Only Look Once") v2 realtime object detection system [11]. 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 the graphics processing unit (GPU) of a laptop (MacBook Pro with AMD Radeon Pro 455), the system is able to analyse footage at roughly 7 fps, sufficient to track changes in a scene for the purposes of controlling music. Figure 3 shows a still from a scene, annotated with detections (purple for people, blue for trains), as shown during the performance.

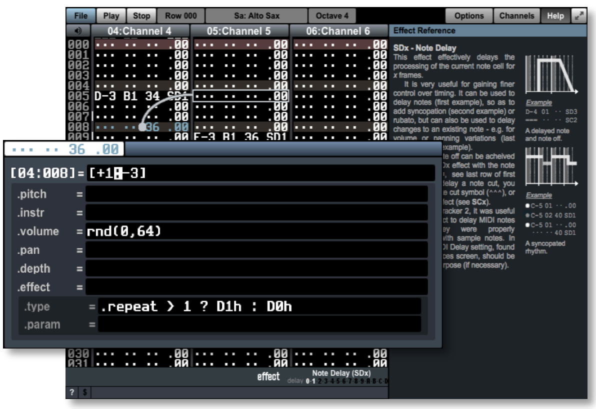


Figure 2. Manhattan music software

2.2 Manhattan Software

Manhattan [1] (pictured in Figure 2) 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. It is designed as 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. [8]

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 immediately apparent. [10] This is additionally beneficial to audiences, who can watch (and see, even if not understand) how the music is manipulated by code and, in this project, helps to visualise to the crowd how scene detections influence what they hear. Realtime highlighting of newly calculated values and data dependency arcs is provides a level of audio-visual synchrony designed to support a basic abstract level of appreciation absent expertise, similar to how an audience would appreciate a virtuosic instrumental performance.

For the artist, the platform balances the manual editing of music (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), including event handling of realtime user and data input. In this project, such flexibility with aesthetic and compositional vs. coding approach proved an important factor in successfully integrating live and dynamic input data from the crowd with a composer's aesthetics and practices.

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 for MacOS and Window, and includes everything required to start writing and programming music (including built-in sounds and synthesisers, plus extensive interactive tutorials).

2.3 Performance Platform

The technology was originally developed to support a public installation for *BBC Music Day 2018* [7], in collaboration with the BBC and Great Western Railways (GWR), 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 & Radio) to get the public 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 railway station into composers.

There were multiple stakeholders in the project. A key research objective was to design an live generative system that composed music, rather than sound art (distinct from most other interactive sonic artworks). Rather than being a bespoke system for single performance, the goal was also to develop the system as a versatile platform, supporting a range of aesthetics or idioms (both art and entertainment music), for use in future projects and by other composers. To this end, two pieces were developed for the installation (outlined in Section 2.5.1 and 2.5.2 respectively).

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

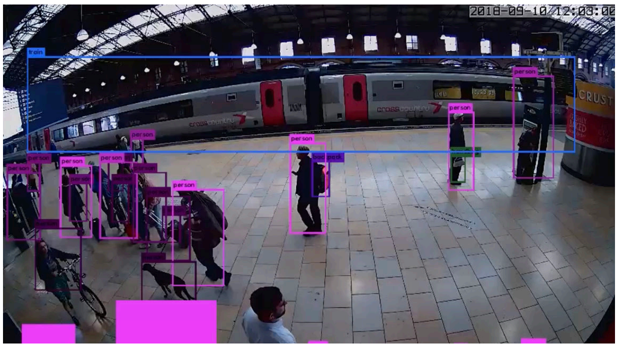


Figure 3. Realtime Object Detection on the Platform

The BBC's goal was to engage as many members of the general public as possible, so accessible and enjoyable music aesthetics were selected – mainstream classical or popular styles, tonal harmonies, conventional rhythms, timbres, and instrumentation. Activities are also intended to reflect the local community and musical culture; for example, events in other regions (including trains and stations) featured local community choirs. To this end, both pieces include vocal parts: one synthesised digitally using concatenative synthesis to sing the neighbourhoods of the city (e.g. "Fren-chay", "Clif-ton", "Tot-ter-down"); another multi-sampling Bristol Drugs Project's *Rising Voices* recovery choir, capturing pitched vowel-sounds and untuned percussive noises and drum timbres. Bristol's music scene, and specifically local artists Massive Attack, also inspired the selection of an electronic trip-hop aesthetic for the second piece.

The hosting environment for the installation dictated other stylistic constraints, chief among them that the music should not unduly interfere with the smooth and safe running of the station. The music could not be overly intrusive or disruptive to commuters or station staff, nor give rise to tension or offence. Loudness and timbre needed to be considerate of staff-customer interactions and station announcements. Per these requirements, the arrangements avoid spoken word and focus on harmonies, rhythms and textures, rather than prominent melodic phrases, or sound effects. At the same time, however, the omission of overt focal elements and distracting motifs also helps prevent the six-hour performance from becoming repetitive.

The following principles were adopted to guide development of the technology, designed to balance the reusability of the system, while focusing the needs of the specific performance:

- ***Agency*** – Members of the public should be able to appreciate their influence on the music.
- ***Accessibility*** – 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.

2.4 Mapping Strategies

The transactional model adopted for mapping scene data to the musical performance was 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. For example, a naïve 1:1 mapping of individuals to separate musical pitches gives everyone a voice but will overwhelm most systems of harmony or counterpoint, creating a cacophony for anything other than sparse, well-behaved scenes.

Pre-processing detected objects using techniques such as clustering (grouping similar objects based on proximity) can be used to extract gestalts that facilitate the mapping process. Carefully designed clustering techniques (e.g. Section 2.4.1) can reduce large crowds to a more manageable limited set 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).

2.4.1 Chaotic Counterpoint

The system's capacity to simultaneously support agency of the crowd, the aesthetic of the composer, adaptability to changing contexts, and remain accessible to a public audience creates a significant mapping challenge. This section describes novel clustering and mapping approaches developed to interpret the disposition of a crowd into coherent musical parts, driving the core harmonic and contrapuntal foundation of the performance. The technique is dubbed "Chaotic Counterpoint" in reference to its extraction of harmony (order) from a noisy system (chaos).

Figure 4 illustrates the process applied to each video frame: people are detected using machine vision, establishing bounding boxes (*x*/*y*-coordinates and size) for each individual in the scene, who are then sorted into five clusters by iteratively merging the two closest individuals/groups in the scene, using *agglomerative hierarchical clustering* based on Euclidean distance between their centre-points. The new, merged group is formed of the weighted average of their members' individual *x*/*y*-coordinates and sum of their weights (area of bounding box). A development of the system also used the y-coordinate in weight calculations to approximate depth, to increase sensitivity to individuals in close vicinity of the system (actively interacting with it).

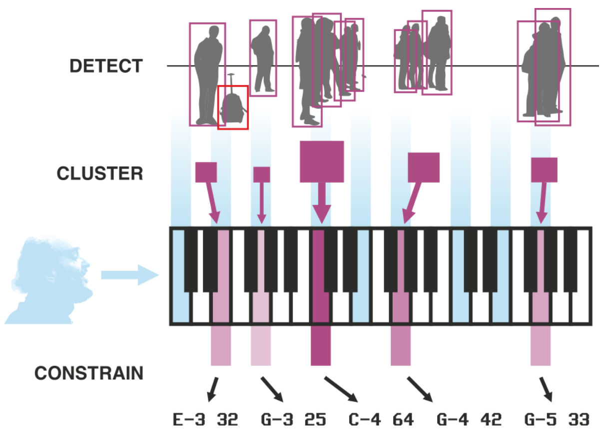


Figure 4. Chaotic Counterpoint technique.

These clusters determine the pitch selection, voicing and note dynamics for a polyphonic musical performance, in which the lateral position of clusters (weighted average *x*-coordinate) map to a note's pitch height and overall cluster weight (~ group size and density) maps to its dynamic (note volume/velocity). Figure 5 illustrates how this weighted cluster-based mapping technique reduces the complexity of the crowd, but ensures individuals (as group influencers) have a perceptible effect on the music.

To avoid the dissonance of a chromatic scale and afford artistic control of harmony and tonality, pitch selection is constrained to a harmonic palette defined by the composer (or code; shown in blue in Figure 4). To achieve a consonant, tonal aesthetic suitable for the general public, the method applied in this performance is based on major or minor triads drawn from a semi-randomised progression through sets of functional chord families (*tonic*, *predominant*, *dominant* [12]). Loosely inspired by Arvo Pärt's *tintinnabuli* technique [4], these pitch classes are then expanded across all octaves, and the nearest pitch selected for each cluster.

2.5 Two Pieces for Crowds and Trains

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, and instrumentation. Objects were chosen to provide elements that support constantly changing contexts over time, at different rates (constant, regular, occasional, rare). Figure 3 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.

2.5.1 Not So Different Trains

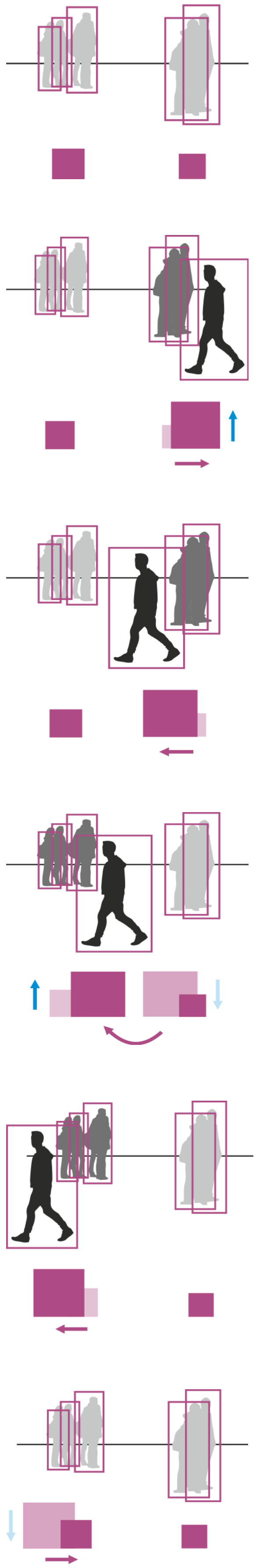
The first piece (Supporting Video 1) used scene data in the live composition of a minimalist piece for piano and virtual choir (singing the suburbs of Bristol using concatenative synthesis), where the distribution of the crowd determined harmonic pitches and voicing using the "chaotic counterpoint" technique (Section 2.4.1); overall crowd size and density (quiet vs. busy) determined tempo, instrumentation, and dynamics; and specific events (e.g. bicycles) varied chord quality (adding a 7^{th} or 9^{th}).

2.5.2 Massive Railtrack

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

2.6 Public Reception

The response to the performance was overwhelmingly positive. Passers-by readily engaged with the music, were able to infer what was going on from the visualisation of the feed and software, and could appreciate their influence on the music, with many attempting to actively interact and "game" the system (sometimes as groups). Both pieces were well-received, with many impressed by the sophistication and approachability of the music generated from the crowd, in contrast to other sound art. Based on comments, the first (classical) piece was perceived as a more successful live soundtrack for the space, but the simpler mappings of the second (trip-hop) piece were easier for people to understand more fully. Many expressed awe in seeing what technology was capable of, notably in the area of machine vision – delighted by its use in art, but disquieted by other possible uses.



1. Initial scene.
Two clusters (and pitches): one left (low pitch), more distant with more people; another right (high pitch), closer with fewer. The groups are stationary.

2. New Person enters.
They join the scene from right, and are clustered with the nearest group: pulling its centre right (raising the associated pitch) and also increasing its weight/size (increasing the dynamic of the associated note – and overall dynamic).

3. Person moves left.
They remain clustered with the right group, still closest (maintaining the dynamic), but the group's mean lateral position is increasingly pulled left (lowering the associated pitch).

4. Person switches groups.
Receding from the right group and approaching the left, the person switches group membership. The clusters move right, raising pitch slightly, but their shifting weights move dynamic emphasis from the higher to lower pitch.

5. Person continues left.
As before, they initially maintain proximity and membership of left group (maintaining the dynamic), but pull its mean lateral position left (lowering the associated pitch).

6. Person exits left.
When the person leaves the field of view, they no longer contribute to the former group's position and weight; returning the cluster and music configuration to before – and decreasing the overall dynamic.

Figure 5. Crowd Control through Clustering.

3. DIGITAL WORKSHOP

This section describes the supporting workshop and practical exercises, first developed for the NIME 2020 conference, later adapted for wider digital delivery using the Manhattan software. The workshop and learning materials support practical explorations of the concepts and technologies presented in this paper, for sessions pursuing the following objectives:

- Discussing and exploring practical applications and expressive opportunities in mapping complex or chaotic systems (e.g. public crowds) to live or prepared musical works.
- Experimenting and developing works of crowd-driven music, based on AI (machine learning) and generative techniques, supported by the Manhattan software.
- Identifying and evaluating effective musical analogies and ontologies for structures and processes in non-musical systems, and the aesthetics they engender.

3.1 Intended Audience

These workshops are suitable for all NIME-related audiences, including digital artists, technologists, and students, especially those with an interest in interactive or generative music. The underlying technology is designed to be accessible to any musician or computer user, with tutorials to develop basic skills in music programming available within the Manhattan software. The workshop would particularly suit those with backgrounds, research interests, or experience in: composition (contemporary, experimental or common-practice), sequencing, programming (usage and semantics), digital notations, live installations or performance, and artistic applications of AI/machine learning.

3.2 Technical Details and Requirements

The digital workshop has only modest requirements, requiring participants to own an Internet-connected, mid-range Windows or Mac computer with speakers or headphones. Some form of MIDI controller is useful, but virtual devices may also be used.

The Manhattan software is freely available, has no specific (e.g. hardware or admin) requirements, and contains all materials required for the session; it is supplied with a library of built-in instruments, but also supports MIDI I/O. The program can either run standalone or as a DAW plugin for integration with other music software. Detailed tutorials on basic syntax and music programming techniques are provided within the environment, supported by video walkthroughs. Participants interested in the crowd-driven musical applications will need to familiarise themselves with the basic programming concepts encapsulated in these tutorials, introductions to which should be provided as part of any hosted session (such as the NIME workshop).

For the purposes of the workshop, recorded footage is used, supplied with metadata containing pre-analysed object detection results, obviating the need for participants' machines to execute processor-intensive machine learning in realtime. Video feeds with metadata are streamed from an online server. Parties interested in live applications should contact the author (see 3.5).

3.3 Representative Session

The following itinerary outlines a suggested structure for short or half-day (2-3hr) sessions, adaptable for offline and online use. A longer full-day (5-6hr) alternative is also outlined (marked *), notably allocating more time for practical development and discussion. Shorter sessions are suitable for introducing concepts or fostering discussion; longer sessions are more likely to lead to exhibitable artworks. Significantly longer, more comprehensive programmes and curricula on music programming are also supported by the software, supporting weekly sessions and project work. The pedagogical development and use of these learning resources will be discussed in future publications.

00:00 – Introductions *(30m, organisers)*
Opening audio/visual presentation introducing relevant concepts and technologies (i.e. *Manhattan* and machine learning), with examples from previous artworks, basic syntax and mapping techniques.

00:30 – 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. data handling, loops, functions, conditions) using musical examples. This group exercise will encourage open discussion of techniques and participant interests and backgrounds, used to help frame and guide subsequent sessions.

01:00 – Music for One *(30m, all with support)*
Mapping concepts are initially explored through event handling of basic live input (using provided MIDI devices), applying changes in repeating musical patterns, in preparation for more complex scenarios.

01:30 – 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 in realtime to streamed detection data, including their position, number, groupings, etc. In each example setting, a diverse range of objects are tracked (e.g. people, vehicles, animals, luggage).

02:30 – Discussion and Questions *(15-30mins, all)*
Review of issues and findings (or goals) that have emerged, and discussion of future work or collaborations. Organisers may wish to discuss and support continued development of pieces or technique.

3.4 Sharing or Exhibiting of Works

Session participants can share works based on provided video examples simply by sharing the project files containing their musical patterns and code (.zip a few kilobytes), executable on any other Manhattan platform with Internet access. The software also supports exporting to audio file (.wav), whilst audio muxed with video can be captured using screen recorders. MIDI output can also be captured for further processing or development in other music authoring packages (DAWs and score editors).

3.5 Live Scenarios

The principles and techniques explored in the workshop easily transfer to applications in live interactive systems, supporting the development of fixed installations or performances, such as that used for *BBC Music Day* [7] (see Section 2.3-2.5). Installations can run for several days unattended (a dedicated machine designed to support a permanent installation is in development).

A live system requires a high-end computer with discrete graphics card (CUDA-based GPU recommended) to support realtime machine learning. The code to analyse the live scene runs as a separate process to Manhattan and delivers data via MIDI SysEx. This may be run on the same machine, but could also be hosted on a dedicated machine that broadcasts over a network. Such configurations could conceivably also support workshops based on developing applications for live scenes.

Live installations require a video feed, for which wired (USB) or network (IP) cameras are supported, where compact USB-powered wireless cameras facilitate easy and flexible placement. Ad-hoc wireless networks should be used to avoid interference with other network traffic. Live presentation to a crowd (*in situ*) requires a TV/projector and loudspeakers, shown in Figure 6, while online or smartphone delivery supports a passive presence.

Significant care should be given to selecting an appropriate space and live scene to ensure an expressive performance. For example, the original setting, a train platform, offered a constant symphony of motion, including proximal and lateral movement of people, groups, and objects, at varying paces, sometimes stationary, fluctuating in density over time (from very quiet to very busy). The system provides considerable flexibility with respect to mapping, able to accommodate complex scenarios, but expressive possibilities are ultimately tied to the entropy of the scene itself, and will suffer from too quiet or uniform a scene.

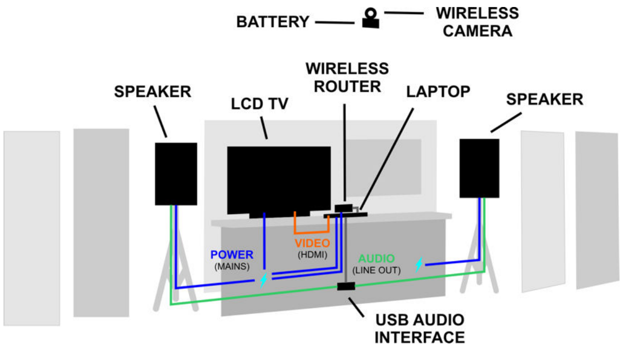


Figure 6. System Overview for Public Installation

4. CONCLUSION

This paper has described a technology platform, and explored artistic methodologies and specific mapping techniques for crowd-driven music, based on machine vision and live music processing (using the Manhattan programming language [5]), discussed in the context of interactive sonic artworks developed for *BBC Music Day 2018* and digital workshop for *NIME 2020.*

The approach and technology were successfully employed in supporting a complex realtime musical performance based on interaction by the public, seeking a balanced integration of *agency* (perceived audience influence on the music), *aesthetic* (artist's expressive control), *adaptability* (musical variation in the context of changing scene dynamics), and *accessibility* (audience's relative engagement and appreciation of the work).

Beyond artistic expression, the installation demonstrated a capacity to engage the public with new concepts in both music and computer technology, and the crowd-driven system has since been exhibited in events such as the *Sofia Science Festival*, a series of public TED-style talks for the *Institute of Engineering and Technology (IET)*, and new public art commissions for UWE Bristol's new Engineering building, as well as local museums.

The versatility of the underlying system was partially explored with the development of pieces in two disparate genres, classical and trip-hop, though there remains considerable scope to explore and evaluate other aesthetics and compositional styles, notably those of more experimental styles, art music, and non-Western cultures. To this end, the paper concluded by detailing a workshop inviting other artists, researchers, and technologists to explore new concepts and applications of crowd-driven music beyond those presented here – and a call for new collaborations.

5. SUPPORTING MATERIALS

The following videos are available online at:

http://nash.audio/manhattan/nime2020

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.

6. ETHICAL STANDARDS

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 workshops or academic settings) has similarly been collected and curated (where appropriate anonymising) to ensure ethical use of data and protection of privacy.

7. ACKNOWLEDGEMENTS

The project gratefully acknowledges the support of BBC Bristol, Robert Sugden and the BBC Music Day team, Network Rail, Great Western Railways (GWR) and Bristol Temple Meads station, Bristol Drug's Project and the Rising Voices Choir. Special thanks to Corey Ford, Alexandre Hurr, and Adrian Hull, for their assistance on the day. This project was principally funded through the award of a *Vice Chancellor's Early Career Researcher* fellowship from the University of the West of England (UWE Bristol).

8. REFERENCES

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

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

[3] E. Flügge. "Soundly Planning – Listening Practically to (Belfast) Sound Spaces." *In Proceedings of the Conference on Invisible Spaces: Sound, Urbanism and a Sense of Place*. Azores, Portugal, 2017.

[4] P. Hillier. *Arvo Pärt (Oxford Studies of Composers)*, Oxford University Press, Oxford, UK: 1996.

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

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

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

[8] 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.

[9] P. Rebelo and M. McKnight. Spaces in Between – Towards Ambiguity in Immersive Auditory Environments. *In Proceedings of the International Computer Music Conference (ICMC) 2019*. New York, USA, 2019.

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

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

[12] D. Terefenko. *Jazz Theory: From Basic to Advanced Study (2nd Edition).* Routledge, 2017.

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