ARTIFICIAL INTELLIGENCE AND THE GENERATION OF EMOTIONAL RESPONSE TO SOUND AND SPACE.

Paul Bavister Flanagan Lawrence / University College London UK
Stephen Gage University College London, UK

1 INTRODUCTION

The enjoyment of sound on its own, or as music, can be considered to be co-dependent on the 'site' of its production; as Wallace Sabine inferred in 1908;

'*Housed or unhoused, dwelling in reed huts or in tents, in houses of wood or of stone, in houses and temples high vaulted or low roofed, of heavy furnishing or light, in these conditions we may look for the factors which determine the development of a musical scale in any race, which determine the rapidity of the growth of the scale, its richness and its considerable use in single-part melody*'.[1]

Appropriately organised physical aspects of a host spaces architecture define the energy fronts that present reflections to the ears in an appropriate order. This allows different sounds to sound 'good' in different spaces, be they auditoria, or private salons. Analysis of successful acoustic phenomena relies on both objective and subjective methodologies for defining the physics of sound, but not the internal emotional impact. If music is defined by the space it was written for, the sizes and key dimensions of concert halls and music venues give a clue to what music would, and would not, be appropriate to play within it. These values have been traditionally defined by popular appeal, and have now set a historic precedent. Such precedents, combined with analysis of successful halls, have given rise to an objective acoustic theory that has defined the architecture of performance spaces since the turn of the 20th century.

This paper explores an artificial intelligence (A.I.) approach to generating sounds that are evolved to fit digital representations of a series of spaces that are recognized as having distinct but differing acoustic properties. Listeners were presented with a series of sounds in simulations of spaces, that were either accepted as successful, or rejected as unsuccessful. (Input from the listener being generated by galvanic skin response (GSR.)) Sounds that generated an 'excited' skin response were deemed to be successful. A 'successful' sound then goes on to seed the next generation of sounds using a genetic algorithm. After a fixed series of iterations, the sound(s) evolving to be considered as a good match between sound space and listener. This process allows for a divergent range of spaces to be considered suitable for listening to music; the question is 'what music?' This iterative process can be seen to be analogous to the 'chance' reciprocation of sound and space that has occurred naturally over the last millennia.

2 BACKGROUND

The enjoyment of Western originated music is often considered to be generated by a combination of the sounds of a musician, and the room that holds the performance; any emotional response created in listening is created by a blend of both architectural and musical qualities. Halls and rooms that garner the correct blend of responses are deemed successful, and rooms that don't are disregarded acoustically. This has led to an iterative development of

host space and music. Key to the success of any given space is the emotional response of the audience. It has been proven that an emotional response can be measured and quantified to a response to music[2], and work has been undertaken in linking emotional response to sound and space[3].

The historical relationship between sound and architecture has been one of conflict and adaptability, leading to a slow evolution of a 'best-fit' criteria defining a space and its occupation. Until early in the 20th century there was no fixed acoustic theory, leading to guesses and approximations of how things will sound[4]. Spaces would stand or fall on their success as an appropriate venue for sound, success being defined by popularity. Should a venue be popular, it would go on to host more events, and be the seed of a new precedent, or typology in venue design. Should a venue not prove popular, it would be removed from the gene-pool of spatial options. Occupation and use of such spaces was commensurate with its host, in that the sounds generated within the spaces would adapt with, and to, the surrounding architecture.

By compressing musical evolution into a series of small and rapid steps, a sequence of evolving sounds are optimised to fit a digital reproduction of a space in conjunction with the subconscious input of a listener. This process allows sounds to be developed that are not deliberately developed to fit a prejudged subjective viewpoint of the creator.

3 METHODS

3.1 Sites / Impulse response & convolution

In the investigation described in this paper, acoustic site data was taken from the following spaces:

- The Turbine Hall, Bankside, London
- The Switch House Foyer, Bankside, London,
- The White Cube Gallery, Southwark, London
- Alms Houses, Bankside, London.




The data was used to generate virtual representations of the host space following research into emotion and auditory environments undertaken by Västfjäll, D., Larsson, P., & Kleiner, M. (2002)[5].

These spaces were all chosen because of a distinct acoustic footprint, that allows for idiosyncratic sounds to evolve within them. Concert halls and specialist spaces were not chosen as the acoustic response to these spaces is already well established.

Sine sweeps were taken in each site, and recorded in B – format, resulting in a series of 4-channel IR's (Executed by Arup) Both source and receiver positions are a 1.5m off the ground, thus commensurate with the human ear, and an acoustic event, such as a human voice or musical instrument played on a platform.

All resultant IR's were loaded into the 'multi-convolve[6]' object for MAXMSP, and played back to the listeners in the Bartlett School of Architectures Sound lab, which currently comprises of 8 speakers in a cube formation, generating 1st order ambisonics.

3.1.1 Sound generation

Single sounds were generated digitally, using algorithms in max/msp. The sounds had to be able to evolve into an acceptable balance between the space and the listener, so complex pre-recorded music would not be suitable. The sounds should also not be too “electronic”, and fall foul of cultural prejudice against synthetic sounds. This ruled out common synthesiser typologies such as additive or subtractive synthesis. Frequency Modulation (FM) synthesis was chosen, because of the simplicity of initial generation of sounds and the complexity[7] that can emerge with a few very simple adjustments to the integers that drive the system.

In operation, FM synthesis works in the following manner; when a carrier oscillator is modulated in frequency by a modulating oscillator, a series of harmonic overtones or side band frequencies are generated over a fundamental tone, these can be sufficiently complex and harmonically rich[8]. When tuned correctly, the output of FM synthesis is complex enough to mimic a grand piano, so has capacity to mimic the complexities of many acoustic instruments and voices. Whilst not perfect, it is complex enough to suit the requirements of the tests that were undertaken.

Commercial applications of the technique use the collectives of carrier and modulator operators in algorithms. There are 32 in the DX7, each of varying complexity. To simplify the approach of the tests the system was restricted a single algorithm, numbers 5 & 6 in a DX7, that has 6 operators stacked in pairs. This generates enough complexity for the tests, and can result in metallic and glass like tones, mimicking sounds heard naturally.

The sound generation patch comprises of the carrier and modulator oscillators being driven by integers from a generative control system. The modulation index and amplitude are controlled by an ASDR envelope function. This enables a sound not only to have complex harmonic qualities, but also to have complexity in shape, thus a sound with a sharp attack and short decay can sound like a click, or snap, this will allow a space to respond clearly and decay in a pleasing way, like clapping in a cathedral, and listening to the echo. Conversely, the ADSR function allows a sound to slowing build, and then fade, and if there is sufficient break points in the envelope then further detailed rhythmic effects can be generated, such as tremolo and vibrato. The test software comprises of four FM ‘voices’ each is independent of each other, and when initially set up, represents four independent sound sources.

3.1.2 Generative systems

The field of artificial intelligence (A.I.) has given us a large tool set in which to operate, and there are positives and negatives with each approach. Machine Learning (ML) is proving a popular tool set to solve highly complex problems, but relies on vast amounts of data sets collated over time. Research undertaken by S., Horner, A., Beauchamp, J., & Haken in 1993[9] used A.I. to try to match FM synthesis algorithms to musical instrument spectra, but this is again seeking to build on and generate existent information, and not searching for new typologies. As the process defined in this paper is a ‘search’ process, a different, more evolutionary algorithm (E.A.) is required.

The test relies on an iterative process of four voices, or sounds, being played into the convolved space, with successful sounds seeding the next generation, thus after a series of iterations of the process, or 'epochs', a series of sounds emerge that reflect a balance between the space and the occupant. For this process, a genetic algorithm[10] (G.A.) a form of evolutionary algorithm produces 'off-spring' from an initial data set of randomly assigned set of integers that go on to modulate differing aspects of the synthesizer engine. The G.A. first seeds the initial integers that defines the voices, and the outputs the corresponding voices, playing them sequentially, with a 9000ms pause between each sound. This lets the sound naturally decay in the room, and allow the listener to appreciate the sounds impact on the space. Should a sound be successful in triggering an emotional response in a listener the next generation of sounds will be based on the sounds parents DNA, thus engendering an 'evolution' of sound in the space. If none of the presented sounds are successful, then the system resets itself with random integers and repeats the process until sufficient diversity allows a response to be generated.

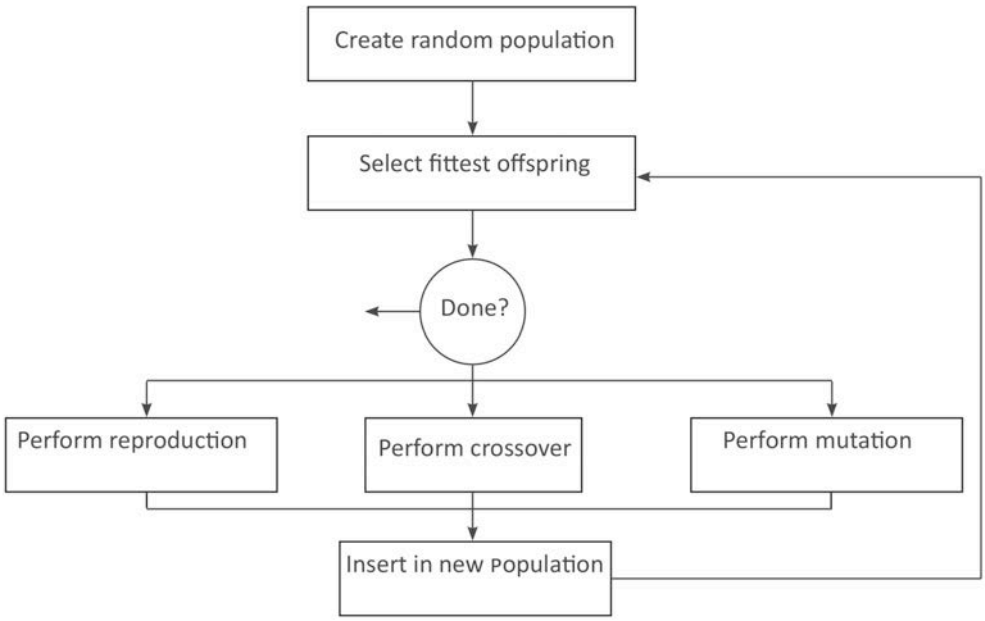

Fig.01: Recursive optimisation

As stated, the initial seeded integers are random, hoping to generate as much sonic diversity as possible. This will increase the chances of one of the voices being successful. Should a voice trigger a response, the range of the crossover and mutation functions are narrowed by a series of exponentially decreasing scaled percentile changes, there are 7, initially 100%, all open, to 50%, 30%, 20%, 10%, 5% and 2%, the last step offering very little variation at all from any parent object. This function seeks to limit the test process to a finite period, short enough to be 'interesting', yet being long enough to generate a clearly evolved outcome.

3.1.3 Biometric input

To signal a response in a listener, GSR sensors were placed on the medial phalanges of the left hands ring and middle fingers. This was identified as the non-working hand and less likely to give false readings. The outputs of the sensors were taken into the test computer via an

Arduino. This allowed the data feed to be collected in max/msp, so to align with the timing of the GA. In order to ensure that the skin Conductance Level (SCL) or tonic level from the skin did not get in the way of clear readings of the Skin Conductance Response, (SCR) or phasic response[11], the analysis software was programmed to constantly check the resistance levels between a fixed point in time, and 3 seconds previous. If there was a difference of +/- 4 *uS* then an emotional response was deemed to have taken place.

The output from the GSR was imported in to MATLAB and de-convolved using Ledelab analysis software to overlay against the evolved sounds.

Other tests involving GSR state very low durational time periods, Patynen & Lokki[3] cite a durational time of 24 seconds for each of the spaces analysed, so as not to be distracted, the imotions guide to GSR recommends a higher time period of 4-5 minutes[11]. In both cases, there is no feedback from the system to the test subject, creating an environment that may possibly lead to distraction.

3.1.4 Test structure

The test was undertaken in a sound lab with 8 speakers in a cube formation. The walls were fully absorbent, with little or no reflections from other surfaces. The listeners were placed centrally in this space. The listeners were given a minutes silence to acclimatize to the situation before the tests started. The listener was advised what was going to happen and why. Information on timing was withheld to stop anticipation of the end of the test from interfering with the results. After each test, the sounds were reviewed with the listener to determine if the results were positive, and well received, or if the emotional response was negative, and the sounds were difficult. This post test interview triangulated the results, and ensured that the outputs would be useful.

4 RESULTS

Over all 6 tests were undertaken, with 3 male and 3 female listeners, some of which were experienced listeners, others were not. Whilst there was no set time limit, the system reached an optimized peak after 2m 30s to 3ms, three key factors emerged:

- There is a clear random starting population pattern in the early part of the waveform analysis, which over the period of the test resolves itself into a repeating ‘optimised’ form. This is generally concurrent with all test results.

- There was generally a peak of dermal activity at about the ¾ mark of the test, as the algorithm starts to point towards a solution. This dies down a bit after saturation and recognition takes place.

- In some instances of high activity, the algorithm reaches its own optimum, and restarts with a new random population, starting the process again, evolving a new form.

One clear repeating element in the results that emerged, was a similarity of sound shape that occurred in all results. *(See fig. 03)* There was a clear split between two differing types of waveform shape; either a very sharp retort, or a long, ‘held’ tone. To borrow from musical parlance; note shape can be defined by the attack, sustain, decay and release in an envelope

position (ASDR). The markedly fast attack envelopes, and a long decay times shown in the results allowed a clear sense of the room to merge with the sound. The two contrasting envelope shapes either allows the sound of the space to ‘bloom' after the initial impact of the leading edge (attack) of the sound, or allowing the running reverberation to colour the sound. *(See fig. 05)* With further testing, this could prove beneficial in generating an acoustic signature for a given space.

Fig 03: Spectrum analysis of rest results

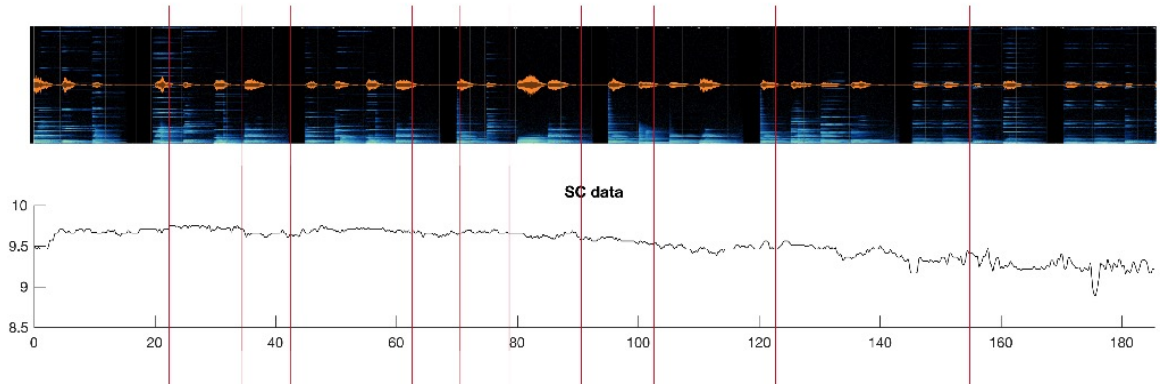

Fig 04: Spectrum analysis and SC data overlaid.

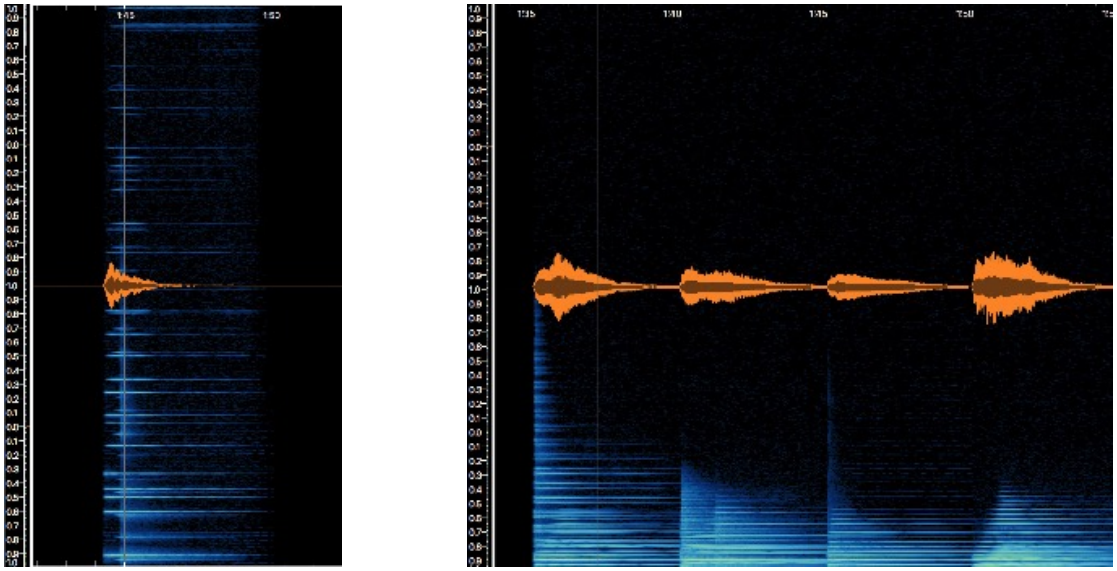

Fig 05: Spectrum analysis of rest results showing short sharp sounds, left, and long 'held' tones on the right

5 DISCUSSION

Previous examples of GSR tests cite a very clear time frame to limit lapses of concentration and fatigue. Sudheesh, N. & Joseph K. limited their experiments to 1000 seconds, but with a highly varied, and user recognisable source input,[2] Patynen & Lokki limited theirs to batches of 24 seconds, but with highly similar sounds[3]. In situations where a test subject is merely responding to an input that is not taking preference into account, it is understandable that a subject may fall tired and think about other things that are not related to the test, giving false readings and incorrect data. However, if the system is feeding back, and generating a clear engagement with the listener, then there is a possibility that the user may be more actively engaged than previously supposed, and the process may continue for longer, generating clearer results. Despite the similarity of source input, the average test here lasted 180 seconds.

In the post test interview, one listener stated that amid the random population of the starting condition sounds, a 'bell' type sound was recognized. We were advised that this was a reminded of home, and was a familiar reminder. As the test ran, this bell sound became more and more prominent, until it was a much more evolved sound, that matched the tonal qualities of a real bell. Subsequent analysis of the dermal results of this test showed that the familiarity of the sound triggered a response, that led to the algorithm focusing on the bell sound, the better it got at it, the greater the dermal response, until it reached a very 'realistic' conclusion demonstrating the spectra of a real bell. The bell sound was 'willed into existence' unknowingly. Whilst this is exciting, the emotional response was not related to the sounds relationship with space, but holds a much more psychological meaning, and as such, should be discarded from the overall output of the test.

6 CONCLUSION

The principles of evolutionary sounds adapting to a space via the agency of a listeners emotional response are unambiguous and are quantifiably demonstrated by this study. A space such as the Turbine Hall having a long reverberation time would be theoretically suitable to long held sounds, or sharp sounds with long room decay. Sounds with both of these qualities were present in all tests. Sounds with a high degree of detail were not successful; although they were 'interesting' they did not generate an emotional response.

Future iterations of the experiment will allow for more 'open' response to mutation and cross over functions; this way a listener can be gently coerced into a recursive path, rather the very focused route undertaken so far. Currently, should a sound generate a response, the system focuses on this single sound in ever increasing detail. This stems from the crossover and mutation functions decreasing in range exponentially, resulting a very sharp determination curve, and a short experiment. If these functions can be made to be more sensitive, and adaptive, responding according to 'strength' of emotional response for example, then the test times can be longer, and with more meaningful results.

It is concluded in this test that the evolutionary adaptation of the presented sounds against a fixed reverberant background increased listening time, whilst actively encouraging the attention of the user.

7 REFERENCES

[1] Melody and the Origin of the Musical Scale Author(s): W. C. Sabine Source: New Series, Vol. 27, No. 700 (May 29, 1908), pp. 841-847 Published by: American Association for the Advancement of Science

[2] Sudheesh, N. & Joseph, K. . (2000). Investigation into the effects of music and meditation on galvanic skin response. *ITBM-RBM*, *21*(3), 158–163.

[3] Pätynen, J., & Lokki, T. (2016). Concert halls with strong and lateral sound increase the emotional impact of orchestra music. *The Journal of the Acoustical Society of America*, *139*(3), 1214–1224.

[4] Blesser B. (2009) Spaces speak, are you listening? MIT Press P90

[5] Västfjäll, D., Larsson, P., & Kleiner, M. (2002). Emotion and auditory virtual environments: affect-based judgments of music reproduced with virtual reverberation times. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *5*(1), 19–32. https://doi.org/10.1089/109493102753685854

[6] Alexandre, P., Harker, A., & Tremblay, P. A. (2012). University of Huddersfield Repository Original Citation Toolbox : Convolution for the Masses . In : ICMC 2012 : Non cochlear Sound

[7] Chowning, J. M. (1973). The Synthesis of Complex Audio Spectra by Means of Frequency Modulation. *Journal of the Audio Engineering Society*.

[8] Rhoads, C (4th April 1996) The computer music tutorial: Curtis Rhoads, MIT Press.

[9] S., Horner, A., Beauchamp, J., & Haken, L. (1993). Machine Tongues XVI: Genetic Algorithms and Their Application to FM Matching Machine Tongues XVI: Genetic Algorithms and Their Application to FM Matching Synthesis. *Source: Computer Music Journal*, *1782238*(4), 17–29. Retrieved from http://www.jstor.org/stable/3680541

[10] Man, K. F., Tang, K. S., & Kwong, S. (1996). Genetic algorithms: Concepts and applications. *IEEE Transactions on Industrial Electronics*. https://doi.org/10.1109/41.538609

[11] imotions: Guide to GSR, (2016)