

3. Music, auralisation, and programming

*At length divine Cecilia came,
Inventress of the vocal frame.*

– Dryden: *Alexander's Feast*, 7

3.1. Introduction

Music is a rich communication medium being able to convey emotion, mood and meaning. The power of music should not be underestimated. Tradition holds that St. Cecilia's mastery of music was so great that an angel was enticed down from heaven by the charm of her melody.

Whilst sound has been used in interfaces, its use has largely been at a primitive level, often involving real-world sound (c.f. auditory icons [66]). Since music offers a highly structured set of mechanisms for communication it is surprising that there have been so few attempts to explore its possibilities.

Most output-oriented activity in HCI research has focused on visual methods and techniques and many investigations involving a large number of output media have been presented. On the other hand, the auditory channel has been exploited only to a limited extent. Since Bly's original thesis [19] on the use of sound in interfaces, the body of research has grown to the point at which there is now a regular international conference dedicated to auditory display. Much of the auditory display work to-date has concentrated on enhancing the graphical user interface with auditory components. The notion of the icon as a visual representation of some object or

task has been extended to provide the user with audio cues and feedback in the form of auditory icons [1, 66, 67, 68, 69] which are based on symbolic sound effects and earcons [15, 16, 17, 32, 34, 35], similar to auditory icons but based on short musical sequences.

Brown and Hershberger used sound to augment visual displays in algorithm visualisation [37] and to assist in physical process sonification [109]. However, the potential for using music as an output medium (the most sophisticated of the auditory media) has hardly been examined at all apart from some use of music in data analysis [117]. Even then the musical mapping consisted merely of quantising data points so that they corresponded to pitches in a 12-tone, or chromatic, scale. This chapter puts forward a case for the serious investigation of music as a useful communication medium, especially in the area of program auralisation.

3.2. A case for music

Alty [3] showed that music offers the potential for providing meaningful program auralisations and concluded: “*We now need some good experimentation to determine what is possible and practicable*” with program auralisation. To this end we developed the CAITLIN system, a musical program auralisation tool [152] for the purpose of carrying out experiments to determine whether music can be usefully employed to assist novices with debugging their programs.

The research was deliberately focused on musical auralisation. The majority of work carried out in the area of program auralisation has dealt with the use of non-speech audio and much has been said about the mappings of sound effects to system events. However, little effort has been directed specifically at using music in the human-computer interface. It has been suggested by Hotchkiss and Wampler [80] that music lends itself well to experiencing data and events subjectively. This, they claim, would give us a greater sense of participation or of being inside a function than is possible using more objective numerical representations. They rather boldly suggest that the sound of an executing program is an “*interesting example of the symphonies that may underlie the running of virtually every computer code.*”

Whilst particular sound effects may be very good at representing system events symbolically, music is an extremely powerful medium for the delivery of large

amounts of data in parallel. Musical techniques such as counterpoint and polyphony which are found in most modern musical styles show how distinct and separate musical ideas can be delivered in parallel without confusion on the part of the listener if certain syntactic and semantic rules are followed. Given that western musical structures, whether by independent evolution or by cultural imposition [128], are widely accepted across the world (even more so across the personal-computer-using world) it makes sense to investigate whether such musical styles can be usefully employed in program auralisations.

Alty [3] put forward a number of reasons why music should be considered for possible use in human-computer interaction. These reasons are herein given and the ideas expanded.

3.2.1. Music as highly structured information

A large-scale musical work contains much information. For example, a symphony will require hundreds of megabytes of storage on an audio compact disc. The information contained in musical works are (normally) very highly organised into hierarchies of complex structures and substructures. Much of this structuring can be made transparent to the listener who is predominantly concerned with enjoying the experience of hearing the music. Regardless of its transparency, the fact remains that this structuring of data is being used quite effectively to convey complex information to the audience.

There is potential here for valid research. A theory proposed by Schenker [143] and developed and studied empirically by others [50, 51, 99, 124, 149] proposes that perceived musical structure is internally represented in the form of hierarchies [51]. According to this model, hierarchical structure is how listeners perceive and represent the structure of tonal⁸ music. As computer programs and many of their associated data structures can be viewed as hierarchical edifices, it would be of interest to see whether the properties inherent in tonal music can be used to convey information about these electronic hierarchies. It would seem appropriate to explore

⁸ In experiments by Dikken [51] tonal music was successfully represented by abstractions or *reductions* which listeners could match with the original. However, such hierarchical representation did not work with *atonal* music (such as that of Arnold Schoenberg).

whether information about program structure can be mapped onto musical structures.

3.2.2. Music in everyday life

Music forms a large part of most people's lives. Melodies in particular are very memorable and durable; it is difficult to forget a (simple) tune once learned. Everyone has had the experience of hearing a tune on the radio in the morning and finding themselves still whistling or humming that tune in the evening. Pop music with its short and highly repeated melodic phrasing appears especially good at being remembered in this way. Once learned even quite complex tunes are very difficult to forget.

3.2.3. Parallel transmission of complex temporal data

Music involves the simultaneous transmission of sets of complex ideas related over time within a semantic framework. A successful composer often uses musical resources and techniques such that the listener is able to disambiguate this information. Devices such as counterpoint and polyphony allow the simultaneous presentation of multiple melody lines and musical ideas by employing different voicings and instruments. As long as the different melodies (which might be considered analogous to data streams) are delivered within a single framework or structure then synergy is achieved. Devices like the fugue, round and canon have a potential counterpart in recursive programming techniques according to Hotchkiss and Wampler [80].

Confusion arises when the different melodies work against each other by, for instance, taking non-complementary musical keys, time signatures or rhythmic patterns. When a common grammar and mode of expression is used by the composer then the musical components work with each other to convey ideas to the listener. This is one reason why the job of a composer is similar to that of the interface designer who must attempt to make all interface components adopt common design principles so that they too work together to the benefit of the user. Brewster et al. [32] demonstrated that earcons could be combined to deliver, in parallel, information regarding multiple interface events.

Music is a temporal medium rather than a spatial one, that is, the time ordering of musical events is what conveys the information. GUIs, on the other hand, are spatial media as it is the geographical placement on the screen of the interface items that governs how the information is presented. Of course, separation of parallel streams in music is assisted by spatial presentation. For example, stereophony which allows left-right placement of sounds in a two-dimensional audio field helps listeners to distinguish between different musical parts. However, the spatial mapping is subordinate to the temporal.

Walker and Scott [154] suggest that sound is more appropriate for processing temporal information whilst the visual modality is better suited to spatially-oriented information. According to Mowbray and Gebhard [121] we are very sensitive to changes in an acoustic signal over time which results in bringing new acoustical events to our attention and also allows us to relegate persistent or uninformative sounds to the background. Wenzel [157] thus suggests that audio is well suited to monitoring state changes over time. Anecdotes abound of people working with early computers who by placing an AM radio on top of the machine could tell from the changes in the radio interference when a particular batch run had finished or when it was stuck in a loop. Cohen [45] talks of “*the warm fuzzy feeling some people got when they heard their hard disc drive purring and knew it was okay [and] the shock and horror when it made unusual sounds*”. Almost everyone has had the experience of noticing the sudden silence when a road drill outside shuts down even though they had tuned the sound out over time to the point at which they were not consciously aware of it.

As program execution is itself a temporal phenomenon it is logical that attempts to render it with temporal channels are made.

3.2.4. Easy-to-use technology

The musical instrument digital interface (MIDI) language and protocols, developed in the 1980s provide interface designers with a straightforward mechanism for controlling electronic musical devices from a computer. It is now much easier for programmers to incorporate music into interfaces and MIDI provides an ideal interface between the computer and musical instruments for such work (pace Smith [146] and Scaletti and Craig [142]).

3.2.5. Widening access to HCI

Music offers a useful communication channel to programmers with serious visual impairments. It is certainly the case that auditory interfaces are assisting the visually impaired with tasks such as word processing (for example see Edwards' *Soundtrack* system [58]). This issue is particularly relevant today because the improvements in visual display technology that have given us the graphical user interface have helped to lower access to computing to blind users. Prior to the GUI, ASCII-based screen readers made it quite straightforward for users with visual difficulties to use common software. However, the GUI with all its sophistication requires a good deal of visual workload to use it. Efforts have been made towards opening up the GUI to the blind, including the *PC-access* system [57], *Audicon* [110] and *Mercator* [122].

Music may open up other areas of human-computer interaction to blind users. Kennel's *AudioGraf* system [88] uses a combination of audio and haptic feedback to enable blind users to identify graphic elements in a diagram. Rigas [7, 136] with his *AudioGraph* system has shown how information about graphical objects can be communicated to blind users via musical sequences which describe the shape of graphical objects.

3.3. Examples of musical data sonification

Brewster's work on earcons has shown that short musical sequences can usefully convey information about interface events to users. However, there is a large gap between using short musical cues to signal events and presenting information about large data sets (which include program state) in a musical framework. The earcon is the best known musical device used in auditory display research, but there are several other noteworthy examples of the use of music in the field.

3.3.1. Structures of chemical compounds

Lunney et al [104, 105, 106] have successfully used musical chords and other audio techniques to assist visually impaired students take part in chemistry experiments. They discovered that procedures such as infrared spectrometry and chemical titration can be carried out by the visually impaired using audio feedback.

Musical chords were used to represent the infrared spectra of chemical compounds. Each spectral peak was mapped to a pitch, and the peak's height (intensity) was mapped to note duration. When the pitches for each spectrum were sounded together as a chord the result tended to be jarring to the ear. However, for simple organic compounds the chords provided sufficient information for chemists to identify the substance. The motivation behind this research was to assist blind chemists in their work. Although the musical mappings are very primitive and take no account of musical style or grammar, it is encouraging that even a crude mapping can convey sufficient detail of some quite complex numerical data for accurate inferences to be drawn. Whilst the chords would probably not allow the listener to describe precise values (such as the absolute intensity of a given spectral peak) they serve to show that musical mappings are able to convey meaningful information about quantitative data.

Yeung mapped vectors of multivariate analytical chemistry data to several dimensions of sound [162]. After training Yeung's subjects achieved accuracy levels of 98% and above when asked to listen to the audio mappings of forty data files. Of particular interest in this study was the detection by one of the subjects of an indexing error in the computer program used. That this error was not detected by the programmers and that it was subsequently brought to light after listening to a sonification produced by the program is an encouraging indicator that debugging with sound is a field that may be worth exploiting. Yeung goes as far as to state that "*audio representation of multivariate analytical data, as shown here, is superior to the known visual methods.*"

3.3.2. DNA gene sequencing

The suggestion that the one-dimensional structure of DNA can be mapped onto musical sequences led to the development of the DNA analysis program *PC/Gene*[™] which utilises the Hayashi and Manakata algorithm [76] to map DNA triplets to music. Subsequent developments for amino-acid gene sequences [75] required four octaves to map the necessary information and, according to King and Angus [89], this led to music with large jumps of pitch that were distracting and discordant. Furthermore, the mapping itself was one-dimensional leading to a monodic musical structure without any accompaniment. A harmonic polyphonic mapping was

achieved by King and Angus [89] who mapped an amino-acid sequence onto a two-part harmonic structure using their own program called *PM* (Protein Music). In this mapping⁹ a DNA nucleotide sequence forms the top, or melody line, and the properties of the sequence's amino-acids make up the bass part. They argued that to achieve an equivalent visual mapping (in terms of information presented) would “*be cumbersome as each position would have to be mapped into seven colours. This ability to display multivariate residue information represents an advance in this work*”.

3.3.3. Chaotic attractors

Meyer-Kress et al. [112] took features of chaotic systems, such as intermittency and self-similarity, and attempted to draw analogies to musical structures. They state that data sonification “*produces abstract sound patterns, and music is created from complex non-linguistic sound structures. It is relevant to investigate cases where the organisational principles of music might help reveal patterns in scientific data*”. By mapping the low-level sequence of system states onto auditory parameters and high-level attributes (such as intermittency and self-similarity) onto polyphonic auditory constructs they were able to use the data generated by a chaotic system and represent it musically. The music thus generated is pleasing for two reasons: at the aesthetic level it possesses qualities that recommend it as a piece of music in its own right— that is, one could actually listen to it for enjoyment. Secondly, by attending to the development of the music over time one can visualise the underlying structure within the chaotic system. For instance, the property of self-similarity comes across in the form of musical phrases which repeat themselves but of which no two are ever exactly alike.

3.3.4. Diagram readers for the blind

Rigas demonstrated with his *AudioGraph* tool [5, 135, 136] that information of a spatial nature can be communicated to the visually-impaired using musical devices. *AudioGraph* allows blind users to manipulate graphical objects on a screen by

⁹ One might question whether such scientific auralisations could ever sound truly musical. However, the mapping described has in fact sold several thousand copies by appearing as the track *S2 Translation* on the Shamen's album *Axis Mutatis* [150]. That the track is both listenable to as a piece of music in its own right and that it successfully aids in biologists' visualisation of gene sequences is very encouraging.

giving information about cursor position in musical tones (mapped onto diatonic and chromatic scales). The system employs musical mappings to describe the outline shape of the simple graphical objects. Experimentation showed that shape recognition using AudioGraph was very high after minimal training. Recognition increased further when hints about the objects represented by the shapes were given, illustrating the importance of context in metaphorical mappings. Facilities for scanning the neighbourhood of the cursor are also provided by AudioGraph.

Encouragingly, Rigas claims [136] that only 20% of blind users who performed the experiment would prefer not to use musical techniques in a general-purpose drawing package.

3.4. Arguments against the use of music

Whilst music offers much potential to be exploited, there are some factors, enumerated by Alty [3], that must be taken into consideration that some might argue militate against its use. These include:

- Whilst music can convey complex information, such information is of an emotional nature and therefore not usually appropriate for interface design.
- Music is too closely linked to local culture and therefore any musical auralisation will have limited appeal.
- Sound is an intrusive medium and so is not suitable for information transfer in situations where computer users share work space.
- Quantitative information cannot be conveyed by sound or music.
- Only trained musicians can appreciate and understand information conveyed musically.

The argument against music on a cultural basis is an appealing one. However, Alty [3] has commented that so-called western music has been well absorbed by many of the world's cultures. The basic musical scales used in western music (chromatic, diatonic and pentatonic) have a higher common usage than either Roman script or Chinese ideograms.

Sound can easily be rendered unobtrusive by the use of headphones, although there are implications for the user in terms of his relationship with the outside world.

The question of representing quantitative data is more problematic. The argument goes that whilst most individuals can tell if a note increases or decreases in pitch, only trained musicians are able to determine exact intervals with any accuracy. However, quantitative information can be meaningfully described and presented in terms of its overall magnitude without needing to know its exact, discrete values. In visualisation the recognition of patterns can be of great value and patterns can be spotted by observers with widely differing visual skills [112]. Thus, it is not always necessary to be able to discern discrete values absolutely.

For instance, just as an unmarked thermometer allows visual judgement of the relative magnitudes of various temperatures, so musical pitch enables aural gauging of the relative values of different data. Interestingly, nature already provides an acoustic thermometer in the form of the striped ground cricket *nemobius fasciatus fasciatus*. By listening to the chirps of this cricket it is possible to predict ambient temperature [130]. When making 20 chirps per-second we know it is 88.6°F; 16 chirps per-second denote an ambient temperature of 80.6°.

Hotchkiss and Wampler [80] argued that although numerical data gives the most objective representation of an event from one moment to the next, the continued increase in size and complexity of data which computers can generate make it increasingly difficult to gain real understanding from a strictly numerical approach. They observe that even graphical representations of functions are not wholly objective because graphs are not intended to give a quantitatively objective value for every point on the curve; instead they aim to give an understanding of the curve in outline. Music, they say, lends itself to subjectivity in which “*one experiences the event in a more personal and intuitive way ... there is a greater sense of 'being inside' a function or participating in it*” [80].

Cambridge University's COAST telescope system illustrates just how useful acoustic feedback can be [13]. The COAST apparatus generates an image by combining and focusing signals from three separate telescopes. To achieve this focus, the three signals must be in phase which entails adding path length to one or two of the signals. In theory, the extra path necessary can be calculated, and simply introduced by moving a mirror to the right position (a movement of a few microns). David Burns of the COAST team explains that [40]:

“... the telescope is subject to thermal drifts etc., and the density of the atmosphere is also constantly changing, hence introducing or subtracting relative path between the telescopes, and the fringes are never quite where you predict. In order to find the fringes amidst a strong background and noise, we cause the positionable mirror to oscillate. The period of oscillation is calculated and set such that if it moves through interference fringes, it does so at a frequency of around 700Hz. We simply then play the output from our detector back through a loudspeaker, and listen for a 700Hz signal. If we do not hear a signal, we conclude that the mirror is not quite in the correct position, and move it slightly, while it is still oscillating.”

Moving the mirror slightly actually entailed adjusting the mirror’s position by only a few microns. It is interesting to note that the COAST array delivers higher-resolution images than the Hubble Space Telescope.

The objection that non-musicians will not comprehend musical auralisations is an appealing one in the same way that the argument that non-programmers are unlikely to understand the significance of different program structures makes intuitive sense. However, just as non-programmers readily use the outputs of and interact with very complex software systems, in everyday life non-musicians happily listen to music and people who cannot sing regularly enjoy musical recordings. Therefore, to suggest that non-musicians are necessarily barred from understanding musical feedback is not reasonable. Brewster et al. [35] showed that not only are earcons (which are musically-structured) an effective means of communicating complex data, but also that non-musicians are as likely to comprehend an earcon’s message as trained musicians.

3.5. The importance of audio in interface design

The audio channel is an important information channel, and has a long established tradition of conveying rich meanings. There is another important reason for encouraging more research in the area. The explosive growth in visual interfaces (such as GUIs) has put visually-impaired people at a considerable disadvantage. Most interfaces now carry an implicit assumption that the recipient has excellent visual acuity. Such a situation is particularly upsetting because blind and partially-sighted users were, until recently, making great strides forward in using computers successfully, (for example, using screen readers). Unfortunately, screen readers cannot easily describe GUIs. In an ideal world one would envisage an interface de-

sign approach that could be accessed by audio alone, by viewing alone, or through a combination of the two (and all under user control).

Another related point is the current overcrowding of visual interfaces. Audio could have a real use here in supplementing the visual interface in a complementary manner.

3.6. Mapping between program domain and music

The key issue in using music is how to map domain entities to musical structures. Experiments by Alty [3] have shown that algorithms such as the bubble sort and minimal-path algorithm can have information about their run-time behaviour communicated successfully through musical mappings. The results suggest that, provided precise numerical relationships are not being communicated, music can transfer information successfully.

Just by using musical tones alone, the AudioGraph system has been used to describe multi-element graphical objects to blind users [5, 7]. After short training periods users were able to use the system to locate, move and modify graphical objects. Alty and Rigas assert that musical messages should be designed within a consistent framework (much as elements of successful graphical user interfaces follow common design principles).

3.7. Using music in program or algorithm debugging

The use of music in algorithm auralisation also suggests another possible use—that of program debugging. In a sense the application is the same, the algorithm being auralised is simply faulty.

The possibility of using musical output for debugging has already been suggested by Francioni et al [62]. They point out that the musical representation can highlight situations which could easily be missed in the visual representation (no doubt there are also cases where the opposite is true). For example, a move of one semitone of a note in a musical chord can change the whole sense of that chord and produce an immediate and compelling effect. This happens when a major triad has its mediant flattened by a semitone to produce a minor chord (see Figure 3.1). A similar movement in the value of one data variable in a graph might not be noticed.

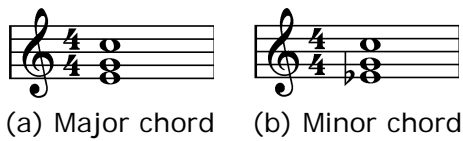


Figure 3.1 A semitone (Trk 01)¹⁰

In (a) we have a first inversion of a Gmajor triad. The lowest note of the chord is an E. In (b), the E is changed to an E flat which turns the chord into a Gminor triad. Although the pitch shift is small (a frequency shift of just under 6%) the effect is very noticeable. This example can be heard on the accompanying CD on track 10.

In the program debugging situation, the richness provided by a musical representation may offer fairly precise bug location possibilities (whether used in isolation or in conjunction with the visual media). One obvious possible mapping is to map the tracing of the execution path through different modules to different instruments. This is an area which is not handled well by visual media, causing frequent screen shifts. Using timbre the switches between modules are capable of being followed easily, allowing the visual sense to concentrate upon program detail. Indeed, in very large programs, visual debugging methods as well as being tedious can be misleading in pointing to the real problem source; giving sound to a running program can give a broader picture of what is happening [80].

3.8. Summary

Within the (still relatively small) research field of auditory display the amount of work involving the study of music as a useful communication medium is scarce. However, the ability of music to convey temporal (and even spatial) information in parallel data streams within a coherent and logical structure and syntax indicates that it offers much potential as a means of conveying information to the computer user. Music has drawn the criticism that certain of its characteristics would militate against its use. However, with careful attention to data mappings and environment these limitations can be overcome to a large degree.

The works of John Cage (see section 2.4) illustrate that meaningful relationships between data sets and music can be described allowing the data-controlled

¹⁰ This example, like many others may also be found on the accompanying audio CD. The number **Trk 01** denotes that this example is on track 1 of the CD. A full track listing can be found in Appendix F.

production of acceptable music¹¹. Such relationships have been utilised in the visualisation of DNA amino-acid sequences and the structures of chaotic systems.

Studies by Rigas have determined empirically that certain musical features can be shown to convey information well. Also, his work highlights a major application area for music-based auditory display, that of assistive technologies for the visually impaired who cannot rely on visual interaction techniques.

¹¹ What is deemed acceptable and what is not is an area of musical criticism that we do not propose to enter in this thesis. For the purposes of this research we will make do with Edgar Varèse's definition of music as *organised sound*.