


This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.




creative commons
COMMONS DEED


Attribution-NonCommercial-NoDerivs 2.5


You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.


 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Can We Use Music in Computer-Human Communication ?

James L. Alty

*LUTCHI Research Centre, Dept. of Computer Studies,
Loughborough University of Technology, Loughborough, UK*

Tel: +44 01509 222648

Fax: +44 01509 211586

Email: *j.l.alty@lut.ac.uk*

The audio channel has been somewhat neglected in Human Computer Interface Design. It is a powerful channel which offers processing options often of a complementary nature to the visual channel. Music makes the most complex and sophisticated use of this channel and has well-organised techniques and structures for disambiguating parallel time-dependent events. This paper examines the contribution music might make to interface design and reports on some preliminary investigations, which indicate that there does seem to be a prima facie case for examining the subject further.

Keywords: Auditory I/O, Audiolisation, Music, Design, Multimodal Interfaces.

1. A Prima Facie Case For Music In Interface Design ?

The visual medium has been extensively exploited in Human Computer Interaction. A large number of output media are available to the interface designer such as text, graphics, animation, pictures or moving pictures (in monochrome or colour and used singly or in combination). The auditory medium, on the other hand, has been exploited to a more limited extent. Although voice output and artificial or natural sounds are now used fairly routinely, the potential of music as an output medium (the most sophisticated of the auditory media) has hardly been examined at all. Gaver (1989) has developed the SonicFinder system which uses natural sound to indicate the state of the natural environment. Some work has been reported on using music in data analysis (Mezrich et al, 1984). Initial attempts have been made to design and build the aural equivalent of visual icons, or "earcons" (Blattner et al, 1989; Gaver, 1986), and some use of music has been made in message transmission in parallel machines (Francioni et al, 1991). The auditory channel

has also been used to augment the visual channel in algorithm visualisation (Brown and Herchberger, 1992) and to assist in physical process audiolisation (Blattner et al, 1992).

One area where the auditory channel is likely to make a significant contribution is for the visually impaired. Edwards (1989), has used musical tones in his Soundtrack Word Processing System for the blind. Auditory objects replace visual objects and the user interacts with an *auditory screen*. Tones are used to identify objects (which can then be identified using synthetic speech if the user requires this). However, in experiments with Soundtrack, although the use of pitch was not found to be intrusive, the majority of users (apart from one trained musician) did not use pitch to encode spatial information. Brewster (Brewster et al, 1994) has reported the use of an auditory enhanced scroll-bar. A tone was used to identify the position in the document which varied in pitch. Increases in pitch identified page boundaries. Scroll bar movement was also identified using pitch. The use of this auditory enhanced scroll bar improved performance time on certain tasks.

Music has been suggested as an input medium (Liebermann, 1986) and some interesting work has also been carried out in using interfaces to enhance musical understanding (for example, HarmonySpace, Holland, 1992) but these aspects will not be discussed further in this paper.

Music involves a highly sophisticated use of the auditory channel and there are some obvious reasons why the use of music should be considered for possible use in Human Computer Interface design:

- The information contained in a large scale musical work (say a symphony) is very large (a typical audio CD contains many hundreds of megabytes). The information is highly organised into complex structures and sub-structures. The potential therefore exists for using music to successfully transmit complex information to a user.
- Music is all-pervasive in life and forms a large part of people's daily lives. It is very memorable and durable. Most people are reasonably familiar with the language of music in their own culture. Once learned, tunes are difficult to forget.
- Music involves the simultaneous transmission of a set of complex ideas related over time, within an established semantic framework. The job of a composer is to use musical resources and techniques to enable a listener to successfully disambiguate such information. There is therefore a strong parallel between the design requirements of the interface designer and those of a music composer.
- The development of the MIDI system (Moog, 1986) has considerably eased the task of incorporating music into computer interfaces.
- Music may offer an important communication channel for blind or partially-sighted users (Mansur et al, 1985).

On the other hand people will argue that music is unsuitable for use in interface design for the following reasons:

- Although it is true that music can convey complex information, such information is of an emotional nature and therefore not usually appropriate for interface design.

- Music is closely identified with local culture. This would mitigate against its use on a wider basis
- Sound, by its very nature is an intrusive medium, and is not suitable for general information transfer.
- Quantitative information cannot be conveyed by sound or music.
- Most people do not have the musical ability to use music effectively as a communication medium

Although these arguments have some validity in certain contexts, many could equally be levelled against the visual medium. For example, paintings and plays are high transmitters of emotional information, but no-one uses this as an argument against use of the visual medium in interface design. Of course, the visual techniques used in interface design are at a much lower level (i.e. less artistic) than those commonly employed by authors, painters or playwrights. But it is highly unlikely that the use of music in interface design would be anywhere near the emotional levels used by music composers either. Interestingly, it could be argued that interface designers, hitherto, have not taken proper advantage of the communication experience of painters and poets, and that many mistakes might have been avoided if programmers had realised both their own limitations, and tapped the experience of others, in grappling with the visual medium. Obviously all musical techniques will not be appropriate for the limited purposes of interface design, but the right combination of techniques might yield a useful communication medium.

Perhaps we are mistaken anyway. Minsky has argued that

"our culture has a universal myth in which we see emotion as more complex and obscure than intellectin fact I think today we actually know much more about emotion than about reason" (Minsky, 1989).

It is true that there is no musical equivalent of standard speech or "non-artistic" written communication (i.e., what the average person does on Email or in letters, or in technical documents, though it could be argued that speech intonation comes fairly close). However, why should there not be a lower "less-artistic" communication medium using music ? Video games seem to use such a primitive form of musical communication in parallel with the visual output.

The local culture objection can likewise be levelled against the visual medium, and, just as different cultures offer different opportunities for visual communication, so this will also be true of music. It might be argued that the so-called "western" musical culture has a wider spread than, say, the use of English as a first language and it is certainly not confined to the Western world. Musical appreciation and usage of such music in Japan, for example, is higher than in many western countries. The standard set of musical notes (Diatonic, Chromatic and Pentatonic scales which are based upon simple ratios) have a higher common usage in the world than either Roman script or Chinese ideograms.

It is true that sound is an intrusive medium, but used through earphones, its use would be no more intrusive to others than the visual medium. The more difficult question is what level of sound would be unacceptable for an individual. There is no doubt that excessive use of sound could be irritating, so an important task is to determine rules for appropriate usage.

The criticism about music's inability to cope with quantitative information is a valid one. Whilst most individuals can tell if a note increases or decreases, only trained musicians are able to determine intervals with any accuracy. However, whilst this criticism provides limits for the application of music, it does not proscribe its use. There are many visual tasks which are difficult to do.

The degree to which human beings can use music for communication is an open question. People can be trained to use music more skilfully and to appreciate more subtle changes of pitch and timbre. To be of any use in human computer interaction however, a basic ability to distinguish different pitches, intervals and rhythmic patterns would have to exist in the "untrained" listener. At present we are trying to determine how general such skills are, both by literature search and by experimentation. For example, it has often been postulated that musicians process music preferentially with the left hemisphere whereas non-musicians use the right hemisphere (Bever, 1980). However, experiments examining the processing of timbre and rhythm by musicians and non-musicians (Prior and Troup, 1988) have concluded that, with stringent experimental conditions, there is little evidence for the hemisphere effect.

The richness of the music channel (it has been claimed that a human being can differentiate 20 dimensions of sound, Yeung, 1980) and its potential for use with the visually impaired, are enough reasons, by themselves, for further research into its use in interface design. However, there does also seem to be a prima facie case for exploring the possibilities offered by music, whether used on its own or in combination with the visual media, for those of us who have normal or near-normal visual capabilities. What has been lacking has been a detailed discussion of the relationship between musical building blocks and interface requirements and experimental investigations into the use of music as an output medium.

2. What Does Music Offer ?

The discussion here will be restricted to so-called "western" musical culture because it is the musical culture which the author knows best, though many of the comments will be valid (perhaps with some transformation) in other cultures. The areas where musical cultures differ significantly may, of course, provide an exciting area for new interface design ideas.

The basic building blocks of music are pitch (i.e. frequency), timbre (i.e. the instrumental sound), loudness, duration, reverberation and location (i.e. stereo placement). Some of these building blocks might be considered to have visual counterparts - timbre (visual texture or colour), loudness (brightness), duration (the same), and location (the x.y axes). Pitch might be thought of as equivalent to colour. However, these mappings are not simple. Sometimes the visual equivalent is richer, sometimes not. Pitch, for example is more versatile than colour for a number of mappings. Using a "height" metaphor, pitch can represent upto 70 or 80 values with a well defined ordering, whereas in colour, such an ordering is absent. Musical timbre is almost certainly more memorable and recognisable than visual texture. On the other hand, location using pixels is far superior to stereo sound location. The next higher level building blocks in music involve chords (harmony) and more complex textures (orchestration) and these might be compared with more subtle use of colour (saturation and hue) and visual composition. One might argue that the basic visual building blocks of music form a richer set than the musical ones.

At this point however, music and the visual media diverge significantly, mainly in their exploitation of time. In music, new building structures are possible - notes played simultaneously in chords (harmony), rhythm (repeated sets notes with different durations), and polyphony (the playing of simultaneous independent musical parts). It is difficult to find equivalents in visual media because the two media use time in a different, and complementary way. The auditory channel, for example, can cope relatively easily with a number of simultaneous events happening in parallel, whereas the visual channel tends to focus in upon a small area (whilst being aware of peripheral events). On the other hand, the auditory channel operates as a continuously moving sampler and cannot easily reflect or replay a recent set of events, whereas the visual channel can efficiently scan recent history and reflect upon it.

Music is a temporal medium and relies on time for many of its effects. The limited time for reflection in the auditory channel has forced musical composers to develop some highly sophisticated techniques for coping with short term memory limitations (Alty, 1995). The repetition of rhythmic phrases and their transformations form the basis of many western musical compositional techniques. One highly sophisticated device is the use of imitative polyphony (or the fugue) in which a set of identical phrases are combined in a staggered sequence commonly known as a "round" (A well known round is "Frere Jacques"). Although the auditory channel, at one level, can "hear" the result as an overall effect, it can also, at will, decide to listen to one particular part. There is no exact equivalent in the visual channel. A person can scan a painting time and time again examining different aspects and relationships, but in the ear this is done in real time and in parallel.

This use of sophisticated techniques to disambiguate simultaneous sequences of events is at the heart of musical composition. In opera, for example, the composer uses a number of techniques to allow a listener to disambiguate an ensemble such as location, timbre (soprano, contralto, tenor, bass), polyphony, harmony and rhythm. Use of these musical techniques might be used to allow observers to disambiguate the vocal utterances of a set of co-operating robots (this example was suggested by Prof. Michael Brady of Oxford University).

At the highest level, music offers a rich set of higher level organising structures to carry larger scale works such as the Sonata, Passacaglia, Variational form, the Partita and the Rondo. Although at first sight there appears to be a huge number of such structures they can be classified into only six different basic types (Kennedy, 1991). Such high level forms may not be of use in interface design though Minsky has compared the structure of a Sonata to that of a Teaching Machine (Minsky, 1989).

One valid application of music which will not concern us in this paper is the use of music as a transformation mechanism to aid understanding from other data forms without the conscious use of structure. An example of this would be Meijer's system which converts a 64 by 64 pixel image into columns of sound with the vertical position being related to a note and the brightness to the loudness (reported by Mestel, 1994). Meijer does not know how successful the brain will be in finding a use for such data input, but the brain is exceedingly versatile and may be able to make sense of the input with training.

The reader is referred to Sloboda (1985) and Deutch (1982) for interesting and readable books on the psychology of music.

3. Using Music In Algorithm Audiolisation

Because of its facility for handling simultaneous events through time, one obvious possible use of music is the auditory equivalent of the visualisation of computer algorithms (the terms Audiolisation (Blattner et al, 1992) or Auralisation (Digiano et al, 1992) have been suggested for this process though they are rather clumsy words). The words are meant to imply a close mapping between what is happening in the algorithm and the sound output. Such Audiolisation can be used on its own or in addition to the visual output.

One example which has been demonstrated is that of the Bubble Sort Algorithm. A combined visual/musical rendering of this algorithm was produced by Brown and Herchberger (1992), using the ZEUS algorithm animation system (Brown and Herchberger, 1991). Their audiolisation was of a preliminary nature, using different instruments to describe the movements of different elements in the list. However, the music was used to enhance the visual activity rather than to describe the activity in isolation

In order to test out if there was a possible use for music in algorithm understanding, two examples were chosen and "audiolised" - the bubble sort algorithm and the travelling salesperson algorithm. Although the bubble sort had been done before by Brown and Herchberger, their implementation did not solely use the audio channel and therefore did not fully exploit the potential offered by a musical approach. Furthermore they did not report any audience reaction to their audiolisation nor the reasons for the mapping chosen. It was therefore decided to re-implement this algorithm, paying careful attention to the musical mapping employed, and testing its effectiveness on people with an average musical ability.

The objectives of the tests for both algorithms were two-fold. Firstly, the actual process of transformation was of interest - how might a designer approach the task of audiolisation? Secondly, how would human beings react to such audiolisations? It must be stressed that both tasks were attempted in order to provide feedback as to whether a more systematic approach might be considered.

Both algorithms were constructed using TurboPASCAL, driving a SoundBlasterPro Card. The timbres were realised through a Roland MT32 multiple timbre device which was also capable of controlling stereo placement.

3.1 Pure Auralisation - The Bubble Sort

The key element in the success or failure of an auralisation is the mapping from the algorithmic elements to the musical elements. The first part of the process identifies the key aspects of the algorithm which need to be communicated. Next the mappings to musical building blocks need to be considered and will usually involve the use of metaphor.

3.1.1 Information to be Communicated

In the Bubble Sort algorithm, a set of numerical values is progressively sorted into either an ascending or descending order. The algorithm progresses through the list of elements to be sorted, swapping any two elements which are out of place. When it reaches the end of the list, it repeats the process until no swap takes place during a pass. This lack of swapping signifies that the sorting is complete. The information to be communicated must therefore include - the current state of the list, the progression of the algorithm through the list, the swap of elements and successful termination.

3.1.2 Musical Mapping

In the auditory mapping of the Bubble Sort, the listener must be able to hear the current state of the partially sorted list and this should be heard before the start of each pass through the list. The most obvious musical equivalent of rising numerical values is pitch, so a metaphor mapping number to pitch was chosen. The value of an element in the list is therefore converted to a pitch based upon middle C using the Diatonic Scale. Thus a 7 is heard as B and a 12 as G'. Four different pitch mappings were tried (Chromatic, Diatonic (Major and Minor), and a scale based upon the Leading Note). The timbre chosen was the celeste.

Next, progress through the list needs to be heard and is simply the current loop integer variable value. This was also mapped in the same way as the list values, but a Harp was used to provide a contrasting timbre (though the harp and celeste are never actually heard together).

The third aspect which needs to be heard is the swapping of elements in the list. This is heard in parallel with the ascending harp. To disambiguate it, a set of trumpets playing a major triad (as a triple with the lowest note as the list element) is played, first on the highest element of the pair, then on the lower, and finally the higher one again.

The three timbre outputs are separated using stereophony, the Harp on the right, the Celeste in the centre and the Trumpets on the left. Disambiguation is thus achieved by:

- Rhythm, using a triple
- Timbre using brass, harp, and celeste
- Harmony using a major triad
- Stereophony

The timing was adjusted to provide an even brisk pace (scherzo) and the end of the sort (in which the celeste is heard rising uniformly, followed by an uninterrupted rising harp) is marked by a standard cadence (i.e. like a sung amen).

3.1.3 Results

This algorithm has been demonstrated to over 600 people, usually in groups of 20 or 30 but occasionally to groups in excess of 100. No visual effects are used (apart from one static overhead transparency briefly describing the algorithm). All groups to which the algorithm has been played expressed a preference for the Diatonic Scale, though all scales worked reasonably well (the Chromatic scale has the advantage of offering many more numerical options). In all cases, groups expressed pleasant surprise at the effectiveness of the demonstration, frequently applauding at the end. The applause is somewhat surprising and may stem from the similarity of the output to a musical performance. The stereo effect, of course, is more difficult to achieve to larger groups.

3.2 Pure Auralisation - Minimum Path Algorithm.

This algorithm is a variant of the travelling salesperson problem - that of finding the minimum (or maximum) path to visit a number of towns exactly once, ending at the starting town. One "satisficing" algorithm that can be used is that of choosing the minimum currently available path

from any town. Thus if there are ten towns, at the start there are nine possibilities. The minimum of these is chosen and the process is repeated (at the second town there are eight possibilities) and so on. At the final town there is only one path choice back to the start. The algorithm is not guaranteed to give the minimum path, but usually provides a very close result. The maximum path is determined in the same way, by choosing the maximum exit path at any town.

3.2.1 Information to be Communicated

For each town, the set of valid output paths needs to be communicated.

The process of choosing the minimum from the set needs to be communicated.

The current state of the minimum path (and its build up) should be communicated.

3.2.2 Musical Mapping

During the process of finding the minimum exit path, the listener needs to hear the current best minimum and the length of the current path being considered. This is therefore done for each path. The same mapping from number to pitch is used to communicate the length of path being considered and the current minimum. Disambiguation is required between the path and the current minimum value. It might be thought that timbre could be used to achieve this (i.e. use a violin for the path and piano for the minimum). However, listeners can experience difficulty in comparing pitches across timbres and because the listener needs to hear the pitch being "transferred" to the minimum value, a change of timbre might mask this (indeed some early experiments showed that this did happen).

Disambiguation between path and minimum is therefore achieved by:

- Stereophony. The current path is played on the right and the minimum value on the left. This has an added advantage that a transfer appears to cross from right to left when a new minimum is found.
- A single note is used for the path, and a shorter triple use for the minimum value.

When the minimum is found, an organ is used to sound the current minimum path (its length increases by one each iteration). The organ is placed centrally between the other two sounds. At the conclusion the organ plays a cadence.

After the minimum path is found, the maximum is also found and the two paths are then played sequentially in two-note chords to show the difference.

3.2.3 Results

This algorithm has been demonstrated to about 50 people. Again, group response to the algorithm was very positive particularly for the Diatonic scale version. The algorithm needed more training however than the Bubble Sort perhaps showing that the musical mapping was not as obvious. The playing of the maximum and minimum paths together baffled groups completely. This is almost certainly because there was a lack of any obvious metaphor in this transformation.

Copies of both musical algorithms are available on cassette tape from the author.

4. A Preliminary Investigation

In order to try to explain why these audiolisations seemed to work, an investigatory experiment was carried out to gauge reaction to tasks which require the interpretation of musical output. In the first task, 12 male and 3 female subjects (aged between 21 and 42) were asked to estimate the "numeric" difference between two tones. The first tone was always middle C and the second tone was taken from the major scale above middle C (one octave). Thus there were eight different tones in all and the maximum interval was an octave. Sampling was limited to one octave because it is known that subjects have exhibited difficulties with intervals of greater than one octave (Deutch has represented this as a spiral, for example (Deutch, 1982). In each case, the major scale was played first (using a clarinet output from an MT32 multitimbre system to ensure consistency of judgement) and then the two tones were sounded three times (separated by 2 seconds). Subjects were asked to estimate the difference, so that their answers would range from 0 to 7. The questionnaire was designed to identify musicians within the subject set.

The results are quite interesting. Figure 1 shows subjects response (Y axis) plotted against the stimulus. All columns do not add up to 14 because one subject did not answer all the items. Two stimuli were presented twice (1 and 6) at the beginning and the end of the session but no learning effect was observed. There is no interval of 3 due to an error in presentation. 62% of answers were correct and there is a positive correlation between stimuli and perceived value (Pearson $r=0.81$)

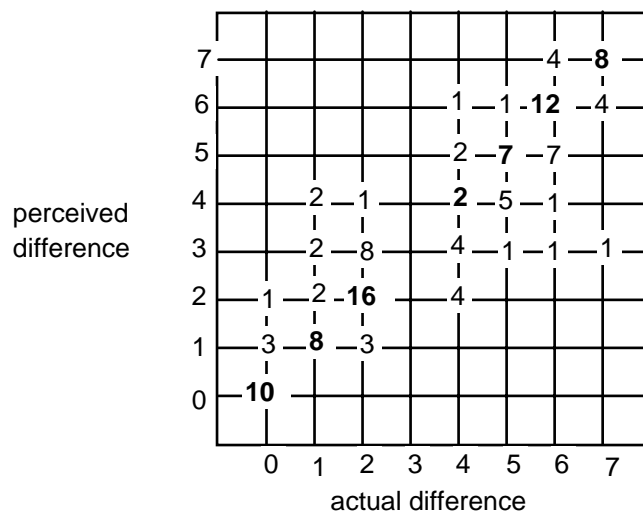


Figure 1 Subjects response to Tonal differences

Not surprisingly, subjects found the middle intervals most difficult presumably because they were using the ends of the scale as reference points. The three musicians in the sample were more accurate, achieving an accuracy of 93%, but their errors were again in the centre (at differences 4 and 5). It can be reasonably concluded that human beings can perceive a numerical difference between two tones, usually to about ± 1 tone within the octave though the accuracy decreases away from their reference point.

Subjects were also presented with different musical shapes, each consisting of 6 notes from the major scale. The scale was always presented first followed by three occurrences of the shapes separated by 2 seconds. Subjects were asked to sketch on lined paper, the shape they heard. Four

of the eight stimuli are shown in Figure 2. The responses of the subjects were averaged and are shown in grey. The mode of the responses is also shown. Although averaging could hide some effects it does show that subjects did pick up the general shape being presented (though an examination of individual results shows a considerable variation in magnitude).

It should be pointed out that all these examples fall into the category of a "tune" and are therefore easier to process than a set of rapidly varying tones. The results indicate that subjects seem to be able follow and remember the general pattern of a tune, but that the magnitude of variation will vary considerably between subjects. It was interesting to note that one subject produced patterns which differed considerably from the stimuli. On further examination it was revealed that this subject could not draw a well-known tune (such as the first line of the National Anthem) even though he was familiar with it. Two skills are therefore being measured in parallel here, and in future tests we will attempt to distinguish between them.

At the end of the tests a number of parameters were varied and subjects were asked to comment on the ease or difficulty of sketching the sequences. These examples were intended to probe for other interesting effects that we might investigate in future tests. Rhythm emerged as an obvious candidate for improving intelligibility, and, interestingly, most subjects suggested that a more rapid presentation of the sequences would have made the task easier.

Finally we presented subjects with sequences of 8 notes, with no preceding rendering of the scale. The six sequences were presented three times. They actually represented a set of notes being progressively sorted in a Bubble Sort. Thus the first sequence was very "spikey" and the last sequence was smoothly rising. Subjects response became progressively more like the stimulus as the experiment progressed, so that, for the last "sorted" sequence, most subjects responded with a uniformly rising sketch (some very accurate).

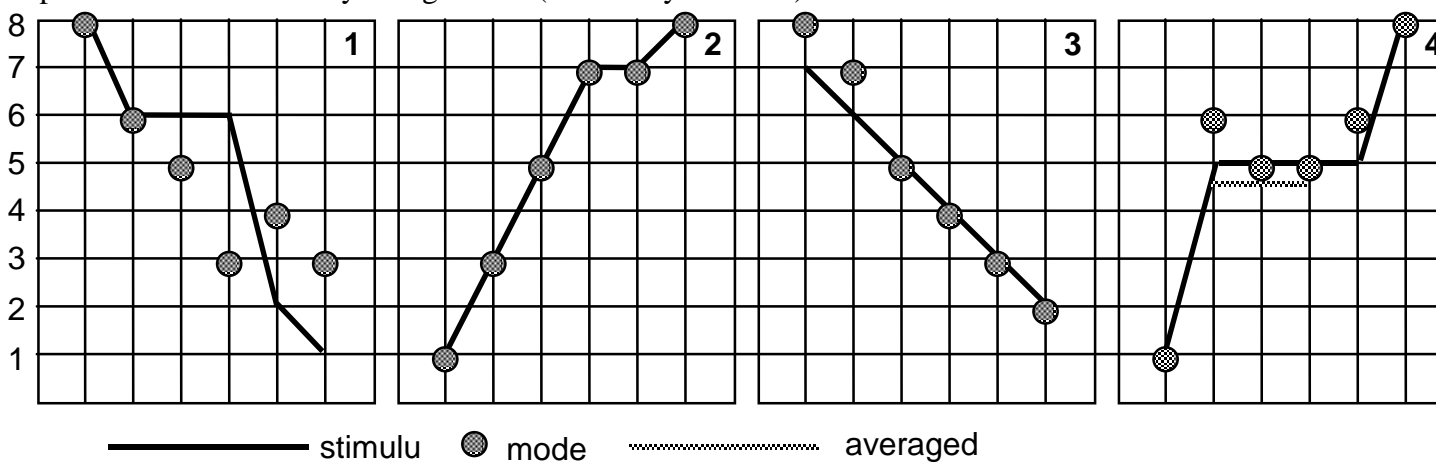


Figure 2 Some Averaged Responses to Musical Sequences

As the sequence approached a "tune" the response became more accurate.

One can perhaps see now why the Bubble Sort worked. Initially, the musical sequence is very disjointed and difficult to follow. However, at this stage, the listeners do not need to be able to sketch the shape, they only need to know that it is "spikey". As we approach the solution, the listener now needs to be able to hear the ordering of the tones, but by this time the notes are organised into a tuneful sequence and the task of analysis is much easier. Thus, not only is the

audiolisation of the algorithm using music for disambiguation, but the output of the algorithm (even though it is changing radically) is capable of being processed at all stages.

5. Future Work - 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. Whilst developing the two previously discussed auralisations of algorithms, the author discovered two programming errors by simply listening to the musical output. In the case of the Bubble Sort, this happened the evening before a major demonstration (at CHI'93) and signalled an error in what was thought to be a fully working program. The second case occurred in the minimum path algorithm when a logic error was "heard". In both cases, the occurrence of a rising note when there should have been a falling note, not only signalled an error, but also pointed to the precise area of code in which the error occurred.

The possibility of using musical output for debugging has already been suggested by Francioni et al (1991) for debugging a parallel processor message system. 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 in the reverse direction). 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. A similar movement in the value of one data variable in a graph might not be noticed.

Digiano and Baecker (1992) have given some suggestions as to how to add aural capability to programming environments using their LogoMedia sound enhancing programming system. However their main suggestions concerned the adding of typical office sounds to the output (ticking, bubbling, fan noise etc..). The reason for choosing such sounds was to provide as unobtrusive a set of background sounds as possible. They identify the phases of program execution, review and preparation as candidates for auralisation, further identifying information values of interest during execution into *Common* - typical information which can be predicted will be of interest to a programmer, *Arbitrary* - unpredictable data and events, and *Internal* - that is, changing internal states of the machine which are program and language independent.

In the programming language debugging situation, the richness offered by a musical representation could offer fairly precise bug location possibilities (whether used in isolation or in conjunction with the visual media). One obvious possible mapping would be 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 would be capable of being followed easily, allowing the visual sense to concentrate upon program detail.

Music could also be used for language debugging. For example, each construct in a programming language (say PASCAL) could be mapped to a particular musical structure. Some mappings would be obvious - a FOR-NEXT loop would be mapped to a rising pitch based upon the loop variable. The REPEAT-UNTIL and REPEAT-WHILE loops would be distinguished by a simple note comparison at either the start or the end of the loop. An IF THEN clause might be mapped to harmonic differences so that the subsequent path is either in a major key or a minor key. Thus, during execution, a programmer could continuously hear

- a REPEAT-UNTIL loop in execution
- within an IF-THEN clause
- in a particular module

This is a nice example of how, by careful mapping, the capabilities of music allow simultaneous representation of a number of different activities to be successfully disambiguated. How well this approach works (and whether the result is useful or just irritating remains to be measured). We are therefore constructing such a debugging system and have given it the name PASsaCagLia. We intend to carry out a set of evaluation studies to find out how and where the musical mapping is most useful.

6. General Design Considerations

A key goal of this work is to produce a set of guidelines to assist others in the use of music in interfaces. However, firstly, a word of warning. The author of this paper is musically trained and has composed music for many years. The musical domain is a special one and the art of musical composition (particularly orchestration) is not a skill that can be easily learned. It is therefore essential that any work carried out in this area is done with the advice and guidance of musicians (if possible composers).

Here are some guiding principles which we have found useful:

- There is a useful metaphoric mapping from a numeric domain to pitch. It can be used to transmit shape and trends but cannot be used for accurate numerical determination. However, comparison of pitch (i.e. with a reference tone) can be used for exact determination.
- A rhythmic figure can be used to render a small numeric figure (upto five).
- Different rhythmic figures can be used to render different information output but they should be kept distinct. For example, two different numeric outputs can be rendered by two sequential distinct rhythmic figures.
- Adding a regular rhythm (for example Dah Di Di) to a sequence of notes can increase its memorability.
- Timbre is useful for distinguishing between different types of information, but people cannot easily distinguish the different timbres in chords. Thus the different timbres should be sounded separately if possible.
- Most people can recognise the difference between the major, minor, whole-tone and pentatonic scales (though they may not know what they are !). These are useful for carrying complex mappings of other musical information (for example they could be used to show the four different paths in a CASE statement in PASCAL whilst pitch rhythm and timbre are simultaneously used for other information).
- Stereophony is a useful additional aid to assist in disambiguation. However it should normally be used in conjunction with another distinguishing feature. This is particularly important if headphones are not being used.

- Timbre is very useful for identification of a particular section of an algorithm or for a distinct program module.

7. Conclusions

Music is a very rich auditory medium. It offers a variety of techniques for resolving ambiguities in information output, particularly where multiple events are happening over time. It may be used on its own or in support of the visual media. Its use could be of great benefit to visually impaired people.

The building blocks already exist, but we need a greater understanding of the mappings from our interface output domains into music, and of the limits of musical processing for people who are not musicians.

There is some prima facie evidence that music can be useful in interface design. We now need some good experimentation to determine what is possible and practicable.

Acknowledgements

Special thanks are due to Simon Holland of the Open University who collaborated with the author in giving a lecture to CHI'93 which looked at both the use of music in HCI and the value of a good interface for musical understanding. During our collaboration he suggested some interesting ideas for using music in interfaces. General thanks are due to the MITS project team for some interesting discussions, and the RACE Directorate for Grant Number R2094 (MITS) under which part of this work was carried out.

References

Alty, J.L., (1995), Navigating through Compositional Space: The Creativity Corridor, in *The LEONARDO Music Journal*, Vol. 28, No. 3, pp 215 - 219.

Bever, T.G., (1980), Broca and Lashley were right: Cerebral Dominance is an Accident of Growth. In *Biology and Language*, Cambridge, Mass., MIT Press.

Blattner, M., Sumikawa, D.A., Greenberg, R.M., (1989), Earcons and Icons: Their Structure and Common Design Principles, *Human Computer Interaction*, Vol. 4(1), pp. 11 - 44.

Blattner M., Greenberg, R.M., and Kamegai, M., (1992), Listening to Turbulence: an Example of Scientific Audiolisation, in *Multimedia Interface Design*, (Blattner, M., and Dannenberg, R.M., eds.) , Chapter 6, pp. 87 - 102, Wokingham: ACM Press.

Brewster, S.A., Wright, P.C., and Edwards, A., (1994), The Design and Evaluation of an Auditory-Enhanced Scrollbar, in *Proc. CHI'94*, Boston Mass., pp. 173 - 179, (1994).

Brown, M.H., and Herchberger, J., (1991), Zeus: A System for Algorithm Animation, *Proc. IEEE Workshop on Visual Languages*, IEEE Press, pp. 4 - 9.

- Brown, M.H., and Herchberger, J., (1992), *Colour and Sound in Algorithm Animation*, IEEE Computer, December, 1992, pp. 52 - 63.
- Digiano, C.J., and Baecker, R.M., (1992), Program Auralisation: Sound Enhancements to the Programming Environment, *Proc Graphics Interface '92*, pp. 44 - 52, (1992).
- Deutch, D., (1982), *The Psychology of Music*, Academic Press, New York.
- Edwards, A., (1989), Soundtrack: An Auditory Interface for Blind Users, *Human Computer Interaction*, Vol. 4, No 1, pp 45 - 66.
- Francioni, J.M., Albright, L., and Jackson, J.A., (1991), Debugging Parallel Programs Using Sound, *SIGPLAN Notices*, Vol. 26 (12), pp. 68 - 75.
- Gaver, W., (1986), Auditory Icons: Using Sound in Computer Interfaces, *Human Computer Interaction*, Vol. 2(1), pp. 167 - 177.
- Gaver, W., (1989), The Sonic finder: an interface that uses Auditory Icons, *Human Computer Interaction*, Vol. 4(1), No. 1, pp. 67 - 94.
- Holland, S., (1992), Interface Design Empowerment: A Case Study from Music, in *Multimedia Interface Design in Education*, Edwards, A.D.N., and Holland, S., eds., *NATO ASI series F*, Springer Verlag, Berlin, pp. 177 - 194.
- Kennedy, M., (1991), *The Concise Oxford Dictionary of Music*, 3rd Edition, Oxford University Press.
- Liebermann, H., (1986), Music as a Command Language, *MIT Artificial Intelligence/Media Laboratory Report*. (unpublished).
- Mansur, D.L., Blattner, M.M., and Joy, K.I., (1985), Soundgraphs: A Numerical Data Analysis method for the Blind, *Proc. 18th Hawaii Conf. on System Sciences*, Vol. 18, pp 163 - 174.
- Mestel, R., (1994), Hearing Pictures, Seeing Sounds, *New Scientist*, No. 1928, pp. 20 - 23.
- Mezrich J.J., Frysinger, S., and Slivjanovski, R., (1984), Dynamic Representation of Multivariate Time Series Data, *J. Am. Statistical Soc.*, Vol. 79, No. 385, pp. 5.22 - 5.28.
- Moog, R., (1986), MIDI: Musical Instrument Digital Interface, *J. Audio Eng. Soc.*, Vol. 34, pp. 394 - 404.
- Minsky, M., (1989), Music, Mind and Meaning, in *Selected Readings from the Computer Music Journal*, Roads, C., (ed.), *The MIT Press*, Cambridge, Mass., pp. 635 - 655.
- Prior, M., and Troup, G.A., (1988), Processing of Timbre and Rhythm in Musicians and Non-Musicians, *Cortex*, Vol 24, pp 451 - 456.
- Sloboda, J.A., (1985), *The Psychology of Music*, Clarendon Press, Oxford.

Yeung, E.S., (1980), Pattern Recognition by Audio Representation of Multivariate Analytical Data, *Analytical Chemistry*, Vol. 52, No. 7, pp. 1120 - 1123.