Designing Earcons with Musical Grammars

John C. K. Hankinson
Department of Computer Science
University of York
York, England, YO10 5DD

Alistair D. N. Edwards
Department of Computer Science
University of York
York, England, YO10 5DD

1   Introduction

Many blind and partially-sighted computer users have come to rely on auditory computer interfaces to interact with their computer. By recognising sounds made by the computer, blind users can extract information from the auditory display which is presented visually to sighted users. Sighted users can also benefit from using an auditory interface in conjunction with their existing visual display.

Earcons have evolved as a means of representing concepts in such auditory displays as short sound motives. These motives usually consist of a small number of notes, with a distinct rhythm and dynamic contour. They are constructed according to rules devised by Blattner [1] and developed by Brewster et al. [2, 3]. Although they appear to be musical fragments, many fundamental musical concepts are not taken into account during their design.

In fact, many early earcon designers tended restrict the maximum number of notes within an earcon to four, as any longer sequence would have a melodic connotation. The presence of a tune was thought to be distracting and hence the use of musical tensions, etc. was avoided. Instead of shying away from the powerful capabilities music affords, it is postulated here that musical structures should be used to our advantage, to remove the tight restrictions placed upon the design of earcons.

It is proposed that the use of a musical grammar during the design of a set of earcons can impose a number of valuable constraints upon the choice of earcon motif. These constraints are advantageous as they allow a basic structure to be built into the surface form of each earcon. This in turn allows only certain combinations of earcons to combine with each other. A grammatical property is therefore imposed upon the interface which provides extra feedback to the user when inappropriate combinations occur.

2   Musical Grammars

The concept of a grammar is not new; grammars have been used to describe language syntax for centuries (Chomsky [4], Holtzman [6]). They have also been effective in capturing the structural properties of numerous types of music at many different levels (Roads [8], Lerdahl and Jackendoff [7]). The abstract nature of grammars makes them easily adaptable to the analysis of both language and music structure.

Grammars come in many shapes and sizes, but are essentially based on one underlying concept: a set of rules describes how basic units of a system can combine to form larger phrases. In a language grammar, these units are often words; the grammar rules describe how they can combine to form sentences. In musical grammars, units can be notes, chords, rhythms, pitch contours or even larger musical phrases (Cope [5], Holtzman [6], Steedman [9]).

To show how a musical grammar can be beneficial to earcon design, a simple grammar will now be constructed, according to simple rules of Western tonal harmony.

2.1   Towards an Auditory Grammar

Given a note of a particular fundamental frequency, a note with half its frequency is regarded as the first overtone (harmonic) of the original note. If this halving of frequency continues, successive notes form the overtone series. The intervals (pitch spaces) between the first few notes of this series (octave, fifth, fourth, major and minor third) are the most important in Western harmony. They can be found in a chord based on the first, third and fifth notes of any diatonic scale. This chord is known as the

triad and is the most stable consonance[1] due to the simple relations between its notes.

The 'colour' of any chord can likewise be determined from the intervals between its constituent tones. Each tone within a chord provides a harmonic function to the overall chord. For example, the tones within the chord below serve the following functions:

Minor Tenth/Third (IIIb)
Minor Seventh (VII)
Fifth (V)
Root (I)
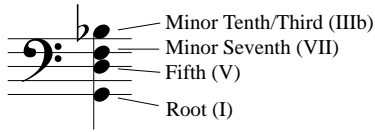
Adding up all the functional constituents gives the type (therefore colour) of the chord (minor seventh chord):

I + IIIb + V + VII = Imin7 chord = Gmin7 (G as tonic)

Extending this idea further, if two chords are played at the same time, the tones from the higher chord will form new relationships with the tones from the lower chord. For example:
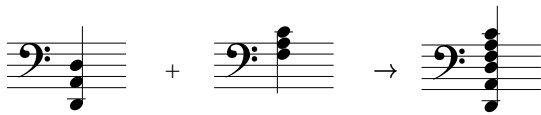
As we can see, in a different harmonic context, constituent tones function differently. That is, the harmonic function of a tone depends upon its relationship with the root of its parent chord. If the root of the chord to which it belongs changes, the tone's contribution to that chord also changes.

In the example above, the new harmonic functions of the second chord's tones fit well with the existing intervals in the first chord. The resultant combined chord is a consonance. However, this is not always the case. Consider the following chordal combination:

This time, the second chord's root note, F, forms a dissonant interval with the new chord's root note. The A also adds to the dissonance of the new chord, as its presence in conjunction with the already present G# forms a semitone clash between the two tones. Consequently, the combined chord is a dissonance.

This is a useful phenomenon to exploit in auditory interface design. If it is possible for a sound to be harmonious in one context and discordant in another, that sound can be used to represent information whose status changes depending on its context. However, before we can make use of such sounds, it is important to be able to describe these sound relationships in a grammar.

To capture this notion of consonance versus dissonance in a musical grammar it is necessary to define the basic units upon which the grammar's rules can act. In this instance, the basic units of the grammar will be chords from a pre-defined set, while each grammar rule will describe a harmonious combination of two of these chords. A combination of basic units which is not listed in the grammar's rules will be viewed as a dissonance. In effect, the grammar classifies combinations of the set of chords as either consonant (grammatical) or dissonant (ungrammatical).

3 Objects and Actions

At this point, it will be beneficial to introduce a sample environment in which a musical grammar of the form discussed above can be designed. For the purposes of this example, a simple object/action model is proposed.

Four types of object have been chosen (disks, printers, files and texts) along with numerous actions that can be performed on each of them (e.g. pause, copy, print, bold). Some of the actions can be performed on more than one type of object; for example, the copy action applies to disks, files and texts. In contrast, the pause action is appropriate only for printer objects. Table 1 lists the acceptable combinations of actions and object categories:

| | | | Object Categories | | | |
			Disks	Printers	Files	Texts
	A1	Pause Resume	✗	✓	✗	✗
	A2	Copy Delete	✓	✗	✓	✓
Action Categories	A3	Print	✗	✓	✓	✗
	A4	Open	✗	✗	✓	✗
	A5	Format	✓	✗	✗	✗
	A6	Bold Italic	✗	✗	✗	✓

Table 1: Object/Action Combinations

---

[1] Certain chords create tensions which resolve to consonances. The more stable a chord is, the smaller the need for resolution to other chords.

These combinations can likewise be described by a set of grammar rules:

$$
\begin{aligned}
I &\rightarrow A2 \circ \text{Disks} \\
I &\rightarrow A5 \circ \text{Disks} \\
I &\rightarrow A1 \circ \text{Printers} \\
I &\rightarrow A3 \circ \text{Printers} \\
I &\rightarrow A2 \circ \text{Files} \\
I &\rightarrow A3 \circ \text{Files} \\
I &\rightarrow A4 \circ \text{Files} \\
I &\rightarrow A2 \circ \text{Text} \\
I &\rightarrow A6 \circ \text{Text}
\end{aligned}
$$

where the first rule states that a legal interaction (I) can be formed ($\rightarrow$) by the combination of ($\circ$) an action from action category A2 and an object from the object category Disks.

Suppose the task is to design suitable audio representations of each of the objects and actions available in this environment. This could be achieved by designing an earcon for each object and action category. It would be advantageous if the earcons maintained the grammatical associations between object and action categories. If a musical grammar of the type discussed previously was used in the design of such earcons, allowable combinations would sound harmonious (therefore correct) whilst illegal combinations of objects and actions would sound discordant (therefore incorrect). To ensure this, our musical grammar needs to be isomorphic to the object/action grammar.

3.1   Audio Representations

Ten chords were chosen for each object and action category such that the combination of the chords was a consonance for legal object/action combinations yet formed a dissonance for inappropriate object/action pairings. A subset of those chords is shown in Table 2.

However, it must be remembered that these chords only represent the *categories* of objects and actions. Played in combination, it is possible to tell only whether the categories fit together. If an interface designer wished users to recognise individual objects and actions, they would need to create earcons for each object and action based on the chord of that object or action's category.

Hence, the chords chosen to fit the underlying grammar provide the framework within which the choice of earcon can be constrained. As long as an earcon's melody be-

longs to its associated category's chord, the earcon will still abide by the rules of the grammar.

Table 2: Some of the object and action chords

To continue the example, a number of melodies based on the chords were chosen for some objects and actions belonging to the defined categories. These are listed in Table 3 below.

Object/Action	Earcon Melody
readme.txt	
"in this example"	
Copy	

Table 3: Earcon melodies for
some of the objects and actions

To reinforce the harmonic structure, the underlying chords can be sounded together with the grammatical melodies. As the melodies are based on their underlying chords, they will not clash. Furthermore, the audio representation of any combination of objects and actions will maintain its grammatical nature regardless of whether melodies and chords or melodies alone are sounded.

4   Sample Interaction

To further illustrate the use of grammatically-defined earcons, a typical interaction is presented next.

A user has recently bought a new printer and decides to print out an old letter written some time ago. They begin by choosing the file in which the letter is stored and by pressing a key to invoke the open command. The system provides auditory feedback:[2]

The user then moves around the file and decides to italicize some text to test the new printer's capabilities. However, after selecting the text, they accidentally hit the open key again. The auditory interface responds with a dissonant earcon pair:
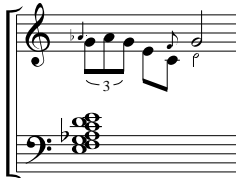
Realising their mistake, the user selects the text and presses the correct italic key. This time, their interaction is accompanied by the combined earcon *Text∘Italic*:
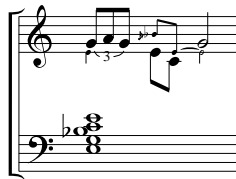
Before the user prints off this revised copy, they decide to print the printer's internal test page. They do this by selecting the printer and pressing print:

Having successfully printed the testpage, the user then prints the document:

---

[2]The smaller notes represent the action melodies.

5 Important Considerations

It is important to remember that the choice of basic units for the auditory grammar was chords based on simple rules of Western classical harmony. The effect upon a listener of correct and incorrect combinations of such chords and subsequent melodies based on those chords, will of course depend upon the musical experience of that particular user of the system. One human's consonance is another's dissonance.

In fact, the perceived well-formedness of a chord also depends upon the context in which it appears. A chord which sounds awkward when it is heard alone, may be more acceptable when it is part of a larger chord sequence. Some chords are known to create tensions (e.g. through the use of suspensions). Others resolve the tensions created by preceding chords. Thus chords are often referred to in terms of their function (Cope [5]). The effect of tensions and resolution plays an important role in music, providing the excitement and suspense necessary for interesting music.

In general, single chords represent a range from very stable consonances (such as a major triad) to extreme dissonances (cacophony). An individual's distinction between consonance and dissonance will lie somewhere along this scale. A user with a different boundary position to that used in the example above might initially classify some object/action combinations incorrectly. However, after a short training period, it is believed that any user should be able to recognise the consonance/dissonance boundary in use by the system.

Of course, the boundary present in the example was a consequence of the particular chords chosen for the system. A different set of chords could easily be used to provide an alternative. For that matter, the decision to use chords as the basic grammatical unit need not have been made; rhythms, pitch contours and chord sequences have all previously been used as basic units in musical grammars. Any potential confusions made by users of the example

system are due to the specific example, not to the technique of using musical grammars *per se*.

In fact, such is the general applicability of grammars for describing structure, that many music theorists have been able to apply grammatical techniques to many styles of music for numerous musical cultures. Consequently, there is no inherent restriction on the musical style of auditory interface that could be developed using musical grammars.

6 Future Work

As the proposed grammatical approach to sound design is closely based on linguistic grammars, it is hoped that basic mappings between similar structures can be produced in future. This in turn should lead to effective methods of communicating more structured information than earcons can presently portray through sound, (such as language-based information).

Furthermore, computers could then be used as an intermediary between two humans who wish to communicate, but who cannot use the traditional auditory methods such as speech or the whistle/drum communication systems used by a large number of tribal communities. The computer could act as a translator between a speech-impaired user communicating with another human using musically-structured sound.

Mappings between music and language using a computer intermediary may also be helpful for other language-based disabilities. A set of training aids could be produced that aim to improve certain language skills by converting them into a musical equivalent that is understandable, thereby reinforcing the language skill. Such aids might include prosodic assistance, semantic and syntactic relatedness and sentence formation tools.

7 Conclusions

It has been shown that constraining the design of earcons with a musical grammar can provide a structured frame-work within which earcon melodies can be composed. This has the advantage that the combination of such earcons sounds consonant or dissonant depending on whether the two earcons are compatible with each other.

Musical grammars are powerful enough to describe more complex structural combinations than those described herein. It is hoped that further research will reveal how these grammars can be put to use to benefit auditory interfaces.

References

[1] Blattner M. M., Sumikawa D. A., Greenberg R. M., "Earcons and Icons: Their Structure and Common Design Principles", *Human Computer Interaction*, Vol. 4, 1989

[2] Brewster S. A., "Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces", DPhil thesis, University of York, 1994

[3] Brewster S. A., Wright P. C., Edwards A. D. N., "Experimentally Derived Guidelines for the Creation of Earcons", in Adjunct Proceedings of HCI'95: People and Computers, Huddersfield: BCS, 1995

[4] Chomsky N., *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press, 1965

[5] Cope D., *Computers and Musical Style*, OUP, 1991

[6] Holtzman S. R., *Digital Mantras: The Languages of Abstract and Virtual Worlds*, MIT Press, 1996

[7] Lerdahl F., Jackendoff R., "Toward a Formal Theory of Tonal Music", *Journal of Music Theory*, Vol. 21, No. 1, 1997

[8] Roads C., "Grammars as Representations for Music", *Computer Music Journal*, Vol. 3, No. 1, 1979

[9] Steedman M. J., "A Generative Grammar for Jazz Chord Sequences", *Music Perception*, Vol. 2, No. 1, 1984