

Musical Phrase-Structured Audio Communication

John C. K. Hankinson

Department of Computer Science
University of York
Heslington
York, YO10 5DD, England
John.Hankinson@cs.york.ac.uk

Alistair D. N. Edwards

Department of Computer Science
University of York
Heslington
York, YO10 5DD, England
Alistair.Edwards@cs.york.ac.uk

ABSTRACT

It has previously been shown that musical grammars can impose structural constraints upon the design of earcons, thereby providing a grammatical basis to earcon combinations. In this paper, more complex structural combinations are explored, based upon linguistic phrases. By mapping between a musical grammar and a linguistic grammar, musical phrases can be generated which correspond to linguistic sentences. A large number of unique meanings can be presented in this way based upon a simple musical vocabulary. This is of great value to auditory designers. A user study has been undertaken which reveals that users can recognise these complex auditory phrases after a small amount of training.

Keywords

music, grammar, earcon, structure, language

INTRODUCTION

Much work has been done on the design of earcons for use in auditory displays (e.g. [1, 2]). For example, guidelines have been developed which try to aid the earcon designer by suggesting how to restrict the choice of sounds to use in an auditory interface, to avoid possible confusions [3]. The basic, restrictive nature of these guidelines has discouraged the use of structured music in auditory design, often due to the perceived complexity of music and its inherent associations and expectations. Rather than reject musical approaches *per se*, it is of interest to explore the use of musically-designed displays as an alternative form of auditory interface.

Although Brewster's guidelines ([2, 3]) are helpful when dealing with isolated low-level sounds, they are too restrictive for structured musical sounds. By exploiting the features of music which earcon design has traditionally avoided, it is believed more expressive non-speech communication can occur at the interface between people and machines. For example, musical grammars have already been used in the design of grammatical earcons [7] and it is thought that a vast amount of more complex, highly-structured information can be easily expressed musically in a considerably more stylised way and with greater design freedom using the approach described herein.

It should be mentioned that recently, a number of researchers have begun to look beyond the limitations of the earcon guidelines. Leplâtre has used simple musical motifs to assist menu navigation [8] and Vickers et al. [9] have pioneered the use of musical *leitmotifs* to represent a number of Pascal programming constructs. Vickers' approach is somewhat different to that outlined here and unfortunately, success has been limited. Other than this, there has been little research into the use of music (and especially its structure) in auditory interface design.

Consequently in this paper, the theory of musical phrase-structured communication is presented, where small meaningful musical units are defined and subsequently combine to form larger musical phrases. A musical/linguistic grammar provides the combinational mechanism. A user study is then discussed which sets out to show that people are capable of recognising such musical phrases. Finally, potential advantages of this novel approach to non-speech audio communication are outlined.

MUSICAL PHRASE-STRUCTURE

Language is a powerful communication tool with which most people are familiar. Linguistic sentences are formed by the appropriate combination of smaller lexical units known as words. As only certain types of word can combine with other words, it is necessary to group together words which have the same syntactic roles. This can be achieved by defining the grammatical category to which each word belongs [4].

Musical phrases can likewise be formed by the combination of smaller musical units (as music is a highly-structured system). This can be achieved by composing musical units which belong to appropriate categories. One way of

Nouns (N)
 Nouns are based upon chord I (in this example in the key of F, nouns are based on F).

Adjectives (Adj)
 Adjectives are based upon chord IV (in this example in the key of F, adjectives are based on Bb).

Verbs (V)
 Verbs are based upon chord V (in this example in the key of F, verbs are based on C).

Figure 1: Small vocabulary of musical words

defining a musical category is to assign a particular chord to each category. Musical units belonging to a category must then be based upon that chord. By assigning a linguistic word to each musical unit, it is possible to create a ‘musical vocabulary’ as shown in Figure 1 [5].

Once the vocabulary is defined, it is possible to describe how the lexical units can combine by using a set of grammar rules. These rules define how larger phrases are generated from smaller phrases and units belonging to certain grammatical categories. An example grammar is shown in Table 1.

Grammar Rule	Explanation of Rule
MP → NP VP	A musical phrase (MP) consists of a noun phrase (NP) followed by a verb phrase (VP)
VP → V NP	A verb phrase (VP) consists of a verb (V) followed by a noun phrase (NP)
NP → AdjP N	A noun phrase (NP) consists of an adjectival phrase (AdjP) followed by a noun (N)
AdjP → Adj AdjP	An adjectival phrase (AdjP) consists of an adjective (Adj) followed by another adjectival phrase (AdjP) or is simply empty
AdjP → ϕ	

Table 1: An example grammar

The grammar’s base units are grammatical categories. Consequently, either linguistic words or musical words can be substituted for each category to generate a surface form utterance. It is in principle therefore possible to communicate through music the same amount of structured information as the linguistic approach permits (Figure 2).

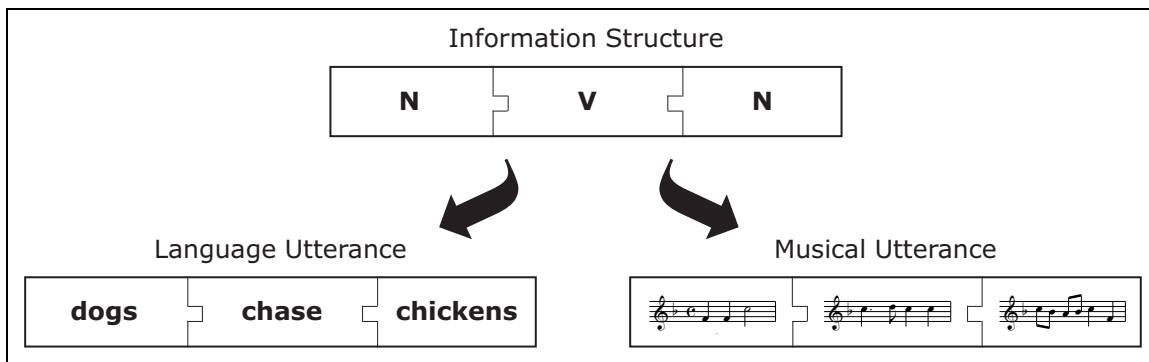


Figure 2: Two examples of structured communication

Figure 3 shows an example musical phrase supplemented with a musical accompaniment. The accompaniment does not add to the meaning of the phrases; it is used simply to fill out the harmony and enhance the musical experience.

The image shows a musical score for three instruments: Vibraphone, Piano, and Strings. The Vibraphone part is in the treble clef with a key signature of one flat and a common time signature. It consists of four measures of music, with the lyrics 'hungry', 'boys', 'eat', and 'chickens' written below the notes. The Piano part is in the treble clef and consists of four measures of accompaniment. The Strings part is in the bass clef and consists of four measures of accompaniment, primarily using chords.

Figure 3: Example musical phrase

It is important to realise that this structured method of combining musical units to form musical phrases does not have to restrict itself to representing language utterances as musical sounds (although this in itself opens up many interesting auditory design questions, especially for the design of disability and communication aids). The approach described here can be used for any information whose structure can be captured by a grammar, whether the surface units of the grammar's vocabulary are 'linguistic words' or other tokens, labelling the underlying base concepts of the information. Therefore it is of broader interest to auditory interface designers. However, for the purposes of explaining the theory and for the user study discussed next, a simple set of everyday words has been used by way of an example.

USER STUDY

A user study has been undertaken to evaluate whether people can interpret the meanings of musical phrases generated as described above. This three-stage experiment firstly needed to determine whether a musical vocabulary could be learnt. Secondly, it was of interest to know whether complex musical phrases could then be interpreted from their component parts. Each stage of the study is described in turn.

Stage 1 - Musical Words and Phrases

In the first part of the study, participants were trained to recognise all three musical nouns (Figure 1). Each musical word was played once, whilst its associated meaning was displayed on screen. These words were then repeatedly presented in a pre-defined order without any meanings being displayed on screen. The task of the participants was to select the appropriate meaning for each musical unit they heard. After an answer was submitted the participant was given appropriate feedback; any two confused words would be replayed with their meanings displayed.

During the following, testing phase, participants had to indicate the meaning of a number of randomly pre-chosen words after hearing each word's melody, this time without feedback. If the number of correct answers was significantly more than would be expected by chance, participants would continue training and testing with an additional two words (the adjectives, Figure 1), otherwise they were retrained and tested on the original three-word vocabulary. Similar retraining and testing would occur if needed for the five-word vocabulary. If the participant's score was still too low after this second attempt, the experiment terminated.

After a short break each musical word/meaning pair from the five-word vocabulary was played to refresh the participant's memory. A selection of musical noun-phrases were then played once. Each musical phrase consisted of either one, two or three words from the five-word vocabulary played serially. A simple accompaniment was also sounded during the playback of musical phrases. To illustrate, typical phrases included the following:

big dogs hungry chickens boys hungry big boys

The participants were asked to select the meaning of the whole phrase by clicking labelled buttons in turn for each of the component units. No feedback was given to the participants and they could only hear each phrase once. Their answers were recorded along with the correct meaning of the phrases. At the end of this test, two new words (verbs) were introduced using the training method as before. A number of musical phrases representing sentences were then played and the participants had to select their meaning. Typical phrases had the following semantics:

dogs chase chickens hungry boys eat big chickens big hungry chickens chase dogs

Stage 2 - MAT Tests

In this stage, each participant took part in a MAT study. MAT tests are a newly developed suite of musical ability tests [6]. Each test is used to measure the extent of a particular musical ability (e.g. meter, chord structure, melody discrimination), of an individual. They give a better indication of participants' strengths and weaknesses at handling and processing musical sounds than an arbitrary classification of 'musicianship'. A general profile of musical abilities can be generated from which it is possible to state that a particular individual has, for example, above-average pitch and rhythm abilities but below-average harmony skills. This profile can then be correlated against the stages' scores to show whether the success of musical phrase recognition depends upon any particular musical abilities.

Stage 3

In the final stage of the study, participants were retrained on three- then five-word vocabularies. They then had to interpret a number of noun-phrases. This time the participants were able to replay the musical phrase as many times as they wished. They were also informed of the correct meaning of a phrase if they submitted an incorrect answer. This process was then repeated without any feedback. After a short break, the two verbs were reintroduced through a further training stage. Two sets of musical phrases were then played (one with feedback, one without) where the phrases were generated from the full vocabulary.

RESULTS OF THE STUDY

Musical Words

The experiment has shown that all participants were capable of adequately recalling the meaning of the first three words. Within the training and testing period, 92% of participants could recall the five-word vocabulary semantics (see Figure 4). From this, it is possible to conclude that the use of grammatical categories to constrain the design of musical melodies does not inhibit a participant's ability to recognise such melodies and to recall associated linguistic word meanings.

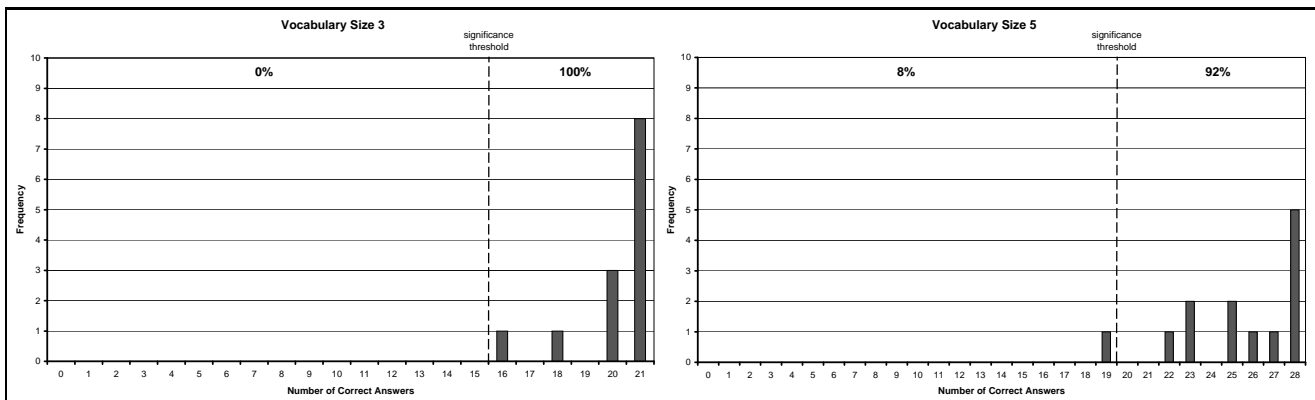


Figure 4: Results of vocabulary testing

Musical Phrases

The results from the first stage of the study show that 46% of participants could understand a statistically-significant number of noun-phrases¹, but only 18% of these participants could adequately understand the seven-word vocabulary sentences (see Figure 5).

However, by the end of Stage 3, 92% of participants were capable of fully recalling the meaning of a statistically-significant number of both five-word and seven-word vocabulary phrases (see Figure 6). This shows that after a suitable training period, most participants could accurately interpret audio messages designed using this musical-structured approach.

Musical Ability

There was no statistically-strong correlation between any of the aspects of musical ability measured with the MAT tests and the results of the study discussed above. The greatest correlation found was between the performance of seven-word vocabulary phrases at the end of Stage 1 and participants' melody discrimination ability ($r_s = 0.437$)².

¹binomial test, 0.05 level of reliability.

²Spearman correlation.

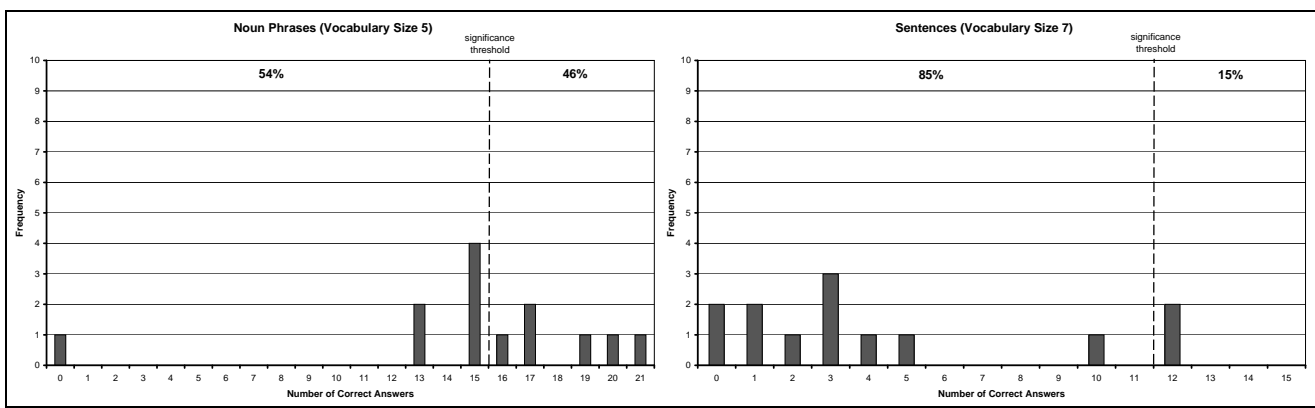


Figure 5: Results of initial phrase testing

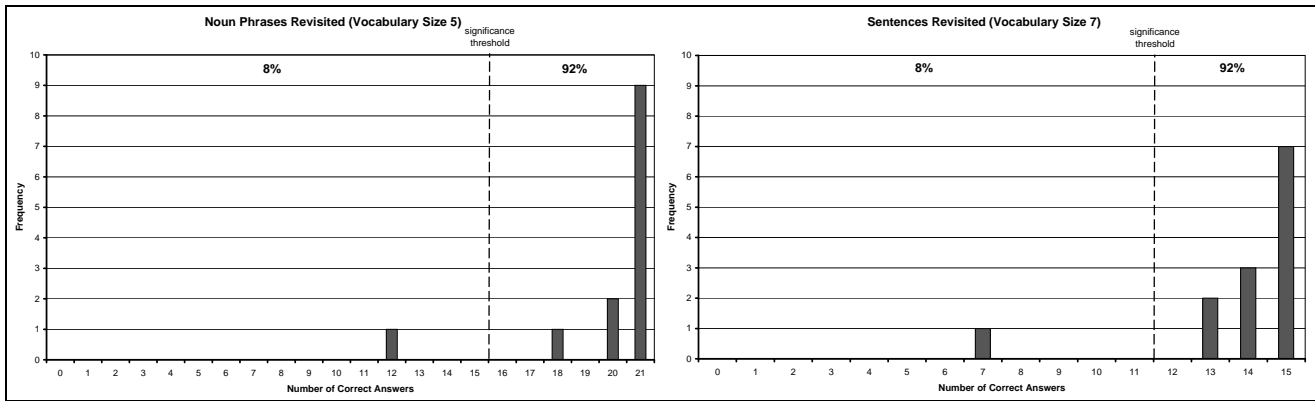


Figure 6: Results of further phrase testing

However, this correlation is not significant enough to suggest that good melody discrimination is necessary for good performance in Stage 1. In fact, it is possible to deduce that successful recognition of the musically-structured sounds given in the study does not depend upon musical ability (as measured by the MAT tests).

ADVANTAGES

Expressive Power

Through learning seven words, participants are capable of potentially understanding any of the 450 possible distinct meanings which can be communicated with the musical phrases. The large number of meanings is due to the structured way in which semantic units are combined. For example, three units have very different meanings when they are combined in different ways. To achieve the same level of expressiveness without using a structured approach would require a unique sound to be assigned to each individual meaning. Therefore, more than four-hundred different distinguishable sounds would have to be designed – which is practically impossible.

If the size of the vocabulary was to increase by two (e.g. one new noun and adjective), the number of meanings which could be presented in audio would increase to over eight-thousand. Further increases in size could similarly increase this figure.

Aesthically-pleasing Audio Messaging and Stylistic Freedom

Although the use of a grammar imposes constraints upon the design of sounds, these constraints aid the designer; they are not a hinderance. They provide a framework within which the designer can compose each auditory unit whilst be reassured that the consistency of all possible musical phrases that can be generated will not be compromised when new words are added to the vocabulary.

The decision to use harmony to define grammatical categories was biased towards a Western musical culture. However it need not be; any musical culture could provide the underlying framework. Also the musical nature of the messages permits an accompaniment to supplement them. The form of the accompaniment is independent of the

melody and therefore can be one of any number of different styles and compositions. Furthermore, the participants readily accepted the combined audio stimulæ as natural music. This is a major advantage of the musical approach; participants stimulate both sides of their brain (the analysing and creative sides) enhancing the learning and retention of presented information.

As the chosen grammatical framework is a harmonic one, other musical considerations are left to the designer; for example, choice of musical style, speed, complexity, duration, pitch and rhythm is left open. Consequently, it is easily possible to create a large number of auditory systems of differing styles providing a substantial amount of customisation to satisfy individual musical preferences.

Delayed Semantic Parsing

As the audio messages are fully-formed musical phrases, participants were often capable of remembering the sound stimulæ as pieces of music. The interpretation of the associated semantics could then be delayed until after the playback of the phrase. This has implications for auditory displays used in stressful environments and overloaded speech communication. It might be possible to play important messages as a piece of music, thereby allowing the listener's musical processing to remember the musical phrase as it is played during a highly linguistically-active time (e.g. during a conversation), and to delay interpretation of the musical message until a less active time (e.g. after a conversation). An opportunity exists for further research on this topic.

Potential Applications

The potential applications of the musical approach outlined here are varied. Language aids could possibly benefit by reinforcing linguistic structure with equivalent musical structures played back to the user. Auditory displays used in noisy and/or stressful environments could benefit from the broader aesthetic dimensions available due to the greater design freedom the musical approach affords. Displays can be customised so sounds are less likely to be masked or to annoy their users. Interfaces which need to present large numbers of related messages can also gain from this method. Finally, auditory interfaces in general can benefit from a new, alternative approach to interface design.

CONCLUSIONS

The user study has shown that people can learn a small musical vocabulary (that is, short melodies with associated meanings) when those melodies are harmonically constrained according to their associated grammatical category. People can also interpret novel grammatically-structured musical phrases built from these musical words given a suitable training period and a replay function. Auditory interfaces can benefit from the huge increase in the number of possible audio messages which can be expressed and the stylistic way in which they can be communicated.

REFERENCES

- 1 Blattner M. M., Sumikawa D. A., and Greenberg R. M. Earcons and Icons: Their Structure and Common Design Principles. *Human Computer Interaction*. Vol. 4, No. 1, pp 11-14, 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. and Edwards A. D. N. Experimentally Derived Guidelines for the Creation of Earcons, in *Adjunct Proceedings of HCI'95: People and Computers*. Huddersfield: BCS, pp 155-159, 1995
- 4 Chomsky N. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA, 1965
- 5 Cope D. *Computers and Musical Style*. OUP, 1991
- 6 Edwards A. D. N., Challis B. P., Hankinson J. C. K. and Pirie F. L. Development of a Standard Test of Musical Ability for Participants in Auditory Interface Testing. *International Conference on Auditory Display, (ICAD)*. April, 2000
- 7 Hankinson J. C. K. and Edwards A. D. N. Designing Earcons with Musical Grammars. *ACM Special Interest Group on Computers and the Physically Handicapped, (ACM SIGCAPH)*. September, pp 16-20, 1999
- 8 Leplâtre G. and Brewster S. A. An Investigation of Using Music to Provide Navigation Cues. *International Conference on Auditory Display, (ICAD)*. November, 1998
- 9 Vickers P. and Alty J. Musical Program Auralisation: Empirical Studies. *International Conference on Auditory Display, (ICAD)*. April, 2000