Examiners' commentary 2017–2018

CO3346 Sound and music – Zone B

General remarks

Overall performance on this paper was reasonably good, with only a few candidates unable to obtain a pass mark, and with a couple of very strong responses in the first class category. What follows is a brief discussion of the individual questions on this paper, with hints towards the answers expected by the examiners.

Comments on specific questions

Question 1

Computational models of music cognition

This question was a popular one, with candidates able to give reasonably good answers.

Part (a) was material directly from the subject guide and candidates could discuss grouping, which is the grouping of musical events such as chords or bass-line and melody, or expectation, which is about predicting what a person will expect next when listening to a piece of music. Examiners expected a justification of the answer regarding modelling using a computer program. It should be possible if you had access to the note sequence data, for example.

Similarly for part (b), simply knowing the material in the subject guide allowed for a good answer. For proximity, the visual aspect may be that things that are physically close together would be perceived as group, while for musical aspects, sounds that are close together in time may be perceived as musical change having not occurred. For similarity, similar looking shapes are grouped while similarly pitched notes are grouped to a musical part, such as bass line. For both, appropriate approaches to developing a model of music cognition were also required.

Most candidates were able to explain, for part (c), that Schellenberg developed a model that incorporated pitch information, as well as the concepts of registral direction (revised) and registral return. Given a pair of consecutive pitches in a melody, $x1$ and $x2$, we can compute the conditional probability $p(x1 — x2)$ as being the number of times that $x1$ is followed by $x2$, divided by the number of times that $x1$ occurs regardless of what follows it.

To implement this, you would need to represent the factors as pairs of notes, or note-sequence and note pairs.

The final part, part (d), required some ability to demonstrate understanding of concepts learned.

Examiners expected a discussion that included the idea of testing the predictions of the model against what actually happens, perhaps using human subjects. This is a phase that happens once the model has been built and used to generate predictions. The experiment most often given was that of a tone probe, where you would play a melody to several people and ask them which note they anticipate next. Key features would include a user interface, data gathering, analysis and comparison of predictions to actual observations. You could use it to generate the most unobtrusive muzak possible for playback in lifts and supermarkets, where the model would be used to generate a predictability score for music, and thereby select the most predictable music.

Question 2

Interactive Sound using Pure Data

This was an unpopular question, chosen by only a few candidates.

For part (a), four different objects might be adc, osc, vcf and line with corresponding descriptions of the purpose of each of these. For example, the purpose of the adc might be for audio input, while for the vcf it is used to remove the partials from a signal line.

Part (b) expected candidates to understand that with sample playback you would start with pre-recorded sound, and then play it, while with synthesis you would synthesise sound from wave-forms. Sample playback is in general more efficient as it can play any sound with the same CPU usage while synthesis will vary dependent on the complexity of the sound.

The main requirement of part (c) was the application of knowledge and techniques, to demonstrate understanding and obtain actual answers. When the slider is moved as described, the table is filled with 100 values in the range 0 : 1. The output of the slider is used for the value and the index. To make `myData` a bigger array, you'd simply edit the value 100 in the table object. Finally, it would take 1 second to play back; the sample rate gives us this answer.

For part (d), one argument for Pure Data might be that it is more intuitive because it has a visual programming paradigm, unlike Java or Processing, though a disadvantage is that it can be quite complicated to wire up all the objects. It was essential to include comparison with other languages that could be used.

For part (e), examiners expected an understanding that frequency modulation involves oscillators modulating the frequency of other oscillators, while additive synthesis involves oscillators being added together.

Finally, for part (f), examiners required demonstration of understanding through explanation and discussion. There are two oscillators added together that have very similar frequencies so they would interfere with each other at a low frequency, causing a 2Hz beating. To play just the first two components of the harmonic series, starting at 200Hz, you would need to set the first frequency (f0) to 200 and the second (f1) to 400.

Question 3

Algorithmic composition and musical interaction

This was a very popular question, chosen by almost all candidates and answered reasonably by most.

Part (a) was a simple question, and most candidates were able to correctly answer that developing an algorithmic music composition involves writing a computer program that writes music, while writing music directly involves a human being creating the music itself.

One good approach to part (b) would be to explain that the music does not need to be repeated and it can be parameterised, so it can vary as the game-play varies. This is not the only response that can be given, and other reasonable ones would be accepted.

Part (c) indicated very specific requirements in the answers expected. Candidates had to identify, for sub-part (i), one each of a characteristic that the system does have and does not have. The system is clearly transparent, as its functionality is visible if one looks at the window. It is not reflective as it does not listen to the human at all. Not all candidates provided justification of this sort for their claims. For sub-part (ii), some examples could be that you could maximise reflection by changing the distribution based on what the human plays (for example the notes or intervals), and you could increase transparency by visualising the distributions and the sampling process. Again, other appropriate examples would be acceptable.

Part (d) expected candidates to explain that the concept of a metrical hierarchy is that there are several metres in a piece of music, for example, the interval between the fastest notes, the interval between start of each bar, and so on. Another explanation could be that different patternings of the notes in time are translated into several grids of events.

A simple answer for part (e) would be to start by taking the note sequence, and find smallest difference between notes; this creates the first layer of the hierarchy. Then keep doubling the note interval distance and gaining a layer until the distance is equal to the original distance between the first and last note.

For part (f), beat detection in this circumstance is because there are many events occurring at the same time and it is tricky to decide which of these dictate the beat.

Finally, for part (g), any coherent and sensible approaches were accepted. One example could be that the process would follow something along these lines:

```
Set interval to first offset value.
Computer predicted offset values from 0 with that interval.
Count how many matches there are to the actual sequences and how
many non-matches.
Increase the interval, repeat.
Select the interval with the highest hit to miss ratio.
```

Question 4

Music Information Retrieval

This question was chosen by almost all candidates, with reasonable answers.

Part (a) was answered surprisingly weakly by some candidates. The task requires the identification of a piece of music, which would include stating the artist, track name, and so on. One approach might be that of Shazam, which uses spectral peaks; other examples include MusicBrainz. Marks were awarded for describing the approach, which not all candidates provided.

For part (b) the most reasonable answer is that it has low specificity as there are many answers to the question 'which music will they like?'.

Part (c) relates directly to material in the subject guide. Content-based similarity involves analysing the audio content of the music and looking for similarities there. This relates to the notes and sound itself. Meta-data based involves using things like title, artist, year plus playlists to look for similar sounds.

The bag-of-frames content based model, for part (d), takes the approach of extracting features from the audio and they are all combined into a single, mega feature, which is the bag.

The most reasonable response for part (e) is that a content-based approach would be more successful, because meta-data does not tell you when the chorus happens.

Part (f), the final part, needed discussion of some aspects of a recommender system design. Sub-part (i) required a diagram, including a database and showing the recommendation process. Sub-part (ii) was about the data such a system would need, and in particular how it would be used in the system. The data could be content-based or meta-data, but it was essential to be explicit about how it would be used. Finally, for sub-part (iii), again, a response showing some technical insights was expected.