# The C@merata task at MediaEval 2016: Natural Language Queries Derived from Exam Papers, Articles and Other Sources against Classical Music Scores in MusicXML

Richard Sutcliffe
School of CSEE
University of Essex
Colchester, UK
rsutcl@essex.ac.uk

Tom Collins
Department of Psychology
Lehigh University
Bethlehem, PA, USA
tomthecollins@gmail.com

Eduard Hovy
Language Technologies Institute
Carnegie-Mellon University
Pittsburgh, PA, USA
hovy@cmu.edu

Richard Lewis
Department of Computing
Goldsmiths, University of London
London, UK
richard.lewis@gold.ac.uk

Chris Fox
School of CSEE
University of Essex
Colchester, UK
foxcj@essex.ac.uk

Deane L. Root
Department of Music
University of Pittsburgh
Pittsburgh, PA, USA
dlr@pitt.edu

## ABSTRACT

Cl@ssical Music Extraction of Relevant Aspects by Text Analysis (C@merata) is a shared evaluation held at MediaEval and this is the third time the task has run. The input is a natural language query ('F# in the cello') and the output is a passage in a MusicXML score which contains this note played on the instrument in question. There are 200 such questions each year and evaluation is via modified versions of Precision, Recall and F-Measure. In 2014 and 2015 the best Beat F (BF) scores were 0.797 and 0.620, both attained by CLAS. This year, queries were more difficult and in addition the most experienced groups from previous years were unable to take part. In consequence, the best BF was 0.070. This year, there was progress concerning the development of the queries, many of these being derived from real sources such as exam papers, books and scholarly articles. We are thus converging on our goal of relating musical references in complex natural language texts to passages in music scores.

## 1. INTRODUCTION

The aim of the Cl@ssical Music Extraction of Relevant Aspects by Text Analysis (C@merata) evaluations is to advance our knowledge of how musical references in texts relate to actual passages within symbolic music scores. The evaluation takes place at MediaEval each year and consists of a series of 200 questions. Each question consists of a short string making reference to a musical feature (A4 sung to the word 'bow') and a score in MusicXML. Participants have to build a system which can answer each question by specifying a set of one or more passages which exactly demarcate the feature in question. In the example above, we wish to know the exact beginning and end of a note of unspecified length which has pitch A4 and which is being sung to the word specified ('bow').

The general organisation of the task has remained the same for three years; this includes the XML format of the question file, the format of the answer file, the type of music scores used (MusicXML) and the means of evaluating answers (Beat Recall/Precision and Measure Recall/Precision). Detailed descriptions of all these can be found in the 2014 [15,16] and 2015 [18,19] overview papers. In this paper, we will start off by summarising the task and the means of evaluation. We then discuss the development of the question types over the past three years and in particular focus on the more sophisticated methods adopted for question generation this year. We will then present the participating systems for this year and discuss the results which they obtained.


*MediaEval 2016 Workshop*, October 20-21, 2016, Amsterdam.

## 2. TASK SUMMARY

Paticipants have to build a system which will take as input an XML file containing 200 questions and produce as output a file containing one or more answers for each of them. Questions are grouped into twenty sets of ten, each corresponding to a particular MusicXML score. There are thus twenty scores in total and each year these are different. The scores for each year are listed in Table 2 (at the end of the paper).

Scores are chosen from public sources, are in MusicXML [9] and are required to fall on a predefined distribution of staves (Table 3). There are two main forms of English musical terminology in use today, European English (crotchet, bar) and American English (quarter note, measure). For this reason, half of the questions in the task each year are in European English and the other half are in American English.

Concerning an answer, we use the concept of a *passage* which starts at a particular beat in a bar and ends at another beat in a bar. Bar numbers are taken from the MusicXML. Beats are measured in *divisions*, a concept taken from the MusicXML standard. A division is an integer and a value of one means we are counting in crotchets; a value of two indicates quavers and so on. Where necessary, a high divisions value (e.g. 12) can be used where we wish to beat in crotchets, quavers, crotchet triplets or quaver triplets for difference answers to the same question. In C@merata, the question always specifies the divisions value to be used for all answers to a particular question. In summary, a passage such as [4/4,1,1:1-2:4] means we are in 4/4/ time, divisions is set to one (crotchets) the passage starts *before* the first crotchet beat of bar one (1:1) and the passage ends *after* the fourth crotchet beat of bar two (2:4).

In evaluation, a passage is beat correct if it starts at the correct beat in the correct start bar and it ends at the correct beat in the correct end bar. So, if the correct answer is a crotchet, the passage must start immediately before the crotchet in question and it must end immediately after it. Similarly, a passage is measure correct if it starts in the correct start bar and ends in the correct end bar. Based on beat correct and measure correct, we can define the following:

**Beat Precision** (BP) as the number of beat-correct passages returned by a system, in answer to a question, divided by the number of passages (correct or incorrect) returned.

**Beat Recall** (BR) is the number of beat-correct passages returned by a system divided by the total number of answer passages known to exist.

**Beat F-Score** (BF) is the harmonic mean of BP and BR.

**Measure Precision** (MP) is the number of measure-correct passages returned by a system divided by the number of passages (correct or incorrect) returned.

**Measure Recall** (MR) is the number of measure-correct passages returned by a system divided by the total number of answer passages known to exist.

**Measure F-Score** (MF) is the harmonic mean of MP and MR.

# 3. QUESTIONS

## 3.1 2014 Questions

For the first year of the task [15,16], questions were chosen from the Renaissance and Baroque periods. There were twenty MusicXML scores chosen mainly from musescore.com. There were six scores containing two staves and six on three staves, four on one stave and two each on four staves and five staves. Composers were Bach, Carissimi, Charpentier, Corelli, Cutting, Dowland, Lassus, Lully, Monteverdi, Purcell, Scarlatti, Tallis, Telemann, Vivaldi and Weiss. Ten questions were posed against each score, i.e. 200 in total. For ten of the twenty scores, English terminology was used ('crotchet' etc) while for the other ten, American terminology was used ('quarter note' etc). The 200 questions were on a predefined distribution: 30 simple pitch, 30 simple length, 30 pitch and length, 10 performance specification, 20 stave specification, 5 word specification, 30 followed by, 19 melodic interval, 11 harmonic interval, 5 cadence specification, 5 triad specification and 5 texture specification. So half of the questions were quite basic and essentially asking for a note, while half the questions dealt with more complex concepts such as intervals, cadences, chords and texture. Questions of type follow asked about one note followed by another.

## 3.2 2015 Questions

For the second year [18,19], questions came from the Baroque, Classical and Early Romantic periods. Once again there were twenty MusicXML scores, mostly from musescore.com. There were now some more complex scores: there were six on two staves and six on four staves, two on one stave and two on three staves, and one each on six, seven, eight and nineteen staves. So the collection now included one full orchestral score. Composers were Bach, CPE Bach, Beethoven, Haydn, Marcello, Monteverdi, Mozart, Purcell, Schubert, Sweelinck and Vivaldi. The distribution of question types of the 200 questions was now: 40 1_melod unqualified, 40 1_melod qualified by clef, instrument etc., 20 n_melod unqualified, 20 n_melod qualified by clef, instrument etc., 20 1_harm, 6 texture, 40 follow and 14 synch. 1_melod were essentially notes while n_melod were specified sequences of notes; 1_harm queries were chords and texture was polyphony etc. Questions of type follow were now more complex because chords and sequences of notes could be followed by other chords and note sequences. The synch questions dealt with one event against another such as specified notes on one instrument being played at the same time as notes on another instrument. Generally the questions were getting more difficult and more interesting; they were closer to the kinds of musical passage which people would actually be interested in asking about.

## 3.3 2016 Questions

This year, questions were chosen from the Renaissance, Baroque, Classical and Early Romantic periods. The twenty MusicXML scores were selected from kern.ccarh.org and from musescore.com. The former scores are from Stanford and have been prepared from various public domain and out-of-copyright sources. They are created in the kern format and are also available in a conversion to MusicXML.

The distribution of scores in terms of staves can be seen in Table 4. It is similar to 2015 except that there are now five scores with eight or more staves: two on eight staves and one each on ten, thirteen and eighteen staves.

The scores themselves can be seen in Table 3, which shows the work, number of staves and scoring (i.e. the istruments used). Composers this year were Bach, Beethoven, Bennet, Chopin, Handel, Morley, Mozart, Palestrina, Scarlatti, Schubert, Vivaldi and Weelkes. There were six works for keyboard (three for harpsichord and three for piano), one Schubert song for voice and piano and two string quartet movements; there were three *a capella* vocal works for SATB and one each for SATTB and SSATB; there were two Vivaldi concertos for strings and continuo and two symphony movements, one by Mozart and the other by Beethoven. Finally there was a movement from Handel's Messiah for SATB and orchestra.

The distribution of question types is shown in Tables 1 and 2. These tables also show several examples of each type. All 1_melod questions are concerned with one note and can me modified by bar/measure ('A#1 in bars 44-59'). Thirty-six of the forty can also be modified by perf ('forte'), instr ('in the violin'), clef ('in the bass clef'), time ('in 3/4') or key ('with G major key signature').

n_melod questions are concerned with a sequence of notes which can be specified exactly ('D4 D5 A5 D6 in sixteenth notes') or inexactly ('two-note dotted rhythm'). They can also be modified by bar/measure, perf etc in the same as 1_melod questions.

1_harm questions deal with single chords ('whole-note unison E2 E3 E4') which can be less specific ('chord of C' or 'five-note chord in the bass'). Once again they can be modified as above ('chord of F#3, D4 and A4 *in the lower three parts*', 'harmonic octave *in the bass clef*'). Note that we allow two notes to be a chord, including octaves etc. This year, references to inversions ('Ia chord') are considered of 1_harm type.

n_harm questions deal with sequences of chords with the usual modifications ('three consecutive thirds in bars 1-43'). Cadences are also included here, since they are sequences of specific chords ('plagal cadence in bars 134-138'). There are also some more complex types here ('A5 pedal in bars 116-138').

Finally, there are texture questions ('all three violin parts in unison in measures 1-59', 'counterpoint in bars 1-14'). Some more complex forms were added this year ('imitative texture in bars 1-18').

Table 2 shows two further forms of question, follow and synch. There are twenty of the former and thirteen of the latter. In a departure from last year, these are not separate types, but range over the queries of type 1_melod, n_melod, 1_harm and n_harm as shown in Table 1. Thus the examples in Table 2 are all within the distribution of query types shown in Table 1. A follow question allows us to specify some passage followed by another passage. Each such passage can be of type 1_melod, n_melod etc.

This allows quite complex sequences to be specified ('D C# in the right hand, then F A G Bb in semiquavers in the left hand', '5 B4s followed by a C5').

Questions of type synch can link two passages which must occur at the same time. In the simplest case, each passage is of exactly the same length ('quarter note E5 against a quarter note C#3'). However, this is not necessarily the case ('C#3 minim and E4 semibreve simultaneously'); here, according to our rules, the whole of the minim must lie somewhere within the duration of the semibreve. The length of the passages need not be specified ('D3 in the bass at the same time as C5 in soprano 1', 'three-note chord in the harpsichord right hand against a two-note chord in the harpsichord left hand in measures 45-52').

When we reach the follow and synch questions, they are starting to become interesting from a musicological perspective as such musical phenomena as these cannot readily be specified except in a natural language. The key advantage of language here is that it can vary in specificity from the constrained to the open; to interpret the open queries requires considerable musical knowledge. Hence, C@merata starts to become interesting and not merely a simple exercise in finding notes.

We will finish this section by summarising how we derived the questions. In previous years, we decided the question types and distribution first; we then selected scores and devised queries by going through them, trying to find passages against which queries could be posed. This reverse-logic approach was a simple development of the one which we had used at CLEF for many years [10,11,12,13,14]. This year we had aimed to generate some of the questions using a more realistic approach. Two suggestions had been made in previous years. The first was to base certain questions on First Species Counterpoint as exemplified for example in Kitson [6] and indeed Fux [2]. Certain suggestions had been made by OMDN participant Donncha Ó Maidín:

- Modes: Dorian, Phrygian, Lydian, Mixolydian, Aeolian, Locrian, Ionian (these would be n_melod queries);
- Melodic intervals: diminished fifth, augmented fourth (n_melod queries);
- Harmonic intervals: perfect concords, imperfect concords and discord (1_harm queries);
- Movement of parts: similar, contrary, oblique and parallel (n_melod against n_melod);
- Special relationships: false relation of the tritone (1_harm)
- exposed fifths and octaves (n_harm).

In the event, we managed queries relating to melodic intervals and movement of parts and we plan to investigate the others in future. The second suggestion was to base questions on music exam papers set in English schools at GCSE level (aged sixteen) and A Level (aged eighteen). This had been proposed by DMUN participant Tom Collins and by co-organiser Richard Lewis. DMUN participation was handed over to Andreas Katsiavalos this year, so Tom Collins joined the organisers and did indeed generate some questions based on a study of exam papers.

The third strand of work was concerned with the derivation of queries from musicological texts. For this campaign, we re-visited some of the texts we studied previously [17] and styled some of the more complex questions accordingly.

# 4. 2016 CAMPAIGN

## 4.1 Participants

This year, ten groups registered for the task, the largest number so far in the three years. Unfortunately, however, only four of these were able to return a run. Participants are shown in Table 5. There was one each from England, Ireland, Poland and Russia.

## 4.2 Approaches used by Systems

The following are short notes on the various systems this year. Full details can be found in the participant papers in this volume.

**DMUN**

This year, the DMUN system was re-written. The group developed a text query parser that, given a sentence such as a C@merata question, generates a script for music operations. The script contains the music concepts and their relations as described in the query, but in a structured form related to SQL in such a way that workflows of specific music data operations are formed. A parser then reads the script and calls the corresponding functions from a framework created on top of music21 [1].

**KIAM**

The KIAM system is written in PHP and is based on regular expressions. Queries are categorised and analysed using regular expressions; answers are then extracted from raw MusicXML files.

**OMDN**

The system used is similar to last year and is based on the author's CPNView system.

**UMFC**

The UMFC system was based on recurrent neural networks which is undoubtedly the most original approach.

## 4.3 Runs and Results

Overall results are shown in Table 6. The best run was DMUN01 which scored BF 0.070 and MF 0.106. These results are very low. Moreover, they are only based on a small subset of the queries. The best scores in 2014 were BF 0.797 and MF 0.854, and in 2015 were BF 0.620 and MF 0.656. In both previous years, the CLAS system was the best but unfortunately they could not participate this year. The questions were more difficult this year but another factor was that three of the four participants had not taken part before. All systems this year were able to answer simple note questions such as 'C# crotchet' but the task has moved on and there are now very few 1_melod questions as simple as that.

Table 6 shows the average results by question type. As expected, 1_melod are the the easiest overall (BF 0.054) followed by n_melod (BF 0.028). After that come 1_harm (BF 0.019) and n_harm (BF 0.013). These questions are progressively more difficult, so this order is to be expected. Texture questions this year could not be answered by any system (BF 0).

Turning to the more complex follow and synch questions (which were a subset of the questions of other types this year), the bottom half of Table 6 shows the average scores for follow (BF

0.078) and synch (BF 0). The follow score is higher than might be expected and this is because KIAM scored well on these questions (see Table 13, KIAM, BF 0.227, the next highest being DMUN with BF 0.044). While BF for synch questions was 0, MF was 0.018, so some systems could at least determine the correct start and end bar, if not the exact beat.

Looking at the results for different systems across question types, the performance of UMFC on 1_harm questions is noteworthy (BF 0.042) this system scored BF 0 on 1_melod and n_melod questions.

Generally, all the systems scored very low on Recall, because they missed out many correct answers. However, they were much better on Precision, because answers returned tended to be correct. Looking at Table 6, average BP over all participants was 0.167 and average MP was 0.447. The best BP was 0.420 and the best MP was 0.640 (both DMUN). These are more respectable looking figures. In fact, looking at MP in Table 6, three out of the four systems scored 0.511 or better. The problem with systems this year was that they were not sophisticated enough to handle complex questions; those that they did answer were often correct.

If we look at the MP figures for the various different questions types (Tables 8-14) we can see the highest values scored on any measure in this year's campaign: For 1_melod, DMUN scored 0.857 and KIAM scored 0.727; for n_melod, DMUN scored 0.706; for 1_harm OMDN scored 0.8; for n_harm, DMUN scored 0.5. Finally, concerning follow and synch, for follow, KIAM scored 1 and OMDN scored 0.333; for synch, OMDN scored 0.333. Of course, in most of these cases, a system was chancing to match passages for the particularl queries and was not in fact attempting the vast majority of questions, leading to very low MR and MF scores.

## 5. Discussion and Conclusions

The main achievement this year was to develop some very interesting questions against some quite complex scores. Questions were a lot more complicated than those of last year which in turn were more complicated than in the first year when many were quite elementary.

Unfortunately, the systems have not kept up with the task. They can answer very few of the questions, though answers when returned are often correct, at least for MF. There is some ambiguity concerning the exact beats of certain questions, in particular those of types follow and synch, and this can partly account for the much lower BF scores than MF scores.

Concerning practical aspects of the organisation, as always there were two problems in finding scores. The first was that there is a limited choice of public domain works, especially ones which are of sufficiently high quality. Over the years we have more-or-less exhausted the musescore archive [8] because most scores there are not licensed to be distributed, only for private use. Moreover, our scores need to fit into a complex distribution of musical periods, numbers of staves and instrumentation. For these reasons we have turned increasingly to the Stanford CCARH Kern scores [5]. There is a very good choice of scores and the level of scholarship and accuracy there are high.

However, this leads to the second problem, which is the quality of the MusicXML. Scores on musescore are produced mainly by amateurs using many different types of software. Conversion to MusicXML is carried out at the end of the encoding process and relies on the extent to which the score writing software supports it. We have found that many MusicXML files do not load in the Musescore software [7] or contain errors or anomalies. For example, an anacrusis bar may be numbered one rather than the more usual zero.

The situation for the kern scores is rather similar. Conversion to MusicXML is carried out by a script at the end; MusicXML is not considered the primary format as it is assumed that kern will always be used. We have found that some of these conversions are inaccurate or crash Musescore. In some cases we were able to re-convert from kern using the latest version of music21 [1]. In other cases, a new score had to be selected for the evaluation.

Finally, looking to the future, there have now been three years where the C@merata query and answer format remained the same: the input was a noun phrase and the output was a passage. Over the years we have also begun to investigate the relationship between our queries and actual musicological texts [17]. This year we looked also at GCSE and A Level exam questions. Naturally, such questions are not in the raw C@merata format and have to be converted in order to fit the task. Sometimes this is not possible because of restrictions in our task. An interesting instance of this is exam questions like 'Which one of the following terms best describes the music at bars x-y? <list of terms>'. This is a sort of ranking task as several terms are specified and we must say which is the most important. The nature of the question implies that all the terms specified apply to the passage to a greater or lesser extent. This is one way of reflecting the intrinsic ambiguity of musical analysis which we could consider for future campaigns.

## 6. REFERENCES


[1] Cuthbert, M. S., and Ariza C. 2010. music21: a toolkit for computer-aided musicology and symbolic music data. In Proceedings of the International Symposium on Music Information Retrieval (Utrecht, The Nethlerands, August 09 - 13, 2010). 637-642.

[2] Fux, J. J. (1725). Gradus ad Parnassum (Practical Rules for Learning Composition translated from a Work intitled Gradus ad Parnassum written originally in Latin by John Joseph Feux). Translated around 1750 by unknown translator. London, Welcker. http://imslp.nl/imglnks/usimg/3/31/IMSLP370587-PMLP187246-practicalrules fo00fuxj.pdf

[3] Huron, D. (1997). Humdrum and Kern: Selective Feature Encoding. In Beyond MIDI, ed E. Selfridge-Field (Cambridge, Massachusetts: The MIT Press, 1997), pp. 375-401.

[4] Huron, D. (2002). Music information processing using the Humdrum toolkit: concepts, examples, and lessons. Comput. Music J. 26, 2, 11-26.

[5] Kern Scores (2016). http://kern.ccarh.org

[6] Kitson, C. H. (1907). The Art of Counterpoint and its Application as a Decorative Principle. Oxford, UK, Clarendon Press. https://archive.org/details/artofcounterpoin00kitsuoft

[7] Musescore (2016). Music Composition and Notation Software. http://musescore.org/

[8] Musescore Music Achive (2016). https://musescore.com

[9] MusicXML (2016). http://www.musicxml.com/

[10] Peñas, A., Forner, P., Sutcliffe, R., Rodrigo, A., Forascu, C., Alegria, I., Giampiccolo, D., Moreau, N. and Osenova, P. (2009). Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation Notebook of the Cross Language Evaluation Forum, CLEF 2009, Corfu, Greece, 30 September - 2 October.

[11] Peñas, A., Hovy, E., Forner, P., Rodrigo, A., Sutcliffe, R., Forascu, C., Sporleder, C. (2011). Overview of QA4MRE at CLEF 2011: Question Answering for Machine Reading Evaluation. Proceedings of QA4MRE-2011. Held as part of CLEF 2011.

[12] Peñas, A., Hovy, E., Forner, P., Rodrigo, A., Sutcliffe, R., Sporleder, C., Forascu, C., Benajiba, Y., Osenova, P. (2012). Overview of QA4MRE at CLEF 2012: Question Answering for Machine Reading Evaluation. Proceedings of QA4MRE-2012. Held as part of CLEF 2012.

[13] Peñas, A., Hovy, E., Forner, P., Rodrigo, A., Sutcliffe, R., Morante, R. (2013). QA4MRE 2011-2013: Overview of Question Answering for Machine Reading Evaluation. Lecture Notes in Computer Science Volume 8138, 2013, pp 303-320.

[14] Peñas, A., Magnini, B., Forner, P., Sutcliffe, R., Rodrigo, A., & Giampiccolo, D. (2012). Question Answering at the Cross-Language Evaluation Forum 2003-2010. Language Resources and Evaluation Journal, 46(2), 177-217.

[15] Sutcliffe, R. F. E., Crawford, T., Fox, C., Root, D. L., & Hovy, E. (2014). The C@merata Task at MediaEval 2014: Natural language queries on classical music scores. In Proceedings of the MediaEval 2014 Workshop, Barcelona, Spain, October 16-17 2014. http://ceur-ws.org/Vol-1263/mediaeval2014_submission _46.pdf

[16] Sutcliffe, R. F. E., Crawford, T., Fox, C., Root, D. L., & Hovy, E. (2014). Shared Evaluation of Natural Language Queries against Classical Music Scores: A Full Description of the C@merata 2014 Task. Proceedings of the C@merata Task at MediaEval 2014. http://csee.essex.ac.uk/camerata/.

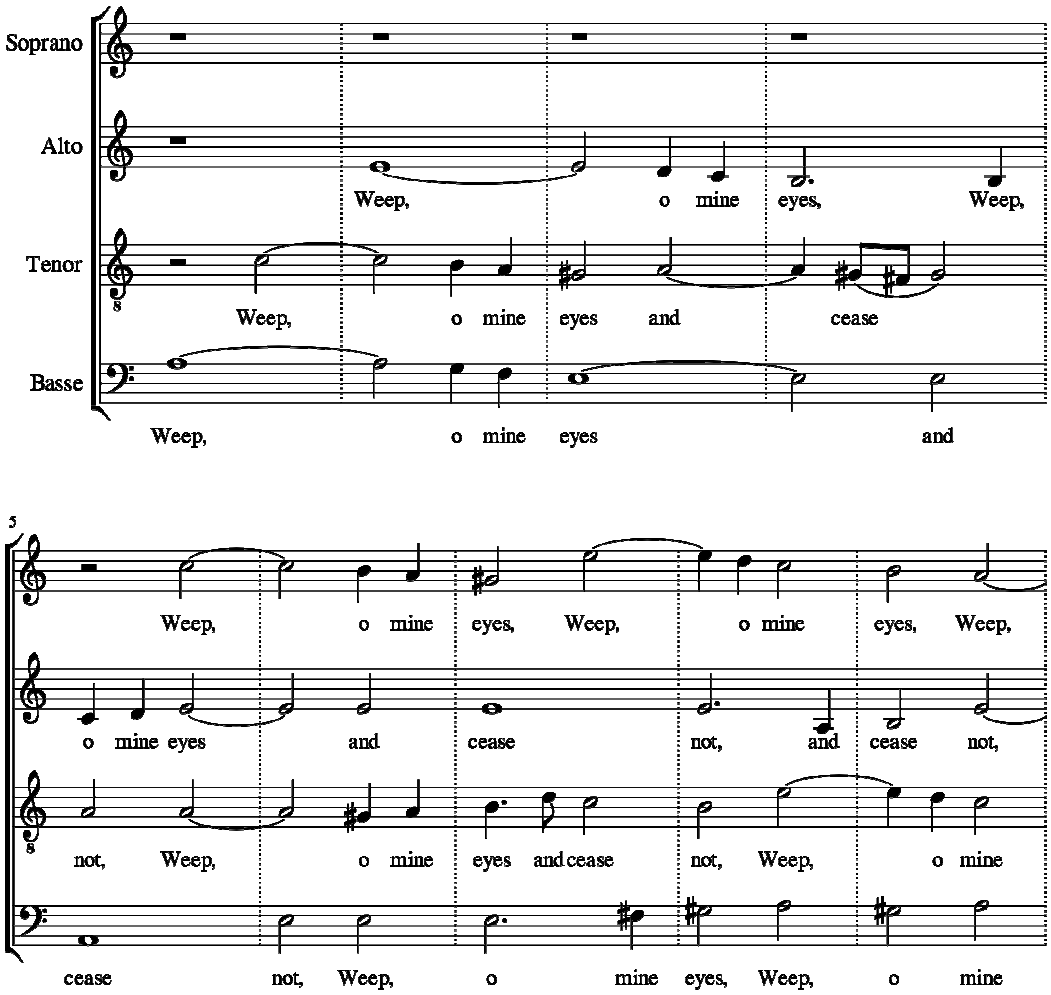
[17] Sutcliffe, R. F. E., Crawford, T., Fox, C., Root, D. L., Hovy, E., & Lewis, R. (2015). Relating Natural Language Text to Musical Passages. Proceedings of the 16th International Society for Music Information Retrieval Conference, Malaga, Spain, 26-30 October, 2015. http://ismir2015.uma.es/.

[18] Sutcliffe, R. F. E., Fox, C., Root, D. L., Hovy, E., & Lewis, R. (2015). The C@merata Task at MediaEval 2015: Natural language queries on classical music scores. In Proceedings of the MediaEval 2015 Workshop, Dresden, Germany, September 14-15 2015. http://ceur-ws.org/Vol-1436/Paper12.pdf.

[19] Sutcliffe, R. F. E., Fox, C., Root, D. L., Hovy, E. and Lewis, R. (2015). Second Shared Evaluation of Natural Language Queries against Classical Music Scores: A Full Description of the C@merata 2015 Task. Proceedings of the C@merata Task at MediaEval 2015. http://csee.essex.ac.uk/camerata/.

# WEEP, O MINE EYES

John BENNET, 1599




**Q31**: *minim on the word "cease"*
**A**: [ 4/4, 1, 7:3-7:4 ], [ 4/4, 1, 9:1-9:2 ]
**Q32**: *note on the word "mine" against a note on the word "eyes"*
**A**: [ 4/4, 1, 3:4-3:4 ]
**Q33**: *semibreve tied to a minim in the Bass clef*
**A**: [ 4/4, 1, 1:1-2:2 ], [ 4/4, 1, 3:1-4:2 ]
**Q35**: *semibreve tied to a minim followed by crotchet, crotchet*
**A**: [ 4/4, 1, 1:1-2:4 ], [ 4/4, 1, 2:1-3:4 ]
**Q36**: *minor third between Basse and Tenor in bars 1-35*
**A**: [ 4/4, 1, 1:3-2:2 ], [ 4/4, 1, 8:1-8:2 ], [ 4/4, 1, 9:3-9:4 ]
**Q37:** *Ia chord in bars 1-10*
**A:** [ 4/4, 1, 2:1-2:2 ], [ 4/4, 1, 5:3-5:4 ], [ 4/4, 1, 8:3-8:4 ], [ 4/4, 1, 9:3-9:4 ]
**Q40:** *counterpoint in bars 1-14*
**A:** [ 4/4, 1, 1:3-14:4 ]

**Figure 1. Sample questions and anwers from 2016 task**

**Table 1. Query Types**

| Type | No | Example |
|---|---|---|
| **1_melod** | 4 | A#1 in bars 44-59<br>quarter-note rest in measures 1-5 |
| **1_melod** qualified by **perf**, **instr**, **clef**, **time**, **key** | 36 | dotted quarter note D6 in the first violin<br>solo C5 in the oboe in measures 32 onwards<br>flute dotted half note only against strings<br>half note on the tonic in the bass clef<br>A4 sung to the word 'bow' |
| **n_melod** | 15 | two-note dotted rhythm in measures 1-24<br>eight note rising passage in quarter notes<br>repeated Bb4 whole note<br>D4 D5 A5 D6 in sixteenth notes repeated twice<br>two tied dotted minims in bars 72-88 |
| **n_melod** qualified by **perf**, **instr**, **clef**, **time**, **key** | 45 | dotted minims C B A in the Bass clef in bars 70-90<br>melodic interval of a minor 7th in the voice<br>rising arpeggio in the left hand in measures 1-10<br>five-note melody in the cello in measures 20-28<br>whole note rest, quarter note in the Violin 4 in measures 1-103 |
| **1_harm** | 17 | 7th triad in measures 1-3<br>Ia chord in bars 1-10<br>chord of C<br>whole-note unison E2 E3 E4<br>chord III in bars 44-59 |
| **1_harm** possibly qualified by **perf**, **instr**, **clef**, **time**, **key** | 23 | chord of F#3, D4 and A4 in the lower three parts<br>harmonic fifth in the oboe<br>harmonic octave in the bass clef<br>harmonic perfect fourth between the Soprano and Alto in bars 1-9<br>cello and viola playing dotted minims an octave apart in bars 40-70 |
| **n_harm** | 25 | interrupted cadence<br>A5 pedal in bars 116-138<br>authentic cadence in measures 14-18<br>plagal cadence in bars 134-138<br>three consecutive thirds in bars 1-43 |
| **n_harm** possibly qualified by **perf**, **instr**, **clef**, **time**, **key** | 15 | consecutive sixths between the Altos and Basses in measures 73-80<br>flute, oboe and bassoon in unison in measures 1-56<br>consecutive descending sixths in the left hand<br>alternating fourths and fifths in the Oboe in bars 1-100<br>Soprano and Alto moving one step down together in measures 1-12 |
| **texture** | 20 | all three violn parts in unison in measures 1-59<br>polyphony in measures 5-12<br>homophonic texture in measures 125-138<br>imitative texture in bars 1-18<br>counterpoint in bars 1-14 |
| **All** | **200** | |

**Table 2. follow and synch Queries within 1_melod, n_melod, 1_harm** and **n_harm**

| Type | No | Example |
|---|---|---|
| **follow** possibly qualified on either or both sides by **perf**, **instr**, **clef**, **time**, **key** | 20 | C D E F D E C in semiquavers repeated after a semiquaver<br>eighth-note twelfth followed by whole-note minor tenth between Cello and Viola<br>D C# in the right hand, then F A G Bb in semiquavers in the left hand<br>B flat in the cbass followed a quarter note later by B natural in the cbass<br>5 B4s followed by a C5 |
| **synch** possibly qualified in either or both parts by **perf**, **instr**, **clef**, **time**, **key** | 13 | quarter note E5 against a quarter note C#3<br>C#3 minim and E4 semibreve simultaneously<br>D3 in the bass at the same time as C5 in soprano 1<br>three-note chord in the harpsichord right hand against a two-note chord in the harpsichord left hand in measures 45-52<br>A#3 in the piano and F#5 in the voice simultaneously |

**Table 3. Scores Used**

| Work | Staves | Scoring | Lang |
|---|---|---|---|
| bach_2_part_invention_no1_bwv772 | 2 | hpd | Eng. |
| beethoven_piano_sonata_no2_m4 | 2 | pf | Amer. |
| beethoven_piano_sonata_no5_m1 | 2 | pf | Eng. |
| chopin_prelude_op28_no15 | 2 | pf | Eng. |
| scarlatti_sonata_k281 | 2 | hpd | Eng. |
| scarlatti_sonata_k320 | 2 | hpd | Amer. |
| schubert_an_die_musik_d547 | 3 | S, pf | Amer. |
| bach_chorale_bwv347 | 4 | SATB | Amer. |
| beethoven_str_quartet_op127_m1 | 4 | 2 vn, va, vc | Eng. |
| bennet_weep_o_mine_eyes | 4 | SATB | Eng. |
| handel_water_music_suite_air | 4 | 2 vn, va, vc | Amer. |
| palestrina_alma_redemptoris_mater | 4 | SATB | Amer. |
| schubert_str_quartet_no10_op125_d87_m3 | 4 | 2 vn, va, vc | Eng. |
| morley_now_is_the_month_of_maying | 5 | SATTB | Eng. |
| weelkes_hark_all_ye_lovely_saints | 5 | SSATB | Eng. |
| vivaldi_conc_4_vns_op3_no10_rv580 | 8 | 4 vn, 2 va, vc, db | Amer. |
| vivaldi_conc_vn_op6_no6_rv239_m1 | 8 | 3 vn, va, vc, db, hpd | Amer. |
| mozart_symphony_no40_m4 | 10 | fl, 2 ob, 2 bn, 2 hn, 2 vn, va, vc, db | Eng. |
| beethoven_symphony_no3_m3 | 13 | 2 fl, 2 ob, 2 cl, 2 bs, 2 hn, 2 tpt, timp, 2 vn, va, vc, db, | Amer. |
| handel_messiah_and_the_glory | 18 | fl, 2 ob, cl, bs, hn, tbn, tuba, SATB, hpd, 2 vn, va, vc, db | Amer. |

**Table 4. Distribution of Scores by number of Staves**

| Staves | Frequency |
|---|---|
| 2 | 6 |
| 3 | 1 |
| 4 | 6 |
| 5 | 2 |
| 8 | 2 |
| 10 | 1 |
| 13 | 1 |
| 18 | 1 |
| **All** | **20** |

**Table 5. C@merata Participants**

| Runtag | Leader | Affiliation | Country |
|---|---|---|---|
| DMUN | Andreas Katsiavalos | De Montfort University | England |
| KIAM | Marina Mytrova | Keldysh Institute of Applied Mathematics | Russia |
| OMDN | Donncha Ó Maidín | University of Limerick | Ireland |
| UMFC | Paweł Cyrta | Fryderyk Chopin University of Music | Poland |

**Table 6. Results for All Questions:** DMUN01 is the best run. BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| DMUN01 | 0.420 | 0.038 | **0.070** | 0.640 | 0.058 | **0.106** |
| KIAM01 | 0.194 | 0.011 | 0.021 | 0.613 | 0.035 | 0.066 |
| OMDN01 | 0.042 | 0.004 | 0.007 | 0.511 | 0.044 | 0.081 |
| UMFC01 | 0.012 | 0.038 | 0.018 | 0.022 | 0.073 | 0.034 |
| **Maximum** | 0.420 | 0.038 | 0.070 | 0.640 | 0.073 | 0.106 |
| **Minimum** | 0.012 | 0.004 | 0.007 | 0.022 | 0.035 | 0.034 |
| **Average** | 0.167 | 0.023 | 0.029 | 0.447 | 0.053 | 0.072 |

**Table 7. Average Results by Question Type:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score. Note that in 2016 **follow** and **synch** questions are across **1_melod**, **n_melod**, **1_harm** and **n_harm**.

| Type | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **1_melod** | 0.232 | 0.044 | 0.054 | 0.520 | 0.101 | 0.129 |
| **n_melod** | 0.125 | 0.016 | 0.028 | 0.384 | 0.051 | 0.086 |
| **1_harm** | 0.076 | 0.023 | 0.019 | 0.300 | 0.033 | 0.035 |
| **n_harm** | 0.063 | 0.007 | 0.013 | 0.128 | 0.032 | 0.030 |
| **texture** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **follow** | 0.317 | 0.047 | 0.078 | 0.458 | 0.076 | 0.126 |
| **synch** | 0.000 | 0.000 | 0.000 | 0.103 | 0.011 | 0.018 |

**Table 8. Results for 1_melod Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.643 | 0.066 | 0.120 | 0.857 | 0.088 | 0.160 |
| **KIAM01** | 0.273 | 0.044 | 0.076 | 0.727 | 0.118 | 0.203 |
| **OMDN01** | 0.000 | 0.000 | 0.000 | 0.474 | 0.066 | 0.116 |
| **UMFC01** | 0.011 | 0.066 | 0.019 | 0.022 | 0.132 | 0.038 |
| **Maximum** | 0.643 | 0.066 | 0.120 | 0.857 | 0.132 | 0.203 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.022 | 0.066 | 0.038 |
| **Average** | 0.232 | 0.044 | 0.054 | 0.520 | 0.101 | 0.129 |

**Table 9. Results for n_melod Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.412 | 0.050 | 0.089 | 0.706 | 0.085 | 0.152 |
| **KIAM01** | 0.000 | 0.000 | 0.000 | 0.333 | 0.021 | 0.040 |
| **OMDN01** | 0.087 | 0.014 | 0.024 | 0.478 | 0.078 | 0.134 |
| **UMFC01** | 0.000 | 0.000 | 0.000 | 0.019 | 0.021 | 0.020 |
| **Maximum** | 0.412 | 0.050 | 0.089 | 0.706 | 0.085 | 0.152 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.019 | 0.021 | 0.020 |
| **Average** | 0.125 | 0.016 | 0.028 | 0.384 | 0.051 | 0.086 |

**Table 10. Results for 1_harm Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.273 | 0.018 | 0.034 | 0.364 | 0.024 | 0.045 |
| **KIAM01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **OMDN01** | 0.000 | 0.000 | 0.000 | 0.800 | 0.024 | 0.047 |
| **UMFC01** | 0.030 | 0.072 | 0.042 | 0.035 | 0.084 | 0.049 |
| **Maximum** | 0.273 | 0.072 | 0.042 | 0.800 | 0.084 | 0.049 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Average** | 0.076 | 0.023 | 0.019 | 0.300 | 0.033 | 0.035 |

**Table 11. Results for n_harm Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.250 | 0.028 | 0.050 | 0.500 | 0.056 | 0.101 |
| **KIAM01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **OMDN01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **UMFC01** | 0.000 | 0.000 | 0.000 | 0.011 | 0.070 | 0.019 |
| **Maximum** | 0.250 | 0.028 | 0.050 | 0.500 | 0.070 | 0.101 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Average** | 0.063 | 0.007 | 0.013 | 0.128 | 0.032 | 0.030 |

**Table 12. Results for texture Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **KIAM01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **OMDN01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **UMFC01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Maximum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Average** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 13. Results for follow Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.500 | 0.023 | 0.044 | 0.500 | 0.023 | 0.044 |
| **KIAM01** | 0.600 | 0.140 | 0.227 | 1.000 | 0.233 | 0.378 |
| **OMDN01** | 0.167 | 0.023 | 0.040 | 0.333 | 0.047 | 0.082 |
| **UMFC01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Maximum** | 0.600 | 0.140 | 0.227 | 1.000 | 0.233 | 0.378 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Average** | 0.317 | 0.047 | 0.078 | 0.458 | 0.076 | 0.126 |

**Table 14. Results for synch Questions:** BP=Beat Pecision, BP=Beat Recall, BF=Beat F-Score, MP=Measure Precision, MR=Measure Recall, MF=Measure F-Score.

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| **DMUN01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **KIAM01** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **OMDN01** | 0.000 | 0.000 | 0.000 | 0.333 | 0.021 | 0.040 |
| **UMFC01** | 0.000 | 0.000 | 0.000 | 0.077 | 0.021 | 0.033 |
| **Maximum** | 0.000 | 0.000 | 0.000 | 0.333 | 0.021 | 0.040 |
| **Minimum** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Average** | 0.000 | 0.000 | 0.000 | 0.103 | 0.011 | 0.018 |