# The C@merata Task at MediaEval 2015: Natural Language Queries on Classical Music Scores

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

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

## ABSTRACT

This was the second year of the C@merata task [16,1] which relates natural language processing to music information retrieval. Participants each build a system which takes as input a query and a music score and produces as output one or more matching passages in the score. This year, questions were more difficult and scores were more complex. Participants were the same as last year and once again CLAS was the best with a Beat F-Score of 0.620.

## INTRODUCTION

The C@merata task is a kind of Question Answering [13,17,2,12,18] combined with Music Information Retrieval [5,6]. The input is a phrase such as 'dotted minim F#4' together with a score in MusicXML [11] and the output is a list of one or more passages in the score each containing such a minim.

There are three main applications for C@merata-type systems. First, we have observed in Grove and elsewhere [7,14,3,8,10] that musicological analyses make references to musical passages. For example, consider 'cellos and basses lead us into the shadows while the upper strings accompany with gently throbbing harmonies' [8, p17]. This refers to a passage in Beethoven's First Symphony, but where exactly?

Second, experts may wish to find a specific passage based on a possibly vague description, for example, 'the Wagner coda from the 7th symphony of Bruckner'.

Third, students of music who are unsure what an 'interrupted cadence' is could benefit from a system which could find examples such as 'The trumpet shall sound' from Handel's Messiah. These three applications motivate our work.

## 1. APPROACH

### 1.1 The C@merata Task

Participants are given 200 questions and twenty scores in MusicXML, ten questions on each score. The task is to find one or more answer passages for each question. Suppose the query is 'dotted minim F#4' against the Andante of BWV 1047 (Figure 1). An answer passage is [ 3/4, 1, 65:1-65:3 ]. This means time signature 3/4, measuring in crotchets, passage starts before the first crotchet in bar 65 and ends after the third crotchet.

The twenty scores were chosen from Baroque, Classical and Romantic composers. They ranged in complexity from one stave up to nineteen staves and from a few bars up to a hundred or more. Query types were different from 2014 (Table 1) and consisted of eight base types which could have certain qualifications. Some were similar to last year ('D4 minim') while others were more complex ('quavers F4 E4 in the oboe followed by quavers E2 G#2 in the bass clef').


Q: dotted minim F#4
A: [ 3/4, 1, 65:1-65:3 ]
Q: F4 crotchet in the oboe
A: [ 3/4, 2, 64:3-64:4 ]
Q: minim A2 in 3/4 time
A: [ 3/4, 1, 62:2-62:3 ], [ 3/4, 1, 64:2-64:3 ]
Q: chord D2 E5 G5 in bars 54-58
A: [ 3/4, 2, 57:1-57:1 ]
Q: quavers F3 A3 followed by crotchet A4 in the violin
A: [ 3/4, 1, 57:2-57:3 ]
Q: four quavers in the violin against a minim in the bass clef
A: [ 3/4, 1, 62:2-62:3 ], [ 3/4, 1, 64:2-64:3 ]

**Figure 1. Extract from Bach BWV 1047 Andante with sample questions and answers**

### 1.2 Evaluation Metrics

A passage is **beat-correct** if it starts in the correct bar at the correct beat and it ends at the correct bar at the correct beat. **Beat Precision** (BP) is 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. Similarly, **Beat Recall**


*MediaEval 2015 Workshop*, September 14-15, 2015, Wurzen, Germany.

(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.

A passage is **measure-correct** if it starts in the correct bar not necessarily at the correct beat and it ends at the correct bar not necessarily at the correct beat. **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.

**Table 1. Distribution of Query Types with Examples**

| Type | No | Example |
|---|---|---|
| 1_melod | 40 | D4 minim; eighth note in measure 9 |
| 1_melod qualified by perf, instr, clef, time, key | 40 | trill on a quaver A; G# in the Cello part in measures 29-39; sixteenth note C# in the left hand; half note E3 in 2/2; sixteenth note G in G minor in measures 1-5 |
| n_melod | 20 | F# E G F# A; Do Mi Do Sol Do Mi Sol Do in bars 1-20; twenty semiquavers; five note melody in bars 1-10 |
| n_melod qualified by perf, instr, clef, time, key | 20 | two staccato quarter notes in the Violin 1; crotchet, crotchet rest, crotchet rest, crotchet, crotchet rest, crotchet, crotchet, crotchet, crotchet, crotchet in the Timpani; melodic octave leap in the bass clef in measures 70-80; G4 B4 E5 in 3/4; rising G minor arpeggio |
| 1_harm possibly qualified by perf, instr, clef, time, key | 20 | eighth note chord Bb, C, E; chord of D minor in measures 109-110; harmonic minor sixth in the Violas; dotted minim chord in the left hand |
| texture | 6 | monophonic passage; homophony in measures 1-14; polyphony in measures 10-14; Alberti bass in measures 0-4 |
| follow possibly qualified on either or both sides by perf, instr, clef, time, key | 40 | quavers F4 E4 in the oboe followed by quavers E2 G#2 in the bass clef; quarter note minor third followed by eighth note unison; C followed by mordent Bb; chord C4 G4 C5 E5 then a quaver; three eighth notes in the Violin I followed by twelve sixteenth notes in the Violin II in measures 87-92 |
| synch possibly qualified in either or both parts by perf, instr, clef, time, key | 14 | four eighth notes against a half note; crotchet D3 on the word “je” against a minim D2; four staccato quavers in the Violoncello against a minim chord Ab3 C4 F4 in the Harpsichord |
| **All** | **200** | |

## 1.3 Gold Standard Queries

200 questions were prepared according to a carefully crafted distribution of query types (Table 1). Answers were identified in the scores and checked by two further experts. The data was used to create the Gold Standard for evaluating results automatically.

**Table 2. C@merata Participants**

| Runtag | Leader | Affiliation | Country |
|---|---|---|---|
| CLAS | Stephen Wan | CSIRO | Australia |
| DMUN | Tom Collins | De Montfort University | England |
| OMDN | Donncha Ó Maidín | University of Limerick | Ireland |
| TNKG | Nikhil Kini | Thane NK Group | India |
| UNLP | Kartik Asooja | NUI Galway | Ireland |

**Table 3. Results by Participant**

| Run | BP | BR | BF | MP | MR | MF |
|---|---|---|---|---|---|---|
| CLAS01 | 0.604 | 0.636 | 0.620 | 0.639 | 0.673 | 0.656 |
| DMUN01 | 0.311 | 0.739 | 0.438 | 0.332 | 0.788 | 0.467 |
| DMUN02 | 0.242 | 0.739 | 0.365 | 0.265 | 0.809 | 0.399 |
| DMUN03 | 0.294 | 0.739 | 0.421 | 0.316 | 0.794 | 0.452 |
| OMDN01 | 0.817 | 0.175 | 0.288 | 0.817 | 0.175 | 0.288 |
| TNKG01 | 0.061 | 0.488 | 0.108 | 0.073 | 0.586 | 0.129 |
| UNLP01 | 0.126 | 0.430 | 0.195 | 0.149 | 0.508 | 0.230 |
| **Maximum** | 0.817 | 0.739 | 0.620 | 0.817 | 0.809 | 0.656 |
| **Minimum** | 0.061 | 0.175 | 0.108 | 0.073 | 0.175 | 0.129 |
| **Average** | 0.351 | 0.564 | 0.348 | 0.370 | 0.619 | 0.375 |

# 2. RESULTS AND DISCUSSION

Five groups from four countries participated, exactly the same as in 2014 (Table 2). The results are shown in Table 3. These were lower than last year but once again CLAS was the best with BF 0.620. This was a great achievement as the questions were generally much harder this year and there were fewer ‘easy’ questions such as ‘crotchet F’ to boost the figures.

Participants generally updated and adapted their 2014 systems. Almost all worked in Python using music21 [4] and parts of the Baseline System from last year [15]. DMUN converted scores from MusicXML [11] to Kern [9] in order to use their pre-existing tools in Lisp. OMDN used their own tools in C++. Only basic NLP was used. Typically, a query was first scanned looking for terms (down bow → down_bow). Some adopted a QA approach and assigned each query to a pre-defined type, each with its method of solution. Others parsed the concepts and converted them to a structured representaton. Some varied the representation of the score according to the question, e.g. using music21 *chordify* for cadence questions. As the amount of data to be searched per query was not large (just one score) no one used any inverted indexing of the music data.

# 3. CONCLUSIONS

This was the second year, and much was learned by participants and organisers alike. All were once again able to produce a working system. Questions were more complex this year and results were lower in consequence. Future campaigns may bring use closer to the examples given in the introduction.

## 4. REFERENCES


[1] C@merata (2015). http://csee.essex.ac.uk/camerata/

[2] CLEF (2015). http://www.clef-initiative.eu/.

[3] Cooke, D. (1995). Bruckner, (Joseph) Anton. In S. Sadie (ed), *New Grove Dictionary of Music and Musicians*, Volume 3, Section 7. Music (p362-366). London, UK: Macmillan.

[4] Cuthbert, M. S., & Ariza C. (2010). music21: a toolkit for computer-aided musicology and symbolic music data. *Proc. International Symposium on Music Information Retrieval (Utrecht, The Netherlands, August 09-13, 2010)*, p637-642.

[5] Futrelle, J., & Downie, J. S. (2003). Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000–2002. *Journal of New Music Research* (32:2), 121-131.

[6] Ganseman, J., Scheunders, P., & D'haes, W. (2008). Using XQuery on MusicXML databases for musicological analysis. *Proc. International Symposium on Music Information Retrieval*, p433-438.

[7] Grove Music Online (2015). http://www.oxfordmusiconline.com/public/

[8] Hopkins, A. (1982). *The Nine Symphonies of Beethoven*. London: Pan Books.

[9] Huron, D. (1997). Humdrum and Kern: Selective Feature Encoding. In '*Beyond MIDI*', ed. E. Selfridge-Field (p375-401). Cambridge, MA: MIT Press.

[10] Kirkpatrick, R. (1953). *Domenico Scarlatti*. Princeton, NJ: Princeton University Press.

[11] MusicXML (2015). http://www.musicxml.com/.

[12] NTCIR (2015). http://research.nii.ac.jp/ntcir/index-en.html.

[13] 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.

[14] Sadie, S. (ed) (1995). *The New Grove Dictionary of Music and Musicians*. London, UK: Macmillan.

[15] Sutcliffe, R. F. E. (2014). *A Description of the C@merata Baseline System in Python 2.7 for Answering Natural Language Queries on MusicXML Scores*. University of Essex Technical Report, 21st May, 2014.

[16] Sutcliffe, R. F. E., Crawford, T., Fox, C., Root, D. L., & Hovy, E. (2014). The C@merata Task at MediaEval 2*014: Natural language queries on classical music scores. In Proc. MediaEval 2014 Workshop, Barcelona, Spain, October 16-17 2014*. http://ceur-ws.org/Vol-1263/.

[17] Sutcliffe, R., Peñas, A., Hovy, E., Forner, P., Rodrigo, A., Forascu, C., Benajiba, Y., Osenova, P. (2013). Overview of QA4MRE Main Task at CLEF 2013. *Proc. QA4MRE-2013*.

[18] TREC (2015). http://trec.nist.gov/.