DMUN at the MediaEval 2015 C@merata Task: the Stravinsqi Algorithm

Andreas Katsiavalos and Tom Collins
Faculty of Technology
De Montfort University
Leicester, UK
+44 116 207 6192
tom.collins@dmu.ac.uk

ABSTRACT

This paper describes the Stravinsqi-Jun2015 algorithm, and evaluates its performance on the MediaEval 2015 C@merata task. Stravinsqi stands for STaff Representation Analysed VIa Natural language String Query Input. The algorithm parses a query string that consists of a natural language expression concerning a symbolically represented piece of music (which the algorithm parses also), and then identifies where in the music event(s) specified by the query occur. For a given query, the output is a list of time windows specifying the locations of the relevant events. Time windows output by the algorithm can be compared with time windows specified by music experts for the same query-piece combinations. Across a collection of twenty pieces and 200 questions, Stravinsqi-Jun2015 had recall .794 and precision .316 at the measure level, and recall .739 and precision .294 at the beat level. The paper undertakes a preliminary analysis of where Stravinsqi might be improved, identifies applications of the C@merata task within the contexts of music education and music listening more generally, and provides a constructive critique of some of the question categories that are new this year.

1. INTRODUCTION

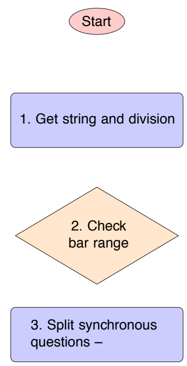
The premise of the C@merata task [1] is that it is interesting and worthwhile to develop algorithms that can (1) parse a natural language query about a notated piece of music, and (2) retrieve relevant time windows from the piece where events/concepts mentioned in the query occur. The premise is strong, if we consider that each year in the U.S. alone over 200,000 freshman students declare music their intended major [2, 3], and that there is a line connecting the types of queries being set in the C@merata task and the questions these students are taught (or, by college, have already been taught) to answer [4]. The C@merata task, apart from posing an interesting research problem at the intersection of music theory, music psychology, music computing, and natural language processing (NLP), could lead to new applications that assist students, and music lovers more generally, in gaining music appraisal skills. Other applications of research motivated by the C@merata task include supporting work in musicology [5], and informing solutions to various music informatics tasks, such as generation of music in an intended style [6] or expressive rendering of staff notation [7], where systems for either task may benefit from being able to automatically extract, say, cadence locations and/or changes in texture.

2. APPROACH

2.1 Overview

The Stravinsqi-Jun2015 algorithm (hereafter, Stravinsqi), which was entered into the C@merata task, is part of a Common

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



Lisp package called MCStylistic-Jun2015 that has been under development since 2008 [8]. The MCStylistic package, free and cross-platform, supports research into music theory, music cognition, and stylistic composition, with new versions released on an approximately annual basis.[1] In addition to Stravinsqi, MCStylistic includes implementations of other algorithms from the fields of music information retrieval and music psychology, for tonal and metric analysis [e.g., 9], and for the discovery of repeated patterns (e.g., motifs, themes, sequences) [10].

A flow diagram outlining the Stravinsqi algorithm is given in Figure 1. The following is a succinct overview of focusing on the differences between this year's (Stravinsqi-Jun2015) and last year's submission (Stravinsqi-Jun2014) [11].[2] Step 1 of Stravinsqi involves extracting the question string and divisions value from

[1] http://www.tomcollinsresearch.net
[2] For more details, please see the six-page version of this paper.

the question file. Step 2 parses the question string for mention of bar restrictions (“minim in measures 6-10”), stores this for subsequent processing (as (6 10), say), removes the restriction from the question string (“minim”), and passes it to step 3.

Prompted by one of the questions from the task description [12, p. 9], Stravinsqi splits queries by synchronous commands first (step 3) and then further by asynchronous commands (step 4). For example, “D followed by A against F followed by F” would emerge as (“D followed by A” “F followed by F”) from step 3, and as ((“D” “A”) (“F” “F”)) from step 4. In general, a question string emerges from step 4 as some nested list of strings $((s_{1,1}\ s_{1,2}\ ...\ s_{1,n(1)})\ (s_{2,1}\ s_{2,2}\ ...\ s_{1,n(2)})\ ...\ (s_{m,1}\ s_{m,2}\ ...\ s_{m,n(m)}))$, where each $s_{i,j}$ is a query element. Examples of query elements include “D”, “A♭4 eighth note”, “perfect fifth”, “melodic interval of a 2nd”, etc.

In step 5, point-set representations of the relevant piece are loaded and possibly restricted to those points that belong to a certain bar-number range. The `xml2hum` script is used to convert each piece from its MusicXML format to kern format [13].[3] Temporarily, in step 6 of Figure 1, each question element $s_{i,j}$ from step 4 is treated independently. A query element $s_{i,j}$ is passed to seventeen music-analytic sub-functions, each of which tests whether $s_{i,j}$ is a relevant query for that function, and, if so, searches for instances of the query in the piece of music. If the query is *irrelevant* to a sub-function, that function returns `nil`.

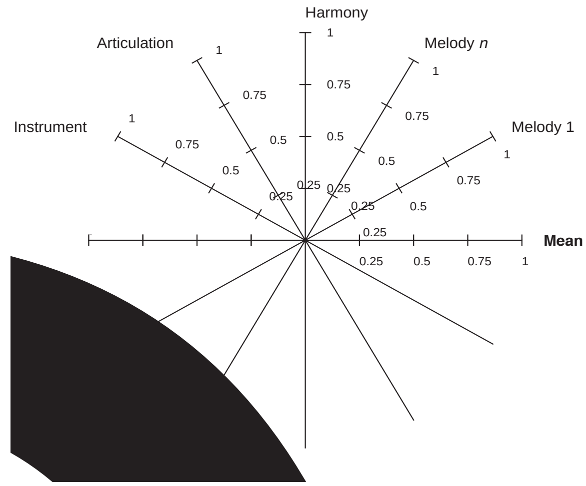
The output of step 6 is a nested list of time-interval sets, $((T_{1,1}\ T_{1,2}\ ...\ T_{1,n(1)})\ (T_{2,1}\ T_{2,2}\ ...\ T_{1,n(2)})\ ...\ (T_{m,1}\ T_{m,2}\ ...\ T_{m,n(m)}))$, one for each query element $s_{i,j}$, some of which may be empty. The purpose of steps 7 and 8 is to determine whether any combination v of these time intervals satisfies the constraints imposed by synchronous and asynchronous parts of the question string (there may be one such v, several, or none). The final step of Stravinsqi, labeled Step 9 in Figure 1, comprises the conversion of the time intervals $v_1, v_2, \ldots, v_r$ into the XML format required by the task.

3. RESULTS AND DISCUSSION

Figure 2 contains a summary of results for the Stravinsqi algorithm across various question categories.[4] The mean measure recall across all 200 questions, indicated by the black line next to the “Mean” label, is .794, and the mean measure precision, indicated by the blue line, is .316. The mean beat recall (green line) and beat precision (red line) are both slightly lower than their measure counterparts (.739 and .294 respectively), but in general it can be assumed that if Stravinsqi returned the correct beginning/ending measure number pairs for a question, then it was also able to identify the relevant beats. Stravinsqi had the highest measure and beat recall of any algorithm submitted to the 2015 C@merata task [1] and the third highest measure and beat F_1 score ($F_1 = 2PR/(P + R)$, where P is precision and R is recall).

Across eight of the eleven question categories shown in Figure 2 (Melody 1, Melody *n*, Harmony, Articulation, Instrument, Clef, Follow, and Synch), Stravinsqi achieves consistently high recall of approximately .75. For the remaining three categories (Time Sig., Key Sig., and Texture) it is less successful. Overall, the results suggest the need to investigate Stravinsqi’s precision being lower than its recall.

We have not yet incorporated in Stravinsqi restrictions to notes occurring after particular clef, time signature, or key signature changes. Currently, a query such as “G4 in the key of G major” would be parsed as though it were “G4”. Therefore, the recall of Stravinsqi remains high for such questions, but the precision will be negatively impacted. The design of Stravinsqi is



motivated more by music-perceptual than typographical concerns, based on the premise that music is primarily an auditory-cognitive phenomenon, and a visuo-cognitive phenomenon secondarily. When music perception and music theory collide, as they do occasionally in the C@merata task and beyond [14], Stravinsqi's precision can be adversely affected. For example, unlike the task description (consecutive elements “must both be on the same stave” [12, p. 7]), Stravinsqi does not require consecutive question elements to be on the same staff, because a staff swap has little (or sometimes no) effect on how the music sounds. Stravinsqi tends to find the correct answers according to the task description, but also some extra answers that involve elements on different staves, which has a detrimental effect on its precision.

4. CONCLUSION

We have provided an overview of the Stravinsqi-Jun2015 algorithm, and described its performance on the 2015 C@merata task. Stravinsqi achieved high recall (approximately .75) in eight of the eleven question categories, and had the highest measure and beat recall of any algorithm submitted to the task [1]. Further analysis of the results is required to determine whether Stravinqi’s precision can be improved while adhering to our general design principle of favouring music-perceptual over typographical concerns. In the introduction, it was remarked that there is a line connecting the types of queries being set in the C@merata task and the examination questions that students of the Western classical tradition are taught to answer. This year’s C@merata test set was lacking cadence and functional harmony queries, which was indicative of a general tendency to replace musically interesting questions (e.g., concerning cadence, triad, hemiola, ostinato, sequence, etc.) with questions that were linguistically challenging to parse but of less musical relevance (e.g., Question 130, “fourteen sixteenth notes against a whole note chord in the bass”). Next year, we would welcome the reintroduction of more musically interesting (if complex) question categories, to re-establish and strengthen the line that connects C@merata queries with concepts that are relevant for music students and enthusiasts.

[3] http://extras.humdrum.org/bin/osxintel64/

[4] Please see [1] for definitions of the various metrics.

5. REFERENCES

[1] Sutcliffe, R. F. E., Fox, C., Root, D. L., Hovy, E, and Lewis, R. 2015. The C@merata Task at MediaEval 2015: Natural language queries on classical music scores. In *MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015*. http://ceur-ws.org.

[2] Kena, G, et al. 2015. *The Condition of Education 2015*. U.S. Department of Education, Washington, DC. Retrieved June 4, 2015 from http://nces.ed.gov/pubsearch.

[3] Eagan, K, et al. 2014. *The American Freshman: National Norms Fall 2014*. Higher Education Research Institute, UCLA, Los Angeles, CA. Retrieved March 15, 2015 from http://www.heri.ucla.edu.

[4] AQA. 2012. *Music in Context: Past Paper for A-level Music Unit 4*. AQA, Manchester, UK. Retrieved January 13, 2014 from http://www.aqa.org.uk.

[5] 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.

[6] Collins, T., Laney, R., Willis, A, and Garthwaite, P. H. 2015. Developing and evaluating computational models of musical style. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*. DOI: 10.1017/S0890060414000687

[7] van Herwaarden, S., Grachten, M, and de Haas, W. B. 2014. Predicting expressive dynamics in piano performances using neural networks. In *Proceedings of the International Symposium on Music Information Retrieval* (Taipei, Taiwan, October 27 - 31, 2014). 47-52.

[8] Collins, T. 2011. *Improved Methods for Pattern Discovery in Music, with Applications in Automated Stylistic Composition*. Doctoral Thesis. Faculty of Mathematics, Computing and Technology, The Open University.

[9] Volk, A. 2008. The study of syncopation using inner metric analysis: linking theoretical and experimental analysis of metre in music. *J. New Music Res.* 37, 4, 259-273.

[10] Meredith, D., Lemström, K., and Wiggins, G. A. 2002. Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *J. New Music Res.* 31, 4, 321-345.

[11] Collins, T. 2014. Stravinsqi/De Monfort University at the C@merata 2014 task. *Proceedings of the C@merata Task at MediaEval 2014*.

[12] Sutcliffe, R., Fox, C., Root, D. L., Hovy, E, and Lewis, R. 2015. *Task Description v2: C@merata 15: Question Answering on Classical Music Scores*. Retrieved June 4, 2015 from http://csee.essex.ac. uk/camerata.

[13] Sapp, C. S. 2013. *Humdum Extras*. Retrieved March 3, 2014 from http://wiki.ccarh.org/wiki/Humdrum_Extras.

[14] Cook, N. 1994. Perception: a perspective from music theory. In *Musical perceptions*, R. Aiello and J. A. Sloboda, Eds. Oxford University Press, Oxford, UK, 64-95.