A NOTE ON CONTENT-BASED COLLABORATIVE FILTERING OF MUSIC

Stuart Cunningham, Vic Grout & Harry Bergen
Centre for Applied Internet Research (CAIR), University of Wales, NEWI
Plas Coch Campus, Mold Road, Wrexham, LL11 2AW, UK
s.cunningham@newi.ac.uk | *v.grout@newi.ac.uk* | *hx.bergen@web.de*

ABSTRACT

Collaborative filters are frequently used in e-commerce to provide a heightened user experience and to tempt users into making purchases by recommending items and drawing the user's attention to additional products. Purchasing of digital media over the Internet continues to be popular and e-commerce giants such as Amazon.com, CDNOW.com and Launch.com heavily employ Automated Collaborative Filtering (ACF).

This paper demonstrates a system for comparing musical compositions and provides an indication of how similar two or more musical pieces are to each other. It is shown that a significant amount of similarity exists between music compositions analysed from within the same genre.

It is proposed that a similarity metric could be incorporated into existing ACF systems to provide a powerful and effective recommendation system that will cater specifically for a user's preferences, and thus encourage purchase.

KEYWORDS

Music comparison, collaborative filtering, recommendation systems, e-commerce, computer music

1. INTRODUCTION

The use of collaborative filtering systems is heavily employed in many well-known e-commerce web sites to attempt to entice the user into making impulse purchases and to stimulate their interest in other products or services. The design of collaborative filters is informed by many parameters based upon data and statistics gained by monitoring users as well as by the sale of and interest in items on the e-commerce web site. The effectiveness of these filters can lead to large increases in revenue and customer good-will towards the particular enterprise. It is therefore vital that collaborative filters are as optimal as possible.

One of the main areas where collaborative filters are of use is in the purchase of digital entertainment media such as music, films and literature. In this paper we focus on analysing the content of music to improve recommendation systems for music. We demonstrate a system to compare musical compositions and show that music within a genre displays significant levels of similarity. Finally, we determine a similarity metric and discuss how this content-based analysis can be combined with Automated Collaborative Filtering (ACF) to produce a powerful hybrid recommendation system.

The classification and filtering of a large media catalogue such as music is no small task, and this is addressed by Billsus and Pazzani (1998) who pertinently describe such a task as being "…*one of the key problems of the information age: locating needles in a haystack that is growing exponentially*". Their work outlines two traditional methods of categorising information and data so that it can be queried to provide accurate searches and useful recommendation systems. Broadly speaking, the two main techniques employed are content-based analysis or feature extraction from the item or media and collaborative methods which rely upon metadata or statistics relating to the item or media.

Chai and Vercoe (2001) describe search techniques for music which rely upon using XML descriptions for efficient searching of music libraries and catalogues, and this is achieved by successfully modelling the user. Such a music information retrieval system could be easily adapted to help search for music on e-commerce sites, and would prove the principle that music can be easily searched and therefore that a similarity measurement can be determined. Their work of modelling users also extends so far as to propose paradigms that would allow the querying of a catalogue or database for this purpose.

The main thrust of our work is based upon searches previously carried out by Cunningham et al. (2005), who employed exhaustive search techniques to compare music across very different genres in an attempt to prove that there is an inherent similarity in all musical compositions. In their work the authors used a form of electronic musical notation, MusicXML (Good, 2001). Using a simple text-based format as the material for analysis, rather than actual audio samples, greatly improves the processing time and reduces the effort required to perform the analysis.

We believe that content-based and collaborative methods can be effectively combined to provide a hybrid model of collaborative filtering. This is recognized as being a key way of improving recommendation systems and reducing the current limitations of existing systems (Billsus & Pazzani, 1998). The advantages of hybrid recommendations are clearly set out in other related work (Chai & Vercoe, 2001, Herlocker et al., 2000).

2. CONTENT-BASED MUSIC COMPARISON

Previous work carried out by Cunningham et al. (2005) established numerical search techniques for music which allow the comparison of different musical pieces. In their work the authors attempt to prove an inherent similarity between all musical compositions and assign a similarity metric. Although the purpose of that work is to measure similarity of music from extremes of musical genres, it is also proposed that there must be a standard baseline similarity in all musical pieces which would allow normalisation of any future similarity measurements.

The following work presented in this section uses similar techniques to assess the similarity between various musical compositions, which are approximately within the same musical genre. Each musical score is then compared on a measure-to-measure basis, and a counter of successful matches maintained. An average similarity measurement is then presented which provides a mechanism to measure the similarity between these songs. The songs analysed in this study and the number of musical measures each song contains is presented in Table 1.

Table 1. Musical Pieces Compared

ID	Artist	Composition	Number of Measures
R_01	AC/DC	Highway to Hell	364
R_02	Beatles	Hey Jude	1710
R_03	Black Sabbath	Paranoid	654
R_04	Jimi Hendrix	Hey Joe	836
R_05	Metallica	Enter Sandman	1176
R_06	Queen	Bohemian Rhapsody	2919

A brief explanation of each of the search methods employed is provided in the following subsections; more detailed explanations of the search techniques can be found in the original paper (Cunningham et al., 2005). Comparisons are mainly focused on sequences of musical notes. Each note has a name and octave value (Károlyi, 1991). A note which has frequency f will have a frequency of $2f$ in the octave above and frequency $f/2$ in the octave below. If we consider the notes in a musical octave N_t where N is the note and t is the octave number, this presents the standard octave format: $\{C_t, D_t, E_t, F_t, G_t, A_t, B_t, C_{t+1}\}$.

2.1 Strict Musical Comparison

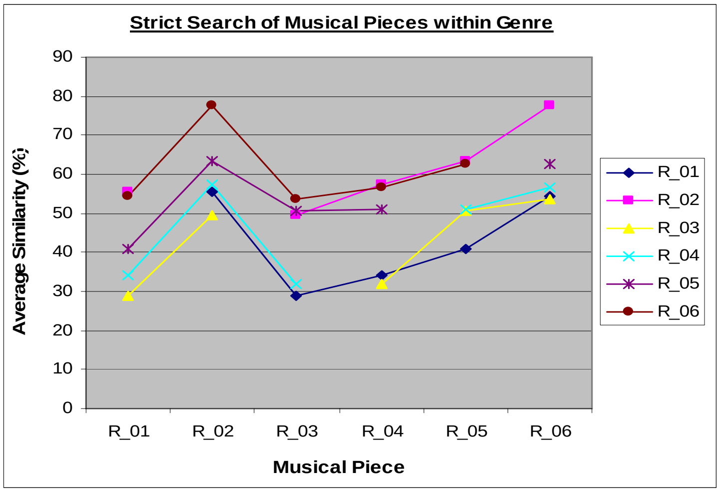
Within the strict comparison method, the measures compared must produce an absolute match to be declared equal and therefore feed into the similarity measurement. For example, the sequence of musical notes A#2-C4-D3 will only match against targets which are exactly the same, for example, A#2-C4-D3 (The capitalised letter represents note pitch, followed by optional flat or sharp alter then the octave number of the note).

The six musical pieces presented in Table 1 are compared with each other on this basis, and the results of the similarity measurement are presented in Table 2 and graphically represented in Figure 1.

Table 2. Strict Search. Average Similarity of Musical Pieces within Genre

ID	R_01	R_02	R_03	R_04	R_05	R_06
R_01						
R_02	55.5					
R_03	29	49.5				
R_04	34	57.5	32			
R_05	41	63.5	50.5	51		
R_06	54.5	77.5	53.5	56.5	62.5	
	Average Percentage of Similarity					

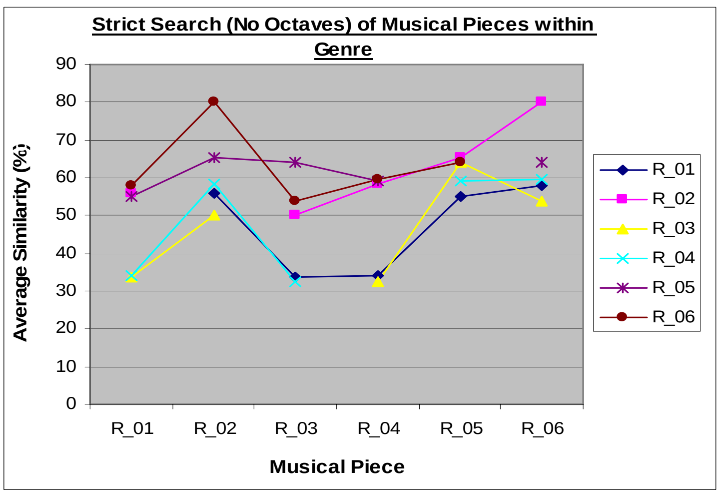
Figure 1. Graph of Strict Comparison within Genre



Subsequently, a modification is made to the strict search parameter which allows the musical octave elements of the music to be ignored. For example, the musical sequence A#2-C4-D3 will now find successful matches for a similar sequence of notes such as A#3-C5-D1 regardless of the octave values.

This has the effect of allowing the search to be more focused on matching the structure and musical notes contained in the melody of a piece, and to be less concerned with absolute matches. This has the general effect of increasing the similarity found between musical pieces, and broadening the range of music which may be deemed to be a suitable match. The results of this search are presented in Figure 2.

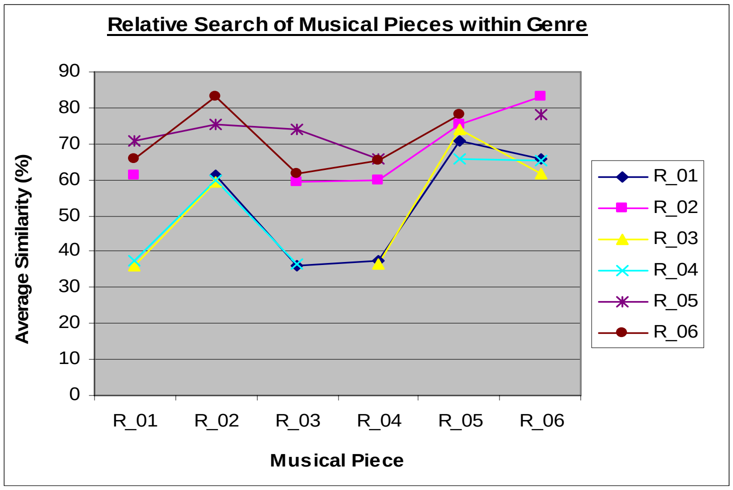
Figure 2. Graph of Strict Comparison (No Octaves) within Genre



2.2 Relative Musical Comparison

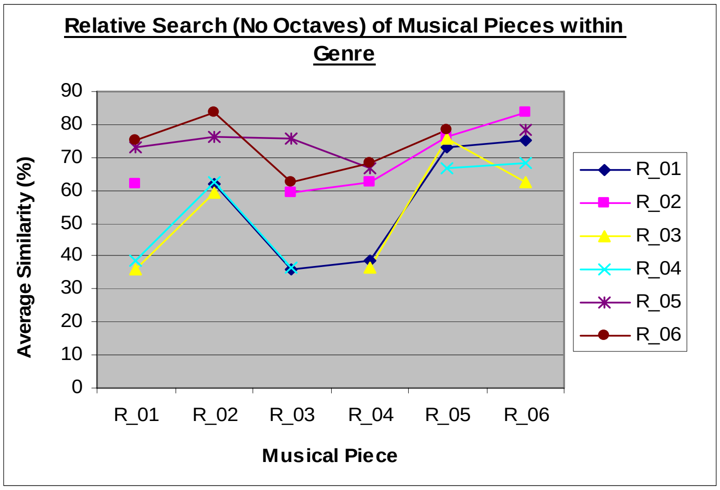
In the relative musical comparison, in order to find a valid match a sequence or measure of notes must each be separated by the same distance, measured by the note's position on the circle of fifths (Károlyi, 1991). That is, within each measure, the distance from one note to the next in the measure is stored and this distance sequence is the material used for comparison. The only limitation upon this relative search is that the notes matched must also have the same octave values. For example, the musical sequence A2-B2-C2 would validly be matched against another musical sequence with identical distance, such as E5-F5-G5. The results of this comparison using this technique are presented in Figure 3.

Figure 3. Graph of Relative Comparison within Genre



As with the strict search method, described in section 2.1 a similar modification is made to the relative search technique to allow the octave values of notes to be disregarded. This now means that a musical sequence such as A2-B2-C2 would be considered a valid match against a musical sequence with identical distance regardless of the octave value. Such a sequence could be B5-C1-D6. The result of the comparison with the new modification is presented in Figure 4.

Figure 4. Graph of Relative Comparison (No Octaves) within Genre



2.3 Results of Comparison Study

The results presented clearly show that there are minimum degrees of similarity of around 29% in all tests with the maximum similarities encountered being around 88%. Across all of the tests carried out the mean similarity is 57.95%. We can also see from the trends in the graphs that there is often a diverse range of values between the highest and lowest similarities discovered.

There is an overall increase across the range of comparison methods used, although these increases tend to be linear across all of the data tested. The methods used become less and less restrictive and therefore the increase in similarity measurements is expected. However, the increase is not so dramatic as to suggest that there are significant differences between the musical pieces within the genre tested. This reinforces the hypothesis that our system is successfully detecting consistent similarities in these songs.

It should be remembered that the musical pieces analysed here have varying content in terms of the measures of music each one consists of. Obviously, a file which has a large number of measures is more likely to have a higher similarity rating to other pieces, especially those which have only a small number of measures. A system of normalising this difference is another avenue for research. It is difficult, however, to see how this can be easily achieved without sacrificing the actual content of the musical pieces in question. After all, the diversity and range of measures within music is entirely responsible for giving a song or composition its feel, its individuality and ultimately, its appeal to the listener or user.

2.4 Comparison with Similarity across Musical Genre

To ensure that the similarity of the musical pieces within a defined category is accurate, we compare our results to those of Cunningham et al. (2005). It is reasonable to expect that there will be a higher amount of similarity present in music within a genre, and this means that a recommendation system would be making the correct decision in recommending suitable products to the user.

The data presented in Table 3 summarises the results of the search undertaken as part of this work where music within the same genre has been searched for similarity. This is contrasted with work carried out by Cunningham et al. (2005) which searched for similarity between musical pieces from two distant genres.

Table 3. Comparison of Similarity from Separate Investigations

	Searching within Genre		**Searching across Genre**	
Search Type	**Average Similarity**	**Standard Deviation**	**Average Similarity**	**Standard Deviation**
Strict	51.20 %	12.92	43.92 %	12.11
Strict (no octaves)	54.90 %	13.07	47.35 %	11.07
Relative	62.07 %	14.85	53.81 %	12.00
Relative (no octaves)	63.63 %	15.41	56.85 %	11.39

As Table 3 demonstrates, music within a similar genre classification has a higher level of average similarity for each search type than music comparisons which expand across two different genres. Interestingly, although larger averages are encountered, the standard deviation of similarity within a musical genre is slightly larger than the comparisons across two musical genres. This could be because the work carried out by Cunningham et al. (2005) was limited to the testing of music from only two distinct musical categories. More reliable figures and sample data sets could have been achieved if their work had spanned a much larger number of musical genres.

To further exemplify the difference between the search within and across musical genres, graphs showing the distribution of similarity comparisons from both studies are presented in Figures 5, 6, 7 and 8. These correspond to the search types presented in Table 3.

Figure 5. Strict Search Distribution

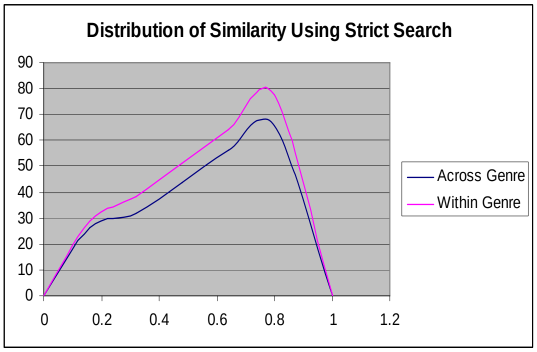


Figure 6. Strict Search (No Octaves) Distribution

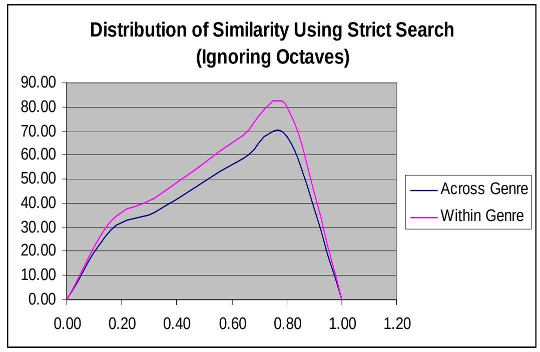


Figure 7. Relative Search Distribution

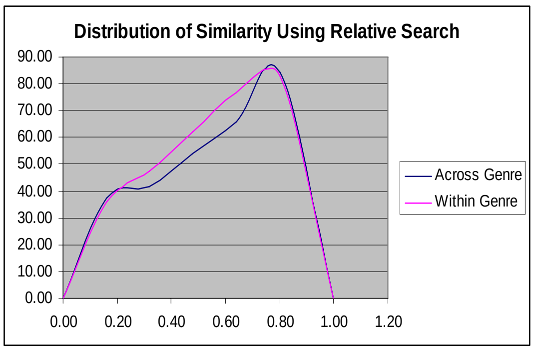
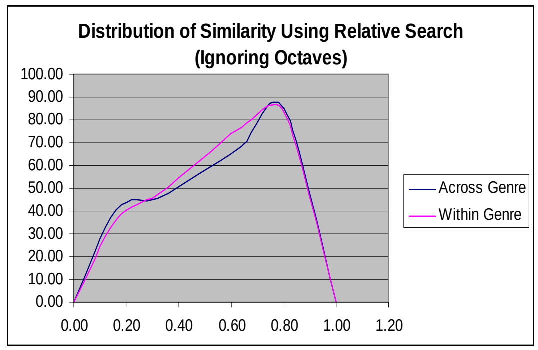


Figure 8. Relative Search (No Octaves) Distribution



This provides a useful visual representation of the data. We can see that the distribution of similarity approximately fits the Poisson distribution model as would be expected from real-world data which is generated by human creativity. We also see that the similarity of music within a genre is clearly higher, although there may be more diversity present within a genre. Yet another explanation for this could be the result of the human creativity factor and the influences around the composer at the time of writing. It is also clear by looking at Table 1 that the musical pieces used in these tests, although roughly classifiable as rock music, also span many years and have a diverse range of sounds and musical layers and harmonies. Therefore, the diversity encountered by the standard deviation indicators is justifiable.

Building upon the suggestion that there must be a baseline or common, quantifiable, similarity between any two musical pieces, it is evident that there is need to establish such a value (Cunningham et al., 2005). Indeed, in order to give more weight to the similarity measurements gained from this work it is useful to establish such a value, and process the previous search results in a suitable manner. Such processing of these results will also provide a more useful metric which can be used to inform existing collaborative filtering algorithms, or even be used as the foundation upon which to develop new, more suitable algorithms specific to music similarity testing.

The similarity index *I* is calculated using simple analysis, described by the following formula:

$$I = \frac{a - \min}{\min} \tag{1}$$

Where *a* is the average of all searches of that type and *min* is the lowest similarity measurement found for that particular search technique. The values used for the lowest similarity measurement have not been formally proven, but have been empirically shown to be a suitable baseline value from the tests conducted. This allows a certain degree of obvious distinction between the ranges of values encountered in the results presented in Table 3. Table 4 shows the modification of the previous results (from Table 3) with the addition of the similarity index value.

Table 4. Comparison of Similarity with Search Index

		Searching within Genre		Searching across Genre	
Search Type	**Min. Value (*min*)**	**Average (*a*)**	**Similarity Index (*I*)**	**Average (*a*)**	**Similarity Index (*I*)**
Strict	20	51.20 %	1.56	43.92 %	1.20
Strict (no octaves)	20	54.90 %	1.75	47.35 %	1.37
Relative	23	62.07 %	1.70	53.81 %	1.34
Relative (no octaves)	24	63.63 %	1.65	56.85 %	1.37

3. MUSIC COMPARISON FOR COLLABORATIVE FILTERING

We propose that such a similarity measuring system could be adapted to inform collaborative filters and recommendations in e-commerce systems. By including this similarity measurement it is clear that any product which has a high level of musical similarity to a product purchased by that user will often be of great interest. This would allow ACF systems to provide much more focused recommendations and not have to rely heavily upon the purchase history vectors of other users and by attempting to match products against other users who have made similar product selections.

Clearly the way in which our similarity measurement should be included into complex filtering algorithms is well beyond the scope of this paper. However, we make a simple suggestion about how to identify and address the salient issues so as to make progress with such a hybrid system.

Linden et al. (2003) present several challenges for collaborative filtering systems. Two of the most significance for the work presented here are:

1. The retailer will often have large amounts of data: millions of customers and millions of items.
2. High quality results are expected from a recommendation system and often have to be provided in real-time.

Clearly, a significant amount of processing will be required to search or catalogue a large volume of music. In the worst case we estimate this to be $O(N^2)$ for an existing music library, where N is the number of music items in the catalog. However, once carried out, there will still be the process time involved in integrating a number (i) of new items into the current music comparison index, although this should be at a significantly lower cost, assuming $i < N$. This is estimated to be $O(Ni^2)$ in the worst scenario. A major issue resulting is the large amount of computation to be carried out; however, there is no reason why such work could not be carried out offline and prior to the integration of these techniques into any existing recommendation system. In this case the only real-time information retrieval would be a simple query to the similarity index, which would be comparable in time costs with any other HTTP, or similar, server request. In fact, the Item-to-Item system employed by Amazon ensures that all heavy computation is carried out offline and only requests for data are carried out in real-time (Linden et al., 2003).

As a starting point for integration into existing recommendation algorithms it is suggested that the similarity index values are used as scalars in existing techniques. By employing such a method carefully, the similarity index can be used to push a potential recommendation above the required threshold and to add increased distance between a collection of items which may have similar ratings.

4. CONCLUSION

We believe that successful integration of the similarity measurements demonstrated in this paper will greatly improve the accuracy and success of collaborative filtering. The combination of the methods described in this paper with existing systems present great opportunities, especially because of the heuristic nature of current ACF systems. A potential set of informative parameters could be adopted in a variety of ways. The optimisation of recommendation systems may never be fully realised, but rather a best fit system may be required to suit various different e-commerce scenarios, depending on the type of products or services which are on offer, and on the nature of the data content which can be analysed to inform ACF systems. Indeed, we have focused only on music in this work, but content analysis or feature extraction could be equally applied to other digital media forms that are significant in e-commerce, such as images and video.

Much work is required in a number of areas to see the concepts presented through to fruition. Further investigation is required to determine if the music similarity system can be modified so that musical pieces of varying lengths can be normalised in some way to allow a fairer and more realistic metric of similarity to be produced.

A significant amount of work is required to see how the similarity metrics can be integrated into existing recommendation algorithms and some new methods might even to be developed in order to make full use of this data.

Further into the future, optimisation of the similarity search technique will help improve adoption and usage of the system proposed in this paper. Ultimately, the computation time required to analyse a large volume of music will be significant and any methods of improving this will doubtless be welcomed.

ACKNOWLEDGEMENT

Thanks to Rich Picking, Mike Headon, Miguel-Angel Sicilia, and Salvador Sanchez-Alonso.

REFERENCES

Billsus, D. and Pazzani, M. J., 1998. Learning Collaborative Information Filters. *Proceedings of 15th International Conference on Machine Learning*. Madison, Wisconsin USA, pp. 46-54.

Chai, W. and Vercoe, B., 2001. Using User Models in Music Information Retrieval Systems. *Proceedings of 2nd Annual International Symposium on Music Information Retrieval*. Bloomington, Indiana, USA.

Cunningham, S., Grout, V. and Bergen, H., 2005. Mozart to Metallica: A Comparison of Musical Sequences and Similarities. *Proceedings of ISCA 18th International Conference on Computer Applications in Industry and Engineering (CAINE-2005).* Honolulu, Hawaii, USA, pp. 332-339.

Good, M., 2001. MusicXML: An Internet-Friendly Format for Sheet Music. *IdeAlliance: XML 2001*. Florida, USA.

Herlocker, J. L., Konstan, J. A. and Riedl, J., 2000. Explaining Collaborative Filtering Recomendations. *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work.* Philadelphia, Pennsylvania, USA, pp. 241-250.

Károlyi, A. 1991. Introducing Music. Penguin, Harmondsworth, UK.

Linden, G., Smith, B. and York, J., 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *In IEEE Internet Computing,* January/February, pp. 76-79.