

A LOOK INTO THE PAST: ANALYSIS OF TRENDS AND TOPICS IN PROCEEDINGS OF SOUND AND MUSIC COMPUTING CONFERENCE

Pratyush¹

Martí Umbert²

Xavier Serra³

Music Technology Group,
Universitat Pompeu Fabra, Barcelona.

¹akpratyush@gmail.com

²marti.umbert@gmail.com

³xavier.serra@upf.edu

ABSTRACT

In this paper we analyze the proceedings of all the past six editions of the *Sound & Music Computing Conference*. The proceedings are analyzed using knowledge based “keywords to text” mapping to discover the overall conference trends. The analysis is done on the basis of number of papers, distinct authors, participation ratio for each relevant topic, interdependence of topics in terms of shared keywords and the overall popularity of keywords. The analysis was done for each conference year as well as for the overall collection of proceedings till date. The objective of the discussed work is to provide an insight over the past six years in the SMC community that was envisioned in the roadmap.

1. INTRODUCTION

Since its first conference in 2004 to Porto’s 2009 conference, the Sound & Music Computing field has traveled a path, which has given it a status that it could be treated as a standalone scientific discipline. The aim of this paper is to analyze the evolution of the publications of the SMC conference that was envisioned in the roadmap of the SMC field [1]. Since these publications were scattered over the Internet in individual websites of each conference year, it was not easy to extract information about the conference.

For the ease of data retrieval & analysis of the publications on the basis of author participation, topic of interests, relationship between different topics & the trend over the years we decided to build a web repository of all the publications of the SMC Conference till date and used it for the analysis reported in this particular paper. Contrary to the other works presented in [2, 3], the current work relies on collaborative effort, knowledge simplification and rule based classification.

In section 2 we describe how this repository was built as well as its preliminary analysis. The methodology used for the analysis of the papers is explained in section 3 while section 4 presents the results of the analyses that were performed and finally in section 5 we discuss the conclusion of the analyses and the work done.

2. THE REPOSITORY

This section discusses the building of the repository straight from the scratch as well as the preliminary analysis results that we obtained.

2.1 Building the repository

Every published paper of the SMC Conference so far was downloaded from each year’s conference’s individual website and then manually entered in a relational database management system. Drupal Content Management Service [4] was chosen as the framework.

2.2 Preliminary Analysis

Preliminary analysis of the built repository was done in order to expose the trend about number of papers & author participation in each edition as well in the overall history of the conference. The results of the preliminary analysis are summarized in Table 1.

Year	No. of Papers	No. of Distinct Authors
2004	46	83
2005	31	68
2006	25	50
2007	60	116
2008	34	70
2009	62	163
Overall	258	482

Table 1. Preliminary analysis of the SMC Publications.

It has to be taken into account that the overall number of distinct authors (482) is not the sum of the individual distribution of distinct authors for each conference year as an author is most likely to participate in two or more editions of the conference. Specifically, over the years, 68 authors have participated in more than one edition of the conference.

High level topics	Middle level topics	
	Name	ID
Processing of sound and music signals	3D sound/music, Sound/music signal processing algorithms, Digital Audio Effects, Musical sound source separation	Topic 1
	Sound synthesis, Spectral modeling synthesis, Physical modeling for sound generation	Topic 2
Understanding and modeling sound and music	Music information retrieval, Musical pattern recognition/modeling, Computational musicology, Technologies for the preservation, access and modeling of musical heritage, Automatic music transcription, Musical sound source separation and recognition	Topic 3
	Music and emotions, Sound/music and Neuroscience, psychology, psychoacoustics, Sound/music perception and cognition	Topic 4
Interfaces for sound and music	Interfaces for music creation and fruition, Gesture controlled audio systems, Mobile music, Interactive performance systems, Musical performance modeling Visualization of sound/music data, Sonic interaction design	Topic 5
Assisted sound and music creation	Web 2.0 and music, Networked music generation	Topic 6
	Computer environments for sound/music processing, Automatic music	Topic 7

Table 2. Hierarchy between High & Middle Level Topics

3. THE METHODOLOGY

This section describes the overall process, which was followed in order to get the analysis result that was envisioned. The process starts with simplification of knowledge to generate “List of Topics” associated with the Sound & Music Computing community, secondly a handful of “Keywords” were created to map each paper to one or the other topics in the list. Finally, statistical analyses of the papers were done.

3.1 Knowledge simplification

To classify all the papers of the previous SMC Conferences on the basis of “Research Topics”, we started with the topics for call for papers for the upcoming Sound & Music Conference. Since these topics were very specific in nature, we classified them into broader topics. The classification, of the original topics into broader ones was done on the basis of “Similarity in Concepts”.

This led to the construction of a hierarchical classification, constituting two levels of simplification:

1. “Middle Level Topics”
2. “High Level Topics”

The mappings of these two levels of topics are presented in Table 2.

3.2 Keyword building

Since the papers presented in the SMC Conference did not have “Keywords” in the full text, there were two options for building a keyword repository for the papers:

1. Automatic extraction of keywords from the abstract of the papers using a probabilistic mixture model as introduced in [5].
2. Use of CSCW methods using Google docs as discussed in [6] to collaborate with other researchers of the field to have a consensus on a set of keyword for each topic.

For the current work, we used the second method for building the keyword list for each middle level topic. The consensus was reached using cross validation of the keywords between each researcher in the second pass of the questionnaires. Furthermore, relevance of each keyword was checked by searching for the keyword in the SMC papers as well as searching for papers using the same keyword in Google scholar.

3.3 Search Mechanism

The search mechanism attempts to assign a topic to each paper, based on the keyword. The entire process can be described as below:

1. Search every keywords of each topic in the abstract & title of each paper.
2. If a particular keyword is present in the abstract or title, we add that keyword & the associated topic as a contender for classifying that paper.
3. Once the keywords are mapped in a particular paper, we count the total number of occurrences of each keyword & the total presence of keywords from each topic.

4. The topic with the maximum number of keywords present in a paper is decided to be the relevant topic for that paper.
5. If more than one topic has equal presence in a paper, we classify the paper as 'Multiple Topic'.
6. If none of the keywords could be mapped to a paper, we label the paper as 'Unknown Topic'.

This process is carried out to classify a paper to a Middle Level topic using the set of corresponding keywords for each topic. For the re-classification of each paper based on High Level Topics, we use the relationship between the Middle Level & High Level topics as depicted in Table 2. Furthermore, if there were discrepancies in the high level classification, we assigned those papers as 'Unclassified'.

Topic	'04	'05	'06	'07	'08	'09	ALL
Multiple	13	4	5	15	8	11	56
No topic	4	0	1	2	3	3	13
Topic 1	3	2	1	3	7	3	19
Topic 2	9	13	7	13	9	10	61
Topic 3	7	0	5	4	1	13	30
Topic 4	2	0	1	0	0	2	5
Topic 5	5	11	4	18	3	15	56
Topic 6	3	1	1	4	3	5	17
Topic 7	0	0	0	1	0	0	1

Table 3. Year wise distribution showing the absolute number of papers for each Middle Level Topic.

Topic	'04	'05	'06	'07	'08	'09	ALL
Unclassified	13	4	5	13	6	10	51
No topic	4	0	1	2	3	3	13
Assisted...	3	1	1	5	3	5	18
Inter- faces...	5	11	4	18	3	15	56
Process- ing...	12	15	8	18	17	14	84
Under- stand- ing...	9	0	6	4	2	15	36

Table 4. Year wise distribution showing the absolute number of papers for each High Level Topic.

4. RESULTS

The results that we obtained after the analysis are presented in this section. For better aesthetics of the plots & charts, we have used aliases for each Middle Level Topic.

4.1 Participation ratio for each relevant topic

Since each paper was classified as either a topic or multiple topic or unknown topic, we can deduce the distribution of each topic in each year's conference as well as in the overall conference till date.

Figure 1 shows the distribution of each Middle Level Topic in the overall conference history, while Table 3 is used to visualize the distribution of each Middle Level Topic in each edition of the conference.

Likewise, Figure 2 displays the distribution of the High Level Topics in all years of the conference taken together, whereas Table 4 is used to show the distribution of these topics in each conference year.

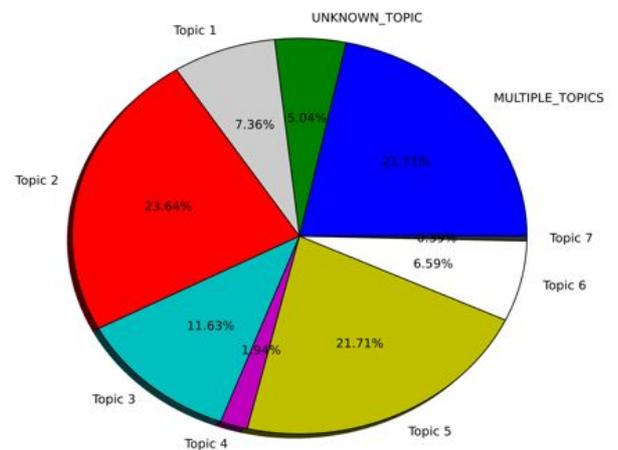


Figure 1. Publication distribution for Middle Level Topics for the overall conference till date.

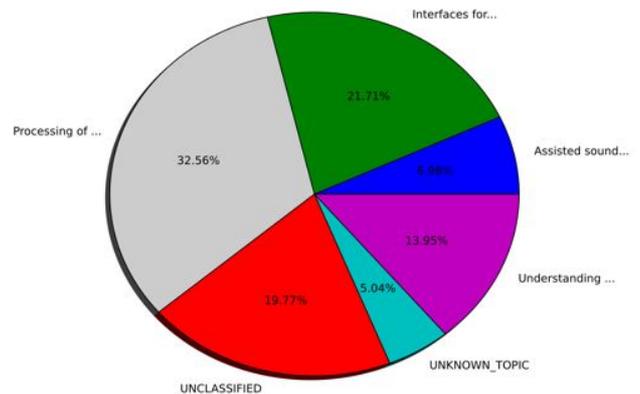


Figure 2. Publication distribution for High Level topics for all years taken together.

4.2 Trends for each level of topics over the entire conference history

The change in the number of papers for each topic over the years is presented both for Middle Level & High Level topics in Figure 3 and Figure 4 respectively.

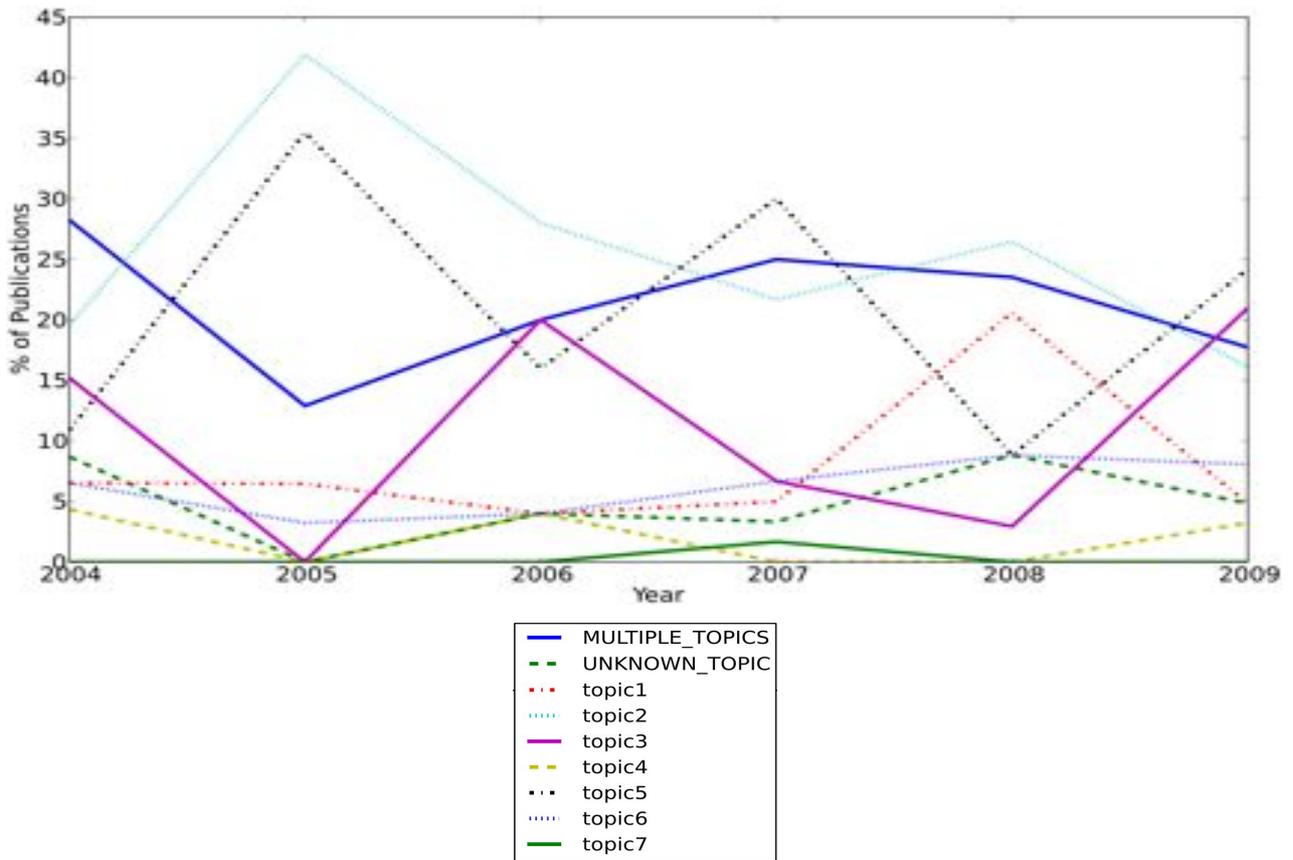


Figure 3. Middle Level Topic trend.

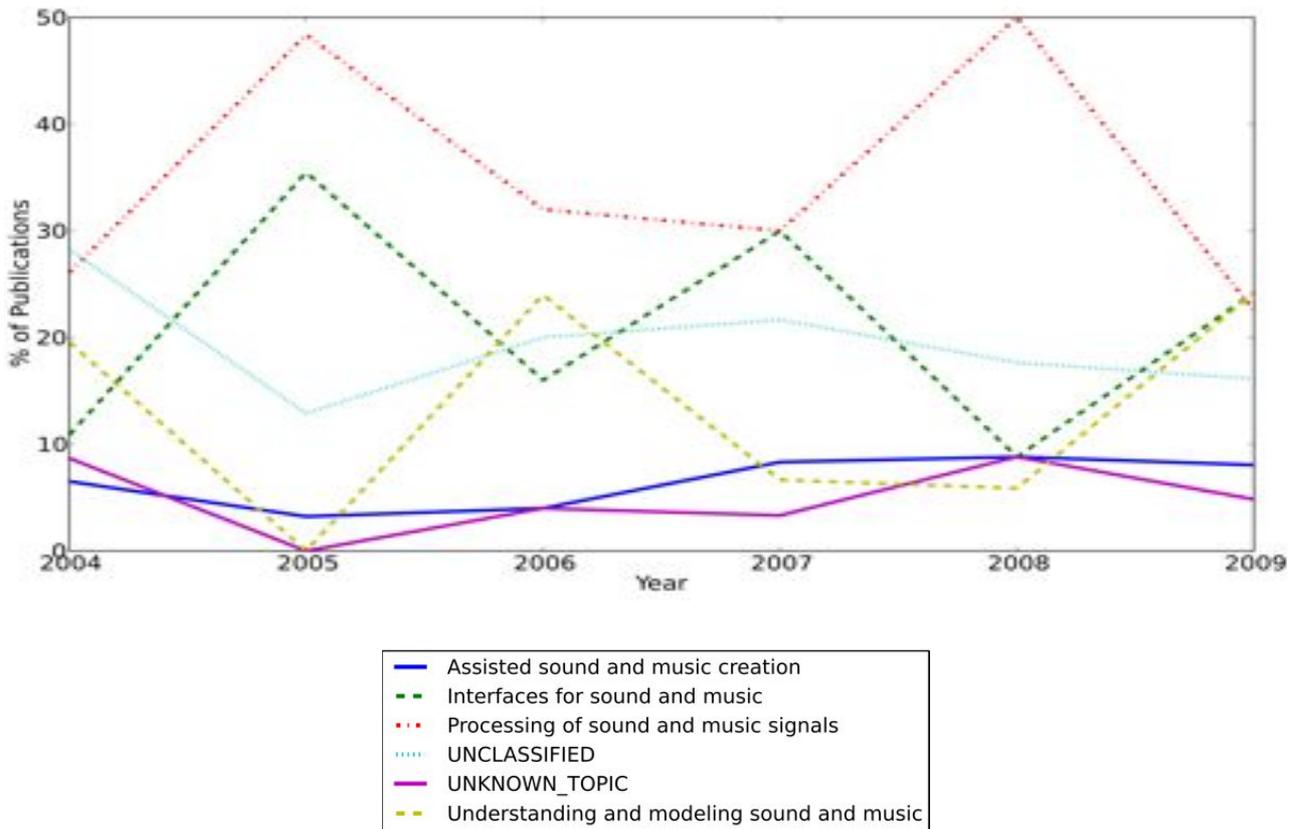


Figure 4. High Level Topic trend.

5. CONCLUSION

In this paper, we analyzed the proceedings of the past SMC Conferences, tried to categorize each published paper into one of the proposed 7 Middle Level & 4 High Level Topics so that the trend of the SMC Conferences could be identified and justified.

To start with we noticed that 482 authors have participated in the SMC Conferences till 2009 and out of those, 68 authors have publications in more than one edition of the conference.

For e.g. we found out that *Topic 2: Sound synthesis, Spectral modeling synthesis, Physical modeling for sound generation* and *Topic 5: Interfaces for music creation and fruition, Gesture controlled audio systems, Mobile music, Interactive performance systems, Musical performance modeling Visualization of sound/music data, Sonic interaction design* remains the most popular topic throughout the conference with a combined share of ~77% in SMC Conference 2005 and ~45% overall. Of all the conferences till date, the share of *Topic 1: 3D sound/music, Sound/music signal processing algorithms, Digital Audio Effects, Musical sound source separation* was highest in 2008 about 20% and *Topic 3: MIR & others* had a considerable share in the 2009's conference with about 21% publications.

From the participation ratio of each Middle Level topic in each year, we find the following trends in the evolution of some topics over the years:

1. Web 2.0 grows since 2005, this can be justified by the fact that web 2.0 evolved a lot since that time, so it attracted much research in the recent years.
2. Sound synthesis/ signal processing has a slight decline in percentage in the recent years this might be because the growing popularity of other fields.
3. Since the theme of the 2008 conference was "Sound in Space", the abrupt increase in the number of publications of the topic "3D Audio" for that year is justified.

From the closeness analysis of each topic Vs the others, we could clearly see that Topic 1 & Topic 2 are closely related to each other, so our classification of grouping them together in the higher level of classification is fairly justified. Although we have grouped Topic 6 & Topic 7 together, this is not fairly justified by the data presented in Table 5. This is due to the fact that there is a hairline difference between the last two high level topics and thus Topic 5 & Topic 6 are also closely related as depicted in the same table. Alternatively, Topic 5, 6 & 7 could be regrouped to a new High Level topic as well.

And finally looking at the Keyword cloud, we could see that the popular keywords from the set we had, are synthesis, analysis, instrument, realtime, voice, net, etc.

4.3 Closeness between the topics

Since we used decision based approach for assigning a paper a relevant topic based on the "presence of keyword", we observed many papers which were classified as particular topic but had a fair amount of keywords of other topics were present as well. This can be used to deduce the closeness of a topic with others. The cross infiltration (presence) of each Middle Level topic in every other for the overall conference publications is showed in Table 5.

Topic	1	2	3	4	5	6	7
Topic 1	58.55	18.42	3.95	0.66	13.82	4.61	0.00
Topic 2	9.37	59.10	7.21	1.26	17.66	5.23	0.18
Topic 3	6.48	10.65	60.19	4.63	16.20	1.85	0.00
Topic 4	12.20	12.20	12.20	43.90	17.07	2.44	0.00
Topic 5	15.43	9.88	3.09	2.78	51.23	17.44	0.15
Topic 6	2.26	15.04	2.26	3.76	24.06	52.63	0.00
Topic 7	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Table 5. Presence of each topic in each other (Middle Level).

4.4 Keywords & their relevance

As the keywords play a pivotal role in the overall procedure that we presented here, we found out the popularity of each individual keyword irrespective of the topic they represent in all the papers published in the SMC Conference till date.

A keyword cloud representing the popularity or presence of these keywords is plotted below as Figure 5. The most frequent 50 keywords are shown with a font size that reflects this popularity. Frequency values range from 6 to 119 occurrences.



Figure 5. Keyword Cloud

Since the overall methodology relied on text mining and knowledge simplification using which we classified a dataset of nearly 31000 words with a keyword set of 117 keywords, the evaluation of the system is tough. Moreover, the evaluation is also hampered by the fact that there were no keywords provided by the authors in the SMC papers to cross check with.

To conclude with, we would like to highlight that only 5% of the papers were of *unknown topic* and about 21% of the papers were of Multiple Topics (*unclassified*), this correlates to the fact that Sound & Music Computing is highly inter-disciplinary in nature. Another point to take into account is that these conclusions have been deduced from the last 6 SMC proceedings, which might not represent enough data to support them.

Also, we would like to continue to explore the presence of research groups of different universities in the SMC Conferences based on publications and how papers, authors & research topics could be classified together on the basis of co-authorship, citations and bibliographic links.

6. REFERENCES

- [1] A Roadmap for Sound and Music Computing. Version 1.0. , 2007. Available at: <http://www.smcnetwork.org/roadmap>
- [2] J.H. Lee, J.S. Downie and M.C. Jones: “An analysis of ISMIR proceedings: Patterns of authorship, topic, and citation”, *10th International Society for Music Information Retrieval Conference*, pp. 57-62, 2009
- [3] G. Widmer et al, : “The ISMIR Cloud: A decade of ISMIR Conferences at your fingertips”, *10th International Society for Music Information Retrieval Conference*, pp. 63-68, 2009
- [4] Official drupal website available at <http://drupal.org>
- [5] A. Velivelli, B.Yu and C. Zhai: “A cross-collection mixture for comparative text mining,” *Proceedings of ACM-KDD*, pp. 743–748, 2004.
- [6] S. Dekeyser and R. Watson: “Extending Google Docs to Collaborate on Research Papers”. *Technical report, The University of Southern Queensland, Australia*, 2006.