THE DISCIPLINE FORMERLY KNOWN AS MIR

Perfecto Herrera, Joan Serrà, Cyril Laurier, Enric Guaus, Emilia Gómez, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{perfecto.herrera, joan.serraj, cyril.laurier, enric.guaus, emilia.gomez, xavier.serra}@upf.edu

1. INTRODUCTION

Music Information Retrieval is a young multidisciplinary endeavour [4]. Even though its origins can be traced back to the 1960's [14], we all probably agree on the capital influence that the International Conference on Music Information Retrieval, started in 2000 as symposium, has exerted on the sense of belongingness to a research community.

Our exploration is not a science-fiction essay. We do not try to imagine how music will be conceptualized, experienced and mediated by our yet-to-come research, technological achievements and music gizmos. Alternatively, we reflect on how the discipline should evolve to become consolidated as such, in order it may get an effective future instead of becoming, after a promising start, just a "would-be" discipline.

Our vision addresses different aspects: the discipline's object of study, the employed methodologies, social and cultural impacts (which are out of this long abstract because of space restrictions), and we finish with some (maybe) disturbing issues that could be taken as partial and biased guidelines for future research.

2. OBJECT OF STUDY

First of all, let's face the misrepresentation that the name of our discipline currently has. From the 3 words used, only the word "music" is totally satisfactory. "Information" is debatable in the sense discussed by Wiering [14]: most of the interesting information that has to be exploited is a-referential (i.e., not about the music itself but about its context, functions, connections, emotions, audiovisual links, etc.). Instead of "music information", we should use "information about music". In fact, what most of the algorithms and systems developed in our community target is "music content", something that can be predicated, in a logical sense, about the music (in other words: knowledge). To conclude, the term "retrieval" just identifies a very narrow topic targeted by some researchers attending to ISMIR and other related conferences. But there is life beyond retrieval, and broad and better suited alternatives for a permanent name would be "processing" or "interaction" or even broader, "research". Therefore, the discipline formerly known as MIR is probably going to be better characterized as "music content processing", "interaction with music information" or (more generally, and if changing the acronym MIR is seen as detrimental) "music information research"[1].

Under our conception of MIR (or whatever other better name we finally find for it) the discipline should be asking profound questions and providing answers about music, about the content treasured in short excerpts, tracks, collections, or datasets, and about the context where it can be experienced. If the developments, devices, systems, experiments and evaluations do not increase our knowledge about music then the field is going to get more and more sterile and it will become just a technology incubator, working just for the sake of technology. We have mixed feelings on the level of the current accomplishment of this quest for understanding music. In the ideal future we foresee a wealth of new knowledge about music and all kinds of interactions with music information, not just a bunch of clever algorithms and nice user interfaces. We will only develop music understanding systems by means of understanding music understanding.

If we consider the Stokes' quadrants [13], our discipline is currently in the so-called Edison's quadrant (pure applied research) but its future lies in Pasteur's quadrant, which corresponds to use-inspired basic research. Complementary pure basic research (Bohr's quadrant) should be generated by computer science, cognitive science, musicology, or any other of the consolidated disciplines that converge in our field (*source disciplines*)[2]. This conceptualization is also paralleled by the consideration of three types of physicists (or scientists we would add) made long time ago by Victor Weiskopf: machine builders, experimentalists and theoreticians. For the sound and safe future of the discipline we need more theoreticians and experimentalists to team up with the excellent machine builders that have

[1] This name is precisely the one adopted by the recently created International Society for Music Information Research.

[2] It could be discussed if *basic research* can be generated in the MIR field.

cropped up in the early years of MIR. In the future, our discipline will be more permeable to researchers grown in the source disciplines, as they can bring theories, experimental methods, and normative data (i.e., data to be used as ground truth) to be exploited in the problems related to music understanding by humans and machines. The transition from Edison's to Pasteur's will be possible only when the above mentioned shift of focus and percolation of human resources is more common than now.

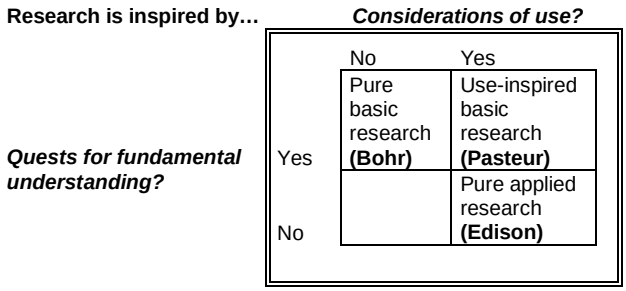

Figure 1. Research quadrants from Stokes [13].

Another feature of our discipline in the future is that work will be predominantly done on the upper rows of Table 1. Understanding music understanding has been mostly pursued bottom up, and this has led to evidencing the existence of a semantic gap [2]. Top-down approaches, whereby a general model based on rational analysis of behaviour or on task knowledge is developed, may overcome this drawback. Examples of fruitful models that are yet to be adapted to the specificities of interacting with music content are, for example, the Information Foraging Theory [10], or Rational Models that connect choice and preference [8].

3. METHODOLOGY

Scientific rigour is not only to be demonstrated in the proper use of mathematical models. It starts with a proper use and acknowledgment of the past research legacy. In our future papers, we will rarely see omissions of basic works and no bias to prefer referencing authors of certain countries or regions will be detected[3]. With the current and future bibliographical research tools, omissions will only be attributable to bad practice, either voluntary or because of lack of care. Student supervisors will be very concerned on playing the role of heritage convey-belts.

[3] This observation arises from our own activity as journal and conference reviewers. It should be substantiated with data that, unfortunately, cannot be made public in most of the cases, or that would require extensive bibliometric analyses.

Generalization poses an important overhead on our research and it has not been seriously taken in many papers, even those published in our most influential journals. The claims written, often being very ambitious, should be accurately substantiated using the appropriate type and amount of instances, the correct type of tests, and a baseline system or theory to compare and judge the claimed improvements or advantages.

Level	Question	Analysis elements
Rational	What problem is solved?	Resources, state and phase dynamics, affordances, constraints
Knowledge	What does the system know?	Goals, preferences, semantic descriptors
Algorithm	How does the system do it?	Front-end, low-level descriptors, classifiers, similarity metrics
Implementation	How does the system physically do it?	System architecture, hardware devices, software functions

Table 1. Levels of explanation adapted from [10]

The future MIR researcher will be aware of experimental design and statistics, way better than the current one. It can be scary to count how many t-tests can be found in the available literature, and how many of these were not justified or not properly done (the same has to be said on the election of evaluation measures). It is also scary to count how many published papers do not include any formal hypothesis test at all (see, for example [3] for some guidelines on good practices). Multivariate statistics, bootstrapping, or Bayesian models will become off-the-shelf tools for the future MIR researcher. In addition, the Cranfield evaluation model (i.e., TREC-like) will be enhanced with alternatives considering cognitive, interactive and relevance issues that are not captured by precision and recall measures [1].

One maturity feature of a discipline is the existence of meta-analysis studies. Meta-analysis is aimed to the accumulation of evidence across different studies targeted to the same problem [11]. There are several musical problems (e.g., genre classification) where the amount of different studies, with different music collections, algorithms, descriptors, classifiers and parameters qualifies them for some meta-analysis integrative effort. In the future, researchers will use meta-analysis as another tool of the trade.

Subjective evaluations will be accomplished following the formal requirements of subject sampling, briefing, debriefing and ethical respect that are followed in other experimental disciplines. A subjective evaluation that is

done as a "quick and dirty" way to demonstrate that "users prefer the developed system" will be dismissed. As an alternative, well-crafted experiments will be the norm, were the effect of a variable (algorithm, system, interface, etc.) is tested on controlled and well-defined indicators of satisfaction, efficiency or task accomplishment.

4. THE FUTURE IS HERE; IT'S JUST NOT WIDELY DISTRIBUTED YET

In the future, most of metadata will be generated in the creation/production phase: This means that we mostly need automatic content processing systems capable of working with old content, and that content description at the very moment of its generation should be more actively facilitated by industrial-strength authoring tools.

We need less than 3000 semantic concepts to make possible flexible description, visualization, retrieval, navigation, etc: Is it possible to approximate, as video retrieval experts did [5], this upper bound for music content semantic description? Which could be the consequences of that?

Everything will be done with social and human computation: Even though automated systems are capable of breaking yet-to-be-discovered glass-ceilings, the most efficient and effective content descriptions are currently (and will be in the future) created by humans. Instead of putting all the effort on the automated analysis let's try to help humans to help other humans.

Music like water? Music as dog!!! Contrasting with the fruitful and dominant metaphor of "Music like water" [6], North and Heartgraves [9] conclude: 'the [pop music] audience places less emphasis on revering the music as high art and more on it as a "friend" that supports them throughout their everyday life, so that pop is less like a god and more like a dog' (p. 358). Do we want to develop research and technology to support this metaphor or, alternatively, to change it?

A digital economy has to be grounded on uncopiable intangibles: Some of them are "personalization", "interpretation" (i.e., helping to enrich, understand or complete the musical context of a work), "embodiment" (e.g., how does it feel as attending a Pink Floyd concert by 1980?), "findability" or "community"[4]. The kernel of the problems about interacting with music content is to be found there, and our theoretical models are still far from explaining and predicting music preference, navigation patterns through collections of musical data or the dynamics of music-related emotions.

[4] http://www.kk.org/thetechnium/archives/2008/01/better_than_fre.php

Dynamics! Dynamics! Time information has been usually given a secondary role in most of the developments of our discipline (although, in other occasions, Hidden Markov Models have been invoked for problems where more parsimonious and simple solutions were available!). The temporal dimension has to be moved to the foreground also when looking at music-related communities: we do not know anything about how they appear, grow and influence the existing ones. Our listeners and users have also been modelled as finite-state machines and approaches based on dynamical systems are still rare [12].

5. REFERENCES

[1] Borlund, P. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8 (3), paper no. 152, 2003.

[2] Celma, O., Herrera, P., Serra, X. Bridging the music semantic gap, *1st International conference on Semantics And digital Media Technology* (SAMT), 2006.

[3] Dietterich, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10* (7) 1895-1924, 1998.

[4] Futrelle, J., Downie, J.S. Interdisciplinary communities and research issues in Music Information Retrieval. Proceedings ISMIR 2002, 215–221, 2002,

[5] Hauptmann, A., Yan, R., Lin, W. How many high-level concepts will fill the semantic gap in news video retrieval? CIVR 2007 *ACM International Conference on Image and Video Retrieval*, Amsterdam, The Netherlands, 2007.

[6] Kusek, D. and Leonhard, G. *The Future of Music: Manifesto for the Digital Music Revolution*, Berklee Press, 2005.

[7] Law, E. The problem of accuracy as an evaluation criterion. ICML Workshop. on Evaluation Methods in Machine Learning, 2008.

[8] Lucas, C., Griffiths, T.L., Xu, F., & Fawcett, C. A rational model of preference learning and choice prediction by children. *Advances in Neural Information Processing Systems* 21, 2009.

[9] North, A., and Hargreaves, D.J. *The Social and Applied Psychology of Music*. Oxford, Oxford University Press, 2008.

[10] Pirolli, P. *Information Foraging Theory: Adaptive Interaction with Information.* Oxford, Oxford University Press, 2007.

[11] Rosenthal, R. and Dimatteo, M.R. Meta-analysis. In Pashler, H. (Ed.) *Stevens' Handbook of Experimental Psychology, Third Edition. Volume 4: Methodology in Experimental Psychology.* New York: Wiley, 2002.

[12] Spivey, M. *The Continuity of Mind*. Oxford, Oxford University Press, 2008.

[13] Stokes, D.E. *Pasteur's quadrant: basic science and technological innovation*. Brookings Institution Press, Washington, D.C, 1997.

[14] Wiering, F. Can Humans Benefit from Music Information Retrieval? In Marchand-Maillet, S., Bruno, E., Nürnberger, A., Detyniecky, M. (Eds.) *Adaptive Multimedia Retrieval: User, Context, and Feedback*, Berlin, Springer, pp. 82-94, 2007.