

# The MIR/MDL Evaluation Project White Paper Collection

## Edition #3

### TABLE OF CONTENTS

#### **PART I.**

#### **Papers Presented at the Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation, 18 July, 2002.**

Who, What, When, Where and Why: Introduction and Acknowledgements (First Edition) <i>J. Stephen Downie, University of Illinois at Urbana-Champaign.....</i>	1
Establishing Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation Frameworks: Preliminary Foundation and Infrastructures <i>J. Stephen Downie, University of Illinois at Urbana-Champaign.....</i>	3
Whither Music IR Evaluation Infrastructure: Lessons to be Learned from TREC [Keynote Address] <i>Ellen M. Voorhees, National Institute of Standards and Technology.....</i>	7
Towards a Workbench for Symbolic Music Information Retrieval <i>David Bainbridge, University of Waikato.....</i>	14
User Studies: A First Step in Designing an MIR Testbed <i>Sally Jo Cunningham, University of Waikato.....</i>	17
Three Criteria for the Evaluation of Music Information Retrieval Techniques Against Collection of Musical Material <i>Joe Futrelle, University of Illinois at Urbana-Champaign.....</i>	20
Music IR for Music Theory <i>Eric J. Isaacson, Indiana University.....</i>	23
Common Music Notation as a Source for Music Information Retrieval <i>Karl MacMillan, Johns Hopkins University.....</i>	27
A Task-Oriented Approach for the Development of a Test Collection for Music Information Retrieval <i>Massimo Melucci and Nicola Orio, University of Padova.....</i>	29
A MIDI Track for Music IR Evaluation <i>José A. Montalvo, Indiana University.....</i>	32
Comparing Aural Music-Information Retrieval Systems <i>Bryan Pardo, Colin Meek and William Birmingham, University of Michigan.....</i>	34
Benchmarking Music Information Retrieval Systems <i>Josh Reiss and Mark Sandler, Queen Mary, University of London.....</i>	37

#### **PART II.**

#### **Papers Presented at the Panel on Music Information Retrieval Evaluation Frameworks at ISMIR 2002, 17 October, 2002.**

Interim Report on Establishing MIR/MDL Evaluation Frameworks: Commentary on Consensus Building <i>J. Stephen Downie, University of Illinois.....</i>	43
Evaluation in Information Retrieval <i>Edie Rasmussen, University of Pittsburgh.....</i>	45
Music GRID - A Collaborative Virtual Organization for Music Information Evaluation <i>Matthew J. Dovey, Oxford University.....</i>	50

Setting Up an Audio Database for Music Information Retrieval Benchmarking <i>Perfecto Herrera-Boyer, Universitat Pompeu Fabra</i> .....	53
What is a Sung Query? <i>Colin Meek, William P. Birmingham, and Bryan Pardo, University of Michigan</i> .....	56
Beyond Recall and Precision: A Full Framework for MIR System Evaluation <i>Josh Reiss and Mark Sandler, Queen Mary, University of London</i> .....	58
Towards Large Databases for Music Information Retrieval Systems Development and Evaluation <i>Gaël Richard, Ecole Nationale Supérieure des Télécommunications</i> .....	64
A Framework for the Evaluation of Content-Based Music Information Retrieval Using the TREC Paradigm (Extended Abstract) <i>Stefan Rieger, Imperial College London</i> .....	68
Evaluating a Music Information Retrieval System - TREC Style <i>Thomas Sødring and Alan F. Smeaton, Dublin City University</i> .....	71
<b>PART III.</b>	
<b>Workshop on the Evaluation of Music Information Retrieval (MIR) Systems at SIGIR 2003, 1 August, 2003.</b>	
MIR/MDL Evaluation: Making Progress <i>J. Stephen Downie, University of Illinois</i> .....	79
Toward Evaluation Techniques for Music Similarity [Keynote Address] <i>Beth Logan, HP Labs; Daniel P. W. Ellis and Adam Berenzweig, Columbia University</i> .....	81
Open Position: Multilingual Orchestra Conductor. Lifetime Opportunity. <i>Eloi Batlle, Enric Guaus, and Jaume Masip, Universitat Pompeu Fabra</i> .....	86
Emphasizing the Need for TREC-like Collaboration Towards MIR Evaluation <i>Shyamala Doraisamy and Stefan M. Rieger, Imperial College London</i> .....	90
If It Sounds As Good As It Looks: Lessons Learned From Video Retrieval Evaluation <i>Abby A. Goodrum, Syracuse University</i> .....	97
Comparison of User Ratings of Music in Copyright-free Databases and On-the-market CDs <i>Keiichiro Hoashi, Kazunori Matsumoto, and Naomi Inoue, KDDI R&amp;D Laboratories, Inc.</i> .....	103
Query by Humming: How Good Can It Get? <i>Bryan Pardo and William P. Birmingham, University of Michigan</i> .....	107
Tracks and Topics: Ideas for Structuring Music Retrieval Test Collections and Avoiding Balkanization <i>Jeremy Pickens, University of Massachusetts, Amherst</i> .....	110
MIR Benchmarking: Lessons Learned from the Multimedia Community <i>Josh Reiss and Mark Sandler, University of London</i> .....	114
Appendix A: Chart of Candidate Music IR Test Collections <i>Don Byrd, Indiana University</i> .....	121
Appendix B: ISMIR 2001 Resolution on MIR Evaluation (taken from <a href="http://music-ir.org/mirbib2/resolution/">http://music-ir.org/mirbib2/resolution/</a> ).....	124
Appendix C: The TREC-Like Evaluation of Music IR Systems <i>J. Stephen Downie, University of Illinois</i> .....	128