# 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.**

*J. Stephen Downie, University of Illinois at Urbana-Champaign* ................................................................. 1  

**Establishing Music Information Retrieval (MIR) and Music Digital Library (MDL) Evaluation Frameworks: Preliminary Foundation and Infrastructures**  
*J. Stephen Downie, University of Illinois at Urbana-Champaign* ................................................................. 3  

**Whither Music IR Evaluation Infrastructure: Lessons to be Learned from TREC [Keynote Address]**  
*Ellen M. Voorhees, National Institute of Standards and Technology* .......................................................... 7  

**Towards a Workbench for Symbolic Music Information Retrieval**  
*David Bainbridge, University of Waikato* ........................................................................................................... 14  

**User Studies: A First Step in Designing an MIR Testbed**  
*Sally Jo Cunningham, University of Waikato* .................................................................................................... 17  

**Three Criteria for the Evaluation of Music Information Retrieval Techniques Against Collection of Musical Material**  
*Joe Futrelle, University of Illinois at Urbana-Champaign* .............................................................................. 20  

**Music IR for Music Theory**  
*Eric J. Isaacson, Indiana University* .................................................................................................................. 23  

**Common Music Notation as a Source for Music Information Retrieval**  
*Karl MacMillan, Johns Hopkins University* ........................................................................................................ 27  

**A Task-Oriented Approach for the Development of a Test Collection for Music Information Retrieval**  
*Massimo Melucci and Nicola Orio, University of Padova* .............................................................................. 29  

**A MIDI Track for Music IR Evaluation**  
*José A. Montalvo, Indiana University* ............................................................................................................... 32  

**Comparing Aural Music-Information Retrieval Systems**  
*Bryan Pardo, Colin Meek and William Birmingham, University of Michigan* ............................................. 34  

**Benchmarking Music Information Retrieval Systems**  
*Josh Reiss and Mark Sandler, Queen Mary, University of London* ................................................................. 37  

## PART II.

**Interim Report on Establishing MIR/MDL Evaluation Frameworks: Commentary on Consensus Building**  
*J. Stephen Downie, University of Illinois* .......................................................................................................... 43  

**Evaluation in Information Retrieval**  
*Edie Rasmussen, University of Pittsburgh* ........................................................................................................ 45  

**Music GRID - A Collaborative Virtual Organization for Music Information Evaluation**  
*Matthew J. Dovey, Oxford University* ............................................................................................................... 50
Setting Up an Audio Database for Music Information Retrieval Benchmarking
Perfecto Herrera-Boyer, Universitat Pompeu Fabra

What is a Sung Query?
Colin Meek, William P. Birmingham, and Bryan Pardo, University of Michigan

Beyond Recall and Precision: A Full Framework for MIR System Evaluation
Josh Reiss and Mark Sandler, Queen Mary, University of London

Towards Large Databases for Music Information Retrieval Systems Development and Evaluation
Gaël Richard, Ecole Nationale Supérieure des Télécommunications

A Framework for the Evaluation of Content-Based Music Information Retrieval Using the TREC Paradigm
(Extended Abstract)
Stefan Rüger, Imperial College London

Evaluating a Music Information Retrieval System - TREC Style
Thomas Södiring and Alan F. Smeaton, Dublin City University

PART III.

MIR/MDL Evaluation: Making Progress
J. Stephen Downie, University of Illinois

Toward Evaluation Techniques for Music Similarity [Keynote Address]
Beth Logan, HP Labs; Daniel P. W. Ellis and Adam Berenzweig, Columbia University

Eloi Battle, Enric Gaus, and Jaume Masip, Universitat Pompeu Fabra

Emphasizing the Need for TREC-like Collaboration Towards MIR Evaluation
Shyamala Doraisamy and Stefan M. Rüger, Imperial College London

If It Sounds As Good As It Looks: Lessons Learned From Video Retrieval Evaluation
Abby A. Goodrum, Syracuse University

Comparison of User Ratings of Music in Copyright-free Databases and On-the-market CDs
Keiichiro Hoashi, Kazunori Matsumoto, and Naomi Inoue, KDDI R&D Laboratories, Inc.

Query by Humming: How Good Can It Get?
Bryan Pardo and William P. Birmingham, University of Michigan

Tracks and Topics: Ideas for Structuring Music Retrieval Test Collections and Avoiding Balkanization
Jeremy Pickens, University of Massachusetts, Amherst

MIR Benchmarking: Lessons Learned from the Multimedia Community
Josh Reiss and Mark Sandler, University of London

Appendix A: Chart of Candidate Music IR Test Collections
Don Byrd, Indiana University

Appendix B: ISMIR 2001 Resolution on MIR Evaluation
(taken from http://music-ir.org/mirbib2/resolution)

Appendix C: The TREC-Like Evaluation of Music IR Systems
J. Stephen Downie, University of Illinois