国際会議参加報告 (ICME 2003)

北原 鉄朗

1 はじめに

本報告書では，International Conference on Multimedia & Expo (ICME 2003) の発表内容や会議全体の様子などについて報告する．ただし，報告内容は，報告者の研究上の興味に多分に依存していることを，あらかじめ断っておく．

2 会議の概要

- **会議名**：International Conference on Multimedia & Expo
- **主催団体**：IEEE
- **開催日程**：2003 年 7 月 6 日〜9 日
- **開催場所**：Baltimore, MD, USA
- **過去の開催実績**：New York, Tokyo, Lausanne を経て，今回が 4 回目
- **投稿件数・発表件数**：投稿件数 760 件中 440 件採録（オーラル・ポスター・スペシャルセッション含む）．上記の他，ICASSP からの振り替え発表 150 件

3 会議全体の分野の傾向

本会議は，マルチメディアに関する学術的な研究や産業的な技術開発すべてを扱う国際会議である．プログラムは，次のセッションで構成されていた：

AIVP: Audio, Image and Video Processing
(9 lecture sessions, 9 poster sessions)
MCN: Multimedia Communication and Networking
(6 lecture sessions, 7 poster sessions)
MD: Multimedia Database
(6 lecture sessions, 4 poster sessions)
MHMII: Multimedia Human-Machine Interface and Interaction
(3 lecture sessions, 1 poster session)
HSMS: Hardware and Software for Multimedia Systems
(2 lecture sessions, 2 poster sessions)
MSCP: Multimedia Security and Content Protection
(2 lecture sessions, 2 poster sessions)
MA: Multimedia Applications
(2 lecture sessions, 1 poster session)
VRI: Virtual Reality and 3D Imaging
(1 lecture session, 1 poster session)
MCSA: Multimedia Computing Systems and Appliances
(1 lecture session)
MSRI: Multimedia Standards and Related Issues
(1 poster session)
ICASSP: ICASSP Presentation
(12 poster sessions)

このように，画像・音響処理からヒューマン・マシン・インタラクション，マルチメディアセキュリティ，バーチャルリアリティまで，カバーする分野は非常に幅広かった．また，上記 11 種類に分類された全 72 セッションには，より詳細な名前がつけられていた．たとえば，音楽情報処理に最も関係の深いであろう **AIVP** セッションにつけられていたセッション名は，以下の通りである．

AIVP-L1: Speech and Audio Processing I
AIVP-L2: Authentication and Recognition
AIVP-L3: Image/Video Rendering/Synthesis
AIVP-L4: Source and Channel Coding
AIVP-L5: Image Coding and Enhancement
AIVP-L6: Image/Video Indexing and Retrieval
AIVP-L7: Video/Image Tracking
AIVP-L8: Face Analysis and Modeling
AIVP-L9: Fast Algorithm for Video Processing
AIVP-P1: Image Processing I
AIVP-P2: Image Processing II
AIVP-P3: Image Compression
AIVP-P4: Coding and Noise Removal
AIVP-P5: Speech and Audio Processing II

AIVP-P6: Image Classification and Detection
AIVP-P7: Image Compression and Modeling
AIVP-P8: Speech and Audio Processing III
AIVP-P9: Motion Estimation

これを見ると, 音響系が 3 セッション (**AIVP-L1, -P5, -P8**) に対して, 画像・映像系が 10 セッション (**AIVP-L3, -L5, -L6, -L7, -L8, -P1, -P2, -P3, -P6, -P7**) と, 発表内容が画像・映像系に偏っていることがわかる. 全体を通しても, 画像・映像系の発表が多く, 音響系の発表は非常に少なかった.

また, マルチメディア・アノテーションやマルチメディア検索などと関係が深いと思われる **MD** セッションには, 以下のセッション名がつけられていた.

MD-L1: Automatic Indexing
MD-L2: Multimedia Learning
MD-L3: Multimedia Retrieval
MD-L4: Multimedia Semantics
MD-L5: Summarization
MD-P1: Content-based Retrieval
MD-P2: Multimedia Indexing
MD-P3: Video Analysis
MD-P4: Segmentation, Summarization and Structuring

以上から, マルチメディアを対象とした検索やアノテーション, インデキシングは, かなり広く研究されていることがわかる. 特に,「Multimedia Semantics」という言葉がセッション名となっていることは特筆すべきことであろう. ただし, **AIVP** と同じように, これらのセッションの多くは画像・映像系の研究であり, 音響系の発表は非常に少なかった.

4 音楽情報処理に関する研究発表

前節で述べたように, 会議全体における音響系の発表は非常に少なかったが, 音響系の発表のなかの音楽関連の発表の割合は, 他の会議に比べて高く, 15 件の音楽関連の発表があった (うち 6 件が ICASSP からの振り替え発表). 以下に, 音楽関連の発表題目を列挙する.

Proceeding Vol. I

MD-L1.3 Anchor Space for Classification and Similarity Measurement of Music (pp.29–32)
MD-L1.4 Automatic Singer Identification (pp.33–36)
MHMII-L1.5 A Hybrid Music Retrieval System using Belief Networks to Integrate Multimodal Queries and Contextual Knowledge (pp.57–60)
AIVP-L1.1 A Statistical Multidimensional Human Transcription using Phone Level Hidden Morkov Models for Query by Humming Systems (pp.61–64)
MA-P1.4* A Fast Search Algorithm for Background Music Signals based on the Search for Numerous Small Signal Components (pp.165–168)
MD-P1.4 Content-based Retrieval of Music in Scalable Peer-to-peer Networks (pp.309–312)
AIVP-P5.4 Inferring Control Inputs to an Acoustic Violin from Audio Spectra (pp.733–736)

Proceeding Vol.II

MD-L3.2 Towards Intelligent String Matching in Query-by-humming Systems (pp.25–28)
AIVP-P8.2 HMM-based Music Retrieval using Stereophonic Feature Information and Frame-length Adaptation (pp.713–716)
AIVP-P8.8 Conventional and Periodic N-grams in the Transcription of Drum Sequences (pp.737–740)

Proceeding Vol.III

ICASSP-9.1* Parametric Vector Quantization for Coding Percussive Sounds in Music (pp.193–196)
ICASSP-10.1* Pitch and Timbre Manipulations using Cortical Representation of Sound (pp.381–384)
ICASSP-10.2* Multidimensional Humming Transcription using a Statistical Approach for Query by Humming Systems (pp.385–388)
ICASSP-10.3* Application of Pitch Tracking to South Indian Classical Music (pp.389–392)
ICASSP-10.8* Musical Instrument Identification based on F0-dependent Multivariate Normal Dis-

tribution (pp.409–412)

* ICASSP からの振り替え発表

5 音響系の発表内容（報告者が拝聴したものを中心に）

ここでは，音響系の研究発表について，報告者が拝聴したものに的を絞り，発表の内容を簡単に紹介する．ただし，ここに書かれた内容は発表を聞いて理解した事柄であり，Proceeding に掲載されている論文を十分に読んだわけではない．そのため，多少の誤解等があるかもしれないことをあらかじめ断っておく．

5.1 音楽関連

AIVP-P5.4 Inferring Control Inputs to an Acoustic Violin from Audio Spectra (Vol.I, pp.733–736)
バイオリンでは，同じ音高でもどの弦のどの場所を弾いたかで音色が変化する．この「どの弦のどの場所を弾いたか」を識別するのがこの研究の目的．非常に難しいようだが，正しい音高を与えればそこそこうまくいくようだ．この技術は，バイオリン演奏指導などに有用と述べていたようだが，具体的な展望は不明．

AIVP-P8.2 HMM-based Music Retrieval using Stereophonic Feature Information and Frame-length Adaptation (Vol.II, pp.713–716)
Query-by-humming システム実現のために，(1) ステレオ音響信号からのメロディの取り出しと (2) メロディの時間軸方向の伸縮による適応を検討．メロディの取り出しは，定位が中央の信号を取り出しフィルター処理によりベースを削除することで実現．「ドライブ中に FM ラジオから聞こえた気になる曲を検索」というシチュエーションを想定しているらしく，携帯電話を通して humming した場合も扱っている．

AIVP-P8.8 Conventional and Periodic N-grams in the Transcription of Drum Sequences (Vol.II, pp.737–740)
ドラムパターンの N-gram によるモデル化を検討．ドラムパターンの周期性を考慮すべく，Periodic N-grams を提案．これは，$L(\geq 2)$ 個間隔で N-gram を計算する（すなわち，$L = 8, N = 3$ なら $p(w_k|w_{k-16}, w_{k-8})$）ものであるが，これでドラムパターンの周期性をとらえるには，1 周期パターンの音符数がかならず L 個でなければならないという制限がある．また，音の順序のみが考慮され，時間的な情報は一切考慮されていない（たとえば小節の先頭に BD が来ることが多いという傾向は，この枠組みでは直接的には表現できない）．この点について意見を求めたところ，「今後の課題」とのこと．この Periodic N-gram と通常の N-gram とを組み合わせることで誤認識を削減したと報告しているが，Fill-in の多い曲に対してはあまり有効には働かない可能性が高いと思われる．

ICASSP-9.1* Parametric Vector Quantization for Coding Percussive Sounds in Music (Vol.III, pp.193–196)
ドラム演奏の音楽音響信号に対する伝送ロスの解消法を提案．あらかじめドラム演奏に対してクラスタリングを行って，各クラスタ（BD，SD などに相当）のコードブック（各クラスタを代表する特徴ベクトル）を転送することで，伝送ロスしたフレームをコードブックで代用する，というのが基本的なアイディア．「楽音が混じった場合に対するアイディアはあるか」と訪ねたところ，「現在はない」とのこと．

5.2 音声と音楽の識別

ICASSP-1.9* A Fusion Study in Speech/Music Classification (Vol.I, pp.409–412)
音声／音楽の識別．「実験にはどのような音楽を用いたか」と訊ねたところ，「クラシックからロックまでさまざま」とのこと．これには歌ものも含まれるようで，歌ものの音楽が音声に誤認識されることもあるとのこと．

AIVP-P5.7 Audio Classification based on Maximum Entropy Model (Vol.I, pp.745–749)
上記と同様に音声／音楽の識別．上記と同様に「実験にはどのような音楽を用いたのか」と訪ねたところ，「さまざまなミュージックビデオから拝借した」とのことで，「さまざまなジャンルが含まれている」とのこと．「歌ものの音楽が含まれると識別は難しいのでは」と訊いたところ，「singing と speech の識別には high zero-crossing rate ratio が効く」とのこと．

5.3 音声関連

AIVP-P5.9* Modeling Prosody for Language Identification on Read and Spontaneous Speech (Vol.I, pp.753–756)

音声から言語を同定する問題を検討．タイトルにもある通り，read speech と spontaneous speech の両方を扱っている．しかし，read speech では，英語・フランス語・ドイツ語・イタリア語・スペイン語の 5 言語（ポスターではこれに日本語が加えられていた）から正しい言語を選ぶという問題設定だが，spontaneous speech では，pair identification（リーグ戦式に任意の 2 言語を識別する）であった．なぜこのような問題設定にしたのかを訊ねたところ，「spontaneous speech で多クラス識別をするのは現時点では難しすぎるから」とのこと．また，「この結果は他の関連研究と比較して，どの程度うまくいっていると判断すればよいのか」と訊ねたところ，「比較は難しい」とのこと．

ICASSP-1.7* Hidden Morkov Model-based Speech Emotion Recognition

音声に含まれる感情の認識を検討．特徴量は，average pitch などの pitch related features と relative maximum of derivation of energy などの energy related features を使用．データベースは acted emotions と spontaneous emotions の 2 種類を用意．acted emotions 対象に約 86%の認識率を実現．spontaneous emotions については「今回の発表までは間に合わなかった」とのこと．spontaneous emotions の認識はかなり難しいようだ．

ICASSP-9.2* HMM-Neural Network Monophone Models for Computer-based Articulation Training for the Hearing Impaired (Vol.III, pp.197–200)

聴覚機能の弱い人のための発音トレーニングシステムについて検討．ここでは，特に母音の発音について検討していた．自分の発声に対する第 1 フォルマントと第 2 フォルマントをリアルタイムに可視化し，簡単に自分の発声の良し悪しを確かめられるデモがあった．「我々の大学の研究者に，日本人学習者のための英語発音教示の研究をしている人がいる」と伝えたところ，「この研究も，そのような外国語学習にも応用できる」とのこと．

6 報告者自身の発表について

報告者は，Musical Instrument Identification based on F0-dependent Multivariate Normal Distribution という題目で発表した (Vol.III, pp.409–412)．これは，SARS の影響により中止になった ICASSP 2003 の振り替えとして，ポスターセッションで発表した．80 分のセッションであったが，質問の多くは，「特徴量は何を使っているのか」「このデータベース (RWC-MDB-I-2001) は，どのような内容か（単音か混合音か，など）」「このデータベースは誰でも入手可能なのか」のどれかであった．

7 感想——まとめに代えて——

本会議に参加してまず思ったのは，発表件数に比べて参加者数がそれほど多くないということである．正確なデータはないが，特にポスターセッションでは純粋な聴衆は少なかったのではないかと思われた．また，オーラルセッション（レクチャーセッション）においても，聴衆こそいるものの議論があまり活発でなかったように見受けられる．中には，聴衆からの質疑が全くなく，座長も質問せずにそのまま発表が終わるケースすらあった．これは，カバーする分野の広さも関係があろう．もう 1 つ感じたことは，上にも述べたが，マルチメディアにおける画像・映像系の偏重である．上でも指摘したように，画像・映像系の発表は非常に多かったが，音響系は少なかった（音響系の発表はデモがしにくいということも関係があるかもしれない．音響系とは異なり，画像検索などの発表では，ポスターに大きく検索結果が印刷され非常にインパクトがあった）．マルチメディアを語る上で音楽が重要なファクターになってくることは明らかであるから，音楽関連の発表も徐々に増えていくことだろう．

本会議は，近年始まったばかりで，会議の性格や方向性が定まっていくのはこれからである．音楽関連の発表が増え，さらに音楽関連の発表を目当ての聴衆が増え，音楽情報科学研究者にとって，本会議が “はずせない” 会議の 1 つとなることを願う．