ICASSP 2004 参加報告

北原 鉄朗

平成16 年 5 月 25 日

1 自分の発表について

AE-P4.10 Category-level Identification of Non-registered Musical Instrument Sounds
T. Kitahara, M. Goto and H. G. Okuno

全体的に「面白い」といってくれる人が多く，比較的好評であったと思われる．質問は，多くが

- 特徴空間は何を使ったのか
- クラスタリングの際の距離尺度は，何を使ったのか
- 識別器は何を使ったのか

の 3 つであった．国内の会議に比べ，興味がより細かい Technical な部分になっていると感じた（音情研では Technical な質問はもっと少ない）．これは，Audience に音声の研究者が多く含まれるからと思われる．PCA や LDA などの次元圧縮手法もほとんどの人が知っていた．また，Martin や Eronen の論文を実際に読んだことがある人がいたことも，国内とは大きく異なる点である．上記以外には，以下のような質問・意見があった．

- クラスタリングの結果は，人間の perception に近くなっているように感じて面白い
- これ（クラスタリング）と同じことを人間に音を聴かせてやってみたか
- アイディアは非常に面白いが，どのようなデータで実験するかで問題の難しさが変わってくるので，もっと多くのデータを使って実験すべきだ．

AE-P5.3 Comparing Features for Forming Music Streams in Automatic Music Transcription
Y. Sakuraba, T. Kitahara and H. G. Okuno

下記のような質問があった．

- Pitch Transition と Pitch Relation Consistency はどう違うのか
- N-gram model は，どうやって（何のデータを使って）作ったのか
- 4 つすべてを使うより 3 つ使ったほうが性能がよい場合があるがどうしてか
- non-crossing grouping が crossing grouing より頻出するという仮定は正しいのか（どうしてこんなことが言えるのか）
- 各楽器は，同時に 1 音しか出さないのか
- 実験ではどんなデータを使ったのか
- ステレオの音響信号を使った研究というのは，他にあるのか

2　他人の発表を聞いて（一部の発表に対する簡単な概要のみ）

SP-P2.5 Feature Space Gaussianization

G. Saon, S. Dharanipragada and D. Povey

特徴ベクトルがガウシアン分布になるように変換する方法の研究である．

SP-P2.6 Online Speaker Clustering

D. Liu and F. Kubala

オンラインで話者をクラスタリングする方法の提案である．ここでの課題は，各フレームの話者ベクトル $s_1...s_{n-1}$が m 個のクラスタ $c_1...c_m$にクラスタリングされているときに，話者ベクトル s_nを $c_1...c_m$のいずれかにクラスタリングするか，新たなクラスタを生成するかである．この課題に対する典型的な解決策はしきい値に基づいて決める方法である．本研究では，このしきい値処理に，新たなクラスタを生成する場合としない場合との尤度比較を組み合わせる．また，この方法は，従来よく用いられる階層的クラスタリングと比べて計算量は極めて小さい．

03-P1.3 ハミング検索のために，オーディオデータからピッチの軌跡を推定する研究である．constant Q transform と cross-correlation を使っている．monophonic music recordings を用いており（ピッチ推定はユーザのハミングに対して行い，楽曲データベースは MIDI ファイルを用いることを想定していると思われる），あまり目新しさは感じなかった．

SP-P3.6 Language Boundary Detection and Identification of Mixed-language Speech based on MAP Estimation

C. Shia, Y. Chiu, J. Hsieh, and C. Wu

一つの文に複数の言語が含まれている発話に対して，言語の境界推定と同定とを行う研究である．こうした発話はアジアでは頻出するため重要な課題である（発表では,「ここから最も近い Starbacks はどこですか」の台湾語を例としてあげていた）．しかし，外国語のなかで出てくる英単語は，その外国語の発音に近くなることが容易に想像される．たとえば，日本語文中の「スターバックス」は本来の「Starbacks」の発音とは大きく異なるはずである．そのため，実際にはかなりむずかしそうに感じた．

SP-P3.12 A Pitch Synchronous Feature Extraction Method for Speaker Recognition

S. Kim, T. Eriksson, H. Kang and D. H. Youn

MFCC 計算時にフレーム長（窓長）をピッチにシンクロさせる（つまりピッチの n 倍をフレーム長とする）ことで，性能改善を図る．この方法は，ピッチ推定の性能に依存すると思われるが，チェックしてないとのこと．実際の話者認識実験により，性能向上が見られた．

SP-P6.2 An Automatic Prosody Labeling System using Ann-based Syntactic-prosodic Model and GMM-based Acoustic-prosodic Model

K. Chen, M. Hasegawa-Johnson and A. Cohen

Pitch Accent と Intonational phrase boundary のラベル付けを扱った論文である．発表中に ToBI という言葉が何回も出てきていた．

SP-P7.5 Acoustic Analysis of Friendly Speech

F. Chen, A. Li, H. Wang, T. Wang and Q. Fang

Normal speech と friendly speech のピッチの違いなどを調べた論文である．かなり中国語に依存した話で（Tone 1〜Tone 4 などによる違いも行っていた），日本語のシステムに応用するのは難しく感じた．

SP-P7.7 Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine - Belief Network Architecture
B. Schuller, G. Rigoll and M. Lang

感情認識を扱った論文である．

SP-P7.9 Yet Another Acoustic Representation of Speech Sounds
N. Minematsu

音声の音響特徴の新たな表現法の提案である．

SP-P7.11 Automatic Emotional Speech Classification
D. Ververidis, C. Kotropoulos and I. Pitas

感情認識を扱った論文である．

SP-L6.4 Robust Speech Feature Extraction by Growth Transformation in Reproducing Kernel Hilbert Space
S. Chakrabartty, Y. Deng and G. Cauwenberghs

ケプストラムに変わる新たな音声特徴の表現法である．MFCC との比較実験で，雑音下での精度が改善されたとのことである．

SP-L6.5 Dimensionality Reduction using MCE-Optimized LDA Transformation
X. Li, J. Li and R. Wang

LDA の欠点を克服する新たな次元圧縮法とのこと．詳細はわからなかったが，基本的には Classification Error を最小化する反復法がベースになっているようだ．

3 音楽関連の研究について

今回，音楽関連の 2 つのポスターセッションがあり，オーラルのセッションはなかった．両方のセッションで自分の発表があったため，他の発表を聞くことはできなかった．以下，楽器音の認識を扱っている 2 つの論文を紹介する．

AE-P4.1 Instrument Recognition in Accompanied Sonatas and Concertos
J. Eggink and G. J. Brown

ピアノ，チェンバロあるいはオーケストラによる伴奏つきの音響信号に対して，メロディの楽器名を同定する問題を扱っている．メロディの楽器名しか同定しないというようにうまく問題を限定することで，単純な手法ながら，比較的複雑な音響信号に対して高い性能を実現している．5 楽器を対象に 86%の認識率を得ている．しかし，単独発音に対する実験結果（57〜76%）に比べて性能が非常に高いのが不思議である．

AE-P5.1 Music Instrument Recognition: From Isolated Notes to Solo Phrases
K. A. G and T. V. Sreenivas

Isolated notes から Solo phrases へのスケールアップを容易にするために，frame-level の特徴のみを使って楽器の認識を行っている．これは，temporal features を使うと，それぞれの音へのセグメントおよび正確なオンセットの検出が必要になるためである．frame-level の特徴のみを使って楽器を認識するため，LSF (Line Spectral Frequencies) を提案している．Isolatee notes に対してはよい実験結果（19 楽器で 77%）が報告されているが，frame-level の特徴のみで楽器を同定することは決し

て簡単ではなく（たとえば，ピアノのスペクトルを持つ定常音をつくってもピアノには聞こえないであろう），この結果は多少疑わしく感じる．

上記以外にもさまざまな研究発表があった．以下，音楽セッションの全発表のタイトルと著者名を列挙する．

AE-P4.1 Instrument Recognition in Acoompanied Sonatas and Concertos
J. Eggink and G. J. Brown

AE-P4.2 Automatic Detection and Tracking of Target Singer in Multi-singer Music Recordings
W. Tsai and H. Wang

AE-P4.3 Time-scale Modification of Music Using a Synchronized Subband/Time-domatin Approach
D. Dorra and R. Lawlor

AE-P4.4 Extraction of Characteristic Music Textures (Eigen-textures) via Graph Spectra and Eigen Clusters
S. Sood and A. Krishnamurthy

AE-P4.5 A Comparison of Human and Automatic Musical Genre Classification
S. Lippens, J. Martens, T. D. Mulder and G. Tzanetakis

AE-P4.6 An Adaptive Learning Approach to Music Tempo and Beat Analysis
S. Gao and C. Lee

AE-P4.7 Using Linear Prediction to Enhance the Tracking of Partials
M. Lagrange, S. Marchand and J. Rault

AE-P4.8 Self-adjusting Beat Detection and Prediction in Music
R. Harper and E. Jernigan

AE-P4.9 An Expressive and Compact Representation of Musical Sound
M. Bocko, O. Altun, D. Headlam and E. Titlebaum

AE-P4.10 Category-level Identification of Non-registered Musical Instrument Sounds
T. Kitahara, M. Goto and H. G. Okuno

AE-P4.11 Recent Improvements of an Auditory Model based Front-end for the Transcription of Vocal Queries
T. D. Mulder, J. Martens, M. Lesaffre, M. Leman, B. D. Baets and H. D. Meyer

AE-P4.12 Automatic Music Summarization in Compressed Domain
X. Shao, C. Xu, Y. Wang and M. Kankanhalli

AE-P5.1 Music Instrument Recognition: From Isolated Notes to Solo Phrases
K. A. G and T. V. Sreenivas

AE-P5.2 Automatic Transcription of Drum Loops
O. Gillet and G. Richard

AE-P5.3 Comparing Features for Forming Music Streams in Automatic Music Transcription
Y. Sakuraba, T. Kitahara and H. G. Okuno

AE-P5.4 Musical Note Segmentation Employing Combined Time and Frequency Analyses
G. Velikic, E. Titlebaum and M. Bocko

AE-P5.5 Bayesian Two Source Modeling For Separation of N Sources from Stereo Signals
A. Master

AE-P5.6 Energy-conserving Finite Difference Schemes for Tension-modulated Strings
S. Bilbao

AE-P5.7 Phonographic Sound Extraction using Image and Signal Processing
S. Stotzer, O. Johnsen, F. Bapst, C. Sudan and R. Ingold

AE-P5.8 Timbral Analogies between Vowels and Plucked String Tones
C. Traube and P. Depalle

AE-P5.9 Separation of Harmonic Structures based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds
H. Kameoka, T. Nishimoto and S. Sagayama

AE-P5.10 Application of the Minimum Fuel Neural Network to Music Signals
A. la Cour-Harbo

AE-P5.11 Baysian Estimation of Simultaneous Musical Notes based on Frequency Domain Modelling
K. Kashino and S. Godsill

AE-P5.12 A Dynamic Programming Approach to Audio Segmentation and Speech/Music Discrimination

4 感想

ICASSP 2004 の参加ははじめて（ICASSP 2003 は SARS により中止）であった．ポスターセッションは，かなり盛り上がっており，あまりの議論の活発さで目の前の人の声が聞き取れないほど部屋はうるさかった．また，質問の内容などから，信号処理に詳しい人が非常に多いように感じた．一方，オーラルセッションはあまり盛り上がらず，質疑が全くない発表があったほどであった．日本からの参加者は，音声の主要な先生方はほぼ参加されていた他，NTT の方々が非常に多かった．日本以外では，中国人などを中心にアジア勢の参加が活発だったように思える．また，海外では音楽の研究が盛んであることも実感した．我々が参考文献として挙げている文献を実際に読んだことのある人がいたり,「僕も同じような研究をやっている．今度の ASA meeting で発表するよ」と言っていた人もいた．また，音楽関連の他の発表にも魅力的なものが多く，気を引き締めて研究を進めなければならないと感じた．