



# SIGMUS 第29回 研究発表会 報告・質疑記録

1999年2月18~19日 NTT 厚木研究開発センター

## (1) 音楽音響国際シンポジウム 98 報告 (ISMA'98 Leavenworth U.S.A.)

中村 勲

記録: 荒木 円博 (豊田中研)

会議の歴史や発表件数などが報告された。また会議中毎晩開かれたコンサートの雰囲気を紹介された。

## (2) 発声者が知覚する音声の音色 — その客観的評価 —

中山 一郎, 岡田 稔枝, 中川 みかほ (大阪芸大)

記録: 増井 誠生 (富士通研)

Q: 足立 (愛知県立大) 自己聴取音がローパスフィルタをかけたような音として聴こえるようだが、その物理的なメカニズムについて説明はできるのか?

A: 耳小骨筋反射という現象に注目して調査を始めたが、実際にそのような周波数特性が出ているかどうかは未確認である。

Q: 物理的ではなく音響心理的に研究しているという段階か?

A: そうだ。

Q: 中村 アーアという母音のみの特性について調べていることについて、言葉が違えば母音もかなり違うはずが、一樣な特性が得られるのか?あるいは言葉によって各々特性は違ってくるのか?

A: 録音再生音に対する周波数特性は母音による違いがあまり見られなかった。ラウドネスは全発声者の全母音における平均だが、統計的有意差はみられなかったため、過少知覚量を約10dbと判断した。音色は、低音の方をより大きく聞いている。これは電話のサイドトーン研究と関連する。自己聴取音には骨導音も含まれるが、骨導音を確認する場合など、耳を塞ぐと鼓膜のインピーダンスが大きく変化してしまうため、今回の実験ではスピーカーボックスを用いた。この場合は骨導音にも母音ごとの大きな特性の変化はみられなかった。

Q: 足立 研究の目的は?

A: 人間の聴覚の根幹に踏み込むような研究を目指す。その手法について、実験システムにおけるディレイの利用は、「コンサートホールのステージにおいて、演奏者が演奏しやすい反射音とは何か?」を考えたことから思いついた。コンサートホールでは、直接音と反射音のバランスについて、演奏者は聴衆に比べて大きな直接音が必要とする。だが、その直接音のレベルを下げれば演奏者と聴衆は同じ聴き方をしていないのではないかと考え、ディレイによって遅延をかけてみよう、と考えたことが研究の発端である。建築音響という立場から生まれたアプローチである。

Q: 小坂 (NTT) 電話のサイドトーンについては、昭和52-57年頃まで電話の品質評価研究として行われた。快適な音声通話のために、気導側音、自然側音、電話機側音がミックスしている必要がある。これらの文献は調べたのか?

A: 論文はフォローしている。ただ、論文を読む限り、聴覚の根幹まで踏み込んでいないように感じられる。

## (3) フランス声楽曲に於ける鼻母音発声の研究

田 大成 (東京芸大)

記録: 増井 誠生 (富士通研)

Q: 桑原 (帝京科学大) ホールのA点(舞台側)に比べてD点(客席後側)では鼻母音のシンギングフォルマントが下がるといふ説明があったが、むしろ通常の母音のフォルマントが上がり、鼻母音については変化していないと解釈すべきではないか?通常の母音のフォルマントがホールの音響特性に対応するのではないか?

A: 相対的な表現であり、指摘の通りである。鼻母音・通常母音ともに変化するが、その変化を相対的に表現しようとした。

Q: 中村 どちらも500Hzで0dbに基準化しているが、その点での絶対的な数値を併記の方が情報量が多く、理解しやすいのではないか?基準とした500Hzの周辺でA点とD点でどう違うか、500Hzの周辺での値と無関係にシンギングフォルマントの変化がわかるとデータがみやすくなるのでは?

C: 白砂 (東京芸大) A点とD点での絶対量を見ればわかりやすくなるはずという指摘だ。

C: 足立 (愛知県立大) これは絶対量の測定法の問題ではなく、結果を提示する際の基準点での絶対差を明記する方がわかりやすいという意見である。

A: 白砂 測定の問題のために基準点での相対化を試みているが、基準点での絶対量ももちろんチェックはしている。

Q: 中山 (大阪芸大) 実験手法について、Unknown Factorが多過ぎると感じる。録音スタジオの特性、スタジオマイクの特性、ホールの特性、スピーカーの特性、客席マイクの特性など。サイエンスであるなら、せめてスタジオ録音は無響室でフラットマイクを使い、再生にはフラットスピーカーを用いるべきだ。それでもホールの特性は強烈である。Unknown Factorを減らすことで、確実な結果を出せるように努力する方がいいのではないか?

A: 研究のアプローチの問題である。フランス語の16母音すべてについて実験環境を描いたので、ホールやマイクの影響は無視できると考えた。

C: 結果をグラフ化するだけではサイエンスにならないと考える。

## (4) 楽曲評価におけるニューラルネットワークモデル

村上 英典 (帝京平成大)

記録: 荒木 円博 (豊田中研)

階層型ニューラルネットワークとバックプロパゲーションを使って、曲の印象を表す形容詞を学習させてみた、といった内容だった。ニューラルネットワークの使い方や評価の方法に関して問題がある旨、コメントがあった。主なコメントを列挙する。

- 過去の Computer Music Journal を読破してこうした目的でのニューラルネットワークの使い方を理解すべきだ。
- 評価データとして学習データ以外のデータを使う必要がある。
- 対になっている形容詞、例えば「明るい」と「暗い」を独立して扱う必要があるのか?
- 通常のニューラルネットワークの使い方と違って、中間層の各ニューロンの役割をあらかじめ決める方法をとっているが、その決め方は妥当なのか?

## (5) 仮想空間におけるピアノ演奏動作の生成と表示

関口 博之, 英保 茂 (京大)

記録: 荒木 円博 (豊田中研)

Q: 原田 (NTT) この研究では楽譜情報を単純にMIDI化したものに基づいて演奏動作を生成しているが、私としてはうまい人の指の動きの再現がみたいので、うまい人の演奏を記録したMIDIデータを元にすることはできないのか?

A: もともとそうした手本のない曲も扱うことを目指して楽譜情報に基づいているため、質問のようなアプローチは考え

ていない。

- Q: 苦手な指を使わない演奏動作の生成も可能か。  
A: その指を使うコストを無限大にすればできる。  
Q: 3本の指に不自然な角度がついている例があったか?  
A: 指の間の従属性を考慮していないのが原因であり、改良したい。  
Q: (不明) 右手のパートと左手のパートが分離されていないデータ(同じチャンネル番号)を与えて、適切に振り分けることは可能か?  
A: 考えていない。それぞれ別のパートになっていることを前提にしている。

(6) 電子楽器 Limber-Row の開発— 技術者としてではなく作曲家として — 岡本 久 (岡本情報メディア研)

記録: 志村 哲 (大阪芸大)

岡本氏自身が開発した新楽器リンバローのシステム解説、デモンストレーション演奏。両手足でコントロールするバーによりベンド情報を発生させる。これを、電子音源のヴォリューム、ピッチ等のパラメータに反映させることができる。今後の展開としては、他の楽器システムとのコラボレーション、インストレーション用の機器開発を考えている。

- Q: 志村 (大阪芸大) 作者の楽器に対する考え方を伺いたい。システムに要求される仕様で「少なくとも2音が同時に発音されること」といっているが、その意味はどこにあるのか。  
A: 私はヴァイオリンが好きだということもあり、その表現力において、4弦が織りなす複音の響きが念頭にある。従来からある考え方もかもしれないが、ハーモニーが作れるということである。また、2音が相互に相対する音として存在させられることがねらいである。  
Q: このようなベンド情報が中心となるシステムにおいて、2音同時に意図通りにあやつった演奏ができると考えているのか。  
A: 身体を半分に分けて、左右で一音ずつコントロールしている。これをうまく使いこなせばできるようになると考えている。  
Q: 小坂 (NTT) 今の質問にも関連するが、この演奏に再現性は求められるものなのか。あるいは、演奏のたびに確率的、状況依存的に変化するものになるのか。  
A: 楽譜はかなりアバウトではあるが、発音開始ピッチ等を規定している。演奏は難しいが、練習すれば曲全体の雰囲気や時間構造はほぼ同じものになる。アドリブはざけたいと考えている。しかし、楽譜に全てを記してあるわけではないので、自分が演奏すれば同じようにいくということである。ただ、将来、他人がやっても同じ演奏ができるかといえ、これは今は何とも言えない。この、楽器自体はその可能性をもっていると考えている。

(7) 最近のマルチチャンネルサラウンド録音制作手法— 音楽ドラマ制作にみる 3-2 サラウンド手法 —

沢口 真生, 深田 晃 (NHK), 高橋 幸夫 (日本コロムビア)

記録: 志村 哲 (大阪芸大)

1970年代のサラウンドシステムへの失望と、こんにち再び脚光をあびることとなったことの、歴史的社会的考察からはじまり、現在の様々な方式の解説がおこなわれた。

- Q: 平賀 (諺) (図書館情報大) ポピュラー曲の場合、ホール録音とヴォーカル用マイクとの兼ね合いはどうなっているのか?  
A: 深田 オーケストラは生演奏で、収録はメインマイクの他、かなりの本数のマイクを用いている。それらを合わせてミキシングし、5チャンネルにまとめている。  
Q: 志村 (大阪芸大) これだけ多くのマイクで収録し、かつこのように良く響く空間において5台のスピーカで再生した場

合、音が濁る場合があると思うが、これに対して何か対策や方法論があるのか。

- A: 深田 マイクの本数が増えれば増えるほど、特にピアノ音楽においては位相関係等が複雑になるので、その要素は増えると思うが、だからといって何でもワイドポイントマイクで収録すればよいというものでもない。対策としては、たとえばホール録音においては、ステージからの直接音が、ステージから離れていくことによってやがて間接音だけになる境目のポジションにメインマイクを置くということが基本になっている。メインマイクは、全指向性のマイクを使用するが、サラウンド録音で響きをとる場合は、フロント用には単一指向性のマイク、そしてあまり離れないところに響き用の単一指向性マイクを反対方向に置いて、できるだけ濁りを少なくしている。  
Q: 小坂 (NTT) 今日の客席のスピーカ配置でのスイートスポットは、どの辺のどの広さだと考えられるか。  
A: 深田 大きい空間で多くの人に試聴してもらう場合、スピーカの位置を上にあげ、距離を長くとることが有効である。いろいろな条件があるので難しいが、後ろの間隔を広くとってやることで後ろの方で聞いている者も違和感なく聞けるようである。本日の配置においては、スイートスポットはセンター付近である。  
C: 平賀 前の方に座っての印象では音像が実際の倍程度の大きさに聞こえた。  
Q: 原田 (NTT) デジタル録音において、量子化ビット数あるいは、サンプリングレートはいくらあれば足りると考えているか。  
A: 沢口 現在、いろいろな試みがあるが、将来的にはフォーマットはどうでも良いが、バケツは大きいほどよいと考える。ただし、音を作る立場としては、現在、スペック的には高い値のものでも機器としてこなれていないものがあるので、その場合は、低いスペックでもこなれた機器の方が良いと考えている。  
A: 深田 現実に、96kHz サンプリング 20bit による録音もおこなっているが、高精度になるとそのずれが全体におよぼす影響に関しても、シビアになる。ただ、沢口も言ったようにバケツは大きい方がよいと考えているし、素材としては録れるものはすべて録っておきたいというのが、現場のエンジニアの考え方である。
- (8) 特徴量に注目した複数楽器の演奏における音源同定処理 木下 智義, 坂井 修一, 田中 英彦 (東大)  
記録: 平田 圭二 (NTT)
- Q: 相川 (NTT) 基本波はどうやって同定しているのか。  
A: 単音形成クラスタリングのステージで、周波数の一番低いものを選ぶ。オクターブの可能性があるので、2fも基本波という複数の仮説を同時に探索する。最終的には、類似度計算のステージで判定する。  
Q: 平賀 (諺) (図書館情報大) データベース中の楽器の実音は、特徴量の選択に影響するか?  
A: 普通に弾かれた楽器音を、考えられる範囲で一通り揃えている。  
Q: ベンチマーク演奏では、各楽器音の立ち上がりは揃っているのか?  
A: ベンチマーク演奏としては、人が演奏したものを与えている。立ち上がりはおおむね揃っており、意図的にズラしたりはしていない。  
Q: 鶴根 (広島電機大) 加算特徴量を減算する際、その特徴量だけのみ減算しているのは何故? データベース中のある楽器に関する特徴量を全部減算した後でもう一度特徴量の再検

査をすれば良いのではないか?

A: そのデータベース中の特徴量データにバラ付きがあり、全部減算するのは危険なため、そうしていない。

(9) 事例ベースで生成された演奏表情を用いて自然な演奏を行なう伴奏システム

西田 深志, 鈴木 泰山, 徳永 健伸, 田中 穂積 (東工大)  
記録: 平田 圭二 (NTT)

Q: 平賀 (譲) (図書館情報大) 事例を聴かせて欲しかった。類似した複数の事例 (case) から演奏の表情付けを生成する時の重み付けは?

A: 事例に付いている重要度で重みを付けている。

Q: 入力と全く同じ事例が存在したら出力はどうなるのか?

A: 重み付けがあるので、事例ドンピシャが出力されることはない。

Q: それで良いのかどうか考える必要があると思う。リハーサルの回数も重み付けに効いてくるのか?

A: 効いてくる。

Q: 山本 (東工大) 自然な演奏に聴こえるために最も重要な要素は何か?

A: テンポだと思う。しかし、テンポと音量の両方で考えるべきだと思う。

Q: 事例ベースを改良して行くとして、どこまで人間のような表情付けに近づくとするか? また、人間と同じような表情付けが永遠にできないとしたら、何が根本的な原因だと思うか?

A: 現在は事例ベースの事例数が少ないので、時々変な結果も出る。もっと事例ベースを充実させて、どこに本質的な問題点があるのかを探りたい。

Q: 平賀 (瑞) (筑波大) 現在の伴奏パートの数は2であるが、この伴奏パート数は増やせるのか?

A: MIDIの制約で15 (= 16 - 1 (ソロ)) である。計算機の速度のため、独奏者の演奏を待つような状況では伴奏4パートが限界である。

(10) くし形フィルタによる実楽器音を対象とした演奏楽器推定

三輪 多恵子, 田所 嘉昭 (豊橋技科大), 斎藤 努 (豊田高専)  
記録: 平田 圭二 (NTT)

訂正: 予稿中 (29) 式の + (プラス) を - (マイナス) に直して欲しい。

Q: 小坂 (NTT) DFTによるピッチ推定を比較データに使っているが、具体的にはどうしているのか? 離散周波数点をそのまま基本周波数と見なしているのか?

A: オーバーサンプリング (54KHz) を用いているので、その離散点はきめ細かいと思う。

Q: 松島 (東邦大) 単音の場合を比較データにすれば良いのではないか?

A: そうかも知れない。

Q: 平野 (神奈川大) ビブラートへの対応は?

A: ボルタメントする音の処理は考察中。ビブラートは将来の課題としたい。

Q: ノッチフィルタでフィルタバンクを組む方式に対する利点は何か?

A: 全ての倍音に対して一気に差分が取れる。計算コストが低い。

Q: 杉山 (会津大) 差分を除去する方式より、全体を最小二乗法で解いた方が安全なのではないか?

A: 最小二乗法の利用は今後検討したい。

C: 平賀 (譲) (図書館情報大) 実際の楽器音には、残響や録音

環境から来るノイズなどが乗っている。

Q: 木下 (東大) この方式の精度を上げるには、実楽器音のテンプレートを実際なく必要とするのではないかと思う。テンプレートを増やす以外の精度を上げる方法は?

A: 倍音の位相情報を利用することも考えられる。

C: 田辺 (カクダソフト) 楽器音の分析にフーリエ解析は適していない。予稿 68 頁にアタック除去と書いてあるが、アタック部分こそが楽器音の最も特徴的な部分である。音が安定した部分のスペクトルは、お茶の出がらしのようなものである。アタック部の非整数次倍音、それらの位相、インパルス、ノイズ等こそ分析すべきである。

(11) 日本語音声教育のための日本語アクセント型判定法  
熊谷 有香, 吉田 奏子, 三輪 譲二 (岩手大)

(12) 音声教育のための声道形の動的3次元表示法  
佐々木 優, 平野 崇, 三輪 譲二 (岩手大)

(13) セグメントの高速探索法  
杉山 雅英 (会津大)

(14) 局所話速変化検出を指向した簡易セグメンテーション手法と実環境音声への適用について  
佐々木 淳一, 広重 真人, 宮永 喜一, 柄内 香次 (北大)

(15) ITU 勧告 G.729 の拡張と品質評価  
林 伸二, 栗原 祥子, 片岡 章俊 (NTT)

(16) 5kHz 帯域低ビットレート音声符号化— 帯域感と主観品質の関係 —  
原田 登, 大室 伸 (NTT)

(17) 低ビットレート音声符号化における背景雑音付加音声の品質向上  
大室 伸, 関野 一則 (NTT)

感想 (11-17): 荒木 円博 (豊田中研)

これらは音声に関する一連の発表である。

この中では (11), (14) がモーラに関わるものであった。(11), (14) で取り上げられたような技術は、村尾らや青野らのモーラに関する研究の中で音声レベルでの分析手段として利用できそうに感じた。

また (17) で背景雑音を付加した方が自然に聞こえるというのは興味深かった。確かにデモを聞いた感触では、背景雑音を付加しない場合、音声がある区間とそうでない区間のつながりが不自然に感じられ結果的に全体を不自然に感じさせていた。