

協調演奏における人間の時間制御モデルの推定*

三井 卓 堀内 靖雄 市川 熹†

tel:043-290-3300 fax:043-290-3269 e-mail:hory@icsd4.tj.chiba-u.ac.jp

千葉大学 工学部 情報画像工学科‡

〒263-8522 千葉県 千葉市 稲毛区 弥生町 1-33

1 はじめに

音楽情報処理研究の一分野として伴奏システムが研究されている。伴奏システム登場以前の計算機を用いた演奏では、計算機の演奏を演奏時に動的に変化させることができず人間の独奏者が計算機に合わせて演奏しなければならなかった。それに対し伴奏システムは独奏者の演奏を入力とし、その演奏と協調してアンサンブルをおこなうシステムである [1]。

これまでの伴奏システムには、出力する演奏が人間にとって不自然に感じられるといった問題点がある。これはシステムがスケジューリングの段階において、「独奏者の直前のテンポで次の発音時刻を予測して、伴奏システムの演奏を決定する」といった単純な処理をおこなっていることに起因していると考えられる。にもかかわらず伴奏システムの人間の演奏スケジューリングに関する研究報告はあまりなされていないのが現状である。中には人間と機械の協調演奏モデルを扱った研究もおこなわれている [2] が、実際の人間の協調演奏からモデルを作成したわけではない。

本研究では、実際に人間がおこなっているスケジューリングを基にシステムのスケジューリングをおこなう必要があるという観点から、二人の人間による演奏を収録し、そのデータから人間の演奏スケジューリングを統計的にモデル化することを試みる。

2 モデルの構成

2.1 データの収録

モデル化に利用するデータは二人の演奏者による演奏を計算機で収録したものである。収録に用いた曲は Hanon 作曲の「ピアノの名手になる 60 練習曲」第 5 番である。この曲は四分の二拍子で 29 小節からなり、全体を繰り返すため 57 小節分演奏することになる。全曲を通じて 16 分音符のみで構成され、右手パートと左手パートがオクターブで平行移動する。収録では右手パートと左手パートをそれぞれの演奏者に単旋律の

曲として合奏してもらった。また演奏者は相手の姿が見えない状況で、互いの発音のみを頼りに演奏をおこなった。演奏者は 3 名で 2 回ずつ演奏をおこなってもらったため 6 通りのデータを収録した。

2.2 線形結合によるモデル

モデルは、人間の伴奏者が自分の過去のテンポ情報と過去の発音時刻のずれを利用していると考え、

$$t_n = \sum_l^L a_l t_{n-l} + \sum_m^M b_m D_{n-m} \quad (1)$$

という形であると仮定した (t は伴奏者のテンポの逆数 (秒)、 D は独奏者と伴奏者の発音時刻のずれ (秒))。テンポの逆数は発音時刻の間隔となる。 L, M はそれぞれの説明変量の情報をいくつ過去のものまで使うかを表している。

係数 a_l, b_m は収録したデータに対して L, M を 0 ~ 4 の間で変化させて重回帰分析をおこなうことで求めた。

2.3 最適な説明変量の組み合わせ

表 1 は求めた重回帰式の残差の平均と標準偏差である。

表 1: 残差の平均 (上) と標準偏差 (下) (単位:秒)

		M				
		0	1	2	3	4
L	0	—	0.3087	0.3073	0.3064	0.3058
		—	0.0329	0.0389	0.0424	0.0446
	1	0.0006	0.0006	0.0005	0.0004	0.0004
		0.0203	0.0178	0.0165	0.0161	0.0161
	2	0.0002	0.0003	0.0003	0.0002	0.0002
		0.0155	0.0137	0.0137	0.0120	0.0119
	3	0.0002	0.0002	0.0002	0.0002	0.0002
		0.0140	0.0127	0.0126	0.0118	0.0116
	4	0.0002	0.0002	0.0002	0.0002	0.0002
		0.0135	0.0122	0.0121	0.0117	0.0116

これらの値が小さいほど回帰式が人間のスケジューリングに近い予測をしているといえる。

*Estimation of Human Performance Control

†Suguru Mitsui, Yasuo Horiuchi, Akira Ichikawa

‡Dept. of Information and Computers Sciences, Faculty of Engineering, Chiba University

表 1 の中でもっともよい予測をしているのが $L = 4, M = 4$ の場合 (伴奏者のテンポ情報を過去 4 つ、ずれを過去 4 つ使う) である。これ以上 L, M を大きくしても平均値と標準偏差はあまり小さくならないと思われるので、このモデルに対して単純なモデルと比較し評価をおこなう。

また、 $L = 0$ の場合はどれも値が大きくなっている。このことから発音時刻のずれのみを使ってスケジューリングすることは妥当ではないことが分かる。

3 評価

作成したモデルと単純なモデルに収録データを入力して、実際に人間の伴奏者が発音した時刻との差を求めた。単純なモデルは $t_n = T_{n-1} + D_{n-1}$ (T は独奏者のテンポの逆数 (秒)) とした。これは、 $T_n = T_{n-1}$ のであると仮定したとき、独奏と伴奏の次の音の発音時刻が一致するようにスケジューリングするモデルである (図 1)。

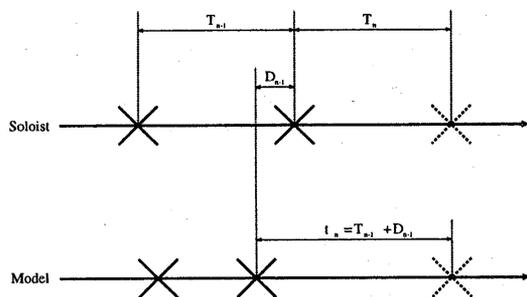


図 1: 単純なモデル

図 2 はモデルを構成するときに用いたデータ (学習内データ) で評価した場合である。縦軸が人間の伴奏者との発音時刻の差 (秒) で、モデルが先に発音した場合に正の値となる。横軸は演奏の拍数である。グラフを見ると作成したモデルの方が人間との差が小さく、標準偏差も単純なモデルが 0.026 なのに対して作成したモデルは 0.012 となっている。このことから作成したモデルの方がより人間に近い演奏スケジューリングをおこなっているといえる。

図 3 はモデルを構成するときに用いたデータ以外で評価した場合 (学習外データ) である。グラフを見ると作成したモデルの方が人間との差が小さく、標準偏差も単純なモデルが 0.020 なのに対して作成したモデルは 0.010 となっている。このことから作成したモデルの方がより人間に近い演奏スケジューリングをおこなっているといえる。

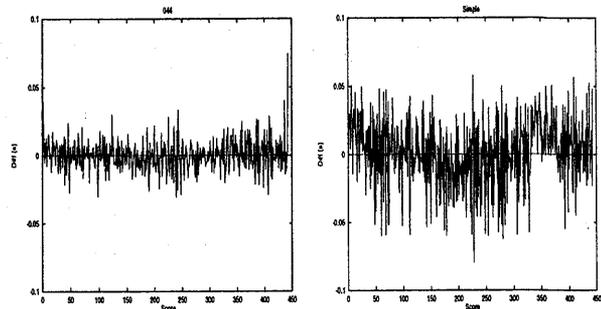


図 2: 学習内データ (左:作成モデル、右:単純なモデル)

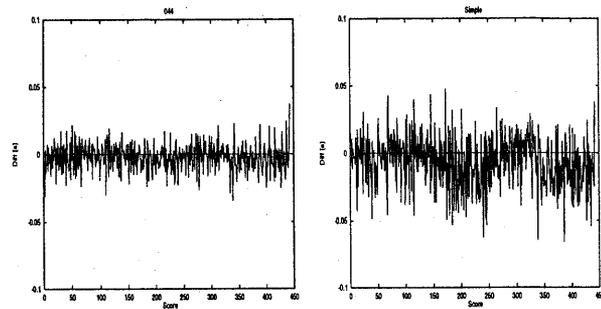


図 3: 学習外データ (左:作成モデル、右:単純なモデル)

4 おわりに

一つの演奏スケジューリングモデル構成法を提案し、その評価をおこなった。今回作成した人間の時間制御を基にした演奏スケジューリングモデルは単純なモデルと比べて、人間に近い演奏スケジューリングをおこなうことがわかった。このモデルを伴奏システムに実装すれば、より人間らしい、自然な伴奏が期待される。

今後の課題としては伴奏システムに実装しての評価実験が挙げられる。しかし、現在の伴奏システムでは発音に遅延が生じるため数ミリ秒～十数ミリ秒の精度での発音時刻、及び演奏者との発音時刻のずれを正確に検証することができない。より精度の高い伴奏システムが望まれる。また、今回はモデルの構成方法として線形結合によるモデルを選択したが、別の手法を用いて構成されたモデルとの比較検討もおこなわなければならない。

参考文献

- [1] 堀内靖雄. 自動伴奏. 長嶋洋一, 橋本周司, 平賀譲, 平田圭二 (編), コンピュータと音楽の世界—基礎からフロンティアまで (bit 別冊). 共立出版, July 1998.
- [2] 澤田秀之, 磯貝昌幸, 橋本周司, 大照完. 音楽演奏における人間と機械の協調動作について. 情報処理学会全国大会講演論文集, 1992.