

## ブレスによる合図を検出する伴奏システム

堀内 靖雄<sup>†1</sup> 西田 昌史<sup>†1</sup> 市川 薫<sup>†2</sup>

従来の伴奏システムでは人間の独奏者と同時に演奏を開始しなければならない楽曲への対応は困難であり、人間の独奏者にとって伴奏システムを使いにくくしていた。本研究では実際の人間同士の合奏でブレスによる合図が重要な役割を演じていることに着目し、伴奏システムが人間の独奏者のブレスによる合図を検出し、人間の独奏者と同時に演奏を開始できる伴奏システムを実現するため、(1) ブレスによる合図と演奏開始のタイミングの分析、(2) 独奏者のブレスによる合図の検出手法の提案、(3) システムへの実装と評価を行った。ブレスによる合図に対するシステムの反応と演奏者の演奏のずれに関して、人間の演奏者が許容できる時間範囲を調べた結果、-60 ミリ秒～97 ミリ秒程度が許容範囲であることが示された。この結果に基づいて、本提案システムの性能を評価したところ、240 データに対し、94.6%が上述の人間の演奏者の許容範囲値内に含まれると考えられ、十分な精度でシステムがブレスの合図により演奏を開始できることが示された。

### Accompaniment System Using Cue by Breath

YASUO HORIUCHI,<sup>†1</sup> MASAFUMI NISHIDA<sup>†1</sup>  
and AKIRA ICHIKAWA<sup>†2</sup>

For previous computer accompaniment system, it was difficult to synchronize with the human soloist at the beginning of some musical pieces where the machine has to begin the accompaniment performance simultaneously with a human soloist. In the actual ensemble by human performers, they use breath as musical cues in general. In this study, in order to develop the accompaniment system using cue by breath, (1) we analyze the timing relation between breath as musical cues and the onset of the first note of the soloist, (2) we introduced the automatic detection method of breath timing, and (3) we implemented the accompaniment system using cue by breath and evaluate the system by human performers. As a result, it was suggested that the tolerated range about timing difference of synchronization between the soloist and the system were from -60 msec in system delay to 97 msec in haste and the system achieved the synchronization in the above range for the 94.6% of 240 experimental data.

### 1. はじめに

伴奏システム<sup>1)-5)</sup>とは人間の独奏者の伴奏をコンピュータが行うもので、人間とコンピュータとのリアルタイム・インタラクティブ・システムの1つとして、音楽情報処理研究だけでなく、人間とのリアルタイム・インタラクションを考えるうえで重要な研究テーマとなっている。伴奏システムでは楽譜情報はあらかじめ独奏パートと伴奏パートの両方がシステムに与えられており、独奏者の演奏と協調するように伴奏システムの演奏タイミングをリアルタイムで制御する。

従来の伴奏システムでは曲の冒頭で独奏者と伴奏システムが同時に音を演奏するような楽曲に対しては対応が困難であり、独奏者がペダルスイッチを踏んで伴奏の演奏をスタートさせたり、独奏の音が演奏されてから即座に伴奏システムの演奏が開始されたりする、等の方法がとられてきたが、これらの方法は人間の演奏者にとって非常に使いにくく、直感的に合図を出すことが困難であった。一方、人間同士の合奏の場合には独奏者は体の動きやブレス(息継ぎ)を用いて伴奏者に合図を送り、両者の演奏を合わせている。そこで本研究では伴奏システムが独奏者のブレスによる合図を理解できるようにすることを目的とする。なお、伴奏システムにあらかじめ与えられている楽譜情報にはデフォルトのテンポが記述されており、そのテンポ情報に基づいて、ブレスの合図から独奏者の演奏開始時刻を推定する。

音楽のブレスに関連する研究としては歌声のブレスを検出する研究がある<sup>6),7)</sup>。これらはブレス音を除去する等の目的のために歌唱中のブレス部分をオフラインで検出するものである。また、歌唱におけるブレス(呼吸)とフレーズを分析している研究もある<sup>8)</sup>。しかしながら、合奏における合図としてブレスを利用した研究はまだ行われていない。また、ブレスによる合図を伴奏システム等のインタラクティブシステムで利用するためには、ブレスをリアルタイムで検出する必要があり、従来のオフラインでのブレス検出手法は適用が困難である。

本論文では2章で独奏者のブレスによる合図と演奏開始のタイミングの時間的関係の分析結果について述べ、3章で独奏者のブレスによる合図の実時間検出手法を提案する。4章

<sup>†1</sup> 千葉大学  
Chiba University

<sup>†2</sup> 早稲田大学  
Waseda University

では本手法を伴奏システムへ実装し、システムと人間の演奏家との評価実験を行った結果について述べる。5章でプレスなしで演奏が開始された場合のシステムの対応方法について述べ、最後に6章でまとめと今後の課題について述べる。

## 2. 合図としてのプレスの分析

人間同士の合奏では曲の冒頭でプレスにより合図を示す場合、演奏開始の1拍前でプレスを行うことが一般的であると考えられる。そのようなプレスが伴奏システムのための合図として利用可能であるかどうかを調べるため、プレスの発音開始時刻と演奏の開始時刻との関係を分析する。なお、伴奏システムに実装する際、曲のテンポは既知であるとする。実際の人間同士の合奏においても、大体のテンポはあらかじめ相談して決めてあり、そのテンポにあわせてプレスの合図を出すことが一般的に行われている。そのため、プレスの開始時刻が分かれば、設定されたテンポから演奏開始時刻を推定することができると予想される。

### 2.1 分析データの収録

本研究では演奏開始時のプレスを分析するためのデータ収録を行った。今回の実験対象とする楽器はフルートとし、フルート用の集音マイクを用いて収録を行った。マイクはクリップでフルートに固定し、集音部は口の前方斜め上、約5cmのところに配置した。マイクの配置においては、息が直接マイクにあたるのを防ぎつつ、フルートの音、ならびに、鼻、口からのプレス音を安定して収録できるように配慮された。

本実験で収録に参加した演奏者は音楽大学フルート専攻の卒業生1名(以下、演奏者A)とフルート経験25年以上のアマチュアのフルート演奏家(以下、演奏者B)の2名である。

データ収録には単純なフレーズ(練習曲)と実際の曲の一部を用いた。収録に用いた楽譜を図1、図2、図3、図4に示す。演奏者には「伴奏者に出だしの合図を示すように明確にプレスするよう」教示した。さらに楽譜1と楽譜2に関しては3種類のテンポ(80, 100, 120 [拍/分])による演奏を収録した。演奏者に3種類のテンポを呈示するため、演奏直前にメトロノームで指定のテンポを2小節間呈示し、メトロノームを停止してから演奏を開始した。楽譜3と楽譜4については演奏者の好みのテンポで演奏していただいた。これら各演奏は5回ずつ収録され、各演奏者ごとに40データが収録された(楽譜1, 2がそれぞれ15データ, 楽譜3, 4がそれぞれ5データ)。

### 2.2 プレスのアノテーション

得られた演奏データに対し、プレス、ならびに、冒頭2拍の音の立ち上がり時刻をアノテートした。アノテーションにはWaveSurfer<sup>9)</sup>を利用した。アノテーション結果の例を



図1 楽譜1 P. Taffanel & Ph. Gaubert : 「17 のメカニスム日課大練習」より  
Fig.1 The score no.1: “17 Grand Exercices” by P. Taffanel & Ph. Gaubert.



図2 楽譜2 P. Taffanel & Ph. Gaubert : 「17 のメカニスム日課大練習」より  
Fig.2 The score no.2: “17 Grand Exercices” by P. Taffanel & Ph. Gaubert.



図3 楽譜3 F.J. Gossec : 「ガボット」より  
Fig.3 The score no.3: “Gavotte” by F.J. Gossec.



図4 楽譜4 J.S. Bach : 「管弦楽組曲第2番よりパディネリ」より  
Fig.4 The score no.4: “Badinerie” by J.S. Bach.

図5, 図6に示す。立ち上がり時刻は人間の手作業によりアノテートされた。その際、人間の聴覚の限界、スペクトログラム読み取りの分解能を考慮し、10ミリ秒単位でアノテートを行った。

図5および図6から分かるように、楽音はきれいな調波構造となっているが、プレス部分は非調波のノイズとなっている。また演奏者によりプレスの音響的特徴が微妙に異なることが分かる。

### 2.3 プレスとテンポの関係

本研究ではプレスがテンポを指示する合図として利用可能であるかどうかを調査するため、上述のように、プレス、冒頭の楽音、1拍後の楽音の立ち上がり時刻をアノテートした。通常、プレスは演奏開始の1拍前を明示するために用いられるため、プレスの立ち上がり時刻から冒頭の音の立ち上がり時刻までの時間長と冒頭の1拍の時間長はほぼ等しいと予想される。そこで、これらの時間長の間の相関関係を分析した<sup>10)</sup>。図7と図8に演奏者ごとの相関を示す。

図7と図8から分かるように、プレスから冒頭までの時間長と冒頭1拍の時間長の間には高い相関(相関係数は図7が0.92, 図8が0.91)が見られる。また、原点を通る回帰直線

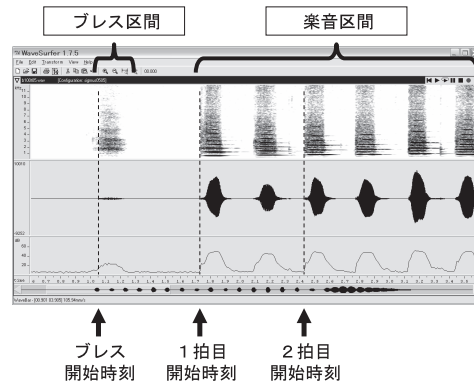


図 5 冒頭のブレスの合図と楽音の音響データとアノテーション結果の例 (楽譜 1, 演奏者 B の演奏). 上からスペクトログラム, 音響波形, パワー波形  
 Fig. 5 Acoustic features of a breath and tones of the beginning part of a piece (score no.1 by performer B) and the result of annotation. From top to bottom, the spectrogram, the acoustic waveform, the power contour.

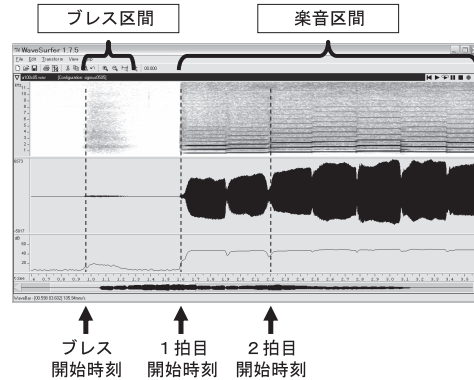


図 6 冒頭のブレスの合図と楽音の音響データとアノテーション結果の例 (楽譜 2, 演奏者 A の演奏). 上からスペクトログラム, 音響波形, パワー波形  
 Fig. 6 Acoustic features of a breath and tones of the beginning part of a piece (score no.2 by performer A) and the result of annotation. From top to bottom, the spectrogram, the acoustic waveform, the power contour.

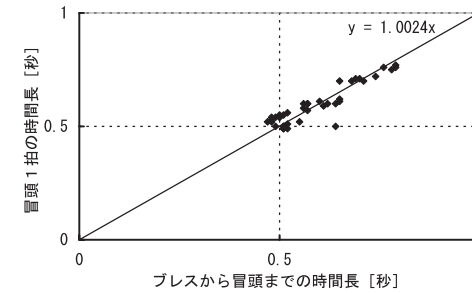


図 7 演奏者 A のブレスから冒頭までの時間長と冒頭 1 拍の時間長との相関  
 Fig. 7 The correlation between the duration from breath to the first note and the duration from the first note to the second note by performer A.

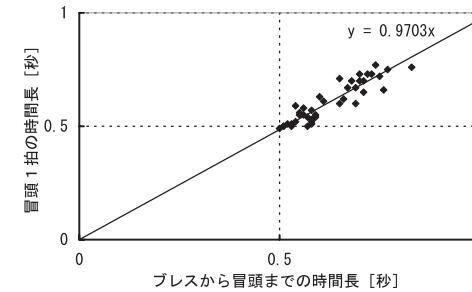


図 8 演奏者 B のブレスから冒頭までの時間長と冒頭 1 拍の時間長との相関  
 Fig. 8 The correlation between the duration from breath to the first note and the duration from the first note to the second note by performer B.

をひいたところ, その傾きは 1.00 (演奏者 A), 0.97 (演奏者 B) となり, ブレスによる合図を検出することにより, 独奏者の演奏開始時刻を推定できる可能性が示されたといえる.

### 3. ブレス検出手法の提案

ブレスの音響的な特徴としては, 以下の 2 点があげられる.

- パワーがある程度の時間持続する.
- 調波構造をほとんど持たない.

本研究ではこれらの特徴を利用して, ブレスのリアルタイム自動検出手法を提案する. 伴奏システムにおけるブレス検出手法の要求仕様をまとめると以下の項目があげられる.

- (1) プレスの開始時刻を検出可能
- (2) プレスと楽音の判別が可能
- (3) 瞬時ノイズに影響を受けない
- (4) 少ない計算量

要求仕様(2)に関しては、曲の冒頭で必ずプレスが行われると仮定すれば、楽音とプレスを区別しなくてもプレスのタイミング検出は可能であるが、プレスなしで演奏が開始された場合への対応(プレスの検出エラーを含む)や、今後、曲中のプレスへの対応のためには楽音とプレスの判別が必須となるため、プレスと楽音の判別を要求仕様とした。なお、プレスなしで演奏が開始された場合のシステムの対応方法は5章で述べる。

要求仕様(4)に関しては、リアルタイムシステムとしての伴奏システム中で利用するため、伴奏処理の妨げとならないよう、計算量の少ない手法により、リアルタイム検出が可能となる必要がある。そこで本システムでは独奏者の演奏認識のために計算されている自己相関値を用いてプレス検出することにより、新たに生じる計算量をできるだけ少なくする手法を検討する。

楽音とプレスを判別する場合、その調波構造の有無を利用して判定することができる。本研究ではなるべく計算量を少なくするため、自己相関値の最大値が自己相関値の中で突出しているかどうかを判別基準として用いる。具体的には、自己相関値をパワーで正規化した後、独奏の対象としている楽器(本システムではフルート)の最低音(約250Hz)と最高音(約2kHz)の間の区間(音域)において最大となる自己相関の値と同音域内の自己相関の平均値との差を求め、それを立ち上がり時刻から累積平均を求めたものを判別パラメータとする。楽音やプレスの場合、ある程度の時間パワーが持続するので、その持続区間において、上記パラメータを計算し、それが閾値以上である場合には調波構造を持っていると考えられるため、楽音と判定し、閾値未満である場合には調波構造をほとんど持っていないと考えられるため、プレスと判定する。

2章で収録した80データに関して、楽音とプレスの場合のそれぞれについて、音の立ち上がり時刻からの時間経過にともなう上記パラメータの時間変化を分析した結果を図9に示す。グラフから、楽音、プレスともに、その立ち上がり付近ではパラメータ値がばらついていて、時間が経過するに従って、プレスと楽音の判別が可能となることが分かる。楽音の場合、その立ち上がり時刻付近では調波構造が安定していないため、上記パラメータではプレスとの判別ができないが、時間が経過するに従って、楽音とプレスのパラメータが明確に分離されていくことが分かる。グラフから、およそ100ミリ秒経過した段階で閾値0.8

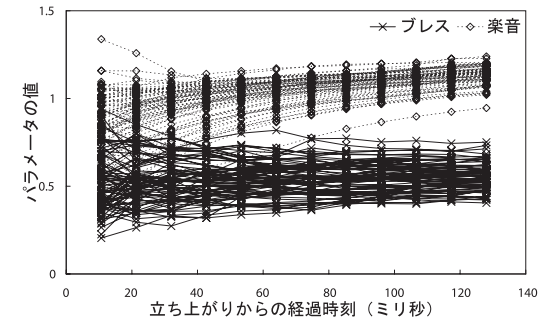


図9 プレス判定用パラメータの時間変化。実線がプレス、点線が楽音の場合

Fig. 9 The transition of parameter for detecting breath. The solid lines mean the cases of breath and the dotted lines mean the cases of tones.

により、楽音とプレスの分離が可能であることが分かる。本システムではこのパラメータを用いてプレス検出を行う。以下、プレス検出アルゴリズムの詳細について述べる。

まず、人間の独奏者との合奏時を想定した場合、音響的な状態は以下のように分類できる。ここで瞬時ノイズは無音状態に含まれるものとする。

- 音が鳴っている状態(有音)
  - 楽器音が鳴っている状態(楽音)
  - プレス中の状態(プレス)
- 無音状態(無音)

プレス検出のアルゴリズムは以下のとおりである。

- (1) 初期状態は無音状態である。
- (2) 音響パワーが閾値  $Th_{power}$  を超えたら(3)へ進む(このとき、有音状態と瞬時ノイズの可能性がある)。
- (3) パワーが閾値  $Th_{power}$  を超えたまま、時間長閾値  $Th_{duration}$  を経過したら有音状態を確定し、(4)へ進む。もし、 $Th_{duration}$  に達する前にパワーが  $Th_{power}$  以下となったら、瞬時ノイズと考え、無音状態として(1)に戻る。
- (4) パワーが閾値  $Th_{power}$  以下となったら(5)へ進む。もし、パワーが閾値  $Th_{power}$  を超えたまま、時間長閾値  $Th_{decision} (> Th_{duration})$  を経過したら、(6)へ進む。
- (5) パワーが閾値  $Th_{power}$  を超えてから、 $Th_{power}$  以下となるまでの区間に対し、上述のパラメータによる楽音/プレス判定を行い、検出処理を終了する。

- (6) パワーが閾値  $Th_{power}$  を超えた時刻から時間長閾値  $Th_{decision}$  までの区間に対し、上述のパラメータによる楽音/プレス判定を行い、検出処理を終了する。

上記のアルゴリズムにおいて、もし仮に(4)の時点で時間長閾値  $Th_{decision}$  の判定を行わない場合、有音区間が長い(楽音だと数秒に及ぶこともある)とその有音区間が終了するまで楽音/プレスの判定を行うことができなくなってしまうため、遅くとも有音区間開始後  $Th_{decision}$  経過した段階で楽音/プレス判定を行うことを意味している。このようにすることで、もし仮にプレスの時間長が  $Th_{decision}$  以上に長い場合であっても、 $Th_{decision}$  経過した段階でプレスと判定できるため、プレスの終了を待たずに演奏開始のスケジューリング処理を行える。すなわち、リアルタイム性を損うことなく、プレスによる合図で演奏を開始することが可能となる。なお、演奏開始のスケジューリング処理では、プレスのパワーが閾値  $Th_{power}$  を超えた時刻をプレスの開始時刻とし、そこから1拍後(その楽曲のテンポから推定)を演奏の開始時刻として伴奏のスケジューリングを行う。

現在のシステムでは  $Th_{duration}$  は80ミリ秒としている。これは収録したプレスの最短時間を参考に決定した。また、上述の図9の結果を参考にして、 $Th_{decision}$  は120ミリ秒、楽音/プレス判定パラメータの閾値は0.8とした。また、有音区間を判定する閾値  $Th_{power}$  については、システムが置かれている実環境の雑音を考慮するため、システム起動時に周辺のノイズのパワーを計測し、その値にマージンを付加することにより自動的に決定している。

上記のアルゴリズムと閾値設定により、2章で収録した80データに対するプレスと楽音の判別は100%成功した。また、後述の評価実験においても、プレスと楽音の判別はすべて成功した。プレスの開始時刻の検出精度に関して、2章で収録した80データに対して評価実験を行ったところ、上記アルゴリズムによるプレス開始時刻の推定値と実測値(手作業でアノテートされた時刻)の誤差の絶対値の平均を求めたところ、演奏者Aに対して11ミリ秒、演奏者Bに対して9ミリ秒、両演奏者の全80データの平均で10ミリ秒となった。これらの結果から本研究の提案手法によるプレスの検出精度は伴奏システムに利用するのに十分であると考えられる。

#### 4. 提案システムの評価実験

提案手法を伴奏システムに実装し、提案システムの評価実験を行った<sup>11)</sup>。

##### 4.1 実験の目的

人間の演奏者がプレスによる合図に関して、どの程度までのずれなら許容できるのか、という時間範囲を調べ、本提案システムの性能が人間の演奏者にどの程度許容されるのかにつ



図 10 楽譜 5 M. Ravel:「亡き王女のためのパヴァーヌ」より  
Fig. 10 The score no.5: "Pavane" by M. Ravel.



図 11 楽譜 6 G. Fauré:「小舟にて」より  
Fig. 11 The score no.6: "En Bateau" by G. Fauré.

いて、その有効性を評価することを目的とする。

##### 4.2 実験参加者

実験に参加した演奏者は音楽大学フルート専攻の卒業生2名(うち1名は2章と同じ実験参加者である。以下、演奏者C、演奏者Dと呼ぶ)とアマチュアのフルート演奏家(2章と同じ実験参加者である。以下、演奏者Eと呼ぶ)の3名である。

##### 4.3 実験に用いた曲

データ収録には実際の曲だけでなく、単調なフレーズからなる練習曲も用いた。収録に用いた楽譜を図10、図11、図12、図13、図14、図15に示す。各楽曲の設定テンポは楽譜5(図10)から楽譜9(図14)まで順に52, 63, 92, 116, 132(それぞれ1分あたりの拍数。単位はbeat/minute)とした。また、楽譜10(図15)に関しては80, 100, 120の3種類のテンポを設定した。これらの楽曲は演奏者に馴染みがあり、演奏が容易にでき、また、幅広いテンポを網羅するように選曲された。

##### 4.4 実験条件

本実験では人間の演奏者がプレスによる合図に関して、どの程度までのずれなら許容できるのか、という時間範囲を調べるため、上記の各楽曲に対し、プレスから推定されるタイミ



図 12 楽譜 7 A. Dvořák : 「ユモレスク」より  
Fig. 12 The score no.7: “Humoresque” by A. Dvořák.



図 13 楽譜 8 F.J. Gossec : 「ガボット」より  
Fig. 13 The score no.8: “Gavotte” by F.J. Gossec.



図 14 楽譜 9 J.S. Bach : 「管弦楽組曲第 2 番よりパディネリ」より  
Fig. 14 The score no.9: “Badinerie” by J.S. Bach.



図 15 楽譜 10 P. Taffanel & Ph. Gaubert : 「17 のメカニスム日課大練習」より  
Fig. 15 The score no.10: “17 Grand Exercices” by P. Taffanel & Ph. Gaubert.

ングどおりに演奏を開始するものに加え、プレスから推測されるタイミングから故意にずらして演奏を開始するシステムを用意した。

具体的には、各楽曲に対し、プレスから伴奏の演奏開始までの時間として、下記の 5 種類を設定した。

- (a) 1 拍より 1/4 拍短い
- (b) 1 拍より 1/8 拍短い
- (c) 1 拍と等しい
- (d) 1 拍より 1/8 拍長い
- (e) 1 拍より 1/4 拍長い

例外として、楽譜 6 ( 図 11 ) については楽曲が 3 連符系であるため、1/4 のかわりに 1/6、1/8 のかわりに 1/12 とした。なお、これらの設定は、あらかじめ予備検討を行い、伴奏の演奏開始が早すぎると感じられるところから遅すぎると感じられるところを含むように設定した。

#### 4.5 実験手続き

上述のようにのべ 8 種類の楽曲 ( 楽譜 10 ( 図 15 ) は 3 種類のテンポがある ) に対し、プレスに対する反応が異なる 5 種類のパターンを用意し、各パターンをそれぞれ 2 回実験に用いた。結果として、演奏者 1 名に対し、1 曲あたり 10 回、合計 80 回の演奏を行ってもらった。各楽譜の順番は、直前の曲のテンポの影響を受けにくいよう、直前の曲となるべくテンポが異なる曲が並ぶように考慮して決定した。また曲中の 10 種類のパターンはランダムに提示された。実験の曲順、データ順を以下に示す。楽譜番号の後ろの括弧内の数字はテンポである。全演奏者に対し、下記の順番で実験を行った。

- 楽譜 10 (80) : e, d, e, c, b, d, c, a, b, a
- 楽譜 8 (116) : c, b, d, c, d, e, a, b, e, a
- 楽譜 6 (63) : e, d, a, c, b, d, c, e, b, a
- 楽譜 10 (100) : d, c, b, a, b, e, a, c, e, d
- 楽譜 5 (52) : c, d, e, b, d, c, e, a, b, a
- 楽譜 9 (132) : b, e, b, d, a, c, e, c, d, a
- 楽譜 7 (92) : b, e, d, b, a, e, d, c, a, c
- 楽譜 10 (120) : c, d, e, b, d, c, e, a, b, a

各演奏において、まず演奏直前に演奏者にメトロノームで指定のテンポを呈示した。メトロノームを停止した後、演奏者の自由なタイミングでプレスの合図をして演奏を開始しても



らった。演奏直後に伴奏が独奏者に対して適切なタイミングで演奏を開始したと感じられたかどうかについて、5段階で主観的に評価してもらった。各評価は以下の5段階から選択してもらった。

- 早すぎる (-2)
- 早いが許容できる (-1)
- 適切である (0)
- 遅いが許容できる (1)
- 遅すぎる (2)

これらは人間の演奏者がプレスによる合図に関して、どの程度までのずれなら許容できるのか、という時間範囲を調べることを目的に設定された。また、伴奏システムとしては上記の評価で -1 ~ 1 の範囲内であればシステムの性能として十分であると考えられる。

#### 4.6 実験環境

伴奏システムは Genuine Intel(R) CPU T2600 @ 2.16 GHz ( 1.00 GB RAM ), Microsoft Windows XP Professional 2002 ( SP2 ) が動作するコンピュータ上で Microsoft Visual C++ .net 2003 を用いて実装された。また、音源として Edirol SD-20、オーディオインタフェースとして Edirol UA-20 を用いた。実験において、演奏者はヘッドフォンにより、伴奏システムの音を聴きながら演奏を行った。

#### 4.7 実験結果 1 (独奏者の許容限界)

各演奏者が伴奏システムの演奏開始タイミングに対して、どのように評価したのかについて分析するため、人間の演奏者が演奏を開始した時刻 (冒頭の楽音の立ち上がり時刻) と実際に伴奏システムが演奏を開始した時刻 (冒頭の楽音の立ち上がり時刻) の時間的なずれと人間の演奏者の評価結果との関係を図 16 に示す。横軸が時間的なずれで、演奏者の演奏よりも伴奏のタイミングが遅い場合を正としてある。グラフは左上からテンポの遅い順に並べてある。

演奏者ごと、楽曲のテンポごとに傾向が若干異なるが、おおむね右肩上がりの分布となっており、演奏のずれが評価値と高い相関関係にあることが分かる。テンポの違いに関しては、テンポが遅い曲の方がその傾きがゆるやかであり、逆にテンポが速い曲ではずれに対して評価が厳しくなっている (時間的なずれに対する許容度が低い) ことが分かる。

また、これらのグラフにおいて、ずれの値が正か負かによって、その傾向が大きく異なることが観察できる。そこで人間の演奏者と伴奏システムがどのようなタイミングで演奏を開始したのかを詳細に分析するため、プレスから推定された演奏開始時刻を 0 とした場合の

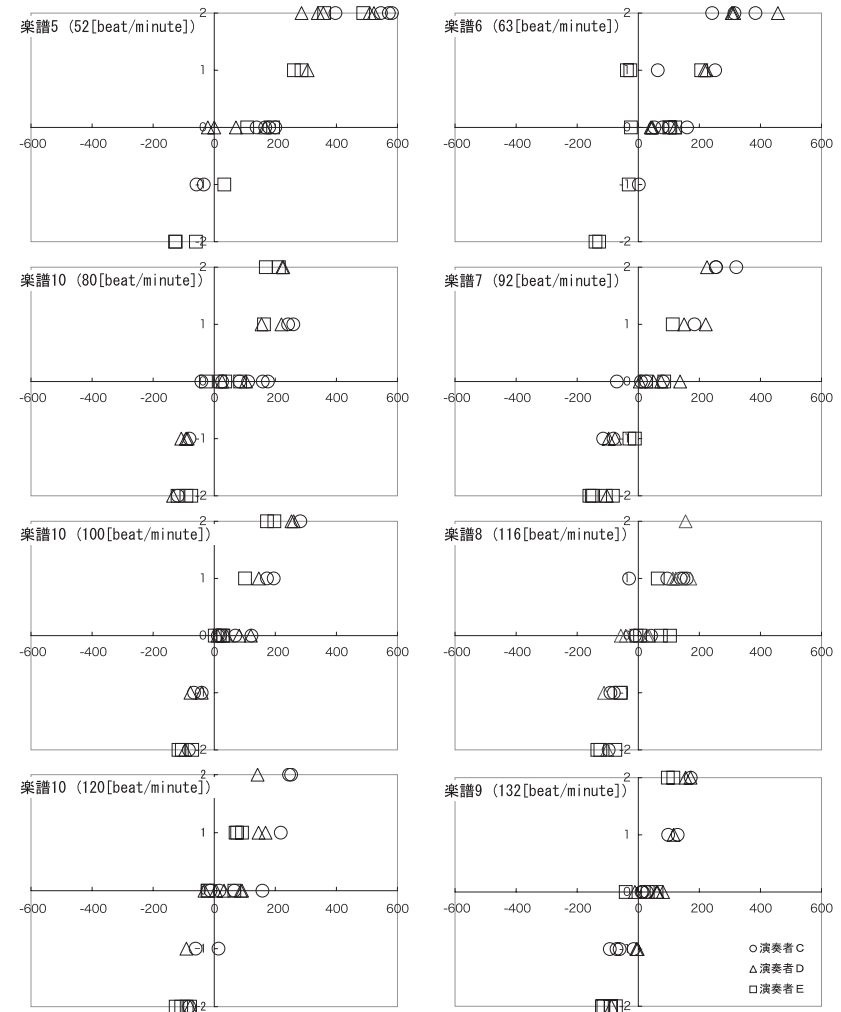


図 16 開始時刻のずれ (横軸) とそのときの評価値 (縦軸) の関係。横軸の単位はミリ秒で伴奏が独奏よりも遅い場合を正とする

Fig. 16 The relation between timing difference in milliseconds and performers' rating results. When the system was delayed, the timing difference is positive.

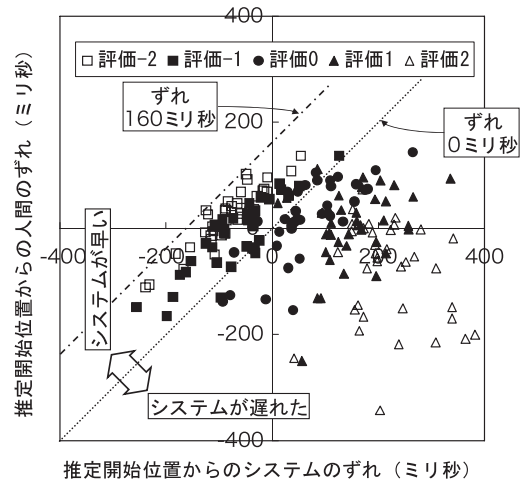


図 17 システムと人間の演奏のずれの分布。原点はプレスから推定された演奏開始時刻。横軸はシステムのずれ (ミリ秒), 縦軸は人間のずれ (ミリ秒) で推定時刻よりも遅れた場合を正とする

Fig. 17 The relation of timing between the system and the human performers. The horizontal axis is the onset time in milliseconds of the system and the vertical axis is the onset time of human performers. The zero in the both axes means the onset time estimated by breath.

人間と伴奏システムのずれの分布を図 17 に示す。横軸は伴奏システムが実際に演奏を開始した時刻であり、正の場合は推定開始時刻よりも遅れた場合を表し、負の場合は推定開始時刻よりも早い場合を表す。縦軸は人間の演奏者が演奏を開始した時刻である。このグラフ上では原点を通り、傾き 1 の直線 (図中に点線で示されている) が人間とシステムが同時に演奏を開始した状況を表しており、その線より右下はシステムが人間よりも演奏開始が遅れる状況、左上はシステムが人間よりも演奏開始が早い状況を表す。システムの演奏は意図的に開始タイミングをずらしているため、横軸方向は値がばらばらになっていることが分かる。一方、人間の演奏はプレスによる合図に対し、縦軸方向おおむね  $\pm 200$  ミリ秒の範囲内におさまっているが、ここで興味深いのは、人間がシステムよりも 160 ミリ秒 (図中に点線で示す) 以上遅れたデータが存在しないことである。このことからシステムが人間の演奏者よりも先に演奏を開始した場合には、人間の演奏者はプレスによる合図で示した演奏開始時刻よりも演奏開始時刻を早めて演奏を開始していると考えられる。しかしながら、そのような状況では伴奏システムの演奏開始時刻は演奏者がプレスによる合図で示した時刻よりも大幅に早かった

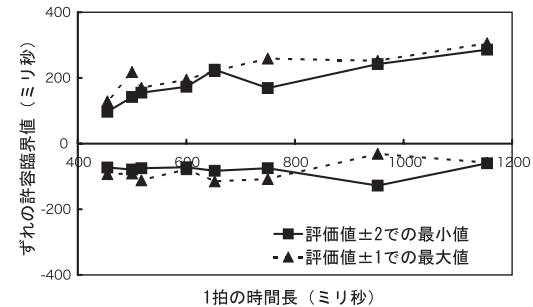


図 18 ずれの許容範囲。縦軸は許容範囲と考えられる時間的ずれ (単位はミリ秒) を表す。評価値 2 および -2 に関して、各ケースの中で最も厳しく評価したものを実線、評価値 1 および -1 に関して、各ケースの中で最も甘く評価したものを点線で示す。横軸はテンポから逆算した 1 拍の時間長 (単位はミリ秒)

Fig. 18 The tolerated range of timing difference. The vertical axis is timing difference in milliseconds and the horizontal axis is the duration of an unit beat in milliseconds (the reciprocal of tempo). The solid lines show the most strict case about the rating 2 and -2 and the dotted lines show the most loose case about the rating 1 and -1.

ため、評価値が下がっているのではないかと考察できる。その結果、図 16 のグラフは左右対称にならず、負の値が 160 ミリ秒程度で頭打ちになっていると考えられる。

システム構築の観点からは -1 ~ 1 の評価は許容可能であり、-2 と 2 の評価値を避ける必要があると考えられる。そこで今回の実験結果について、-2 および 2 と評価されたケースに対し、各曲中で最も厳しく評価された時間的ずれを許容臨界値と考え、その臨界値と楽曲のテンポとの関係を図 18 の実線に示す。また、参考として -1 と 1 と評価されたケースに対して、各曲中で最も甘く評価された時間的ずれを図 18 の点線で示す。点線がおおむね実線の外側に出ていることから、実線で示される許容範囲は演奏者によっては -1 または 1 の評価を受けることもありうるが、それ以上に厳しい評価基準での許容範囲を示していると考えられる。

図 18 からシステムの演奏開始が遅れる場合にはテンポが速いほど許容臨界値が短くなるのが分かる。一方、システムの演奏開始が早い場合には許容臨界値がおおむね 70 ~ 80 ミリ秒で一定であることが推測される。テンポと許容臨界値の関係を明らかにするため、図 18 の縦軸を 1 拍の時間長に対する比率で表現したグラフを図 19 に示す。図 19 を見るとシステムの演奏開始が遅れる場合、その許容臨界はテンポに関係なく、ほぼ 0.25 であることが分かる。これは 1 拍の 1/4 に相当し、四分の四拍子であれば十六分音符 1 つ分に相当する。

これらの結果から、人間の演奏者の許容臨界値はシステムと人間の演奏開始の前後関係に



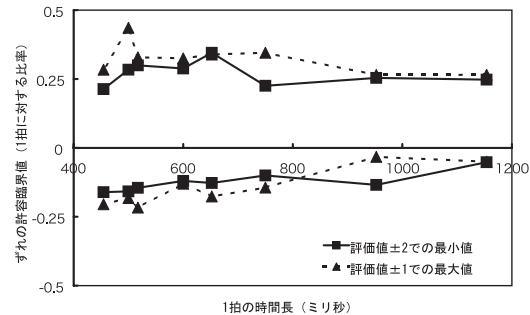


図 19 図 18 の縦軸を 1 拍の長さに対する比で表したグラフ

Fig. 19 The tolerated range of timing difference. The vertical axis is the ratio of timing difference to the duration of a beat and the other features are the same as Fig. 18.

よって異なることが推測される。すなわち、システムが人間よりも遅れる場合、1/4 拍程度までのずれが許容されるのに対し、システムが人間よりも先に演奏を開始してしまった場合、70~80 ミリ秒までのずれしか許容されないということを示している。

#### 4.8 実験結果 2 (システムの性能評価)

本実験ではシステムは意図的にずれたタイミングで演奏を開始したケースがほとんどであるが、もし仮にシステムが推定どおりのタイミングで伴奏を開始したと仮定した場合、人間の演奏家がどのような判断を下したと推定されるかについて検討を行う。実験データは前節と同じデータを用いた (演奏者 1 名あたり 80 データ、全 240 データ)。

そこでまずシステムがプレスを検出してから計画した時刻に正しく演奏を開始できるかどうかについて評価を行う。図 20 にシステムがプレスから演奏開始を計画した時刻 (意図的に開始時刻をずらした場合にはそのずらした時刻) とシステムが実際に演奏を開始した時刻とのずれの頻度を示す。なお、システムが実際に演奏を開始した時刻とは伴奏システムの冒頭部の演奏のパワー情報から閾値処理により求めた。

図 20 からシステムの演奏開始のタイミングが全体的に予定時刻よりも遅れており、また、そのずれについてもばらついていることが分かる。これは OS の負荷や音源の発音遅延等の理由でシステムが予定した時刻どおりに発音がなされなかった結果であると考えられる。

図 20 のデータと上述の人間の演奏家の許容臨界値 (図 18 の最も小さい値として -60 ミリ秒から 97 ミリ秒の間とする) とを比較してみると、全体の 85.0% が許容臨界値内に含まれており、比較的高い精度でシステムが演奏を開始できることが示された。

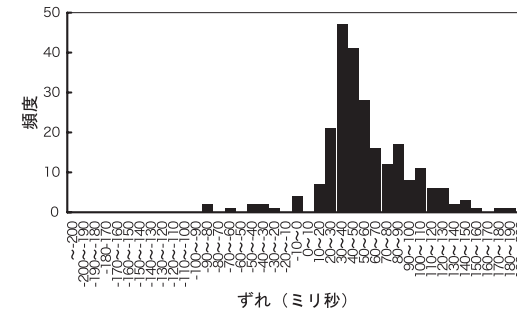


図 20 システムの演奏のずれのヒストグラム

Fig. 20 The histogram of timing error at the onset of the first note by the system.

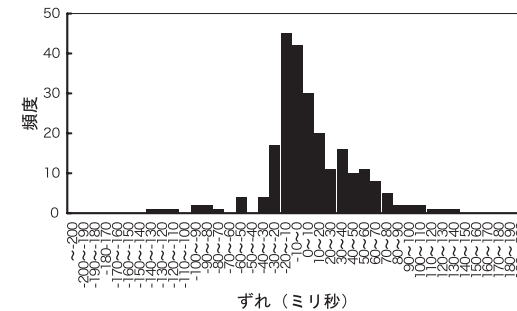


図 21 中央値を用いて補正したシステムの演奏のずれのヒストグラム

Fig. 21 The histogram of timing error at the onset of the first note by the system after the compensation for the constant delay.

さらに図 20 の結果から、全体的に伴奏の演奏開始が遅延していることが分かるため、これらのデータの中央値 (47.9 ミリ秒) の分だけ早く出力するようにシステムを修正することにより精度の向上が期待できる。図 21 に上記補正を行った仮想実験結果としての頻度分布を示す。このようにシステムを修正したと仮定した場合、全体の 94.6% が上記の許容臨界値内に含まれ、推定ではあるが、非常に高い精度でシステムが演奏を開始できる可能性が示されたといえる。

## 5. 合図のプレスがない場合への対応

本研究で提案したプレス検出手法は楽音とプレスの判別を行っている．これはプレスなしで演奏が開始された場合（プレスの検出エラーを含む）や，今後，曲中のプレスへ対応するために必要となる．本システムは曲の冒頭でプレスによる合図で演奏を開始するが，もし，独奏者がプレスなしで演奏を開始した場合や，システムがプレスの検出に失敗した場合には，演奏開始の楽音の情報を用いて演奏を開始することが可能である．具体的には，検出された楽音の開始時刻を用いて伴奏スケジューリングを開始する．ただし，楽音の開始時刻の検出には 120 ミリ秒を要するため，冒頭で同時に鳴るべき音に関しては若干の遅延が生じるが，それ以降の音のスケジューリングは独奏の開始時刻に基づいて行われるため，その遅延は解消されることとなる．

本システムにおいて，プレスがない場合の対応はエラー対策であるため，とくに評価実験等は行っていないが，プレスなしで演奏を開始しても伴奏システムが演奏を開始することが可能であり，冒頭の音以降では独奏者の演奏開始タイミングにあわせて伴奏制御を行っていることが確認されている．

## 6. おわりに

本研究では伴奏システムが人間の演奏者のプレスによる合図を検出し，同時に演奏を開始できる手法を提案・実装した．評価実験の結果，本システムのプレスにタイミングを合わせる性能は人間の演奏者のずれの許容範囲に対し，十分な精度を有していることが示された．

今後の検討課題として，上述の補正をシステムに実装すること，また，演奏のずれの原因を追求し，可能であれば，より正確に演奏を出力できるようにシステムを改良したいと考えている．また，冒頭以外の箇所においてもプレスによる合図は重要であるため，そのような利用方法も検討していきたい．その際，楽曲途中のプレスをどのように伴奏制御に利用できるのかについて，実際のデータ収録等も含めた検討が必要になると考えられる．本研究では合図としてプレス情報のみを利用したが，人間同士の合奏では体の動きも非常に重要な合図となるため，体の動きとプレスの両方を用いた手法も検討したいと考えている．

謝辞 本研究を進めるにあたり，当研究室の卒業生・学生の飯塚涼子氏，増田香織氏，大三川晴香氏，ならびに，実験に参加いただいた演奏者の方々に謹んで感謝の意を表す．

## 参考文献

- 1) 堀内靖雄：自動伴奏，コンピュータと音楽の世界—基礎からフロンティアまで，長嶋洋一，橋本周司，平賀 譲，平田圭二（編），pp.252–269，共立出版（1998）.
- 2) Dannenberg, R.B.: An On-Line Algorithm for Real-Time Accompaniment, *Proc. International Computer Music Conference*, pp.193–198 (1984).
- 3) Vercoe, B.: The synthetic performer in the context of live performance, *Proc. International Computer Music Conference*, pp.199–200 (1984).
- 4) Raphael, C.: Music plus one: A system for flexible and expressive musical accompaniment proceed, *Proc. International Computer Music Conference* (2001).
- 5) 武田晴登，西本卓也，嵯峨山茂樹：HMM による MIDI 演奏の楽譜追跡と自動伴奏，情報処理学会研究報告，Vol.2006, No.90, pp.109–116 (2006).
- 6) Ruinskiy, D. and Lavner, Y.: An Effective Algorithm for Automatic Detection and Exact Demarcation of Breath Sounds in Speech and Song Signals, *IEEE Trans. Audio, Speech and Language Processing*, Vol.15, No.3, pp.838–850 (2007).
- 7) 中野倫靖，緒方 淳，後藤真孝，平賀 譲：無伴奏歌唱におけるプレスの音響特性と自動検出，日本音響学会 2008 年春季研究発表会講演論文集，pp.265–268 (2008).
- 8) 中村敏江：音楽における「間」と呼吸について，日本音響学会音楽音響研究会資料 MA94-16, pp.19–26 (1994).
- 9) <http://www.speech.kth.se/wavesurfer/>
- 10) 堀内靖雄，飯塚涼子，西田昌史，市川 薫：伴奏システムでのプレス情報利用に関する検討，情報処理学会研究報告，Vol.2005, No.45, pp.13–18 (2005).
- 11) 堀内靖雄，増田香織，西田昌史，市川 薫：プレスの合図を認識する伴奏システムの実装と評価，情報処理学会研究報告，Vol.2007, No.81, pp.1–6 (2007).

(平成 20 年 5 月 14 日受付)

(平成 20 年 11 月 5 日採録)



堀内 靖雄（正会員）

1967年生。1990年東京工業大学工学部情報工学科卒業。1995年同大学大学院理工学研究科情報工学専攻博士課程修了。博士（工学）。1995年千葉大学工学部助手。2000年千葉大学大学院自然科学研究科助手。2002年同助教授。2007年同大学院融合科学研究科准教授。音楽情報処理，音声言語処理，福祉情報工学に関する研究に従事。1992情報処理学会全国大会奨励賞，1995人工知能学会研究奨励賞，1996人工知能学会全国大会優秀論文賞。人工知能学会，日本音響学会，ヒューマンインタフェース学会，ソフトウェア科学会，社会言語科学会，日本手話学会各会員。



西田 昌史（正会員）

1997年龍谷大理工学部電子情報学科卒業。1999年同大学大学院修士課程修了。2002年同大学院博士後期課程修了。博士（工学）。2002年4月～2003年6月科学技術振興事業団さきがけ研究21「協調と制御」領域博士研究員。2003年7月～2007年3月千葉大学大学院自然科学研究科助手。2007年4月より同大学院融合科学研究科助教。音声情報処理，話者認識，音声認識，ヒューマンインタフェース，福祉情報工学に関する研究に従事。電子情報通信学会，日本音響学会，人工知能学会各会員。



市川 薫（正会員）

1964年慶應義塾大学工学部卒業。日立中央研究所を経て1992年千葉大学工学部教授，1998年同大学大学院自然科学研究科教授。現在，早稲田大学人間科学学術院客員教授，千葉大学名誉教授。工学博士。信学会フェロー。音声対話理解，手話，指点字，福祉情報機器等に取り組み，特に対話言語のプロソディ情報に関心を持つ。最近は言語獲得や発達障害の課題にも眼を向けている。信学会理事，音声研委員長，手話工学研副委員長，福祉情報工学研委員長，音響学会評議員，人工知能学会理事，言語・音声理解と対話処理研主査，JEITA アクセシビリティ標準化対応専門委員長等を歴任。他に電子情報通信学会，IEEE，ISCA，言語処理学会，HI学会，音声言語医学会，手話学会等の会員。昭和64年信学会論文賞，平成8年人工知能学会研究奨励賞，平成17年総務大臣賞等受賞。