

## 映像と音の呈示に対する応答のモデリング

### Human response modeling when images and sounds are shown simultaneously

神谷 達夫

#### 要旨

視覚や聴覚の相互作用の存在はよく知られている。複数の感覚に刺激を与えることによって応答時間が短縮される感覚モダリティ間の促進効果は機器操作に重要な影響を及ぼす。しかし、映像と音を同時に発生させ人の反応を測定する装置を構成するためには正確な同期が必要となり、実験を困難なものにしている。本論文では、映像による視覚刺激と聴覚刺激を同期させて発生することのできる測定システムを構築し、視覚刺激と聴覚刺激の指示により演奏する者の演奏精度の変動を AR モデルに同定した。

本システムは発生する視覚刺激と聴覚刺激を容易に変更でき、刺激に応答して動作する人の動作メカニズムを解明するために有用である。

キーワード: 視覚刺激, 聴覚刺激, 演奏, AR モデル

Keywords: Visual stimulus, Auditory stimulus, Musical performance, AR model

#### 1. はじめに

近年、ビデオゲームに代表される映像と音の同時呈示といったマルチモーダル情報をユーザに呈示するインタラクションシステムが急速に普及してきている。このようなシステムは、ユーザの複数の感覚モダリティに対しての刺激を発生するため、感覚モダリティ間の相互作用の影響を受けることになる。発光や音の発生に対する人の応答時間が、発光と同時に音も発生させるとそれぞれを個別に与えた場合よりも短くなるという現象<sup>(1)</sup>が感覚間の促進効果として古くから知られている。このような効果が存在することは、ビデオゲーム機器のように複雑な表示装置を持ったシステムから発せられる感覚に対する刺激を設計する際に、個別の感覚モダリティに対して発生させる刺激の線形結合として考えることができないことを示している。つまり、高度なインタラクションシステムの発生する刺激

を設計するためには、複数の感覚モダリティに対しての人間の応答に対するモデリングが必要である。

感覚モダリティ間の相互作用を示す現象は幾つか知られている。視覚と聴覚の相互作用の場合、口の動きの映像が音声の聞こえ方に影響する McGurk 効果や音源定位における腹話術効果はよく知られていて、視覚が聴覚に影響力を与える例である<sup>(2)</sup>。さらに、複数の感覚モダリティによって応答時間の短縮される感覚モダリティ間の促進効果が確認されている<sup>(3)</sup>。また、同時に与えられる異種感覚刺激に対して、感覚の共通する属性が同じ方向に変化する共鳴現象も報告されており<sup>(4)</sup>、オーディオ・ビジュアルメディアにおいて画面サイズを大きくすると聴感上の迫力を増大させるという効果<sup>(5)</sup>は感覚モダリティ間共鳴現象の例である。さらに、聴覚の方が時間分解能に優れ、時間領域の知覚においては聴覚の方が視覚に対して優位に働くという報告もある<sup>(6)</sup>。

感覚間促進効果により反応時間が短縮される原因は、刺激のエネルギー加重の効果<sup>(7)</sup>と刺激が複数存在すると検知される可能性が増えることにより反応時間が短縮する統計的促進モデルが<sup>(8)</sup>考えられる。しかし、Gielen<sup>(9)</sup>の実験では視覚刺激と聴覚刺激による反応を腕と手首の動きを測定することにより、統計的促進モデルで予測される反応時間よりも大きく反応時間が短縮することが報告されている。これは、統計的促進モデル以外にも何らかの感覚間促進効果が存在することを示しており、感覚間促進効果の原因にはエネルギーの加重効果と反応準備時間の増加<sup>(10)</sup>の両方<sup>(11)</sup>が原因となっていると考えられている。Schmidt<sup>(12)</sup>は Gielen と同様の装置により、感覚間促進効果の原因を検討していて、複雑な反応動作であれば加重効果の検出されないことを報告している。つまり、複雑な反応動作の場合、反応準備時間の増加が感覚間促進効果の原因である可能性が高いとされている。

これまでの複数の感覚モダリティに対する刺激への応答時間や感覚の時間分解能測定の実験は、単純な感覚刺激呈示装置(視覚に対してはランプ点灯、聴覚に対しては発音など)を用いた実験によりなされており<sup>(1,5,6,8,10-15)</sup>、この成果を直接ビデオゲーム機のようなシステム的设计に応用することはできない。このような用途にはパーソナルコンピュータで発生させた映像を用い、MIDI 音源などの出力やデジタル録音された音をそれに同期させて出力する方法<sup>(16)</sup>を用いる場合があるが、実験システムの遅延を考慮する必要がある。特にパーソナルコンピュータから発生されるアニメーションやビデオ映像を用いる場合、パーソナルコンピュータに採用されているオペレーティングシステムがリアルタイム性を考慮していないため、レイテンシが安定せず、10[ms]程度のレイテンシ変動がしばしば発生する。しかし、人の応答の変化はランプ点灯実験などから、数ミリ秒の変化を検出する必要があることが知られており、測定装置には1[ms]程度の時間分解能が必要となる。

本論文では、実験に必要な時間分解能を得るため、映像と音を同期させて被験者に呈示した場合の応答を測定するシステムを構築した。さらに、このシステムを使用することにより、映像による視覚刺激と聴覚刺激を用いた動作指示に従って演奏した被験者の演奏精度の変動をARモデルに同定した。

### 3. 測定システムと測定精度

#### 3.1 測定システムの概要

本測定システムはデジタル VTR と視覚刺激表示のための CRT、聴覚刺激発生のためのヘッドフォン、実験データを記録するためのデータレコーダ、垂直同期信号変換機(VSC: Vertical Synchronous signal Converter)で構成されている(Fig.1)。視覚刺激と聴覚刺激はコンピュータ上で合成された後、ビデオテープに格納される。実験時には、このビデオテープを再生することにより刺激が発生される。使用した CRT は 28 型である。

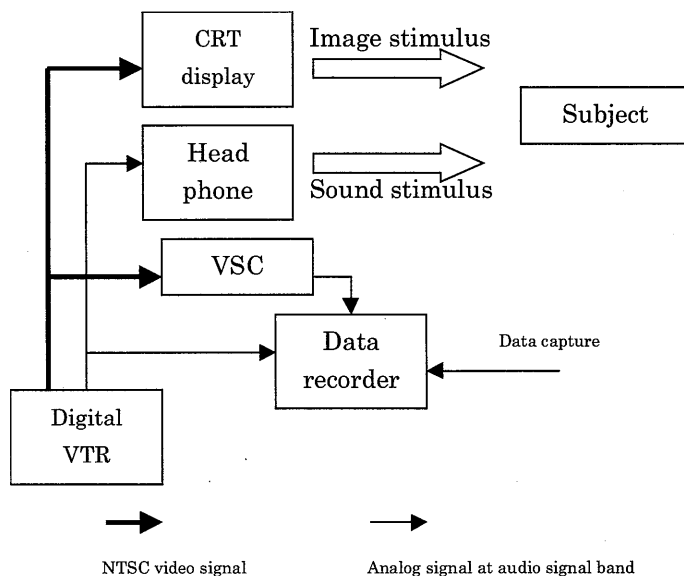


Fig. 1 Arrangement of measurement system

デジタル VTR は映像を 1 秒間に 29.97 枚表示するドロップフレーム映像として再生するため、映像を作成する際には、1 画像の表示時間が  $1/29.97[s]$  であるとして処理される。デジタル VTR の出力する NTSC 信号はインターレース方式であり、偶数フィールドと奇数フィールドの出力により、1 枚の画像が構成される。本システムでは画像表示開始の基準として偶数フィールド表示の開始を採用した。

市販のアニメーション作成ソフトウェアを用いて生成されたアニメーション映像は 1 秒間に 30 枚の静止画を使用するノンドロップフレームで映像が作成され、音付き映像に含まれる映像と音の同期に異常の発生する場合がある。しかし、本システムの場合、静止画と音の PCM データを音声サンパ

リング周波数単位で同期させているため、デジタルVTRから発せられる映像と音のずれは1フィールドの表示時間より大きくなることはなく、現実にはさらに小さなずれ量となる。

VSCはデジタルVTRから発せられるNTSC映像信号の垂直同期信号を検出し、垂直同期信号に同期した繰り返し波形を発生する。この繰り返し信号波形は $1/29.97$ [s]を1周期とする波形であり、その信号レベルにより偶数フィールドか奇数フィールドかを検出することができる。また、VSCの出力波形は可聴周波数に帯域制限されおり、アナログ信号としてデータレコーダで記録することができる。

### 3.2 映像と音の同期

本システムにおいて視覚刺激は画像を連続に表示することにより、映像として呈示される。視覚刺激が $n$ 枚の画像により構成されているとすると、視覚刺激 $\mathbf{G}$ は素画像 $g_i$ の順列として式(1)のように表現できる。

$$\mathbf{G} = (g_i) = (g_1, g_2, \dots, g_n) \\ i \in \{1, 2, 3, \dots, n\} \quad (1)$$

本測定システムでは素画像 $g_i$ をコンピュータ上の演算により生成するため、素画像 $g_i$ を変更することにより、各種視覚刺激を呈示することができる。

聴覚刺激は音声PCMデータをD/A変換し再生することにより、音として呈示される。本測定システムにおいて、聴覚刺激は素画像の表示期間と同じ単位で管理されている。視覚刺激を構成する素画像数と同じ長さの聴覚刺激を $\mathbf{S}$ とすると、 $\mathbf{S}$ は素画像 $g_i$ の表示期間と同じ期間の音となる素音 $s_i$ の順列として式(2)のように表現できる。

$$\mathbf{S} = (s_1, s_2, \dots, s_n) \\ i \in \{1, 2, \dots, n\} \quad (2)$$

しかし、素画像の表示期間は $1/29.97$ [s]であり、素画像表示時間単位での聴覚刺激発生時間の制御では測定システムに必要な分解能である $1$ [ms]に対し十分でない。本システムでは、聴覚刺激にオフセット音を加えることにより、素画像表示時間以内の聴覚刺激発生制御に対応した。オフセット音は無音のPCM信号であり、オフセット時間を $t$ 、聴覚刺激のサンプリング周波数を $f_s$ とすると、オフセットデータ長は $n_o$ となり、オフセット音 $s_o$ は式(3)のように表現される。本測定システムにおいて、サンプリング周波数 $f_s$ は $48$  [kHz]である。

$$\begin{aligned}
s_o &= (o_1, o_2, \dots, o_i, \dots, o_{n_o}) \\
i &\in \{1, 2, \dots, n_o\} \\
o_i &= 0 \\
n_o &= t f_s \\
0 \leq t &< \frac{1}{29.97} [s]
\end{aligned} \tag{3}$$

実際に本測定システムから呈示される聴覚刺激  $\mathbf{S}'$  は  $\mathbf{S}$  にオフセットデータを付加した素音の順列となり、式(4)のように表現できる。

$$\begin{aligned}
\mathbf{S}' &= (s_o, s_1, s_2, \dots, s_n) \\
i &\in \{1, 2, \dots, n\}
\end{aligned} \tag{4}$$

聴覚刺激はオフセット時間により発生時間を制御できるため、聴覚刺激の発生が音声サンプリング時間単位で制御できることになる。被験者に呈示する刺激は聴覚刺激と視覚刺激を合成した複合刺激となる。視覚刺激と聴覚刺激の時間軸を合わせて合成する演算を とすると、複合刺激  $\mathbf{C}$  は  $\mathbf{G}$  と  $\mathbf{S}'$  の合成となり、式(5)のように表現される。

$$\mathbf{C} = \mathbf{G} \oplus \mathbf{S}' \tag{5}$$

デジタル VTR より出力される複合刺激は同期信号を伴った映像と音声で構成されており、その映像信号には同期信号が含まれる。本測定システムでは、VSC により映像信号中に含まれる垂直同期信号が分離され、方形波として分離された垂直同期信号は音声帯域に制限される。この音声帯域に制限された垂直同期信号は聴覚刺激信号と同時にデータレコーダに記録される。映像と音声の同期ずれ時間は、この記録された垂直同期信号と聴覚刺激信号のずれ時間を測定することにより、測定可能である。400 回の感覚刺激の呈示に相当する同期確認実験において、映像と音の時間的ずれは平均 0.0633[ms]で標準偏差が 0.0564[ms]、最大のずれ 0.147[ms]となった。この結果から、実験に必要な分解能に対して十分な精度の維持されていることが判明した(Fig.2)。ずれ発生の主な原因はデジタル VTR 内での音声復号のレイテンシ変動と、測定信号に含まれた雑音によるものと推測される。

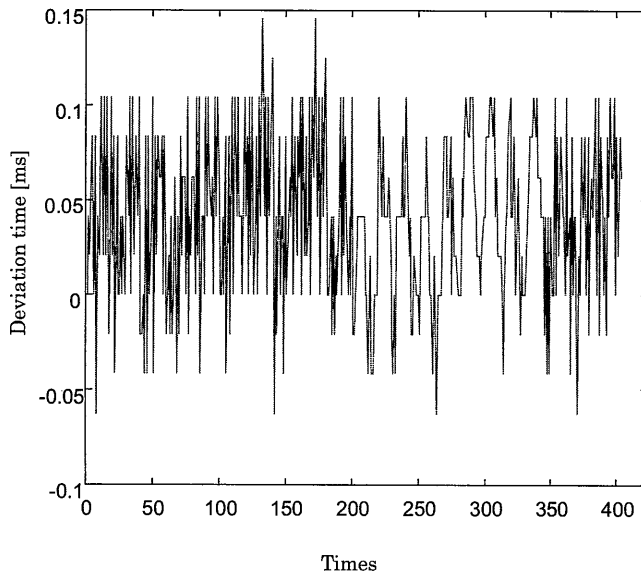


Fig. 2 Deviation time between V-sync signal and Auditory stimulus

## 4. 人の応答モデルの同定

### 4.1 移動指標

本測定システムによる実験の対象として、市販の音楽ゲームに採用されているユーザーインターフェースを簡略モデル化した画面上を移動する矩形の指標による動作指示を用いてドラムを演奏した場合の人の応答を測定した。被験者が電子ドラムを打鼓すると、電子ドラムがドラムの打鼓面の振動に応じた電圧を発生し、データレコーダによりその信号を記録する。

視覚刺激には画面上を上から下に移動する矩形(指標;Marker)と横線(バー;Bar)、Vertical line が含まれる (Fig.3)。本論文では指標がバー上に達した時を演奏点とし、被験者に対し演奏点でドラムを打鼓するよう教示した。

CRT に表示される画像は縦 480[pixel]、横 720[pixel]である。Vertical Line は $(x, y) = (360, 0)$ から $(x, y) = (360, 479)$ に太さ 1[pixel]で描画される。バーは $(x, y) = (0, 402)$ から $(x, y) = (719, 402)$ に太さ 12[pixel]で描画される。指標の長辺は 60[pixel]であり、 $x=330$  から描画される。指標の表示位置を素画像ごとに変化させ、素画像を連続的に表示することにより、画像は滑らかに指標の動く仮現運動として被験者に知覚される。

$i$  番目に表示される指標の上側の座標を  $y_{ij}$  とすると、指標の上側の座標は式(1)で与えられる。ただし、視覚刺激の先頭からの画像番号を  $j$  とする。1 曲分の視覚刺激は  $N$  枚の画像で構成されるため、

$1 \leq j \leq N$  となる。

$$y_{i,j} = \text{round} \{ \alpha (j - a_i) \} + \beta \quad (1)$$

$\alpha$  は pixel/frame 単位の移動速度で、 $\beta$  は指標の初期位置を示す。指標の短辺が 12[pixel]であるため、 $\beta = -12$  としている。ただし、 $a_i$  は  $i$  番目の指標が画面上に出現する画像番号である。また、 $y_{ij}$  が画面に表示されない範囲となる条件では指標を描画しない。

画像の表示間隔を  $F$ 、1 曲の音符数を  $M$  とすると、 $a_i$  は式(2)のように表現される。刺激の呈示には NTSC の表示機構を使用するため、 $F=1/29.97$ [s]である。 $a_i$  は第 1 番目の指標が出現する直前までの画像数で、操作に支障のない値に設定される。

$$a_i = \begin{cases} a_1 & \text{where } i = 1 \\ \text{round} \left( a_{i-1} + \frac{b_i}{F} \right) & \text{where } 1 < i < M \end{cases} \quad (2)$$

式(2)の  $b_i$  は前の音価の持続時間を秒単位で表している。

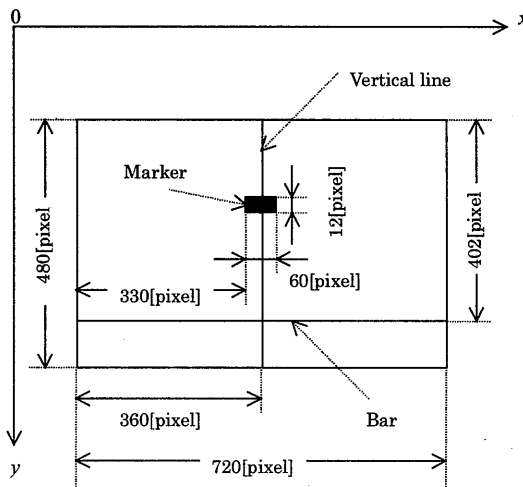


Fig. 3 Primitive Image

#### 4.2 視覚刺激

実験の視覚刺激には何も表示されない画面となる視覚刺激  $G_0$  と背景が表示されていて突然演奏点に指標が表示される  $G_1$ 、画面上の指標移動速度の異なる 4 種類の視覚刺激  $G_2, G_3, G_4, G_5$  の合計 6 種類の映像を用いた。視覚刺激  $G_0, G_1, G_2, G_3, G_4, G_5$  は各刺激の発生間隔をランダムに設定して、1 人の被験者に対して 60 回ずつ呈示した。また、等間隔に発生される刺激に対する応答を測定する目的で、視覚刺激  $G_1$  を一定間隔 533.9[ms]で表示する視覚刺激  $G_0$  と視覚刺激  $G_3$  を一定間隔 533.9[ms]で表示す

る視覚刺激  $G_7$  を用いた。CRT の画面サイズは 28 型であり、画面より被験者までの距離は 1.0[m] と設定したため、被験者から見た指標移動角速度は  $G_2, G_3, G_4, G_5$  においてそれぞれ 5.9、12、24、47[degree/sec] となる。

### 4.3 聴覚刺激

実験に用いた聴覚刺激は  $S_0$  と  $S_1$  である。実験中被験者はヘッドフォンを装着し、聴覚刺激はこれを通して被験者に呈示される。 $S_0$  は無音の聴覚刺激であり、 $S_1$  は指標が演奏点に表示される瞬間に  $f = 400$ [Hz] の減衰する正弦波音(聴覚プロンプト)の出力される聴覚刺激である。視覚刺激によって指標が表示されない場合、聴覚プロンプトの出力される時が演奏点となる。

聴覚プロンプト信号には 1 画像表示時間の 1/29.97[s]後に振幅が -40[dB]以下となる時定数を用いた。聴覚プロンプトの振幅は、予備実験より被験者が最適な状態に設定したと判断した状態で固定した。本論文ではさらに  $S_1$  を 533.9[ms]ごとの等間隔に発生する聴覚刺激  $S_2$  を使用した。聴覚刺激の発生時間は指標の表示開始に合わせてオフセット量を調整した。

時定数を  $\tau$ 、サンプリング周波数を  $f_s$ 、1 つの聴覚刺激波形のサンプル数を  $v$  とすると次の式(6),(7)(8)となる。

$$v = \text{round}\left(\frac{f_s}{29.97}\right) = 1602 \quad (6)$$

$$\tau = -\frac{v-1}{\log 0.01} \quad (7)$$

聴覚プロンプト は次の式(8)のように表現される。 $A$  は信号の最大振幅である。

$$\eta(x) = A \cdot e^{-\frac{x}{\tau}} \cdot \sin 2\pi f \cdot \frac{x}{f_s} \quad (8)$$

$$x \in \{0, 1, 2, \dots, v-1\}$$

聴覚刺激の音量は、数人の被験者が試し聞きやすいと判断した状態に決定した。そして、全ての被験者に対して同じ設定を用いた。

本論文では以後視覚刺激と聴覚刺激を同期させて合成した複合刺激を  $\{G_0, S_0\}$  のように 2 つの刺激の集合として表記する。さらに、この場合視覚刺激の表示開始時間と聴覚プロンプトは音声サンプリング周波数単位で映像との同期が取れているものとする。



#### 4.4 刺激に対する応答

健康な男性 16 名、女性 2 名の 18 名(平均年齢 22.4±1.09 歳)を被験者として、各刺激を呈示した場合の応答を測定した。実験の結果、 $\{G_1, S_0\}$ を呈示した場合の応答時間の平均値は 324.6[ms]で、 $\{G_1, S_1\}$ を呈示した場合の応答時間の平均値は 264.3[ms]であり、応答時間は平均で 60.30[ms]の短縮が確認された<sup>(17)</sup>。さらに、聴覚プロンプトのみの呈示となる $\{G_0, S_1\}$ による応答時間の平均値は 295.0[ms]となり、視覚刺激  $G_1$ との両方を呈示した場合より平均で 30.72[ms]応答が遅延した。また、 $\{G_3, S_1\}$ に対する応答時間は平均-9.363[ms]となった。負方向は演奏点以前に打鼓していることを示している。 $\{G_1, S_1\}$ を用いた場合の応答時間は平均 264.3[ms]であったため、 $\{G_3, S_1\}$ を用いた場合は 304.4[ms]先に打鼓していることになり、応答時間の短縮が確認された。また、 $\{G_3, S_0\}$ への応答時間は-16.17[ms]となった。これらの $\{G_3, S_0\}$ 、 $\{G_3, S_1\}$ による実験結果を一元配置分散分析によって検定した結果、危険率 5%で有意差が検出されなかった。つまり、視覚刺激  $G_3$ が存在した場合、被験者の応答が聴覚プロンプトの有無によって影響を受けないことが判明した。

本論文ではばらつきの少なくなる、すなわち応答時間の標準偏差の最も小さくなる条件が人の動作指示にとって最適であると考え、打鼓の応答時間の標準偏差を応答精度と定義した。各複合刺激に対する応答精度を比較すると $\{G_4, S_1\}$ のが最も良く、34.17[ms]であった。さらに、各刺激における聴覚プロンプトの有無と視覚刺激の種類についての二元配置分散分析の結果、 $G_3$ と $G_4$ の間には、聴覚プロンプトの有無と視覚刺激の種類、交互作用の全てで有意差が検出されなかった。したがって、最適な指標移動速度は $G_3$ と $G_4$ の近辺、すなわち被験者から見て 12[degree/sec]から 24 [degree/sec]近辺に存在すると考えられる<sup>(17)</sup>。

通常の演奏の場合、テンポにはゆらぎが検出され<sup>(18, 19)</sup>、短周期のゆらぎほど大きさの小さくなる傾向があり、長周期のゆらぎはテンポのゆらぎの影響を強く受けている<sup>(20)</sup>。また、被験者にテンポを誘導するような合図を用いない条件で均一に拍打ちさせる場合、0.1[Hz]より高い周波数成分ではホワイトノイズ状の分布を示し、それより低い周波数成分では 1/f ゆらぎの傾向を示すことが報告されている<sup>(21)</sup>。これに対して、訓練を積んだピアノ演奏者が芸術的表現を取り入れず意識的に音価通りに演奏したメトロノーム的演奏のスペクトルの傾きは 0.1 程度であることが報告されている<sup>(22)</sup>。本実験の応答時間は、通常の演奏と異なり訓練を積んだピアノ演奏者のメトロノーム的演奏に近いゆらぎ特性を示している。

刺激呈示からの打鼓応答時間は演奏点からの偏差であり、偏差のスペクトルはゆらぎの状態を示すため、打鼓応答時間のスペクトルの傾きを求めることによって打鼓の時間的ゆらぎの傾向を知ることができる。最小 2 乗法によって求めたスペクトルの傾きの絶対値は最大 0.1020 であった。さらに、標準偏差が平均値に対して大きく 95%信頼区間が平均値の絶対値よりも大きいため、スペクトルの傾きが特定の傾向を示していない。したがって、応答時間のスペクトルはホワイトノイズ状の分布と

なっていて、本実験での応答時間はランダムにゆらいでいる(Fig.4)。

等間隔に発生された刺激に対しては被験者の応答間隔を測定した。測定の結果、等間隔に発生した刺激に対する応答間隔は刺激の種類に依存しないことが確認された(Table 1)。応答間隔の平均値ほどの刺激に対しても534[ms]近辺となっており、刺激発生の周期である533.9[ms]と一致している。被験者と刺激の種類による二元配置分散分析の結果、被験者と刺激種類、交互作用に有意差が検出されなかった。

身体運動のリズムや視覚的リズムの領域において、人間が身体運動や映像などの刺激を能動的に知覚することでリズムを心的に構築するという報告がある<sup>(23)</sup>。さらに、この等間隔打鼓は独立した打鼓とは違い、能動的にリズムを構築する自発的リズムで打鼓しているため、ランダムな発生感覚を持つ刺激を与えた場合の34.17[ms]に比べ演奏精度が向上したと考えられる。

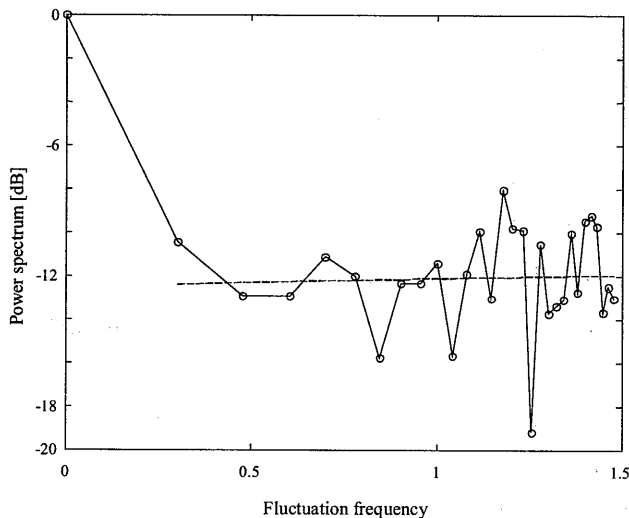


Fig. 4 Power spectrum of response time

Table 1 Statistic of response time

Stimulus	{S <sub>2</sub> ,G <sub>0</sub> }	{S <sub>2</sub> ,G <sub>6</sub> }	{S <sub>0</sub> ,G <sub>6</sub> }	{S <sub>2</sub> ,G <sub>7</sub> }	{S <sub>0</sub> ,G <sub>7</sub> }
Average[ms]	534.0	533.9	533.5	534.0	533.9
Std. dev.[ms]	14.17	14.26	15.03	15.84	17.07

#### 4.5 演奏モデルの同定

前節の実験により、視聴覚刺激を用いた演奏指示により、演奏精度を向上させることが可能となることが分かった。本節では、この結果を基に、実際の曲に合わせて視聴覚刺激を発生し、それに合わ

せて打鼓した場合に音楽演奏が成立するのを実験により確認した。実験では 5 名の大学生(平均年齢 22.7±1.85 歳)の被験者に本システムの演奏指示により曲を演奏させ、音楽の演奏精度を測定した。さらに、音楽演奏に対する演奏精度変動のモデルを検討した。

被験者は本測定システムによって呈示される視覚刺激に従い、ドラムパッドを打鼓するように教示される。被験者に演奏させた曲は、我々が作曲した合奏試験曲である<sup>(24)</sup>。この曲は 4 拍子の曲で、5 パートに分かれており、独立した各パートの演奏を演奏後に合成することによる楽曲の成立を確認するために作曲された。演奏速度は MM.=112 であり、パート 1 からパート 5 の音符数はそれぞれ 70、85、44、88、51 である。被験者には全てのパートを独立して演奏させた。指標の移動速度は全てのパートを前節で検討した最適指標移動速度範囲である 9[pixel/frame]に設定している。ただし、演奏中に演奏音を発生させず、被験者にはドラムパッドを打鼓する行為が楽曲の演奏であることを教示しなかった。

各音符に対する演奏精度を演奏点となった時間と実際に打鼓した時間の差である  $I_i$  の標準偏差とし式(9)のように定義した。

$$y_i = \sqrt{\frac{1}{q-1} \sum_{k=1}^q (I_{i,k} - \bar{I}_i)^2} \quad (9)$$

$i = \{1, 2, 3, \dots, M\}$

ただし、 $M$  は音符数、 $q$  は同一被験者の同一刺激での演奏回数、 $I_{i,k}$  は  $k$  回目の演奏の  $I_i$  であることを示す。

この演奏精度が過去の演奏にどのような影響を受けているかを解析するため、演奏精度の変動を自己回帰モデル(Auto-Regressive Model, AR モデル)でモデリングした。本論文で作成した演奏精度の変動の AR モデルは、過去  $m$  個の演奏精度  $y_{i-1}, \dots, y_{i-m}$  と白色雑音  $v_i$  で表現され、式(10)となる。

$$y_i = \sum_{j=1}^m a_j y_{i-j} + v_i \quad (10)$$

ただし、ただし、 $m$  は次数、 $a_j$  は自己回帰係数であり、また  $v_n$  は平均が 0、分散が  $\sigma^2$  である  $y_n$  の過去と独立な正規白色雑音である。

本論文ではユールウォーカー法により AR モデルを決定した。最適 AR 次数の選定には情報量基準 AIC(Akaike's Information Criteria)を用いた。AIC はモデルに含まれる白色雑音の分散と次数  $m$  で決定されるが、分散は  $m$  によって決定されるため、 $m$  の関数となる。AIC の値を  $A(m)$  とすると式(11)となる。

$$A(m) = N \log 2\pi + N \log \sigma^2 + N + 2(m+1) \quad (11)$$

AIC の値が最も小さいものが最適な AR 次数となるため、最適次数  $m_0$  は式(12)を満たす。

$$A(m_0) = \text{Min } A(m) \tag{12}$$

合奏試験曲の演奏結果を AR モデルでモデリングした結果、最適 AR 次数は 2 次以下となり、さらに 2 次となるパート 1 でも 1 次の係数の絶対値が 2 次の係数の倍近くとなるため、全パートで  $a_1$  の係数が支配的であることが判明した(Table 2)。

Table 2 AR modeling of playing accuracy

	AR order	$a_1$	$a_2$	$\sigma^2$
part 1st	2	0.5116	-0.2530	158.8
part 2nd	1	0.2300		337.0
part 3rd	1	0.4064		117.1
part 4th	1	0.2174		210.5
part 5th	1	0.3258		345.0

自発テンポによる等間隔打鼓の場合、過去 20 打鼓を限度とした過去の影響を受けることが報告されている<sup>(25)</sup>。これに対して、移動指標による打鼓指示の場合、視覚刺激による指示が自発テンポ形成の影響より強く作用していることが演奏精度の AR モデルへの同定によりわかる。また、移動指標による動作指示を用いると打鼓のゆらぎがランダムになるという結果は、自発テンポによる等間隔打鼓とは異なり、刺激指示に応じた単独の打鼓の重ね合わせとなることを示していると考えられる。

実験によって取得されたデータを用いて MIDI 音源を駆動し、音を発生させたところ、十分音楽として認識することができた<sup>(24)</sup>。

## 5. まとめ

本論文では、映像による視覚刺激と聴覚刺激を人に与え、人の応答を同定するための測定システムを構築した。そして、そのシステムを用いてビデオゲームに採用されている移動指標による演奏動作指示システムを単純化した視覚刺激を用いて人の応答を計測し、人の応答特性と刺激による演奏指示に対する演奏精度の変動を AR モデルに同定した。

実験の結果、視点からの角速度 12[degree/sec]から 24 [degree/sec]近辺に打鼓に最適な指標の移動速度が存在し、この条件では聴覚プロンプトの有無に影響を受けないことが確認された。また、刺激が時間的に等間隔に発生される場合は刺激の種類によって応答精度が変化しなかった。したがって、等間隔に発生される刺激に対しては自発テンポの影響を受けていると考えられる。

実際に曲を演奏した演奏実験の結果、演奏精度の変動を AR モデルに同定することができた。さらに、AR 次数とモデルの係数を検討した結果、移動指標を用いると自発テンポによる等間隔打鼓の場

合と異なり、過去の演奏動作の影響を受けず、指標による刺激に対する独立した応答となることが明らかになった。

本論文で構築した測定システムは、映像による視覚刺激と聴覚刺激を容易に変更可能であり、様々な刺激条件で人の応答を測定することができる。したがって、本測定システムは、同期して発生される映像による視覚刺激と聴覚刺激に対する応答に関する各種実験への応用が期待できる。

《参考文献》

- (1) M. Hershenson, "Reaction time as a measure of intersensory facilitation," *J. Experimental Psychol.* 63, pp.289-293 (1962).
- (2) 丸山欣哉, 佐々木孝之, "視覚と聴覚間の相互作用諸効果," *音響学会誌*, 52, pp.34-39 (1996).
- (3) 岩宮眞一郎, "オーディオ・ビジュアル・メディアによる音楽聴取行動にける視覚と聴覚の相互作用," *音響学会誌*, 48, pp.146-153 (1992).
- (4) 中野良樹, "視覚反応時間に対して聴覚刺激の付加が及ぼす促進効果," *心理学研究*, 68(2), pp.140-145 (1997).
- (5) 岩宮眞一郎, "オーディオ・ビジュアル・メディアを通しての音楽聴取における大画面再生映像が音楽再生音及び映像の印象に与える効果," *音響学会誌*, 51, pp.123-129 (1995).
- (6) 下條信輔, クリスチャンシャイア, ロミニジャワン, ラダンシャムズ, 神谷之康, 渡辺克巳, "知覚モダリティを越えて: 視知覚に及ぼす聴覚の効果," *音響学会誌*, 57, pp.219-225 (2001).
- (7) Donald Reynolds, "Effects of double stimulation: Temporary inhibition of response," *Psychol. Bul.* 62, pp.333-347 (1964).
- (8) David H. Raab, "Statistical facilitation of simple reaction times," *Trans. N.Y. Acad. Sci.* 24, pp.574-590 (1962).
- (9) S.C. Gielen, R.A. Schmidt, P.J. van den Heuvel, "On the nature of intersensory facilitation of reaction time," *Percept. Psychophys.* 34, pp.161-168 (1983).
- (10) Raymond S. Nickerson, "The effect of preceding and following auditory stimuli on response times to visual stimuli," *Acta Psycho*, 33, pp.5-20 (1970).
- (11) 丸山欣哉, 中野良樹, "視覚と聴覚の相互作用による反応時間の変動," *感性情報処理の情報学・心理学的研究(重点領域研究平成5年度成果報告書)*, pp.150-151 (1994).
- (12) R.A. Schmidt, S.C. Gielen, P.J. van den Heuvel, "The locus of intersensory facilitation of reaction time," *Acta Psychol.* 57, pp.145-164 (1984).
- (13) Francis B. Colavita, "Human sensory dominance," *Percept. Psychophys*, 16, pp.409-412 (1974).
- (14) Francis B. Colavita, Daniel Weisberg, "A further investigation of visual dominance," *Percept. Psychophys*, 25, pp.345-347 (1979).

- (15) 丸山欣哉, 中野良樹, “感覚相互作用についての ERP を座標とした研究,” 感性情報処理の情報学・心理学的研究(重点領域研究成果報告書), pp.158-159 (1993).
- (16) 長嶋洋一, “音楽的ビートが映像的ビートの知覚に及ぼす引き込み効果(2),” 電子情報通信学会技術研究報告ヒューマン情報処理, 104(168), pp.45-50 (2003).
- (17) 神谷達夫, 矢野学, 松田稔, “打楽器ゲームにおける視覚・聴覚刺激に対する応答,” 日本生理人類学会誌, 8(4), pp.83-91 (2003).
- (18) 吉田友敬, 中西智子, “音楽におけるリズムのゆらぎ,” 情報文化学会誌, 3(1), pp.43-50 (1996).
- (19) 吉田友敬, 中西智子, “リズムのゆらぎと情報,” 情報文化学会全国大会講演予稿集, 3, pp.42-49 (1995).
- (20) 池内智, 佐々木実, 北村音一, “リズム並びにテンポのゆらぎの数量化に関する研究, あるギター曲のメロディーを例にとった場合,” 音響学会誌, 40, pp.228-234 (1984).
- (21) T.Musha, K.Katsurai, and Y.Teramachi, “Fluctuations of Human Tapping Intervals,” IEEE Transactions on Biomedical Engineering, 32(8), pp.578-582 (1985).
- (22) 佐々木實, 石川智子, 山田真司, “等時価からなる楽譜のピアノ演奏における時間的ゆらぎ,” 音響学会誌, 45, pp.743-747 (1989).
- (23) 堀内靖雄, 財津茜, 市川薫, “人間の演奏制御モデルの推定,” 情報処理学会研究報告, 99(51), pp.59-64 (1999).
- (24) 山炭昌司, 矢野学, 神谷達夫, 松田稔, “動的電子楽譜を用いた個別合奏,” ゲーム学会第3回全国大会講演論文集, pp.53 (2004).
- (25) M. Yamada, “Temporal control mechanism in equaled interval tapping,” Appl. Hum. Sci, 15, pp.105-110 (1996).