

ビデオ教材の画像抽出法の検討

— タイムサンプリングによるショット抽出の理論 —

三 尾 忠 男¹⁾・藤 田 恵 璽²⁾

Time Sampling of Pictures from Video Programs

Tadao Mio & Keiji Fujita

Abstract

In order to analyze, compare and evaluate a number of educational video programs, it is necessary to have a sampling technique which provides us with samples of representative pictures summarizing the visual information of the programs. Fujita, K. (1989) reported that most shot length distributions of video programs can be described by a compound exponential distribution. On the basis of his previous study, we theoretically estimated the percentage of shots sampled by time sampling with various time intervals. The agreement between the estimated and the observed values was exceedingly good for all of the 53 educational video programs examined. It was found that for most videos more than 95 % of the shots were sampled by the time sampling method with 5~10 second time intervals. A variety of applications of this method to educational and technical use are discussed and demonstrated.

キーワード

映像教材 画像抽出法 タイムサンプリング 映像多重提示

1. はじめに

教育利用を目的とした映像の制作は非常に盛んであり、放送教材とともにビデオ教材は、さまざまな教育活動の場面で利用されている。ところが、映像教材そのものに関する分析研究は、十分なされていない。それは、映像が印刷教材のように“とばし読み”や“見当をつけた検索”などができないため、継時的な視聴になり時間軸に拘束されている。そのため映像の分析作業に時間がかかるだけでなく、提示される画像および音響情報の内容を

“記述したり比較する”ことが非常に困難である。したがって、映像の構造要素と時間系列を手段としてビデオ番組を分析するためには、まずはじめに測定の「単位」を設定することが必要となる。映像は、“異なる画像もしくは異なるショットが時間的に連続したもの”とみなすことができる。そこで、藤田は、ショットを測定の単位として、その「提示時間」と「構成要素 Person/Media/Setting 軸」(人物／使用メディア／状況軸)を用いて映像教材の記述をする“ショット分析法”を開発した。放送大学の映像教材を中心に、ショットの提示時間を測定し、統計的な構造の特徴を解析した。その結果、それらの統計情報がビデオ教材の構造的特徴を示すことを発見し、ショットの時間分布の理論モデルの開発をおこなった(藤田1989)。その後、国内外のビデオ教材についてもショット分析を続け、計53番組の測定を終えている。さらに、ショット測定と分析研究の支援環境として、パーソナルコンピュータとビデオデッキを連動した探索的分析システム E.V.A.S. (Exploratory Video Analyzing System) を開発している(三尾・藤田1990)。

測定単位であるショットを映像から切り取るために不可欠である“ショットの切れ目”の識別は、ビデオ教材のサマリー作成を目的とした画面輝度差測定による自動検出が試みられている(菊川ほか1988)。この方法では、編集技法の1つであるオーバーラップなどによりショット間の画像が物理的に連続している場合、ショットの切れ目の検出が困難である。しかし、映像音響資料データベース用のサマリー作成が目的であり、画像だけでなく音響情報も抽出しているため、データベースの利用者の立場からは有効な手段であり(川淵ほか1990)、実用の段階にある。一方、我々のように映像教材を視聴者の側に立っておこなう分析研究では、ショットの切れ目は肉眼的識別によらざるを得ない。

測定の単位としてショットが有効であることを示した我々の研究は、国内外、過去現在にわたって幅広く、さらに多くの映像教材についてショット分析をおこなうとともに、単位であるショットの内部構造の分析を行なう段階にきた。これらの研究では、これまでどおり人間の肉眼に頼ってショット計測することは非能率的であるとともに、ショット内構造のカテゴリー開発やショット間関連の分析において適切な測定方法とはいえない。したがって、実際の映像を時間軸に束縛されて継時的に視聴するのではなく同時的に画像を提示する方式の開発が不可欠である。

2. 目 的

本研究の目的は、ショット画像のサンプリングについて時間間隔 time interval を変化させて抽出ショット数の割合を理論的に求めた結果と実際に測定した値を比較し、検証することと、このサンプリング法を使用する場合の可能性を探ることである。

3. ショット分析

ビデオ画像は1つの長い“絵巻もの”と考えることができる。すなわち、1まとまりの連続した画像が多数、継続的に提示されたものとするのである。1つの画像が連続して

いるときそれをショットと呼び、これを分析の単位とした。われわれの行なったショット分析の結果では、45分～60分の教育番組は、100～200個のショットで構成されているものが多かった。また、ショット時間を1秒単位で測定し、その分布を調べてみると、図1に示すような逆J字型の分布を示し、1ショットの平均時間は15秒～30秒位のものが多かった。この分布の形は広い範囲の番組データで一定しており、その分布の密度関数は2つの母数をもつ複合指数分布で示されることが Fujita, K. (1989) によって明らかにされてい

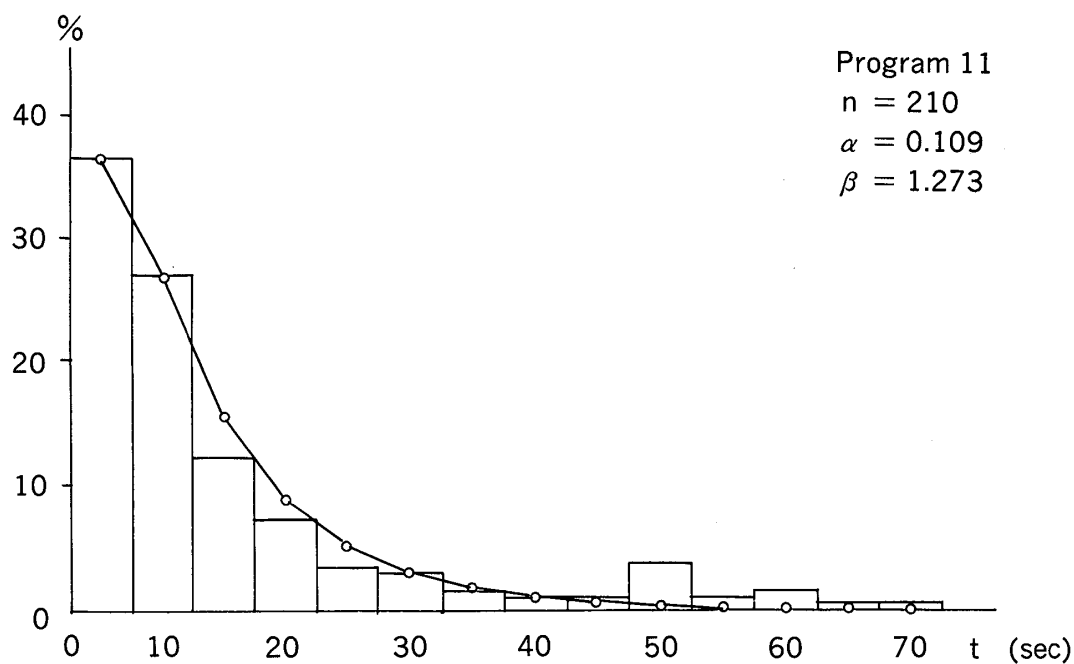
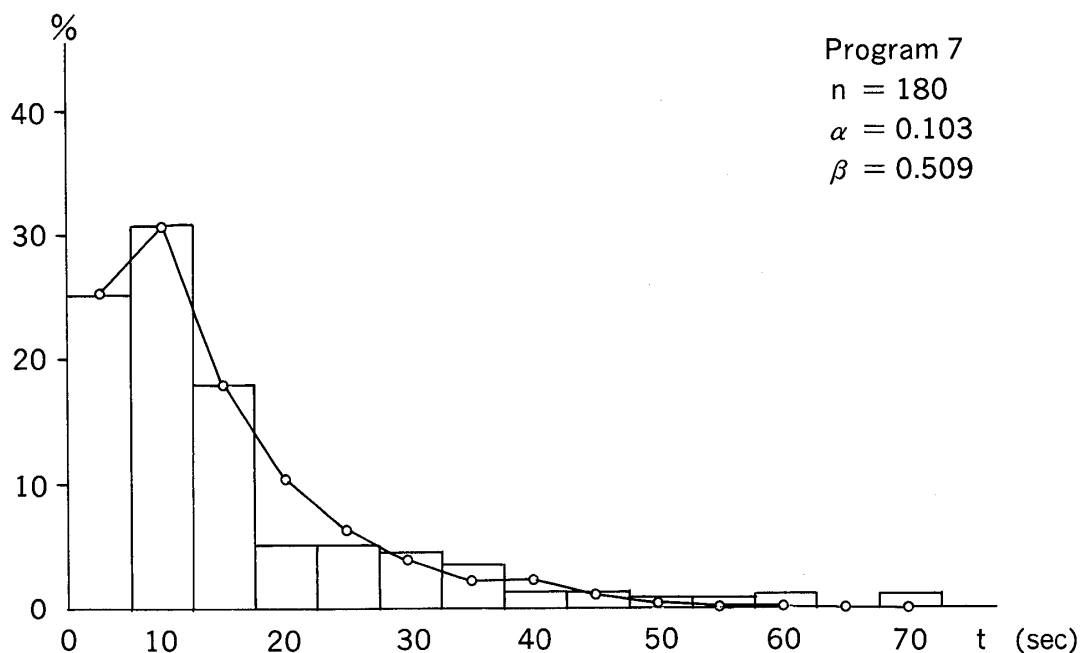


図1 ショットの時間分布に対する理論分布の適合
 (ヒストグラム：実測値の分布、折れ線グラフ：理論分布)

る。すなわち、このショット時間 t は 2 つの時間 t_1 と t_2 の和で構成されていると考え、

$$t = t_1 + t_2 \quad (1)$$

t_1 、 t_2 はそれぞれ独立に指数分布しているとすれば、

$$p(t_1) = \alpha e^{-\alpha t_1} \quad p(t_2) = \beta e^{-\beta t_2} \quad (2)$$

ショットの時間 t の密度関数は、つぎのように求められる。

$$p(t) = \frac{\alpha\beta}{\beta - \alpha} (e^{-\alpha t} - e^{-\beta t}) \quad (3)$$

この関数は初め、McGill (1963) が神経生理学的研究のデータを用いて反応時間分布に適合させたが、その後、人間の学習反応時間の分布を数理的に解析するモデルとして用いられるようになった (Fujita, K. & Naruse, M., 1977)。反応時間は情報刺激の「受け手」の側に起こる反応であるが、ショット時間は「送り手」の側に生じる情報提示の時間反応である。Fujita (1989) の研究は、この 2 つの時間反応が同じメカニズムによって説明されることを示したものである。実際、放送大学の番組のショットの時間分布にこの複合指数分布を適合させると、図 1 のようによく適合することが分かる。

さらに α 、 β 値によりその番組のショットの時間分布の形状が特長づけられ、平均と標準偏差も理論的に求められる。それによって、映像教材が、講師のトーキングヘッドやパネルによる解説のシーンのように単調な画面でショット時間の長い番組なのか、それとも多様な資料映像を短いショットで頻繁に提示するような番組なのかを区別することができる。

ところで、これまでのような肉眼によるショット分析では、番組の映像をはじめから最後まで視聴せねばならず、測定と分析に多くの労力と時間を要する。そこで、作業時間の短縮の方法として、一定の時間間隔で画像を時系列的に抽出するタイムサンプリングにより、ショット画像を取り出す方法が考えられる。

4. タイムサンプリング

一本のビデオ番組からショット画像をタイムサンプリングする過程を図 2 に示す。

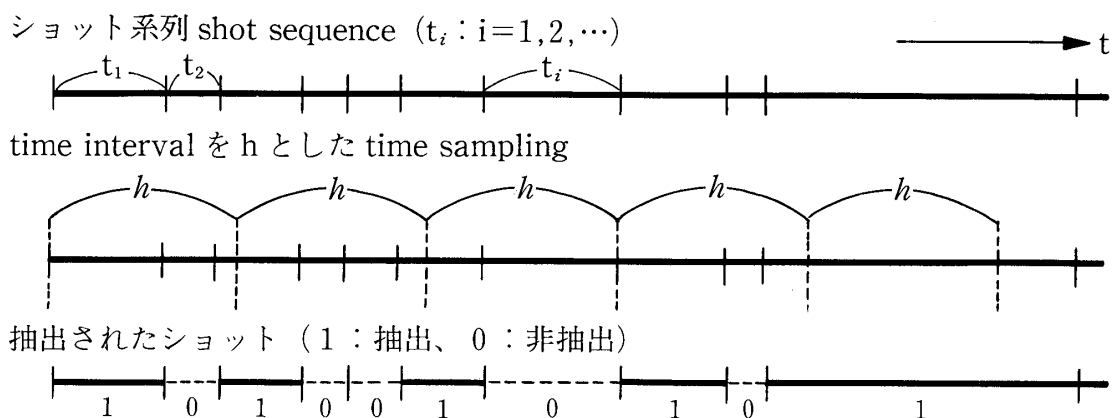


図 2 ショット系列のタイムサンプリング

program name 0U-P3

Shot no.	Time (sec)	Length (sec)	shot	Shot picture		
				Interval time		
				5 sec	13sec	20sec
1	0	4				
2	4	5				
3	9	5				
4	14	10				
5	24	67				
6	91	13				
7	104	14				
8	118	1				
9	119	7				
10	126	10				

図3 ショット画像とタイムサンプリングによる抽出画像の比較 (写真)

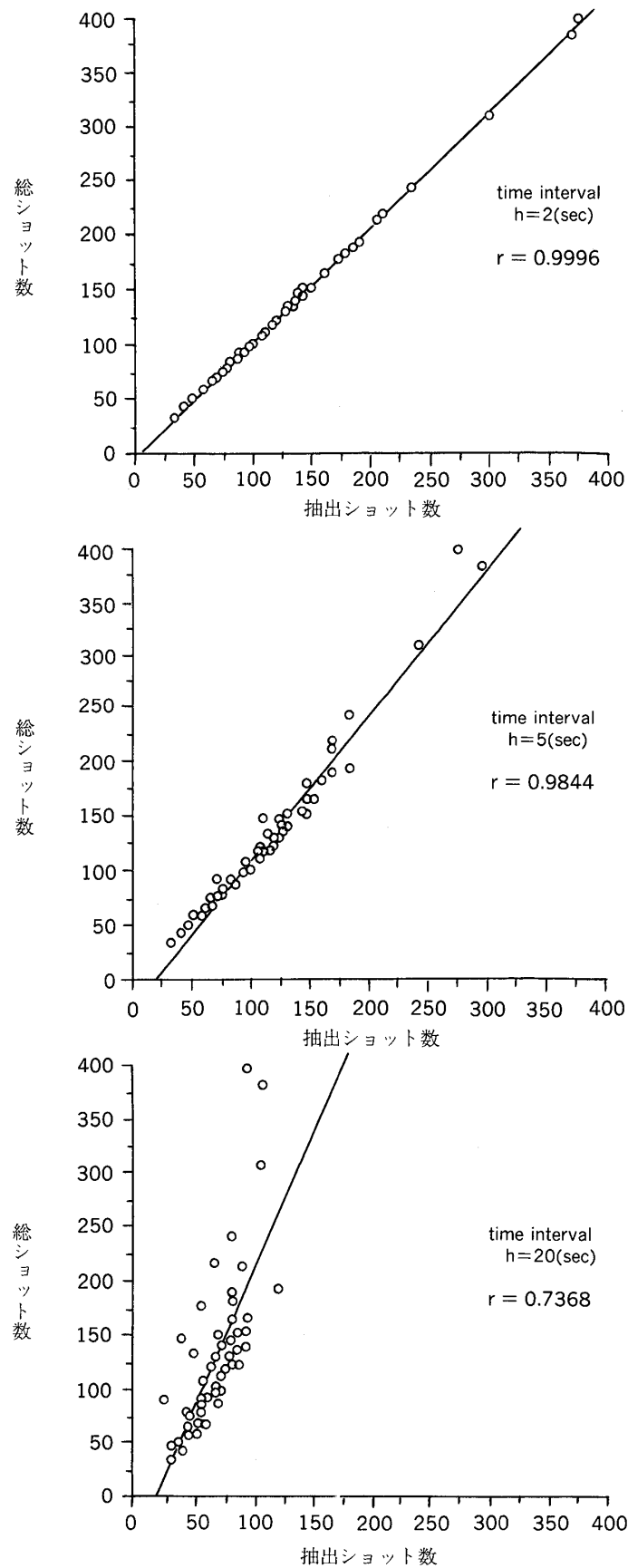


図4 抽出ショット数と総ショット数の間の相関係数 (53番組)

ショット系列 shot sequence とは、番組に登場する順 $i=1, 2, \dots, n$ にショットの時間 t_i を連続して並べたものである (図2上段)。

サンプリング作業を開始するビデオテープの最初のショット (図2中 t_1) は、必ず抽出する。次のショットからは、時間間隔 h (秒) で抽出していく。図下段の破線で示したショットは、抽出されなかったものである。なお、一度抽出されたショットは2度目以降は、画像を抽出しないこととする。

このようにして、実際にショット画像 shot picture を抽出した例を図3に示す。これは英国の公開大学のあるビデオ教材のはじめの136秒間の10ショットについて、5秒、13秒、20秒の3種類の時間間隔でタイムサンプリングしたときの抽出ショット画像の例である。時間間隔5秒では、ショット番号8は抽出されていない。同様に図中、空白のショットは抽出されていないことを示す。ショット番号4と9は、時間間隔13秒では抽出されないが20秒では抽出されている。サンプリングの時間間隔によって、同じショットでも抽出される場合とされない場合があることを示す例である。

このように時間間隔を何秒に設定するかによって、全体の何パーセントのショットが抽出できるかを分析した。抽出されたショット数と総ショット数との相関係数を53本の教育番組について、時間間隔を変えて求めた。図4は、時間間隔が2, 5, 20秒のときの相関である。時間間隔2秒での相関係数は0.9996であり、ほとんど全てのショットが抽出されていることがわかった。5秒では0.9844、20秒で0.7368と時間間隔をひろげると相関係数は低くなる。

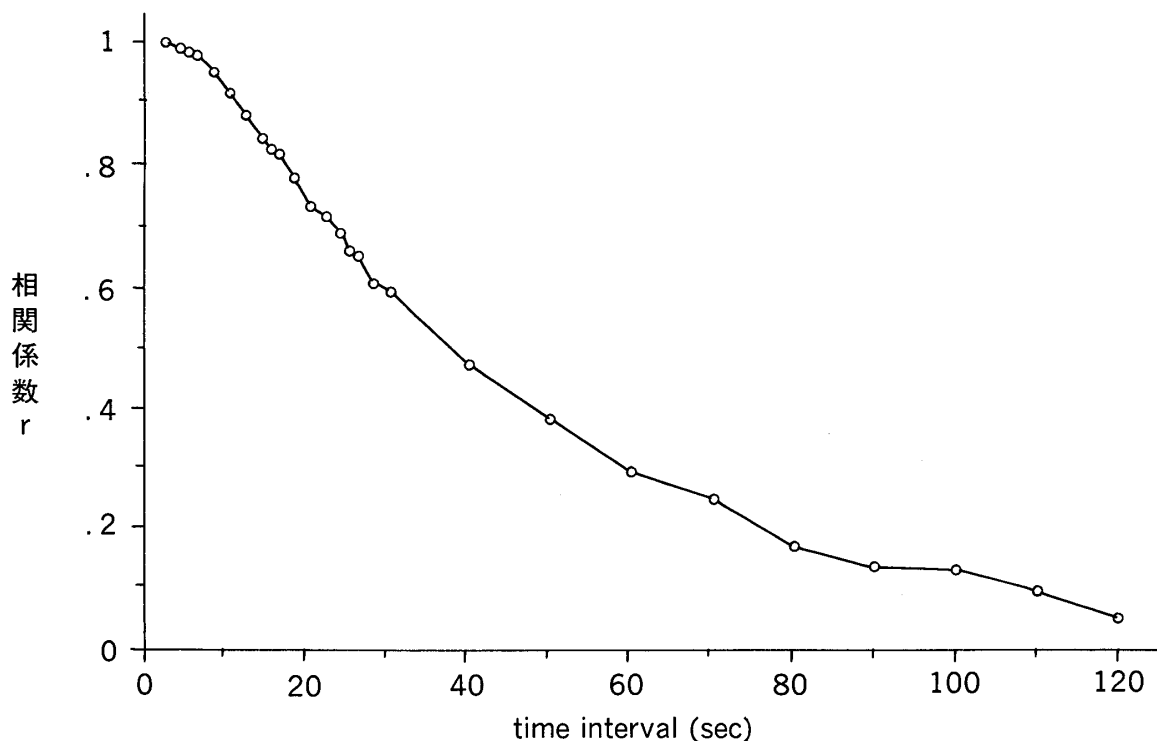


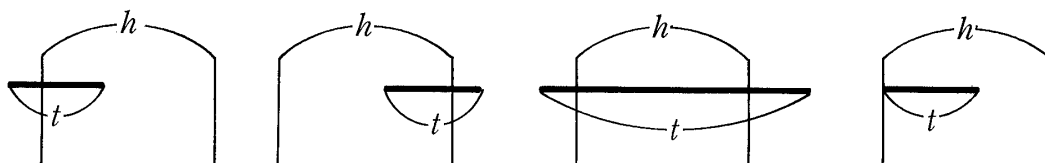
図5 抽出ショット数と総ショット数の相関係数 r のサンプリングの時間間隔による変化 (53番組)

時間間隔を2秒から120秒間まで27通りに変化させて、抽出ショット数と総ショット数との相関係数の推移を求め、グラフに示したのが図5である。これにより、5～10秒の時間間隔の抽出では、相関係数が0.916以上でショットが抽出できることが分かった。

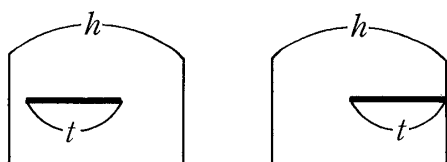
5. タイムサンプリングの理論

ショット時間の分布密度関数のパラメータ α 、 β をもちいて抽出できるショット数を推定する理論式を導き、前項のタイムサンプリングの結果に適合させてみよう。

抽出される場合



抽出されない場合



S : 抽出される場合 $P(S | t \geq h) = 1$

$$P(S | t < h) = \frac{t}{h}$$

図6 ショットが抽出される確率と時間間隔

ショット時間 t と設定した時間間隔 (time interval) h の関係をみると、図6に示すように、ショット時間 t が時間間隔 h より大きい場合は、そのショットは確率1で抽出される。

$$P(S | t \geq h) = 1 \quad (4)$$

一方、ショット時間 t が時間間隔 h より短い場合に抽出される確率は、その幾何確率をとって t/h とする。

$$P(S | t < h) = \frac{t}{h} \quad (5)$$

すなわち、ショット時間の理論分布(図7)において、時間間隔 h よりショット時間 t の長いショット群(破線の右側の部分)は必ず抽出され、短いショット(斜線の部分)は、 t/h の確率で抽出されることになる。

そこで、時間間隔 h でサンプリングしたときに抽出されるショット数の総ショット数に対する比率 $P(S | h)$ は次のように求められる。

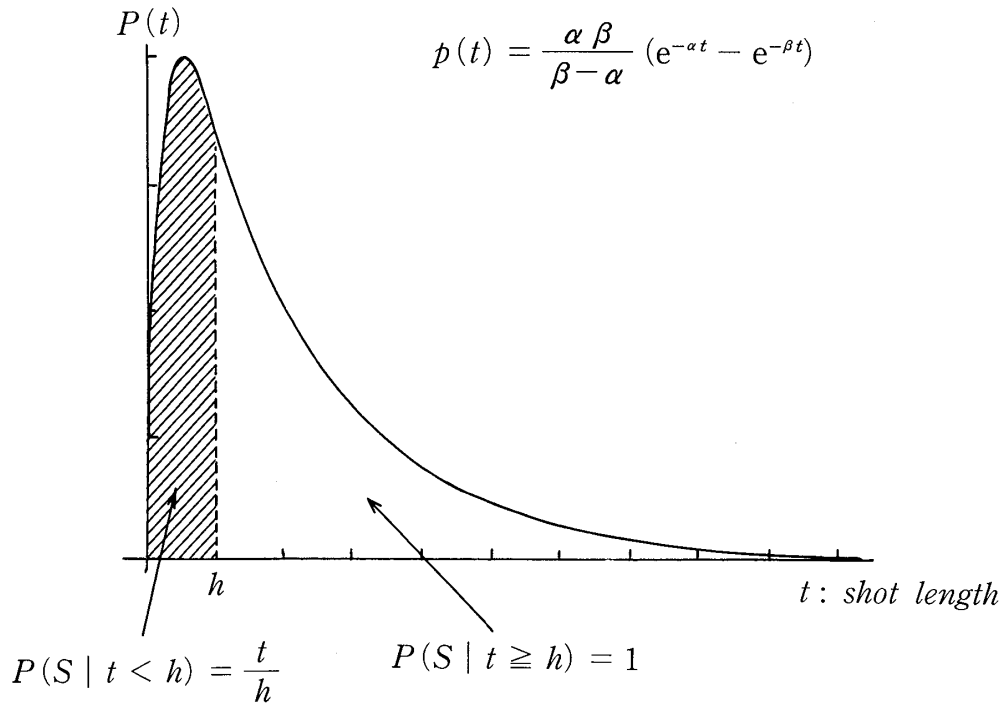


図7 ショット時間の理論分布と抽出確率

$$\begin{aligned}
 P(S | h) &= \int_0^h P(t) P(S | t < h) dt + \int_h^{+\infty} P(t) P(S | t \geq h) dt \\
 &= \int_0^h \frac{t}{h} P(t) dt + \int_h^{+\infty} P(t) dt \\
 &= \frac{\alpha \beta}{\beta - \alpha} \int_0^h \frac{t}{h} (e^{-\alpha t} - e^{-\beta t}) dt + \frac{\alpha \beta}{(\beta - \alpha)} \int_h^{+\infty} (e^{-\alpha t} - e^{-\beta t}) dt \\
 P(S | h) &= \frac{\beta^2(1 - e^{-\alpha h}) - \alpha^2(1 - e^{-\beta h})}{\alpha \beta h (\beta - \alpha)} \quad (6)
 \end{aligned}$$

この理論式によって推定するショット数の比率と実測した抽出ショット数と総ショット数との比率を比較すると、ショット分析を行なった国内外の53に及ぶすべての教育番組について、非常に高い一致度が得られた。図8はショット系列や分布の条件の異なる番組4本について、その適合度を示したものである。理論値と実測値の差異は一般に時間間隔が長くなる右端に起こる傾向があるが、総ショット数が著しく小さい番組を除けば、その差は5%を越えることはなかった。

このように映像教材のショットの時間分布から得られるパラメータ α 、 β によって式(6)を用いて、タイムサンプリングにおける抽出ショット数の総ショット数に対する比率が、高い精度で推定できるといえる。

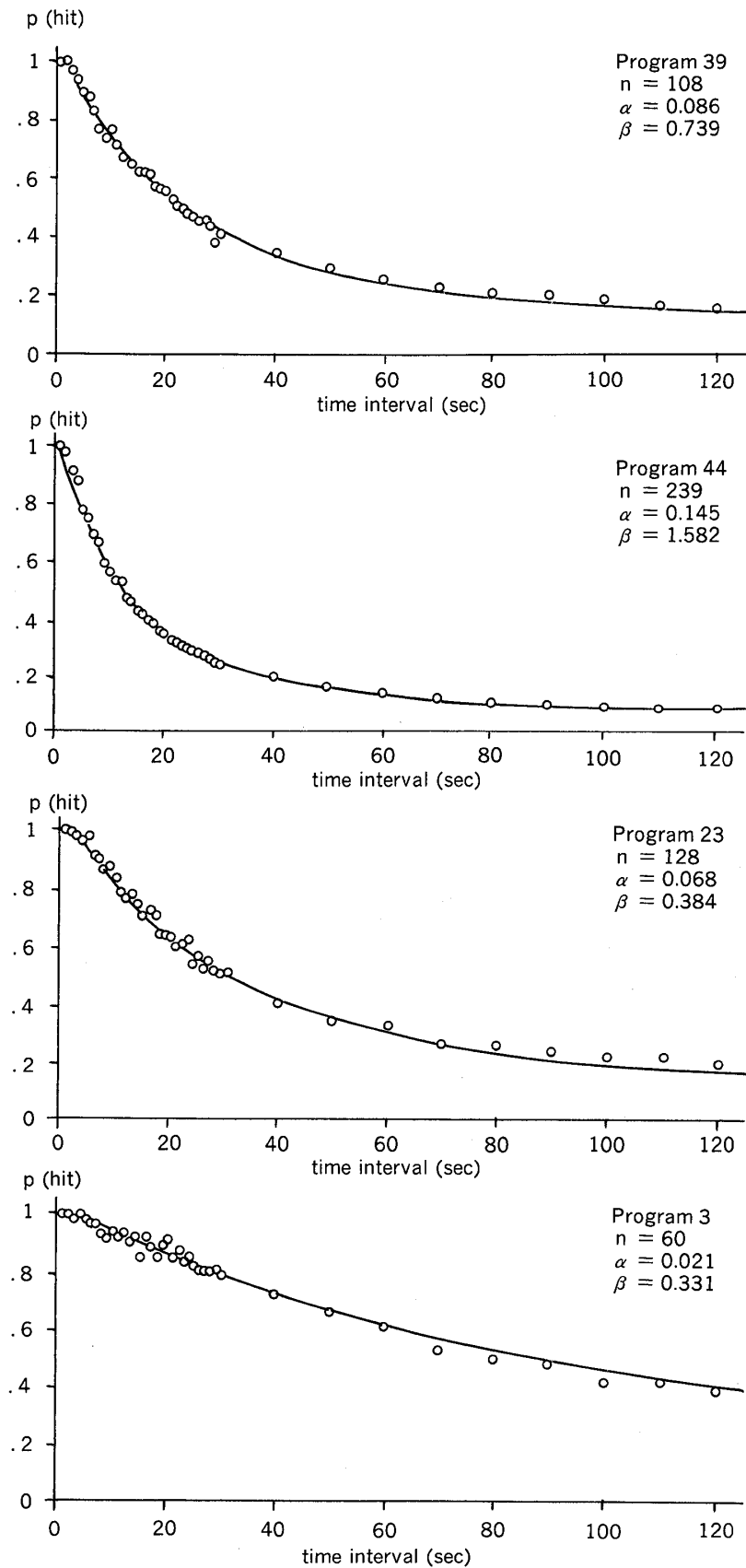


図8 抽出ショット数の理論値（実線）と実測値（点）の比較

6. タイムサンプリング法の応用例

このタイムサンプリングの方法は、教育活動の次のような場面での応用が可能となるであろう。

- 1) ビデオ教材の記述：現在、ビデオの画像はビデオプリンターなどで容易に印刷できる。ビデオをTVモニターで視聴するのではなく、シートに印刷したショット画像を印刷物のように平面的に提示し、閲覧することが可能になった。実際のシートの作成作業には、シートのコストなど経済的な理由により抽出ショット数の制限および作成作業時間（通常、1シートの印刷に2～4分かかる）の制約がある。ショット画像の抽出率（いいかえれば、総ショット数の何割の画像が欠損するのかを知ることができる）の推定値を考慮しながらタイムサンプリングの作業時間を調整することができる。
- 2) 要約された画像情報：ビデオ教材の記述の場合と同じ方法で画像を抽出するが、シートへ印刷するのではなく、その画像を再び映像としてビデオテープなどに記録する。これは、番組を要約した画像情報として利用できる。このとき、サンプリングの時間間隔により、元のビデオ教材からのショット抽出率が明示できる。
- 3) ビデオ教材の選択：視聴者が、映像ライブラリーなどからビデオ教材を選択する場合、ビデオ教材をシートに記述したものや要約された画像情報の映像を視聴することで、検索作業時間を大幅に短縮できる。この時、利用者へ元の番組と比べてどの程度のショットが抜けているかを数値で示すことができる。
- 4) 映像制作現場において、編集者がショットの提示時間を決定するための1方法：ショット提示時間の算出のための作業時間がこれまでより短縮できると予想できる。したがって、映像の制作のうち編集作業において、「長いショットを連続してつなぎ合わせていないか」など演出上好しくない編集をしていないかの確認をとることが容易になる。
- 5) ビデオ教材中の画像検索：ビデオ教材から欲しい画像を検索するとき、はじめは大きな時間間隔でサンプリングされた画像を試写する。この時点で、欲しい画像が発見できない場合、時間間隔を小さくして抽出ショット数を増やして再度サンプリング画像を試写する。この際、視聴者に時間間隔に対応して視聴できるショットの抽出率を明示できるため、検索作業量の目安として使用できる。

以上は、映像教材全体から画像を抽出する場合の例であるが、ショット分析作業においても、ショット計測の作業時間の短縮がはかれるとともに、ショット内の画像をタイムサンプリングすることで、提示時間の長いショット内での画面の変化を測定し特徴を分析することができる。また、ビデオ番組のショット時間の分布密度関数パラメータ α 、 β を知ることによって、ショットの時間分布、平均、標準偏差に加え、タイムサンプリングの画像抽出率まで推定することができ、多くの番組の特性を記述、比較することができる。

7. 結 論

これまでの映像教材の構造に関する実験的測定と理論的分析によって得られたショット時間の分布の理論に基づいて、開発した画像のタイムサンプリングによる抽出率算出の理論は、ビデオ教材の研究および活用に非常に有効であることを示した。

本研究で開発し、検討した画像のサンプリング法は、それに具体的に対応する映像提示装置の開発によってはじめて広く実用化される。現在、設計・開発中の映像多重提示システム（三尾1993）を用いた実証的検証を計画しており、その成果は次回報告する予定である。

〈参考文献〉

- 1) Fujita, K. and Naruse, M. 1977 An analysis of test item response times: Their relationship to correctness and confidence, *Educ. Technol. Res. Vol ; 1.5-15.*
- 2) 藤田恵璽、1990：「番組分析の構想とショット分析」放送教育開発センター研究報告45、第1章2-15。
- 3) 藤田恵璽、1987：「放送教材の番組分析（Ⅰ）－ショット時間分布－」日本教育工学会第3回全国大会発表論文集、267-270。
- 4) Fujita, K. 1989: Shot Length Destributions in Educational TV Program, *Bulletin of The Natonal Institute of Multimedia Education, No. 2, 107-116.*
- 5) McGill, W. J. 1963 Stochasitic latency Mechanisms. In Luce, R.D. and Galanter, E. (Eds) *Handbook of Mathematical Psychology, Wiley, New York, Vol. 1, 309-360.*
- 6) 三尾忠男・藤田恵璽、1990：「探索的ビデオ分析システムの開発（Ⅱ）」日本教育工学会第6回全国大会発表論文集、381-382。
- 7) 菊川 健・川淵明美、1988：「映像音響資料データベースシステムの開発Ⅲ－映像教材の自動サマリー作成システム－」教育工学関連学協会連合第2回大会講演論文集、633-634。
- 8) 川淵里見・菊川 健、1990：「映像資料サマリー作成の研究－カット対応型と時間分断型の比較－」日本教育工学会第6回大会講演論文集、419-420。
- 9) 三尾忠男、1993：「映像多重提示システムの開発－設計と教育活動への応用の検討－」日本教育工学会研究報告集 JET93-1、39-44。

（1）研究開発部助手）

（2）研究開発部教授）