言語解析を応用した映画におけるショット時間長の文脈構造解析
Film Context Analysis of Shot Length based on Language Processing

大河原 一輝[†‡]，大山 潤爾[†]
Kazuki Ookawara, Junji Ohyama
[†]産業技術総合研究所，[‡]筑波大学
National Institute of Advanced Industrial Science and Technology, Tsukuba University
ookawara.kazuki@aist.go.jp

Abstract
ヒトは時間変化する情報を、物理的な時間とは異なる単位で知覚している。映画におけるショット長の遷移パターンをヒトの時間知覚の観点から見ることにより、視聴者の映画の印象や評価に影響するストーリー展開の文脈構造を分析できる可能性がある。この仮説を検証するため、本研究では時間知覚特性に基づいてショット長を符号化し、言語解析手法によりショット長の遷移パターンの傾向を分析した。分析対象は、1973 年から 2013 年までの邦画 20 作品（平均上映時間 115.34 分、SD=16.12 分）とした。分析の結果、ショット長の遷移パターンと映像作品の評価値の間には相関があり、ショット長の遷移パターンが映像の印象を構成する要因の一つである可能性が示唆された。

Keywords — 映像情報処理，映像文脈，時間遷移パターン，時間知覚

1. はじめに

映像は非言語情報であり、時空間的に変化する視覚情報である。映像作品において、一台のカメラで撮影されて編集されていない連続した複数のフレーム画像で構成される映像の最小単位は“ショット”と呼ばれ[1]、複数のショットを時系列に配置することにより、一連の場面展開や映像全体のストーリー（時間文脈構造）が表現される。

我々が映画から受ける印象が映像の中の各ショットの時間長（以下“ショット長”と呼ぶ）と時系列配置（以下“遷移パターン”と呼ぶ）によって異なることは経験的に知られている。例えば、映画の予告は本編の映画に使われているショットを本編と異なるショット長と遷移パターンに編集したものであるが、実際の映画と異なる印象であった、ということはしばしば経験されることである。これは、ショットの遷移パターンを編集したことにより、我々が知覚する映像の印象が変化したことを意味する。映像情報は、表示される各フレームの画像に含まれる空間情報と、連続したフレームで表現される空間情報の短い時間変化を含むショットの時空間情報と、各ショットのショット長と遷移パターンに含まれる時間文脈構造で構成されていると言うことができる。しかし、映像情報におけるショット長やショットの遷移パターンの構造解析や、映像内容の理解や印象との関係は解明されていない。

映像には文法的な構造があり、その時間文脈がヒトの中でテキストや音声のように符号化および構造化されて認知されている可能性が報告されている[2]。これは、映像全体の構造を言語構造のような符号列として解釈することができ、その時系列構造に対する言語構造解析が適用できる可能性を示唆している。

また、予期しない視覚刺激によってヒトが知覚する時間情報に変化が生じること[3]や、聴覚刺激のテンポによってヒトが感じる印象が変化すること[4]が知られていることから、映像作品のショット長と遷移パターンは、そうしたヒト（視聴者）の認知特性に合わせて編集され、映像全体のストーリー展開や、そこから得られる作品の印象評価などに関わる時間文脈情報が反映されている可能性が考えられる。

そこで本研究では、映画のショット長の遷移パターンが視聴者の映画の印象評価に与える影響を調査するため、ショット長の遷移パターンの傾向と性質の分析を行う。符号列としてみたショット長の遷移パターンに対して言語解析手法を適用することにより、遷移パターンの規則性および映画作品の評価との関連性を分析する。言語解析手法により見出されたショット長の遷移パターンが、視聴者の映画評価にどのような影響を与えるのか把握することにより、映画におけるショット長の遷移パターンの役割について検討を行う。

1.1. 映像データセットの構築

解析対象となる映画作品は、1973 年から 2013 年の邦画 20 作品とする。作品に関する情報は、オンラインの映画情報サイト Yahoo 映画[5]からレビュー評価を収集した。レビュー評価は、評価値 1 から評価値 5 のそれぞれに対するレビュー件数である。

本稿では映画作品のショット長の時間分脈構造を解析するため、作業者が、動画編集ソフトを用いて 1 フレームずつ確認しながら、邦画 20 作品のショットの切り替わりを判断してショット単位に分割した。各ショ

ット長はショットのフレーム数から機械的に算出し、1作品ごとにショット長の遷移パターンのデータを作成した。今回対象とする、邦画20作品における作品の平均上映時間は115.34分（SD=16.12分）であり、1作品に含まれる平均ショット数は704.75であった。

2. ショット長解析

ショットは多様なショット長で表現される情報であるため、ヒトがショットから受け取る情報はショット長に依存して異なると考えられる。本章では、ヒトの時間に関する知覚特性の観点から、心理空間におけるショット長を定義し、その特性について分析を行う。

2.1. 時間知覚特性

ヒトは時間変化する情報を、物理的な時間とは異なる単位で知覚している[6]。Kristofferson による時間量子説[7]では、ヒトは時間変化する情報を離散的な情報として知覚しており、情報の時間長が長くなるほど時間分解能（単位時間）が荒くなることを実験的に示している。具体的には、100ms 以降の単位時間は 13ms であり、200ms 以降では 25ms、400ms 以降では 50ms、800ms 以降 1480ms までは 100ms であった。同様な実験を行なった Jingu の実験[8]においても、単位時間の詳細は異なるが、時間長が長くなるほど単位時間が大きくなる特性を確認している。

Fetterman et al. による実験[9]では、ヒトが時間情報から受け取る情報量が、時間長に対して対数変換を施した量（一般化ウェーバーの法則）に漸近することが示唆されている。時間長を対数変換した単位は、時間長が長い情報ほど単位時間が大きくなるという特性を表現しており、これは前述の時間量子説と一致する。

この時間知覚特性をショット長に適応することにより、心理空間におけるショット長の特性や遷移パターンの把握が可能となる。本稿では、物理的な時空間におけるショット長t_{real}に対応する、時間知覚特性を考慮したショット長t_{mind}を、一般化ウェーバーの法則を用いて、次式のように定義する。

$$t_{mind} = log_{10}\ (t_{real}) \qquad (1)$$

2.2. 方法

時間知覚特性を考慮したショット長が、映像作品においてどのような分布をとるか調査を行った。映画データセットに含まれる全映画作品に対して、時間知覚特性を考慮したショット長の分布と、そのヒストグラムを算出した。ヒストグラムの刻み幅は、0.10s から 1000s までのショット長が対数空間上で20等分されるように決定した。刻み幅は対数空間で 0.2 であった。

ショット長の分布に対して、分布の対称性を表す歪度、正規分布と比べた分布の鋭さを表す尖度を求め、分布の傾向について調査を行った。

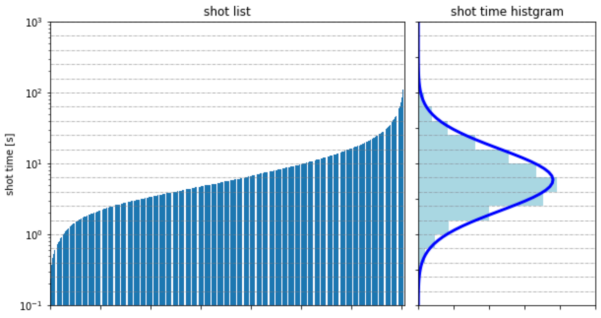


図1 時間知覚特性を考慮したショット長の分布（左）とそのヒストグラム（右）。縦軸はショット長を表す。左図の横軸に全作品のショット長を並べている。右図の横軸はショット長の頻度を表す。右図の青線はフィッティングした正規分布曲線を表す。

2.3. 結果

図1は、時間知覚特性を考慮したショット長分布と、ショット長のヒストグラムを示す。ショット長の頻度分布は、対称性を持っており（歪度 ±0.0）、正規分布より少し鋭い構造となること（尖度 0.78）が示された。尖度は正規分布（尖度 0.0）とロジスティック分布（尖度 1.2）の区間内であることから、正規分布に近い分布構造をとることがわかった。ショット長の分布に対して、正規分布曲線のフィッティングを行った結果、フィッティングした正規分布曲線において、ショット長の最頻値は 5.82 秒であった。

2.4. 考察

ショット長の分布は対数空間で正規分布に近い形であった。対数で対称性を持つ分布を示したことから、映画監督はショット長を対数尺度で捉えており、分布中央値の周辺のショット長を基準に、ショット長を設計している可能性が考えられる。視聴者が時間情報に対して知覚する情報量が、ショット長を対数変換した単位に漸近することを考慮すると、映画監督は視聴者の時間知覚特性を考慮してショット長を設計している可能性が示唆される。

ショット長の分布最頻値は 5.82 秒を示した。映画監督が視聴者の知覚特性を考慮してショット長を設計している可能性から、このショット長には、意識的か無

意識的かは不明だが、映画監督の何らかの意図が含まれている可能性が考えられる。映画は数時間に渡る文脈情報を視聴者に伝える必要があり、視聴者の認知的な負担も高いと考えられるため、ショット長あたりに含められる情報量や視聴者の認知負荷量などのバランスが考慮された基準値である可能性が考えられる。

ショット長のヒストグラムは、山なりの分布を示しており、フィッティングした正規分布曲線の傾向をよく表現していることが読み取れる。

3. n-gram 構造解析

2 章の解析から、映画作品に利用されるショット長が対数空間で正規分布に近い形をとることが示された。一方、ショット長は正規分布の出現確率の中で、ランダムに設計されているのか、それとも複数のショットのショット長遷移パターンに特徴があるのか、は不明である。そこで、本章では、ショット長の遷移パターンについて分析評価するために、言語解析における n-gram 構造[10]をショット解析に応用し、n-gram 構造の観点からショット長の遷移パターンにどのような傾向があるか明らかにする。

3.1. 時間長ラベル

言語解析を適用するための準備として、ショット長の符号化を行う。3 章では、ショット長のヒストグラムが正規分布曲線の傾向をよく表現していた。そこで、ヒストグラムの刻み幅を 1 つのラベルとした時間長ラベルを定義し、以降では、時間長ラベルの遷移パターンを用いてショット長の性質の解析を行う。一定の刻み幅で区切られた、i番目の時間知覚特性を考慮したショット長$t_{mind,i}$を用いて、時間長ラベルtは次式で与えられる。本稿では、刻み幅は 3 章のヒストグラムと同様の 0.2 とした。

$$t = \arg_i \Delta t_{mind,i}$$
$$\Delta t_{mind,i} = t_{mind,i+1} - t_{mind,i} = const. \quad (2)$$

3.2. n-gram 分析

言語解析では、構文の文法構造を解析するために、単語列や品詞列などの符号列を一つの単位として扱う[10]。単語や品詞は文法に基づいて並べられるため、部分的な符号列にも文法的な意味合いが含まれている。n 個の符号列で表現される単位は n-gram と呼ばれ、例えば、3 つの連続する単語からなる単位は単語 3-gram と呼ぶ。(赤い, 林檎, と, 黄色い, 梨) という例文を単語 n-gram (n=1, 2, 3) の遷移として表現すると以下のようになる。

1-gram (赤い、林檎、と、黄色い、梨)
2-gram (赤い林檎、林檎と、と黄色い、黄色い梨)
3-gram (赤い林檎と、林檎と黄色い、と黄色い梨)

この例文において、単語 1-gram では「赤い」のが「林檎」と「梨」のどちらであるかわからない。単語 2-gram では「赤い林檎」は存在するパターンであり、「赤い梨」は存在しないパターンであるため、統計的性質を用いて「赤い林檎」が妥当な単語列であると判断することができる。n-gram 構造の傾向は遷移パターンの解釈に利用可能である。

n-gram 構造を映画のショット長に応用することにより、短いショット長の遷移パターンを解析することが可能となる。本稿では、n個の連続するショット長で構成される単位をショット長 n-gram と定義する。例えば、時間長ラベル集合T、任意の時間長ラベル$t \in T$、映画作品のショット長$t_1 t_2 \dots t_h$が与えられたとき、映画作品のショット長の 3-gram の遷移は$t_1 t_2 t_3$, $t_2 t_3 t_4$, …, $t_{l-2} t_{h-1} t_h$となる。便宜上、本稿ではショット長 n-gram における 1 番目、2 番目、…、n 番目のショットの時間長ラベルを 1st shot、2nd shot、…、nth shot と記載する。

3.3. 方法

ショット長の遷移パターンの傾向を把握するため、2 つの n-gram 構造解析を行った。

(1) 映画作品の 2-gram のショット長の変化量について調査し、任意の時間長ラベルからの遷移パターンの傾向をみた。本稿での“ショット長変化量”とは、連続する 2 つのショットの間のショット長の差であり、差分を計算する際の 2 つのショット長は、離散符号化した時間長ラベルの数値での時間の差である。

(2) 映画作品の 3-gram のショット長変化量について調査した。1st shot と 2nd shot の変化量と、2nd shot と 3rd shot の変化量を求め、2 つのショット長変化量の関係性を検討した。

3.4. 結果

図 2 は 2-gram におけるショット長変化量を示す。横軸は 1st shot、縦軸は 1st shot と 2nd shot の時間長ラベルの変化量を示す。1st shot の時間長ラベルが大きくなるほど、2-gram のショット長変化量は減少する傾向 (r=0.95、p<0.001) が見られた。変化量の回帰直線

では、1st shot が 10.35 のときショット長変化量が 0 であり、分布最頻値（時間長ラベル 9.82）の周辺で変化量の正負が逆になることがわかった。これは、1st shot が分布最頻値より小さいときは、ショット長が長くなるように遷移する傾向があること、分布最頻値より大きいときは、ショット長が短くなるように遷移する傾向があることを意味しており、ショット長が分布最頻値に向かって遷移する傾向を示している。

図 3 は 3-gram におけるショット長変化量を示す。横軸は 1st shot と 2nd shot の変化量、縦軸は 2nd shot と 3rd shot の変化量を示す。第 1 象限は 1st shot、2nd shot、3rd shot の順に、ショット長が徐々に長くなるパターン（up-up）、第 2 象限は 2nd shot で短くなり 3rd shot で長くなるパターン（down-up）、第 3 象限は徐々に短くなるパターン（down-down）、第 4 象限は 2nd shot で長くなり 3rd shot で短くなるパターン（up-down）の分布を表現している。

3-gram の全分布のうち、16.1%が up-up、34.7%が down-up、14.0%が down-down、25.2%が up-down に分布しており、up-up と down-down の出現回数が、up-down と down-up の半数にも満たないことがわかった。

全分布に対して主成分分析を行った結果、第 1 主成分は 135 度の方向（水平成分 −0.704、垂直成分 0.709）を示しており、第 2 主成分は 225 度の方向（水平成分 −0.709、垂直成分 −0.704）を示した。第 1 主成分の寄与率 0.789 は、第 2 主成分の寄与率 0.210 の 3 倍以上の値を示していることから、135 度の方向への分布の広がりが大きいことがわかった。これは、up-down および down-up におけるショット長変化量が多様であることを意味する。

3.5. 考察

2-gram の解析では、ショット長の分布最頻値に近づくように、ショット長が遷移する傾向が示された。これは、ショット長の分布最頻値の周辺での遷移を繰り返す傾向があることを意味しており、映画監督が特定のショット長を意識的または直感的に調整して並べている可能性が考えられる。

3-gram の解析では、ショット長が連続して長くなる（短くなる）パターン A はショット長変化量の標準偏差が小さく、長短を繰り返すパターン B は変化量の標準偏差が大きいことが示された。セリフやアクションなど、特定のショット長を有する演出に合わせてショット長が調整されるのであれば、演出の多様性に応じて標準偏差は一様になると考えられる。しかし、パターンによって標準偏差が異なっていることから、映画監督の意図が、ショット長の遷移パターンに反映されている可能性が示唆される。特に、変化量の標準偏差が小さいパターン A は、セリフやアクションなどの演出とは独立した、別の意図の下で構成されている可能性が考えられる。

以上から、ショット長の遷移パターンには映画監督の意図が反映された文法規則があり、演出内容とは独立した役割を持つ可能性が示唆される。

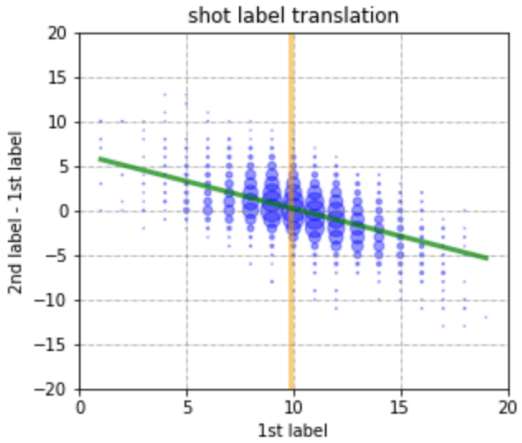


図 2　2-gram 構造におけるショット長変化量。青点は全映像作品における変化量のプロットであり、その大きさは頻度を表す。青点は任意の 1st shot における平均変化量であり、緑線はその回帰直線を表す。黄線はショット長の分布最頻値を表す。

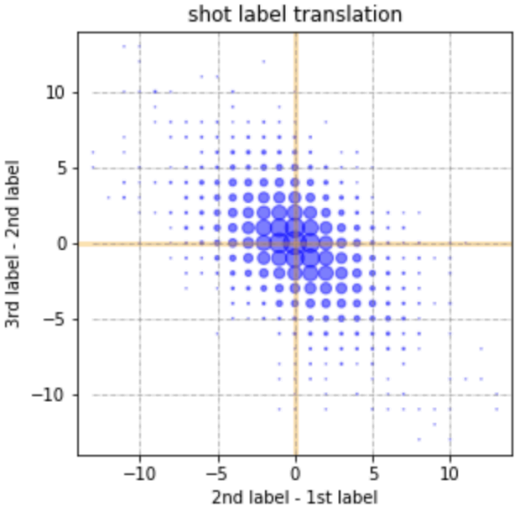


図 3　3-gram 構造におけるショット長変化量。青点の大きさは頻度を表す。

4. IDF 構造解析

3 章の解析から、ショット長の遷移パターンには文法規則があり、映画監督が遷移パターンよってショット長を使い分けている可能性が示唆された。しかし、異なる映画監督が共通して利用する文法規則がどのようなものであり、それが視聴者の印象評価にどのような影響を与えているか、は不明である。そこで、本章では、任意の n-gram がどのくらいの映画作品に共通するパターンなのかを分析評価するために、言語解析における逆文書頻度（Inverse Document Frequency、IDF）[11]をショット解析に応用し、複数の映画作品に共通するショット長の遷移パターンとその役割を明らかにする。

4.1. 逆文書頻度（Inverse Document Frequency）

言語解析では、文書間の単語 n-gram の利用率を測定する手法として逆文書頻度（Inverse Document Frequency、IDF）[11]が利用される。IDF では、多くの文書に現れる単語 n-gram ほど小さく、特定の文書に現れる単語 n-gram ほど高く採点される。高い IDF 値を持つ単語 n-gram は特定の文書を表現するための特徴と考えることができる。言語解析における IDF では、助詞や代名詞といった複数文書に共通する語句を低い値として検出することが可能である。

言語の IDF を映画のショット長に応用することにより、任意のショット長の n-gram がどのくらいの映画作品に共通するパターンなのか把握することが可能である。時間長ラベル集合Tの任意の時間長ラベル$t \in T$、映画作品集合Mの任意の映画作品$m \in M$、映画のショット長の符号列$t_1 t_2 \dots t_h$が与えられたときの、i番目のショット長の n-gram の IDF 値は次式のように表される。ここで、Nは映画の作品数$N = |M|$であり、$|\{v \in V : ngram_i \in v\}|$は$i$番目の n-gram が現れる映画作品数である。

$$IDF(gram_i^n) = log \frac{N}{\left|\{m \in M : gram_i^n \in m\}\right|} \quad (3)$$

$$gram_i^n = t_i t_{i+1} \dots t_{i+n-1}$$

4.2. 方法

IDF によるショット長の構造解析では、次の 3 つの解析を行った。

(1) IDF 値に対する 3-gram の頻度分布、種類分布、時間ショット長分布を調査した。頻度分布および種類分布は分布の合計が 1 になるように正規化した。ショット長分布は 3-gram における 1 番目から 3 番目のショットの時間長ラベルの値を個別に集計した。

(2) IDF 値の頻度分布の中央値で 2 つ（高 IDF 群、低 IDF 群）に分割し、各群に属する 3-gram の時間長ラベルの変化量の傾向を調査した。具体的には、ショット長 3-gram における 1 番目から 2 番目の時間長ラベルの変化量と、2 番目から 3 番目の時間長ラベルの変化量を求め、2 つの変化量の関係性について調査した。

(3) IDF 値で表現された 3-gram が映画作品においてどのような役割を持つか調査するため、映画作品の IDF による情報と映画作品情報との関係に対してピアソンの積率相関分析を行った。IDF 値による情報は、3-gram の作品内の平均および標準偏差とし、作品内の全パターン（all）、up-up、down-up、down-down、up-down の 5 群に対して求めた。

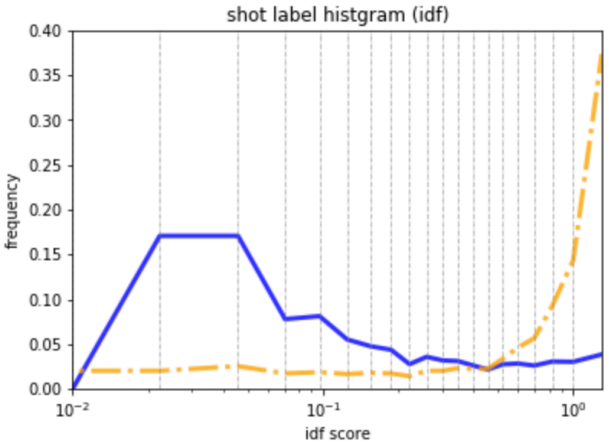


図 4 IDF 値に対する 3-gram の頻度分布と種類分布。青実線は頻度分布を表し、黄破線は種類分布を表す。

4.3. 結果

図 4 は、IDF 値に対する 3-gram の頻度分布、種類分布を示す。横軸は IDF 値、縦軸は分布割合を示す。頻度分布は、IDF=0.02、0.05 の 2 点で最頻値をとり、IDF 値が増加するにつれて頻度は減少する傾向（r=−0.52、p<0.05）がみられた。種類分布は、IDF 値が大きいほど種類数が多くなる傾向（r=0.53、p<0.05）が示された。IDF 値が小さい 3-gram は、種類は少ないが多く利用されることがわかった。

図 5 は、IDF 値に対する 1st shot、2nd shot、3rd shot の時間長ラベル分布を示す。3 つのグラフは上から順に 1st shot、2nd shot、3rd shot を示す。横軸は IDF 値、縦軸は時間長ラベルを示す。1st shot、2nd shot、3rd shot の全てにおいて、IDF 値が高くなるほどショット長のとる時間長ラベルの区間が大きくなる傾向（1st shot

は r=0.917、p<0.001、2nd shot は r=0.917、p<0.001、3rd shot は r=0.907、p<0.001）が示された。

図 6 は、高 IDF 群と低 IDF 群における、3-gram のショット長変化量の分布を示す。横軸は 1st shot から 2nd shot へのショット長変化量、縦軸は 2nd shot から 3rd shot へのショット長変化量を示す。低 IDF 群に比べて、高 IDF 群は広い範囲の分布を取っており、高 IDF 群に含まれるパターンは前後のショット長の差が多様であることがわかった。一方で、低 IDF 群のパターンはショット長の差が小さいことがわかった。

表 1 は、映像作品の IDF による情報と映画作品情報との相関分析結果を示す。all、up-down、down-up の標準偏差が、映画作品の評価値 1 との相関傾向が示唆された。all、up-up、up-down、down-up の標準偏差が、映画作品の評価値 2 と相関傾向が示唆された。また、分布の標準偏差と、評価値 4 との相関傾向が示唆された。

4.4. 考察

ショット長分布では、1st shot、2nd shot、3rd shot の全てにおいて、IDF 値が小さくなるほど狭い区間のショット長をとる傾向が示された。そのショット長はショット長分布最頻値を含む区間に分布していることが読み取れる。また、ショット長変化量の分布では、低 IDF 群のパターンはショット長の変化量が小さく、高 IDF 群のパターンはショット長の変化量が多様であることが読み取れる。このことから、複数作品に共通するパターンは、狭い区間のショット長の組み合わせとして表現されやすく、作品固有のパターンは、分布最頻値から離れた区間のショット長の組み合わせを取りやすいことが考えられる。また、異なる映画監督が共通したショット長を考慮している可能性を示唆する。

一方で、評価の高い映画作品は、遷移パターンの IDF 値の標準偏差が小さく、評価の低い映画作品は、遷移パターンの IDF 値の標準偏差が大きいことが示された。IDF 値が高いほどパターンの変化量が大きい傾向にあることから、多様なショット長の変化を伴う映画作品は低評価を招く傾向があることも示された。時間情報（楽曲）のテンポ感とヒトの好みの関係性について、Holbrook の調査[4]では、極端なテンポより中庸なテンポが好まれやすいことが報告されている。ショット長が多様に変化する映画作品は、ヒトのテンポ感の知覚に好ましくない影響を与えている可能性が考えられる。ショット長の変化が大きいパターンは作品個性を強調

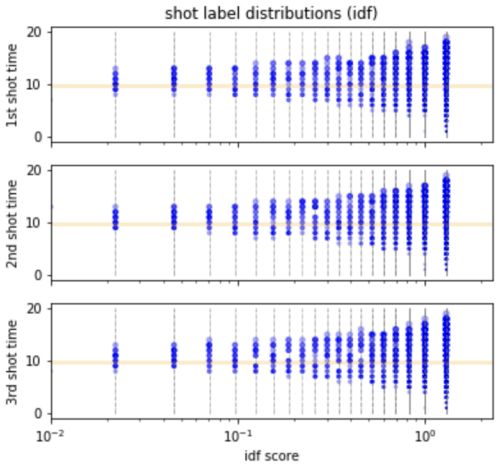
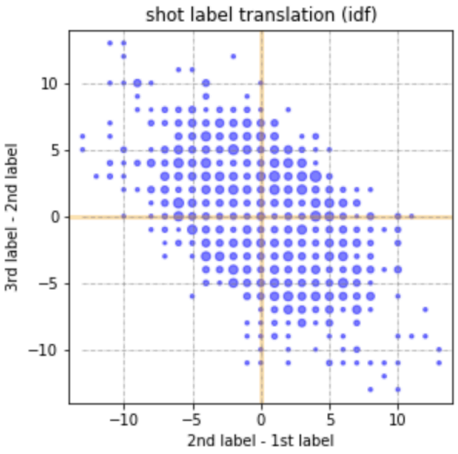
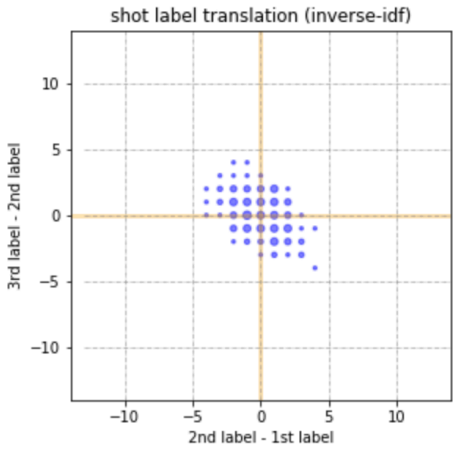


図 5 IDF 値に対する 3-gram の 1st shot、2nd shot、3rd shot のショット長分布。青点はショット長（時間長ラベルの値）、黄線はショット長の分布最頻値を表す。



(a) 高 IDF 群におけるショット長の変化量



(b) 低 IDF 群におけるショット長の変化量

図 6 高 IDF 群と低 IDF 群における 3-gram のショット長変化量。青点の大きさは頻度を表す。第 1 象限は up-up、第 2 象限は down-up、第 3 象限は down-down、第 4 象限は up-down の分布を表す。

表１　IDF 値と作品評価の相関分析結果（*p<.05、**p<.01、***p<.001）。

作品評価	IDF 値（平均）					IDF 値（標準偏差）				
	all	*up-up*	*down-up*	*down-down*	*up-down*	*all*	*up-up*	*down-up*	*down-down*	*up-down*
評価値５	−.10	−.19	−.16	−.01	−.12	−.19	−.22	−.31	−.05	−.12
評価値４	−.37	−.40	**−.45***	−.38	**−.46***	**−.67****	**−.59****	**−.73*****	**−.55***	**−.64****
評価値３	−.02	.09	.01	−.08	−.04	−.04	.05	.01	−.11	−.16
評価値２	.35	.44	.43	.30	.42	**.63****	**.59****	**.73*****	.39	**.57****
評価値１	.26	.27	.34	.22	.34	**.50***	.41	**.61****	.39	**.50***

することにつながるが、それを多用することにより映像の不快感を招きやすいことが示唆される。

以上より、作品共通のショット長の遷移パターンは、ショット長の変化量が小さい傾向にあり、視聴者の映像作品の低評価を退ける役割があることが示唆される。

5.　MLE 構造解析

4 章の解析から、映画作品に共通するショット長の遷移パターンが映画作品の高評価につながることが示唆された。しかし、作品単位の出現確率に基づく傾向では、ショット長の遷移パターンがどのくらい扱われるものであるか、また多く扱われるパターンが視聴者の印象評価にどのような影響を与えているか、は不明である。そこで、本章では、任意の n-gram がどのくらい映画作品に現れるパターンなのか分析評価するために、言語解析における最尤推定（Maximum Likelihood Estimation、MLE）[12]をショット解析に応用し、よく利用されるショット長の遷移パターンとその役割について明らかにする。

5.1.　最尤推定（Maximum Likelihood Estimation）

言語解析では、構文の単語分割や品詞検出に単語の遷移確率が利用される。単語の遷移確率を計算する手法には最尤推定（Maximum Likelihood Estimation、MLE）[12]がよく用いられる。言語における MLE は、先行する n-1 個の単語列から次に現れる単語の遷移確率を計算する。そのため、n-1 個の符号の遷移パターンに依存して、どのような符号が利用される傾向にあるか把握することが可能である。

言語の MLE を映画のショット長に応用することにより、n-1 個のショット長の遷移パターンから、どのようなショット長が利用される傾向にあるか把握することが可能である。時間長ラベル集合Tの任意の時間長ラベル$t \in T$、映画のショット長の符号列$t_1 t_2 \dots t_h$が与えられたときの、i番目の n-gram の MLE 値は次式で表される。$count$は全映画作品における頻度を表す。

$$MLE(gram_i^n) = P(gram_i^n | gram_i^{n-1}) = \frac{count(gram_i^n)}{count(gram_i^{n-1})} \quad (4)$$

$$gram_i^n = t_i t_{i+1} \dots t_{i+n-1}$$

5.2.　方法

MLE によるショット長の遷移パターンの解析では、次の 3 つの解析を行った。

(1) MLE 値に対する 3-gram の頻度分布、種類分布、時間長ラベル分布を調査した。頻度分布および種類分布は分布の合計が 1 になるように正規化した。時間長ラベル分布は 3-gram における 1 番目から 3 番目の時間長ラベルを個別に集計した。

(2) MLE 値の頻度分布の中央値で 2 つ（高 MLE 群、低 MLE 群）に分割し、各群に属する 3-gram のショット長変化量の傾向を調査した。具体的には、3-gram における 1 番目から 2 番目のショット長変化量と、2 番目から 3 番目のショット長変化量を求め、2 つの変化量の関係性について調査した。

(3) MLE 値で表現された 3-gram が映画作品においてどのような役割を持つか調査するため、映画作品の MLE による情報と映画作品情報との関係に対してピアソンの積率相関分析を行った。MLE 値による情報は、3-gram の MLE 値の作品内の平均および標準偏差であり、作品内の 3-gram の全パターン（all）、up-up、down-up、down-down、up-down の 5 群に対して求めた。

5.3.　結果

図 7 は、MLE 値に対する 3-gram の頻度分布、種類分布を示す。横軸は MLE 値、縦軸は分布割合を示す。頻度分布より、MLE=0.2 で最頻値をとり、99.0%の 3-gram が MLE=0.4 以下に分布していた。種類分布は、MLE=0.4 以下の範囲において、値が小さいほど種類数が多い傾向（r=−0.880、p<0.001）が示された。このことから、比較的少ない種類の 3-gram が多く利用さ

れる傾向にあることが読み取れる。

図 8 は、MLE 値に対する 1st shot、2nd shot、3rd shot の時間長ラベル分布を示す。3 つのグラフは上から順に 1st shot、2nd shot、3rd shot を示す。横軸は MLE 値、縦軸は時間長ラベルを示す。MLE 値が高くなるほど、3rd shot は狭い範囲の時間長ラベルをとる傾向（r=−0.618、p<0.01）が示された。値に対する 1st shot、2nd shot の分布の標準偏差には関係性が見られなかった。3-gram の MLE 値は、3rd shot に依存する傾向にあることがわかった。

図 9 は、高 MLE 群と低 MLE 群における、3-gram のショット長変化量の分布を示す。横軸は 1st shot から 2nd shot へのショット長変化量、縦軸は 2nd shot から 3rd shot へのショット長変化量を示す。高 MLE 群では、9.9%が up-up、36.5%が down-up、12.7%が down-down、40.9%が up-down に分布していた。低 MLE 群では、16.9%が up-up、32.2%が down-up、16.9%が down-down、33.9%が up-down に分布していた。高い MLE 値を持つパターンが、down-up と up-down に多く分布していることがわかった。

表 2 は、映像作品の MLE による情報と映画作品情報との相関分析結果を示す。all と down-up の平均および標準偏差が、作品の評価値 2 と相関傾向にあることが示唆された。特に、down-up の標準偏差と評価値 2 との間に強い相関傾向（r=0.64、p<0.01）がみられた。評価の低い映画作品は、down-up の MLE 値の標準偏差が大きい傾向がみられた。

5.4. 考察

時間長ラベル分布では、高い MLE 値の 3rd shot が、ショット長の分布最頻値の周辺に分布する傾向が示された。MLE による遷移確率の計算は、直前のショット長の遷移パターンから、次のショット長の出現確率を求めるため、3rd shot が分布最頻値に収束するパターンの出現確率が高いことは、妥当な確率計算の結果であると考えられる。また、高い MLE 値の 1st shot と 2nd shot がとりうるショット長の範囲が広いことから、様々な遷移パターンがショット長の分布最頻値に向かうように並べられている傾向が示唆される。

3-gram におけるショット長変化量の分布では、高い MLE 値を持つパターンが up-down と down-up に多く分布することがわかった。これは、1st shot から 2nd shot にかけてショット長が増加する（減少する）と、2nd shot から 3rd shot にかけてショット長が減少する（増加する）ようなパターンの出現確率が高いことを

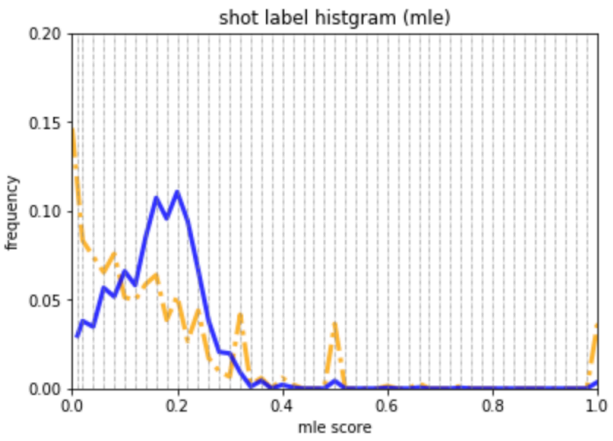


図 7 MLE 値に対する 3-gram の頻度分布と種類分布。青実線は頻度分布を表し、黄破線は種類分布を表す。

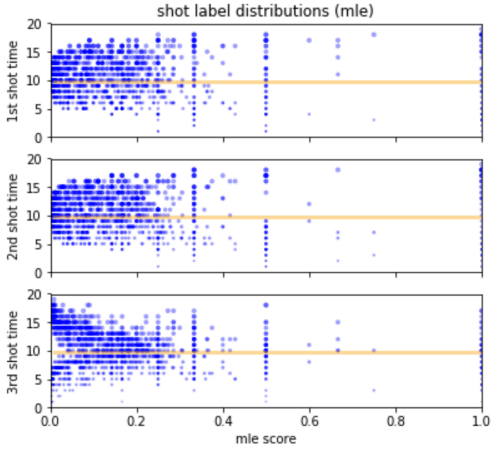


図 8 MLE 値に対する 3-gram の 1st shot、2nd shot、3rd shot のショット長分布。青点はショット長（時間長ラベルの値）、黄線はショット長の分布最頻値を表す。

意味する。ショット長の遷移パターンは、ショット長の長短を繰り返しながら並ぶ傾向があると考えらえる。

遷移パターンと映画作品の評価との相関分析では、評価の高い映像作品は、down-up が特定の MLE 値に偏る傾向があり、評価の低い映像作品は、down-up が様々な MLE 値となる傾向が示唆された。また、評価の低い作品は、down-up の MLE 値の平均が高い傾向も示唆された。このことから、遷移確率が高いとされるパターンを扱うことが視聴者の高い評価につながるわけではなく、遷移確率が均等になるようにショット長を構成することが効果的であると考えられる。また、遷移確率の低いパターンを多用する映像作品は少ないと考えられるため、好ましい遷移確率は MLE 値の分布最頻値（本稿では MLE=0.2）の周辺であると考えらえる。

表 2　MLE 値と作品評価の相関分析結果（*p<.05、**p<.01、***p<.001）。

作品評価	MLE 値（平均）					MLE 値（標準偏差）				
	all	*up-up*	*down-up*	*down-down*	*up-down*	*all*	*up-up*	*down-up*	*down-down*	*up-down*
評価値 5	−.43	.08	−.27	−.22	−.24	−.29	.00	**−.45***	−.15	−.14
評価値 4	−.05	.16	−.04	−.18	−.08	−.31	.02	−.37	−.31	−.28
評価値 3	.19	−.21	−.01	.19	.18	.13	−.13	.23	.09	.04
評価値 2	**.49***	−.01	**.46***	.23	.22	**.48***	.10	**.64****	.25	.37
評価値 1	.20	−.10	.14	.20	.14	.27	.00	.38	.26	.18

以上より、ショット長の遷移パターンは、ショット長の分布最頻値が基準となって構成されており、遷移確率が均等になるようにショット長を並べると、映画作品の高い評価につながる傾向が示唆された。

6.　まとめ

映画の時間変化に含まれる文脈情報は、視聴者の映画評価に関わる一要因になると考えられる。また、ショットを時系列に並べた情報であるため、特定のショット長の遷移パターンから映画の評価に関わる情報を獲得できることが期待された。上記の仮説を検証するために、本研究では 4 つの分析から次の考察を得た。

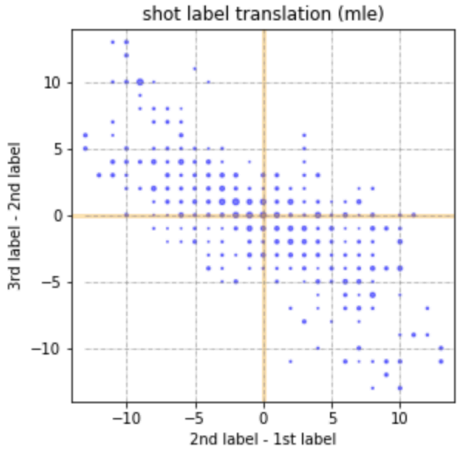
(1) ショット長の分布解析では、対数変換したショット長の分布が正規分布に近いため、ショット長にヒトの時間知覚特性が考慮されていることが示唆された。

(2) ショット長の n-gram 構造解析では、ショット長の分布に対応した遷移パターンが示され、ショット長の並びに文法規則がある可能性が示唆された。

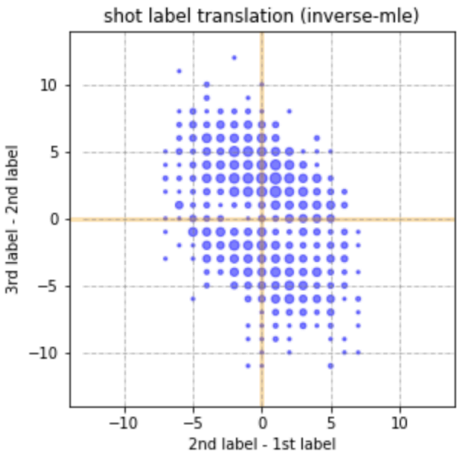
(3) 言語の IDF を映画に応用した解析では、多くの映画作品に共通するショット長の遷移パターンは、ショット長の分布最頻値付近のショット長で構成された、ショット長変化の小さいパターンであることが示された。また、作品共通の遷移パターンを多く利用するほど映画作品の高評価につながる傾向が示唆された。

(4) 言語の MLE を映画に応用した解析では、よく利用されるショット長の遷移パターンは、長短を繰り返すようなパターンであることが示された。また、MLE による遷移パターンが均等になるようにショット長を並べると、映画作品の高い評価につながる傾向が示唆された。

ショット長の分布解析と n-gram 構造解析では、ショット長にヒトの時間知覚特性が考慮されている可能性があり、その遷移パターンに特定の規則性を見ることができた。このことから、映画監督は何らかの意図に基づいてショット長とその時系列配置を設計しており、その意図には視聴者の時間知覚特性への配慮が含まれていると考えられる。また、IDF 構造解析と MLE 構造解析では、異なる遷移パターンの性質を採点しているが、両解析とも特定の遷移パターンの利用と視聴者の映画作品の評価に関係性が見られた。このことか



(c)　高 MLE 群におけるショット長の変化量



(d)　低 MLE 群におけるショット長の変化量

図 9　高 MLE 群と低 MLE 群における 3-gram のショット長変化量。青点の大きさは頻度を表す。第 1 象限は up-up、第 2 象限は down-up、第 3 象限は down-down、第 4 象限は up-down の分布を表す。

ら、監督が意図して配置したショット長の遷移パターンが、視聴者の印象評価に影響を与える一要因になっていることが考えられる。

本稿では、言語解析手法をショット長の解析に応用することにより、ショット長の遷移パターンに特定の規則性があること、特定のショット長の遷移パターンが映画作品の評価に影響を与えている可能性があることを見出すことができた。ショット長の遷移パターンには、監督の意図や視聴者の視聴特性が反映されている可能性があり、ショット長の遷移パターンから、映画の時間変化に含まれる文脈情報を映画監督または視聴者の観点から分析できる可能性が示唆された。本研究は、言語の文章構造解析が、映像のような非言語情報の時間文脈の特徴分析に応用可能であること、映像の時間文脈構造解析が、映像内容の評価や推定に利用可能である可能性を示唆する。

参考文献

[1] Chasanis, V. T., Likas, A. C., & Galatsanos, N. P., (2009) "Scene detection in videos using shot clustering and sequence alignment", *IEEE transactions on multimedia*, Vol. *11*, No. 1, pp. 89-100.

[2] Ohyama, J., & Watanabe, K., (2010) "Contextual segmentation in the visual stream produces a temporal synchronization effect on visual perception", Psychologia, No. 53, Vol. 2, pp. 125-137.

[3] Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P., (2004) "Attention and the subjective expansion of time", Perception & psychophysics, Vol. 66, No. 7, pp. 1171-1189.

[4] Holbrook, M. B., & Anand, P., (1990) "Effects of tempo and situational arousal on the listener's perceptual and affective responses to music", Psychology of Music, Vol. 18, No. 2, pp. 150-162.

[5] Yahoo Japan Corporation, (2018) "Yahoo!映画", https://movies.yahoo.co.jp/.

[6] Ornstein, R.E. (1969), "On the experience of time", Harmondsworth: Penguin.

[7] Kristofferson, A. B., (1984) "Quantal and deterministic timing in human duration discrimination", Annals of the New York Academy of Sciences, Vol. 423, No. 1, pp. 3-15.

[8] Jingu, H., (1989) "Numerical processing of duration in temporal tracking behavior", Japanese psychological research, Vol. 31, No. 4, pp. 169-178.

[9] Fetterman, J. G., & Killeen, P. R., (1992) "Time discrimination in Columba livia and Homo sapiens", Journal of Experimental Psychology: Animal Behavior Processes, Vol. 18, No. 1, pp. 80-94.

[10] Cavnar, W. B., & Trenkle, J. M., (1994) "N-gram-based text categorization", *Ann arbor mi*, Vol. *48113, No.* 2, pp. 161-175.

[11] Robertson, S., (2004) "Understanding inverse document frequency: on theoretical arguments for IDF", *Journal of documentation*, Vol. *60, No.* 5, pp. 503-520.

[12] Chen, S. F., & Rosenfeld, R., (2000) "A survey of smoothing techniques for ME models", *IEEE transactions on Speech and Audio Processing*, Vol. *8, No. 1*, pp. 37-50.