

修士論文

コメント付きユーザ配信型ストリーミングの
「面白い部分」の抽出によるアーカイブ化

**Extraction of “Enjoyable Parts” from Commented
User-delivered Streaming for Archiving**

2010 年 2 月 9 日

指導教員 相田 仁 教授



東京大学大学院

工学系研究科 電気系工学専攻

37-086467 大泉 拓

内容梗概

ニコニコ動画のようなコメントがつけられる動画サービスにおいて長時間の動画、シリーズ物の動画が増えている。一方ユーザー配信型の生放送サービスも普及しているが、生放送はフルサイズでアーカイブされてもそれを試聴することは時間的な負担が大きい。そこでこうした番組の面白い部分、話の流れの中で必要と思われる部分のみを抜き出したようなアーカイブを作ることによって視聴時間の短縮を測ることができないだろうかと考えられる。本研究では動画、番組に対して時間ごとに「面白さの値」というものがあると仮定し、これを用いて内容を圧縮したアーカイブを作成する事を試みた。この「面白さの値」を計算する方法、「面白さの値」から動画の編集領域を計算する方法を提案する。

目次

第1章	序論	1
1.1	はじめに	2
第2章	研究背景	3
2.1	関連研究	4
2.1.1	コメント情報を用いたダイジェスト動画作成支援	4
2.1.2	ニコニコ動画における映像要約とサビ検出の試み	5
2.1.3	映像と音声特徴に基づいた映画のダイジェスト生成	5
2.2	ユーザ配信型生放送	7
2.2.1	ニコニコ生放送	7
2.2.2	ニコニコ生放送の拡大	8
2.3	投稿型動画との比較	10
2.4	研究目標	11
第3章	提案手法	12
3.1	面白さの値によるアーカイブ化	13
3.1.1	編集構造	14
3.2	面白さの値	14
3.2.1	コメントによる面白さの値の計算	15
3.2.2	畳込みの概念	15
3.2.3	窓関数	17
3.2.4	意味タグ	20
3.2.5	音声による面白さの値の計算	24
3.2.6	映像による面白さの値の計算	24
3.2.7	オプティカルフロー	24
3.2.8	映像と音声による面白さの値について	27
3.3	クリッピング	27
3.3.1	音声によるクリッピング	27
3.3.2	映像によるクリッピング	28
3.4	編集領域	28
3.4.1	閾値を用いた編集領域の決定法	28
3.4.2	クリップ位置を用いた編集領域の決定法	28
3.4.3	編集領域の探索	28

3.4.4	アプリケーション	30
第4章	評価	32
4.1	実験	33
4.1.1	実験の設定	33
4.1.2	アーカイブの作成方法	33
4.1.3	アンケート	34
4.2	結果	35
4.2.1	アンケート結果	35
第5章	考察	38
5.1	編集領域の決定法	39
5.2	面白さの値	39
5.2.1	コメントによる面白さの値	39
5.2.2	映像と音声の面白さの値	39
5.2.3	面白さとは	40
5.3	ユーザと放送主のインタラクション	40
5.4	結論	41
	参考文献	43
	発表文献	45
	付録A ニコニコ動画のコメントXML	46
	付録B アプリケーションの設定方法	47
	付録B.1 畳み込みの設定	47
	付録B.2 編集領域の設定	49
	付録B.3 クリッピングの設定	50

目次

2.1	オプティカルフロー	6
2.2	ニコニコ生放送	7
2.3	生放送の来場者数のヒストグラム	8
2.4	生放送のコメント数のヒストグラム	8
2.5	コメントと来場者数の関係	9
2.6	一週間の生放送数	10
3.1	アーカイブの流れ	13
3.2	編集構造	14
3.3	窓幅ごとの畳込み	16
3.4	畳込みのユーザ耐性	18
3.5	畳込みのコメント数耐性	19
3.6	ガウス窓	20
3.7	三角窓による畳み込み	21
3.8	“w”の窓関数無しの畳込み	22
3.9	“w”の窓関数有りの畳込み	22
3.10	“w”のコメント数と文字数の関係	23
3.11	音声振幅の畳込み	25
3.12	オプティカルフローの畳込み	26
3.13	閾値の変化による編集領域割合の変化のグラフ	29
3.14	閾値の変化による編集領域割合の変化	30
3.15	アプリケーション	31
4.1	続きの動画が必要か	36
4.2	動画は面白かったか	36
4.3	繋ぎは不自然だったか	37
4.4	動画の長さは短くした方がいいか	37
5.1	面白さとは	41
.2	畳み込みの設定	47
.3	窓関数の設定	48
.4	編集領域の設定	49
.5	クリッピングの設定	50

表目次

2.1	作成時間とマウスロガーの結果	4
2.2	ニコニコ生放送の一週間の番組数	9

第1章

序論

1.1 はじめに

YouTubeなどの動画共有サイトの普及後、ニコニコ動画等のコメントをつけられる動画サイトが普及した。ニコニコ動画では現在多種の動画が投稿されているが、近年は各メディアの公式コンテンツが配信されたり、ユーザによる番組形式、長時間の連載物の動画もアップロードされるようになった。またニワンゴは2008年末よりニコニコ生放送というユーザ配信型の生放送のサービスを開始した。ニコニコ生放送はこの一年で急激に成長し、現在では一日に5万件以上もの番組が放送されるようになった。しかしこのユーザ配信型の生放送は1素人であるユーザが配信する放送であるため冗長性がとても高い。また実況プレイ放送等を毎日している放送主もいるため、一日放送を見逃しただけで話についていけなくなっている。ニコニコ生放送では2009年末よりすべての放送の1週間のフルアーカイブを提供している。これにより見逃した番組も見れるようになり、必要であればアーカイブを録画しそれを別に保存しておくことも可能となった。しかしこうした放送の話題についていくために毎日数時間も放送された動画を見るのは時間を大量に消費するため困難である。一方、動画やユーザ配信型の生放送は日々増加しているため、ユーザは大量に放送されている動画の中から自分に合った放送を実際に見て判断するのが難しい状況も生まれている。そして両者に言えることであるがコンテンツの氾濫によりユーザの選択肢は増えたもののコンテンツ選択が困難な状況が生まれている。

そこでこうした番組のおもしろい、もしくはダイジェスト部分のみを抜き出したようなアーカイブを作成することによって、この状況を解決できるのではないかと考えた。そこで本研究ではニコニコ動画に代表されるようなコメント機能を用いてアーカイブを作成することを目標とした。

また単にアーカイブを作成するだけでなく、発展途上であるユーザ配信型のストリーミングにおいて本研究から逆説的に提案できる番組の形態の研究、放送支援も目標とした。

第2章

研究背景

2.1 関連研究

2.1.1 コメント情報を用いたダイジェスト動画作成支援

ニコニコ動画に関する先行研究として磯貝ら [1] はコメントを利用したニコニコ動画のダイジェスト動画を作る研究を行っている。

磯貝らは被験者に動画のダイジェストシーンとおもしろいと思うシーンを選んでもらい、その選択件数とコメントの量の相関係数を得るという実験を行った。ここでのダイジェストシーンとはその動画の内容を把握するに当たって必要とされる部分である。被験者は学生20人で3つの動画を対象とし、それぞれの動画から10,000件のコメントを抽出し各時間におけるコメント量のグラフを作成する。そして1シーン3秒として5箇所のシーンを選んでもらう。選ばれたシーンはまた3秒に分割されそれぞれの時点において選択係数をカウントする。この実験をもとに相関係数を計算すると選択係数とコメント量は相関が低いことがわかった。これは視聴者がコメントを打つのに数秒時間がかかっているからと述べている。そこで磯貝らはコメント量を0秒～12秒の間でずらし、それぞれの場合において相関係数を取得しなおした。これによるとダイジェストシーンの選択係数とコメント量の相関係数は対象とした3つの動画いずれの場合においても「3秒間」ずらした際に相関係数が最大となることがわかった。同様におもしろいシーン選択係数とコメント量の相関係数は1つの動画では5秒、二つの動画では3秒ずらすことによって相関係数が最大になることがわかった。これによりコメントを打ちこむのには約3～5秒かかっていると考えられる。

次に磯貝らはコメントの量をもとにダイジェストシーン編集の補助をするという実験を行った。コメント量の多い部分のみを抽出すると偏りによっては局所的な抽出しかできないため、動画を一定間隔で区切り、その中でコメントが多かったシーンを抽出する。前の実験を元にコメントは全て4秒間ずらして計算されており、抽出されたシーンはサムネイルとともに編集者に提示される。具体的な時間間隔の設定や抽出方法は述べられていない。

そしてこのシステムとWindowsに付属しているWindows Movie Maker(以下WMM)単体での編集作業との比較をおこなった。マウスロガーによる作業量の観測と作業時間の計測を行い、被験者にアンケートを行った。被験者は学生10人である。

実験による作業量の比較は表2.1のとおりである。Timeは作業時間。Clickはマウスのクリック数、Moveはマウスの移動量である。作業時間および、作業量は磯貝らの提案手法

Table 2.1: 作成時間とマウスロガーの結果

WMM			磯貝らのシステム		
Time(sec)	Click(回)	Move(m)	Time(sec)	Click(回)	Move(m)
1138.4	891.4	56.9	327.2	202.2	21.9

を用いることで減少していることがわかる。被験者のアンケートの結果によると、作りやすさについては比較的作りやすいという意見が報告されている。しかしコメント量によってシーンを抽出するのでダイジェスト動画として必要なその動画の要点が得られないという意見もあった。

この研究では編集はあくまでユーザが行っており、コメントの利用も単位時間におけるコメント量を比較したものである。我々の研究は生放送も対象としているため、投稿型のサービスとは違いコメント量が少なく単純にこのシステムを流用することは難しい。

2.1.2 ニコニコ動画における映像要約とサビ検出の試み

青木らは[2] コメントを用いた動画のサビ検出に関して、動画の種類ごとに検証を行っている。青木らは各種の動画(音楽、ゲームのプレイ動画、ニュース、料理、エンターテインメント)において動画のサビ部分でどのようにコメントがなされているのか調査し、コメント量に基づいたサビ抽出を検証している。まず音楽分野についてはサビの抽出がその歌の特徴に左右されると報告している。音楽の動画におけるコメントは印象的な歌詞の部分に集中される。そのため、サビの部分にキャッチーな歌詞がない音楽動画に関してはサビの抽出がうまくいかないことがあるとされる。しかし楽曲構造と完全に異なるピークを持つ例は少なく、サビの冒頭、キャッチーな部分にコメントが集中することが報告されている。他種の動画に関しても同様の調査を行ったところ、広義の面白い部分が抽出されるため要約映像に取り入れる価値は高いが、面白さを理解するためのコンテキストが他にある場合は十分に面白さが伝わらないことがあるとされている。しかし要約されたことによって新しい面白さが生まれる可能性があり、新たな動画形態の提供できる可能性を述べている。

しかしニコニコ動画の要約については需要について見直す必要がある。例えば音楽の動画はフルサイズでもせいぜい5分程度であり、それほど時間的な負担ではない。その他ニコニコ動画で人気のあるコンテンツについて調査をしたところ、2010年1月25日現在、再生数TOP100位にある動画の平均時間は7分30秒であった。しかしこの中には曲の詰め合わせなどがあり、これらは極端に時間が長く、また動画の趣旨として要約する対象とならない。これらの動画を除いて一つ作品として加工編集されたものに限れば平均時間は5分30秒程度である。つまり人気のある作品はそれほど長いものはない。そもそも投稿された動画は投稿者が編集したもの、つまり一度投稿者によって要約されたものであり再度要約することの必要性が薄れる。

2.1.3 映像と音声特徴に基づいた映画のダイジェスト生成

画像処理、音声認識の分野では過去より動画の要約に関する研究がなされているが、その中でも本研究にとって有用だと思われる鹿嶋ら[7]の研究を紹介する。

鹿嶋らは映像、音声から映画のダイジェストを生成する研究を行っている。まず音声については映画がステレオで作られていることを利用して、左右の音量の差が大きいところを抜き出す。これは左右の音量の差が激しいところでは動画に何らかの展開がある部

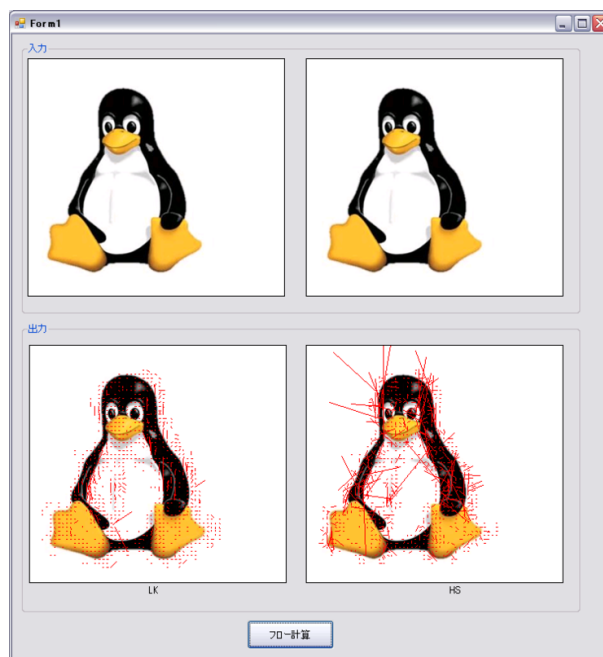


Fig. 2.1: オプティカルフロー

分であるという考えによるものである。またフーリエ変換による周波数解析によって会話部分を抽出する。会話部分は他の部分 (BGM だけの部分等) に比べて低い周波数にスペクトルが集中する傾向があるため、これを利用して抽出する部分を決定する。

映像に関しては Fig.2.1 オプティカルフロー [5]、[6] を用いたフロー解析を元に抽出を行う。オプティカルフローは映像解析分野ではよく用いられる手法であり、鹿嶋らの利用している勾配法では時間的、空間的な輝度勾配の拘束方程式を用いてフローを求める。鹿嶋らは処理時間を考慮し、オプティカルフローの計算では LK 法を用いている。映像、音声から求められた抽出領域を比べて同シーンであった場合は映像フロー解析の結果を利用する。この手法を用いた 4 つの映画のダイジェスト動画に関してダイジェストになっていたかどうかのアンケートを被験者 20 人に行った。結果によると 5 段階評価で 4 以上をダイジェスト動画として適当であるとした場合、非視聴者では 50% ~ 60% の割合で適正という結果になった。一方映画を既に視聴していた人々では 4 以上の割合が多い結果となった。つまり映画を見たことがない人にとっては内容を把握しきれなかったということであり、鹿嶋らの手法では映画の見所を抽出するのには向いているがダイジェストには向いていないためであると報告している。

鹿嶋らの手法は映像、音声の意味を解釈しているわけではないのでダイジェスト動画を生成するためには足りない情報が多い。しかし、一度視聴した人間にとっては内容を復習できる範囲でダイジェスト作られていると考えることもできる。

ニコニコ放送 ログインユーザー: blaze* プレミアム会員です! [ログアウト]

NICO NICO LIVE

2009年07月30日 10:00 からスタートしています (放送者 わろかいさん)
適当にゲームをやろう!!!
 みんなで楽しくゲームやりませんか?
 ひとりやってつまらないと感じても皆でやったら案外楽しいものですよ
 主のキャラ
 キャラ名 [21].blaze*
 KD 56%
 階級 少尉4
 所属クラシ ヨウの死者ヨウ
 コテハンは換装程度でお願いします。
 コメントはサブPCで読めます。
 ちなみに主もいつも以上に今日は下手ですのでご注意ください。
 主のブログ<http://bureizusablog73.fc2.com/>
 twitter <http://twitter.com/blazewarosu>
 質問が多すぎる、いろいろ漏れ入は即NG突込みますお！
 13時からサドンアタックやります。

2009-06-11 01:07:56作成
 わろかいの生放送部
 とりあえずは放送は不定期です。
 SAの実況をするかCSSの実況をするか……。主のスペック 自宅警備員15歳 特技 FPSゲーム 苦
 レベル:14/参加人数:130

この放送をしているコミュニティに参加しますか?
このコミュニティに参加する

ニコニコ情報局 ニコニコ動画からのお知らせや、最新ニュースなどの情報をお伝えします!

ニコニコ掲載! 産経にWEB面登場

口ってるなあ 延長しろwww
 しむらー！うしろー！

最新コメント

時刻	コメント	コメント番号
12:53	crash the castleくりあしたー	1979
12:53	ミキサーきれたな	1980
12:54	俺はSF→SA→ペーパーマン→マインクラフト	1981
12:54	今更クラッシュがキャスルやってきた	1982
12:54	それなんていうゲーム?	1983
12:54	40秒で設置しな	1984
12:54	3分でキュービー3分クッキングを終わらせる(本当)	1985
12:54	しくらぐらい?	1986
12:54	あと6分ですね。	1987
12:54	ソースうま(口)	1988
12:54	もうちょっとでSAできるぞー↑	1989
12:54	俺主の部屋に行くわ	1990
12:54	値段いくらですか?	1991
12:54	グロってるなあ	1992
12:54	しむらー！うしろー！	1993
12:54	延長しろwww	1994

経過時間 174:42

来場者数 370

総コメント数 1,985

ニコニコ放送 座席表
 co36907-14番

Fig. 2.2: ニコニコ生放送

2.2 ユーザ配信型生放送

ユーザ投稿型の動画が普及するとともに、携帯電話、携帯ゲーム機、その他新しいディスプレイ付きのガジェットの普及により人々はどこでも動画がみられるようになった。このような再生環境の普及により今後も動画の需要は高くなると思われる。

一方でニコニコ生放送に代表されるようなこうした生放送分野は未だ成長段階にあり、これまでの動画技術をそのまま利用していいものか調査する必要がある。

2.2.1 ニコニコ生放送

ここでは本研究の主な実験対象とするニコニコ生放送 (Fig.2.2 参照) について説明する。ニコニコ生放送では放送主は 30 分の放送の権限を放送前に得て放送することが可能で、現在は最大で 1800 枠の番組が一度に放送できる。実際には運営公式放送の定員にも左右

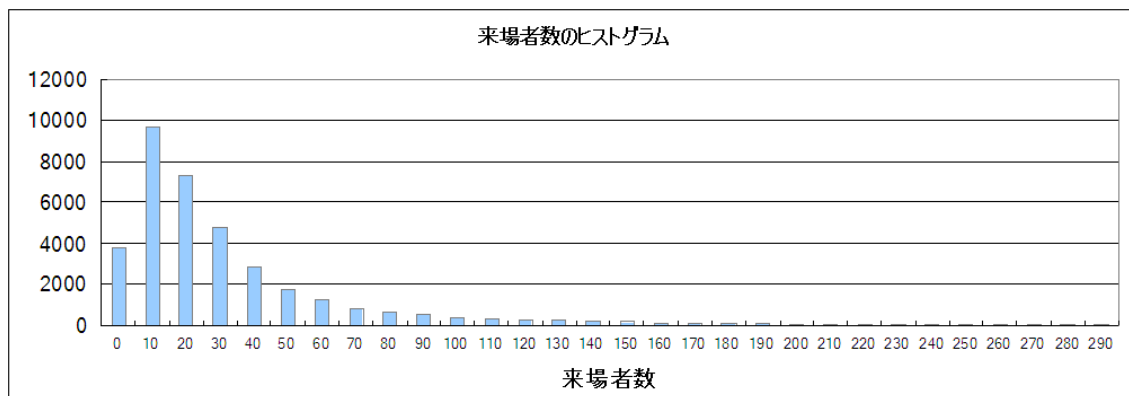


Fig. 2.3: 生放送の来場者数のヒストグラム

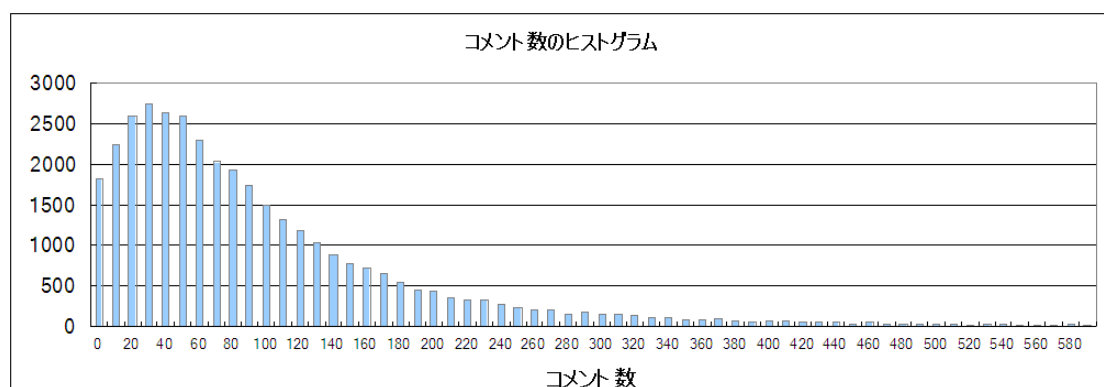


Fig. 2.4: 生放送のコメント数のヒストグラム

されるため必ずしも 1800 枠になることはなく、時間によっては放送番組数がこの 1800 未満になることもある。予約をすることによって時間を指定することも可能であり、放送終了時間前に延長をすることで最大 300 分まで放送可能である。

2.2.2 ニコニコ生放送の拡大

11 月 30 日 13:00~12 月 1 日 12:59 の間に放送されたニコニコ生放送の 30 分番組について調査した¹。Fig.2.3 は来場者数、Fig.2.4 はコメント数のヒストグラムである。来場者数は放送中にその番組を視聴した視聴者の総人数であり、最大人数などの瞬間値については取得ができない。来場者の平均値は 49、中央値は 26、コメント数の平均値は 119、中央値は 76 である。来場者数が 200 人以下の番組が 96% で、コメント数は 400 以下の番組が 95% であった。このように生放送で得られるコメント量は投稿型のストーリーミングサービスに比べて少なく、磯貝らの研究のように 10,000 件のコメントを用いることはできない。

¹この頃はまだ放送枠は 30 分 1500 枠であった。

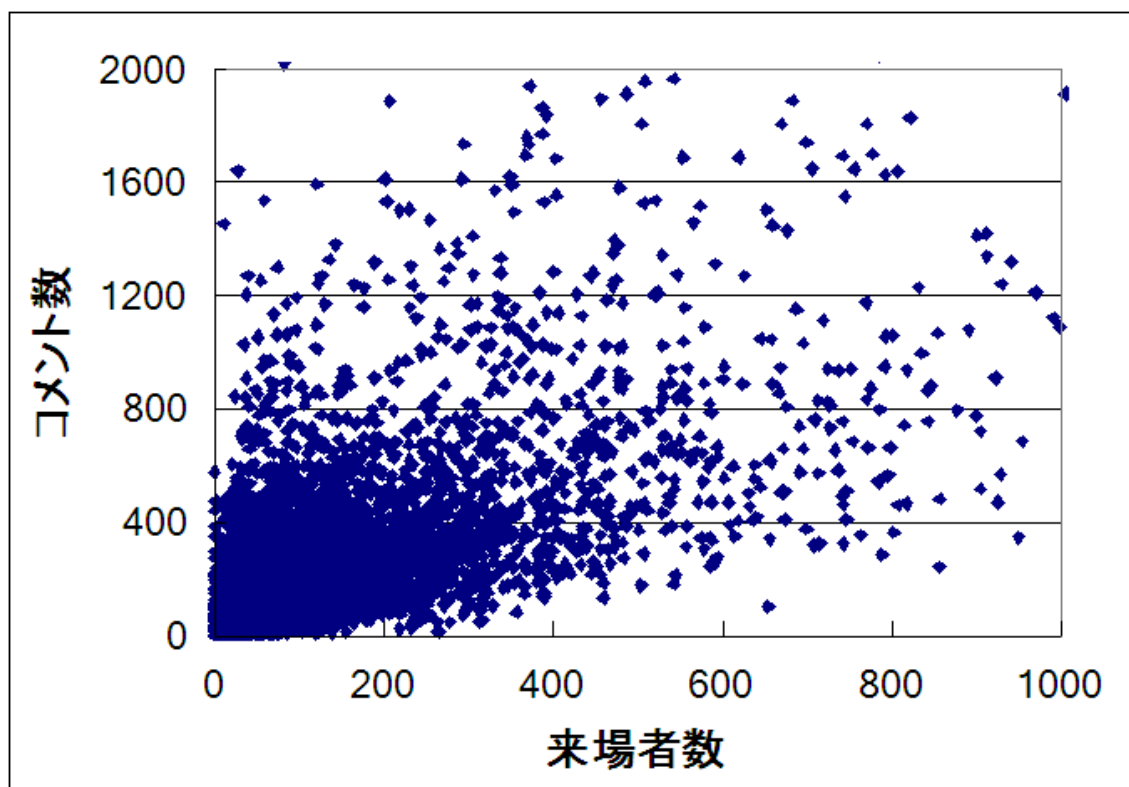


Fig. 2.5: コメントと来場者数の関係

コメントと来場者の関係はFig.2.5で、来場者数とコメント数の相関係数は0.6966であった。あまり強い相関はあるとは言えないが、おおむね来場者数が増えるに従ってコメント数は増える傾向が読み取れる。

Fig.2.6は11月21日から11月27日の間に放送されたニコニコ生放送の番組数を一時間ごとにプロットしたものである²。具体的な番組数は表2.2のとおりである。

現在は番組数がこの頃よりも増えており、1日で50,000番組以上放送されている。中には1日に何度も放送をする人や、実況プレイ放送³などの続き物の放送をする人は多い。このような放送主は毎日のように放送することも多く、視聴者も継続して見ることが多い。現在はアーカイブが一週間限定で配信されているが、話題を得るために30分以上の動画を見返さなければならないのは視聴者には時間的な負担が要求される。また動画もすべ

Table 2.2: ニコニコ生放送の一週間の番組数

11月21日(土)	22日(日)	23日(月)	24日(火)	25日(水)	26日(木)	27日(金)
49100	49950	49340	40910	41560	40600	44350

²木曜日にある空白の時間は恒例メンテナンスの時間である。

³ゲームをプレイしながらコメントと対話する放送。ニコニコ動画で人気があるコンテンツである。

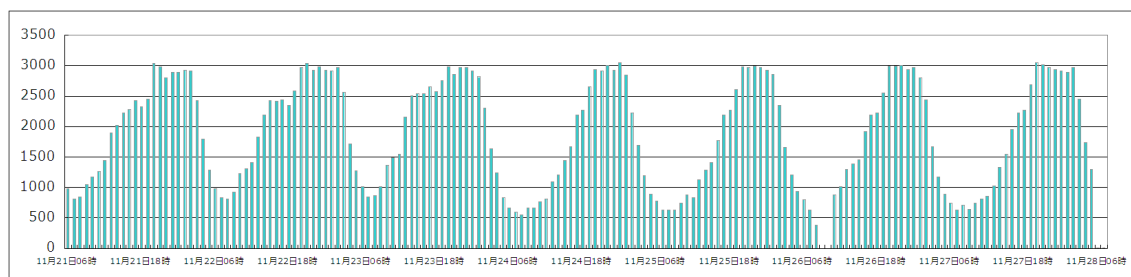


Fig. 2.6: 一週間の生放送数

てが面白い部分であるとは限らず、アーカイブにしてしまった時点でその放送のリアルタイム性、インタラクションというコンテンツの魅力が欠けている状態であり、それを意欲的に見ることは難しいと考えられる。放送枠が増えるにしたがい視聴者が多くの放送を把握するのは困難になっており、おもしろい放送があるのか実際に自分で見て確認するのは不可能になりつつある。

もし番組が短く編集されてアーカイブされていたとしたら、放送主が新しい視聴者を獲得するために過去の放送を参照することによって話題作りが容易になる可能性がある。視聴者も新しい番組を見始めるに当たって短いダイジェスト動画を見ることが可能ならば、自分が好みの番組であるかどうかを簡単に判断することができる。

2.3 投稿型動画との比較

現在ニコニコ動画では総動画数は360万、総コメント数は23億で、一つの動画につき平均638のコメントがあることになる。そのため生放送に比べるとその動画の視聴者の反応を伺いやすい。

この中でもニコニコ動画では人気を集めている実況プレイ動画という動画に注目する。実況プレイ動画はニコニコ動画では40万件以上⁴投稿されているカテゴリで、割合でいうと最も多い部類にある⁵。実況プレイ動画はゲームを実況しながらプレイして録画したものを投稿しているコンテンツである。近年はプレイしたものを投稿主独自の編集を施して投稿されている動画もあるが、基本的にはプレイした内容がそのまま投稿されている。そのため一つの動画の長さが20分を超えることが多く、パート数も数10～数100に及ぶものが多い。こういった特性は30分の番組を数回繰り返すような生放送やラジオなどと非常によく似ている。またニコニコ動画では各メディアの公式コンテンツの配信も開始しており、ラジオ等も配信されている。ラジオは編集されて配信されている事が多いが、回数が非常に多く毎週大量に配信されるコンテンツを複数個試聴するのは非常に困難である。本研究ではこうした生放送と似た性質をもつコンテンツも研究の対象とする。

⁴現在ニコニコ動画の総動画数は360万程度

⁵ただしゲームプレイ実況自体は著作権的にはグレーな存在である。例えば新作のゲームの実況動画は削除されることが多く、すべての実況動画は権利者に黙認、または放置されているだけの状態である。本研究は研究対象としてゲームプレイ実況を利用するが、その正当性を主張することは意図していない。

2.4 研究目標

以上の背景を踏まえた上で本研究の目的を明確化する。

まず要約の対象とする動画の時間は20分を超えるような長時間なもの、もしくは連載していて長期的に放送、投稿されているものに限る。また複数の動画及び楽曲等を単純に繋ぎあわせた内容の動画に関しては動画要約の必要性が薄いため除外する。生放送の種類の中の動画のリクエスト垂れ流し放送⁶も同様の理由で除外する。これはあらかじめ編集されたもの、圧縮されているものはそもそも再編集の必要性が薄いためである。

次に主に要約の対象とする動画内容に関してはパーソナリティが存在し、会話及びストーリー性のある物に限る。カテゴリとしては実況プレイ、ラジオ、雑談等が該当する。また以上の定義に含まれる生放送以外の投稿動画に関しても研究の対象とする。これはインタラクション性を低いことを除けば、生放送よりも定性的な結果を得やすいといった特徴があり、コメントの特性をつかむ上では重要であるためである。

しかし以上の定義に含まれなかった動画に関しても要約以外の考察においては積極的に利用する。そしてこのデータを元にどのような放送をするのが効果的であるのかを検証することも本研究の目的の一つとする。

また抽出すべき編集領域を総じて“面白い部分”と呼び定義する。この“面白い部分”は広義として「視聴者が興味を抱いている部分」、そして「興味を抱くために必要な部分」を含む領域と定義する。この定義によって本研究によって作られたアーカイブが、単純に視聴者が興味を抱いた部分だけではなく、ダイジェスト動画として利用できるようにするためである。

⁶リスナーから動画のリクエストを得てその動画を流す放送のこと。

第3章

提案手法

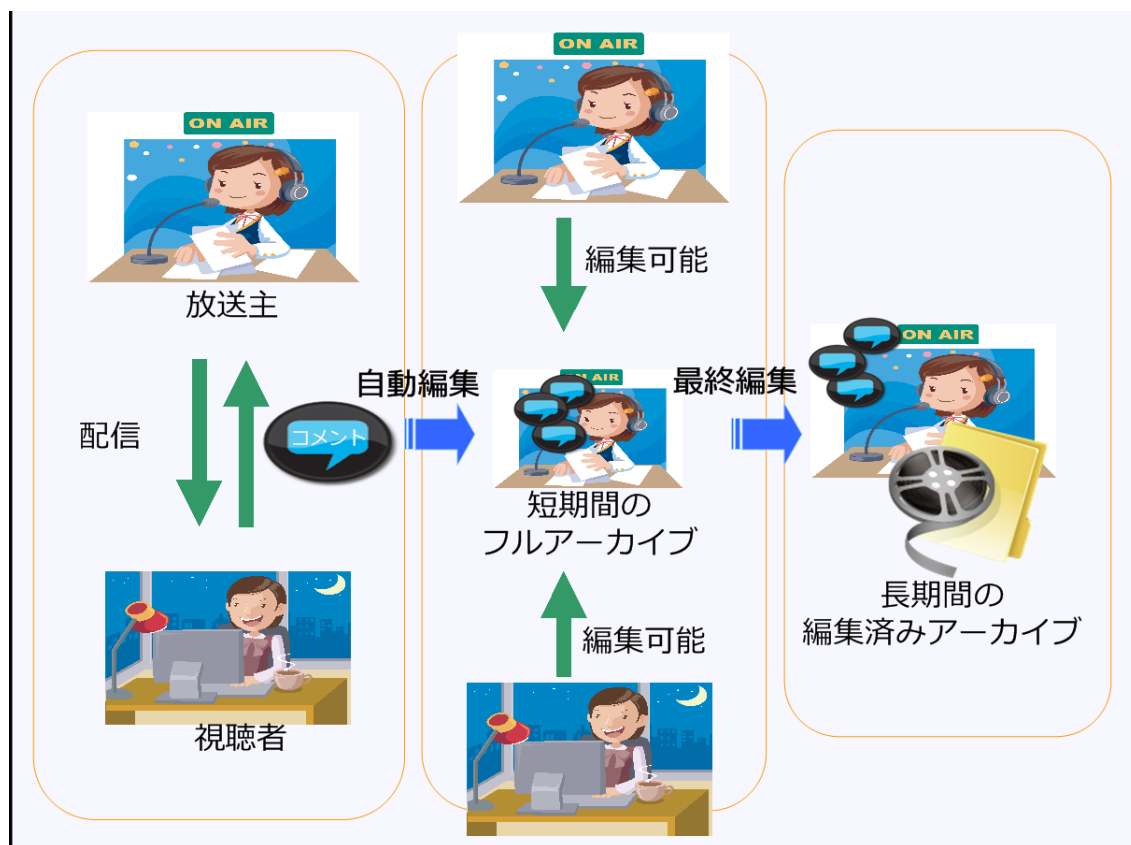


Fig. 3.1: アーカイブの流れ

3.1 面白さの値によるアーカイブ化

Fig.3.1 は提案手法におけるアーカイブ化の流れを表した図である。アーカイブの作成にあたって生放送のフルサイズの録画とそのコメントの短期的な保存を仮定する。これは前章で紹介したようにニコニコ生放送が1週間のフルアーカイブ化を開始したため技術的に可能である。

次にフルサイズの録画を元に自動編集機能を用いて短縮された仮編集動画を放送主、視聴者に公開する。このフルサイズの録画保持期間では編集領域を手動でも変えられるようにする。そしてフルサイズの録画期間の終了前に放送主は編集された動画を自分の動画サーバーにアップロードする。もちろんサービス側がフルアーカイブ可能であれば、再エンコードをする必要がなく編集領域だけを再生するようにすればよい。またその場合、仮編集状態を常に保つことによって放送主、および視聴者は自分の好きなようにカスタマイズすることが可能である。

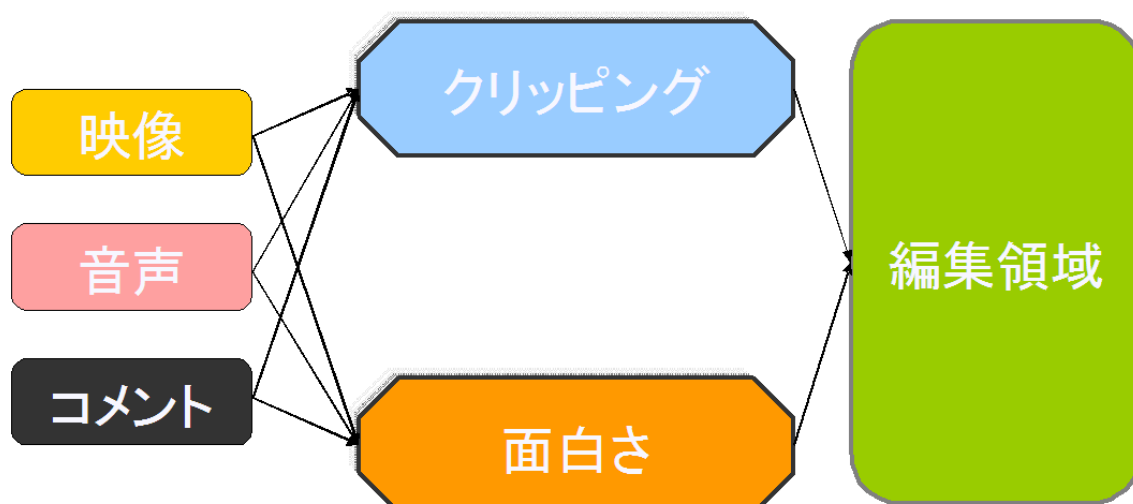


Fig. 3.2: 編集構造

3.1.1 編集構造

次に提案手法の編集構造 Fig.3.2 を説明する。まず元の動画の映像、音声、そしてコメントからクリッピング領域を計算し、同様に映像、音声、コメントから面白さを計算する。そして求められたクリッピング領域、面白さの値から編集領域を決定する。ここでコメントを元の要素として用いているがここにあたるものはなにもテキスト化された視聴者の反応だけではない。今後視聴者の反応、意見が反映されるようなサービスが新しく追加された場合はもちろんそれを利用してもよい。

3.2 面白さの値

磯貝らの研究ではニコニコ動画のダイジェスト動画を作成する際に 10,000 件のコメントを用いているが、生放送等のコメントは時間が制限されているためそのまま利用することが不可能である。また前述のように生放送のコメントは 30 分で 100 件やそれ以下のオーダーがほとんどである。このコメント量で各単位時間におけるコメントを見ても目で見てわかるほどの差異を感じづらい。そこで各単位時間においてコメント量の畳み込み演算などを試みるが、その前に動画の編集領域の決定方法を説明する。動画のどの部分を編集するかを決める基準のようなものがあると便利であると考え、ここで以下の P という値を定義する。

$$P = P_v + P_s + P_c \quad (3.1)$$

P_v 、 P_s 、 P_c はそれぞれ、映像、音声、コメントから求められる面白さである。 P は映像、音声、コメントすべてを考慮したその場の面白さの値である。以降それぞれの P について解説する。

3.2.1 コメントによる面白さの値の計算

Fig.3.3の窓幅1は各時点におけるコメント量をグラフにしたものである¹。以後、動画のある瞬間的な時間を“時”と呼ぶ。これを見るとコメントが分散しすぎてどの部分にコメントが集中しているのかわかりにくく、時間的な推移も把握しづらい。そこでまずコメントを畳み込むことによって視覚的にわかりやすくことにした。 x を放送時間内における時間的位置 t 、 N_C を各時間におけるコメント数とする。また本研究では時間の最小単位を1秒とし、以後断らない限り時間の分解能は1秒である。

$$Pc(x) = \sum_{t=\tau}^{t=\tau+W} G(t)N_C(x+t) \quad (3.2)$$

この畳み込み計算においては、ある一定の窓幅 W を持つ畳み込み窓を、どれだけずらすか τ という考えのもと定義されている。現段階では G は1として単なる矩形窓とした。これを元に畳み込みを行ったのがFig.3.3である。

窓幅が1の場合がその時点で投稿されたコメントのみを表す。最大でも6のコメントしかなくやはりこれによってシーンの優劣を決めるのは難しい。順に窓をふやして行くとコメントの密度がどのように変化しているのかが視覚的にわかる。ここで窓幅が10と窓幅が60の場合を見てみると、断続的なピーク値の変動がなくなり長期的な山ができていくことがわかる。他の放送のコメントを用いてもこの現象は窓幅が30から60になる時点で見る事ができた。

以後この畳み込みで得られた値をコメントによる面白さの値とする。

3.2.2 畳み込みの概念

各時点付近のコメント数を畳み込みすることによって視覚的に情報を得やすくするという考えから畳み込みを行った。しかしこの考え方は視覚的な根拠だけに基づいているわけではない。ある時点の映像、音声に対する視聴者の反応はその時点以後現れる。その反応は各視聴者ごとに様々であり、直後に反応する人もいれば若干時間がかかる場合もある。また通信網、サーバなど利用者ごとに決まった遅延時間も存在する。そのためある時点における反応を見るためには一定の窓幅の畳み込みを行う必然性がある。

動画に対する反応はこの積み重ねであり、いくつかの山が連なって最終的なコメント群となる。それでは動画におけるコメントはある時点の内容に対する反応だけであるかというところではない。動画には全く関係ないコメントがなされることもあれば、コメントに対するコメント等もありそれらを一度に畳み込みを行うと純粋に内容に対する反応とすることはできない。

しかしこれらを含めて動画の面白さであると定義するとコメント頻度の高い領域を抽出することは目的のアーカイブ化にとっては有用と考えられる。

¹<http://www.nicovideo.jp/watch/sm7497478>

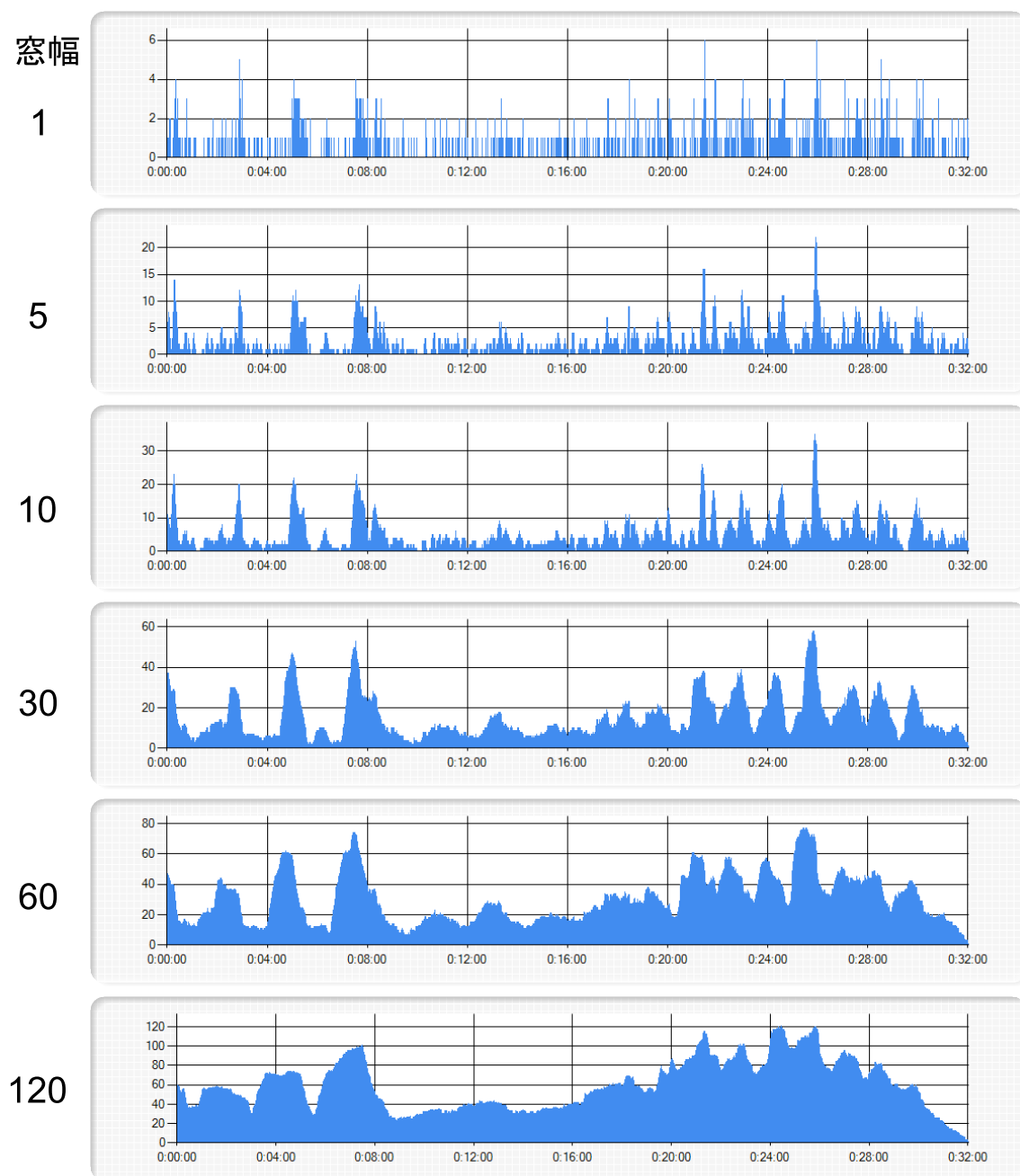


Fig. 3.3: 窓幅ごとの畳込み

3.2.2.1 畳込みの耐性

畳込みの耐性とは、一つの動画に対するコメントの畳込み結果が時間の推移、ユーザ、コメント数によって変化するかどうかということである。ここコメントを被らないように1000件ずつ取得しそれらを畳込みによって図示したFig.3.4。これを見るとこのコメント数と窓幅ではほぼ同じ結果が見られるが、詳細な点を見ると異なる点がある。つまりコメントの少ないコンテンツでは必ずしもコメント量によって定常的な結果を得ることが可能とは限らないことが読み取れる。さらにそのコメント数の違いによってどのような差異が生じるかも調査した(Fig.3.5)。これを見るとコメント数が50くらいの場合ではあまり山の外形が把握しづらいものの、200からはほぼ同じような外形が得られることが読み取れる。ここに生放送でのコメントの少なさに対しての解決を見出した。つまりよほど少ないコメント数でない限り、生放送であっても同様の手法で計算が出来るということである。

3.2.3 窓関数

ある時点以後のコメントを畳み込むとしたがこれは完全な畳込みの定義ではない。畳込みはもとのデータに対して時間変化する窓を用いて全時間領域で畳み込むことも可能である。そもそもある時点における面白さは過去の時点における反応も影響するはずである。つまりコメントが集中している部分が存在し、その直近の過去においてもコメントが集中している場合、その場所同士の面白さは関連があると考えられる。そうした場合窓幅が十分に大きくないとそれぞれの関連性を考慮することができない。ここまで用いた矩形窓は端と中心で同じ重みでコメントを畳み込んでいたため時間的な推移を考慮できていない。ここで代表的な窓の特徴と利用例を説明する

3.2.3.1 ガウス窓

ガウス窓は信号処理の分野において理想的な窓とされるが、本研究の対象は時間変移するユーザの反応であり波のようなものではない。またガウス窓は端が0でないため有限域の畳込みの値を求める窓としてあまり有用でない。しかしある点をピークとして持ち、ガウス分布に従って係数が減少するという特性は感覚的に本研究の目的に沿うものである。そこで本研究では特に断らない限りガウス窓の二次近似である二次関数を用いる。

$$\omega(x) = \exp\left(-\frac{x^2}{\sigma^2}\right) \quad (3.3)$$

3.2.3.2 三角窓

パートレット窓とも呼ばれる三角形の窓である。三角窓は紹介されることが多いが、特に自然界の現象において利用することの根拠に乏しい。Fig.3.7は三角窓を適用した例である。

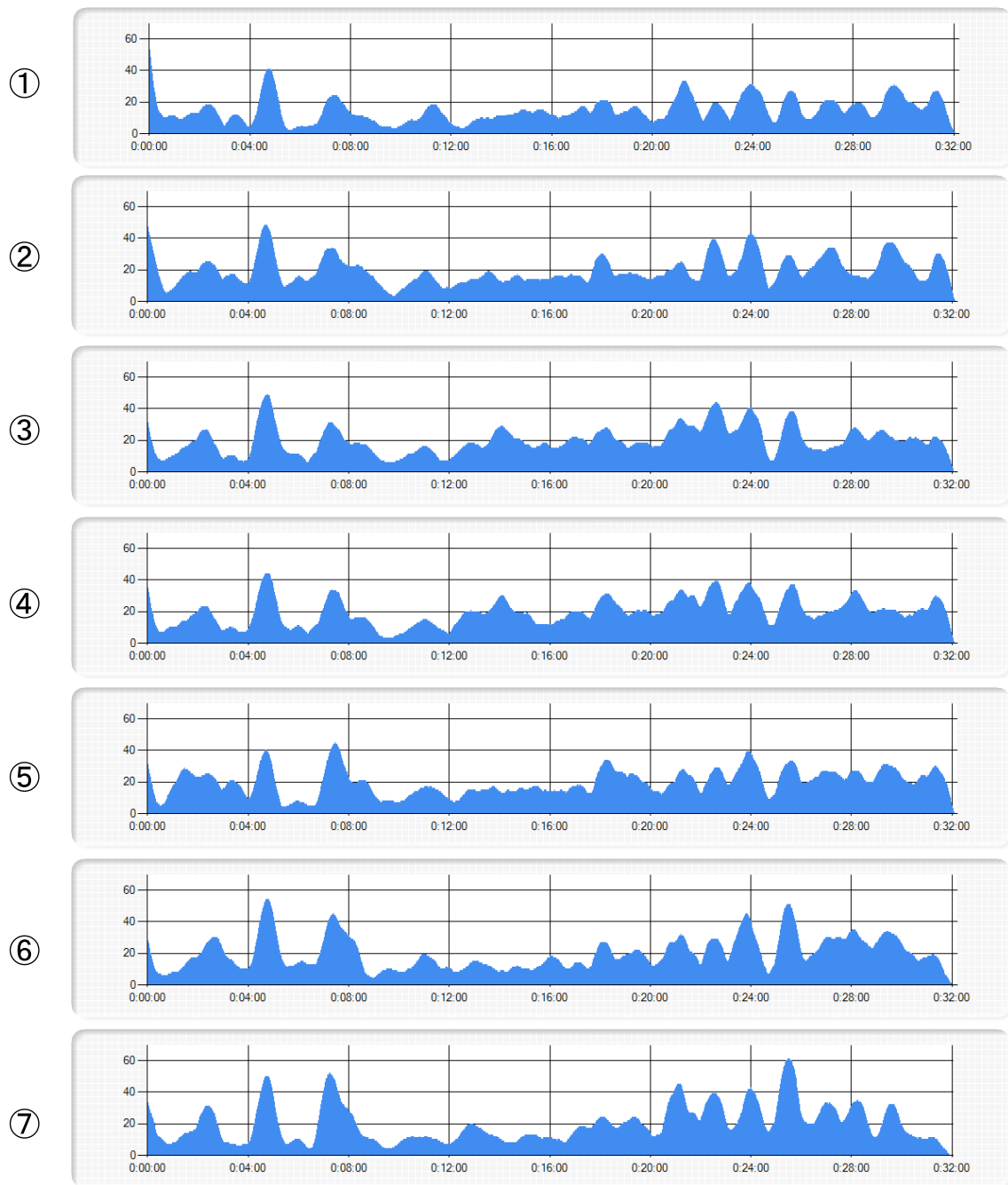
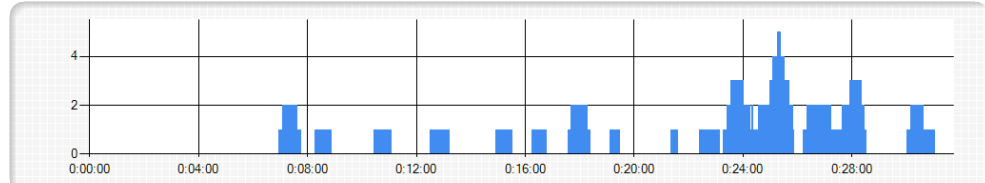


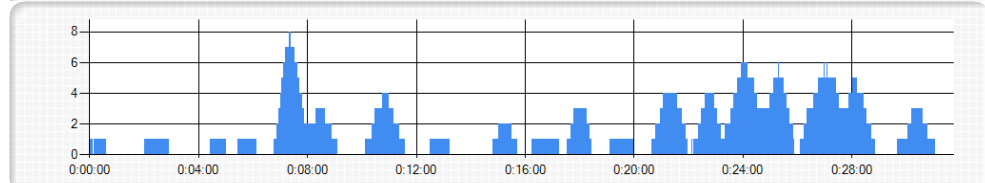
Fig. 3.4: 畳込みのユーザ耐性

コメント数

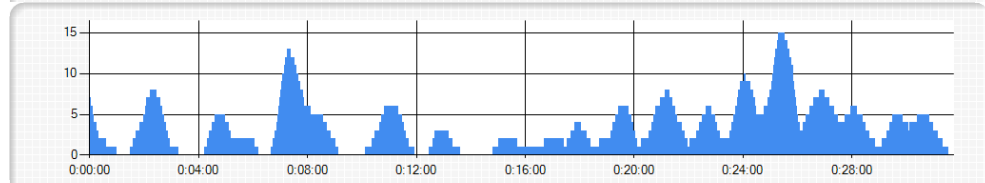
50



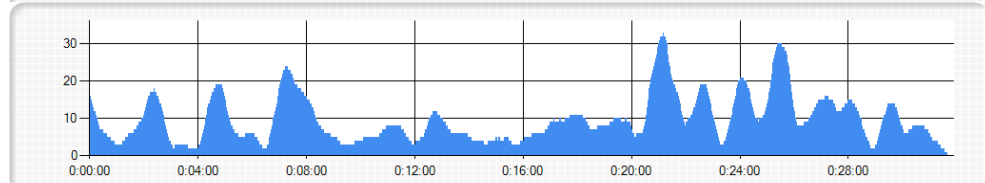
100



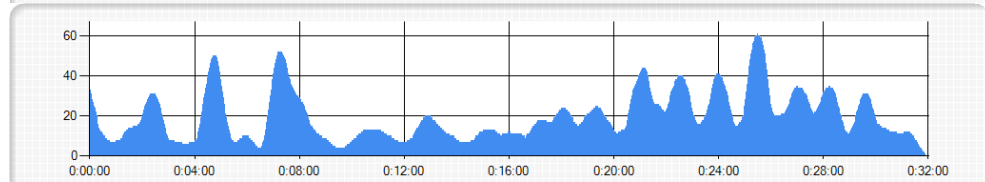
200



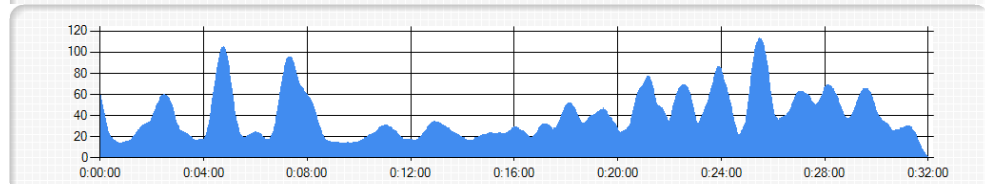
500



1000



2000



5000

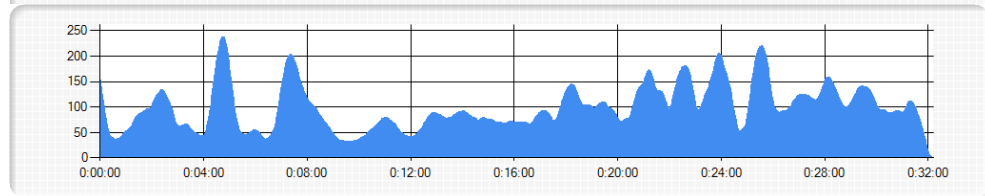


Fig. 3.5: 畳込みのコメント数耐性

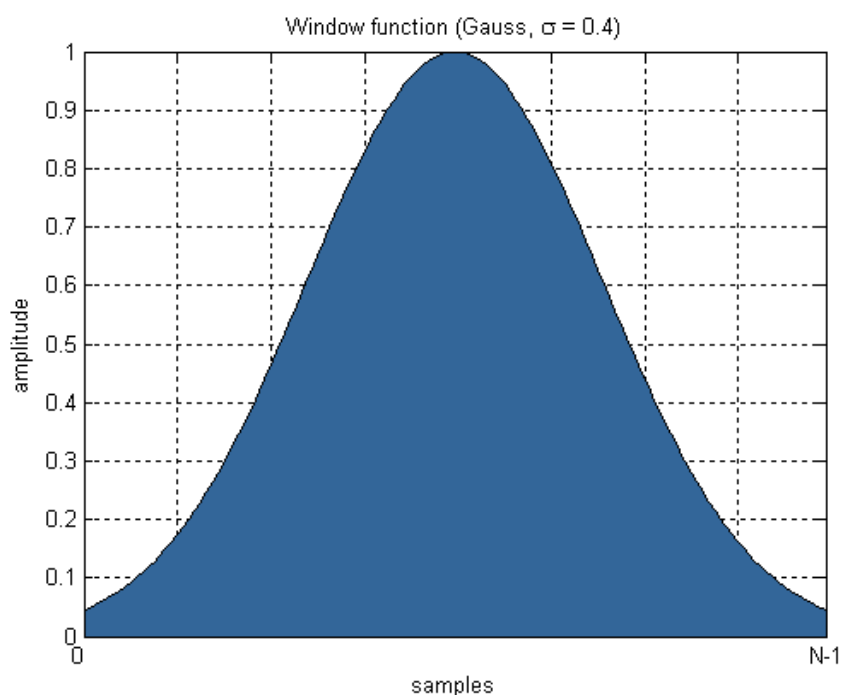


Fig. 3.6: ガウス窓

$$\omega(x) = 1 - 2|x - 0.5| \{0 \leq x \leq 1\} \quad (3.4)$$

3.2.3.3 矩形窓

窓関数における基本的な窓である。一定時間内のコメント量をそのまま反映することが可能であり、ピーク値を求めるには適している。しかし目でわかる欠点としてはコメントの差に左右されやすく、非連続的な値をとってしまうことである。

3.2.4 意味タグ

コメントによる P の定義をしたが、これはコメントの量というコメント内容によって左右されない情報である。そのため単にコメントの多いところで編集領域が決まるが実際に放送のおもしろいところや話題となるところで必ずしもコメント量が多くなるとは限らない。またコメントの内容を解析することによって利用者が希望する種類の内容のみを抽出することが可能と考えられる。

そこで文章や、単語に対して意味を付与する意味タグ[15]を利用する。日本語の意味タグとしては以下のような物があげられる。

これらを一つのコメントごとに定義し、それらをもとに前節で説明した定義を更新する。しかしこうした意味タグの付与にはそれぞれの単語や文との対応リストが必要とな

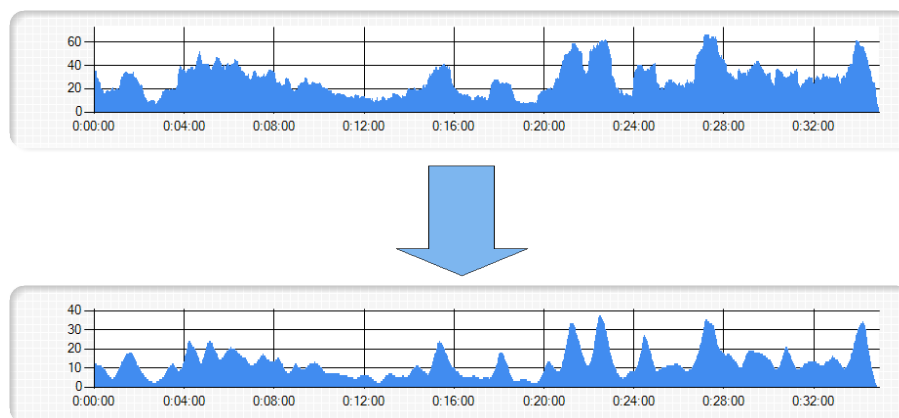


Fig. 3.7: 三角窓による畳み込み

る。だが対象としているユーザ配信型のライブ放送におけるコメントが従来の言語体系、利用法に則っているかどうかはわからない。まずは生放送でよく見かける表現を元にどのような意味タグのリストを作成すべきかを検証する。

そこでコメント内の“w”に注目する。“w”は日本語圏内のインターネットにおけるスラングであり、(笑)を省略したものである²。

3.2.4.1 “w”の畳み込み

Fig.3.8、3.9はこの“w”(または全角の“w”)の数を同じ窓幅と窓の開始位置でそれぞれ窓有り、窓無しで畳み込みしたものである³。

この図よりこの放送では“w”の量が多いところと少ないところで極端に分かれていることが読み取れる。ここで元の定義の場合との比較をしてみると両方の定義でピーク領域がほぼ一致する場所があるが、逆にコメント量ではピーク値となっていない部分でも“w”の量ではピークになっている部分、そしてその逆の部分があることがわかる。また、“w”を畳み込みした場合では窓関数を持つ場合と持たない場合で結果に差が現れる。これによって単なるコメント量の多いところでは抜き出せない部分の抽出が可能であり、この量を用いることによってフィルタのようなものを作ることが可能であると考えられる。現在ニコニコ動画、およびニコニコ生放送では“w”の利用がとても多いため、面白いと考えられる部分を抽出するに当たってこの量を用いることは効果的だと考えられる。

そしてFig.3.10はこの“w”のコメント数と文字数の関係を示したものである。これを見るとコメント数と文字数は比例していることが伺える。よってコメント数が少ない場合は文字数をカウントすることによって編集領域の分解能を上げることが可能である。また各部位で特定の傾きを持つ直線を成していることがわかる。調べたところこの傾きは時間軸上のグラフにおける一つの山での一定の傾きを持つことがわかった。

²[http://ja.wikipedia.org/wiki/\(笑\)](http://ja.wikipedia.org/wiki/(笑))

³<http://www.nicovideo.jp/watch/sm4511948>

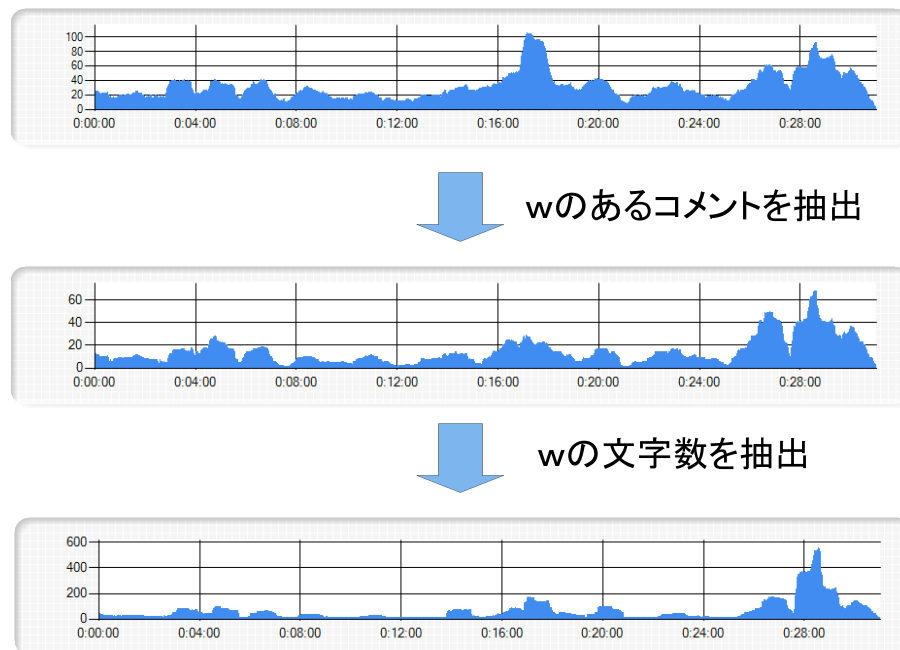


Fig. 3.8: “w” の窓関数無しの畳込み

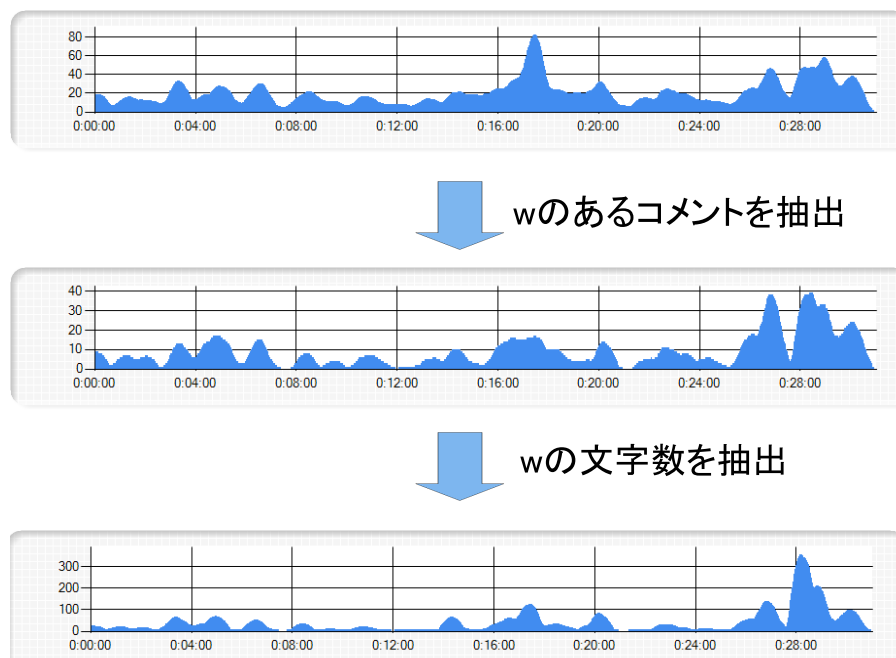


Fig. 3.9: “w” の窓関数有りの畳込み

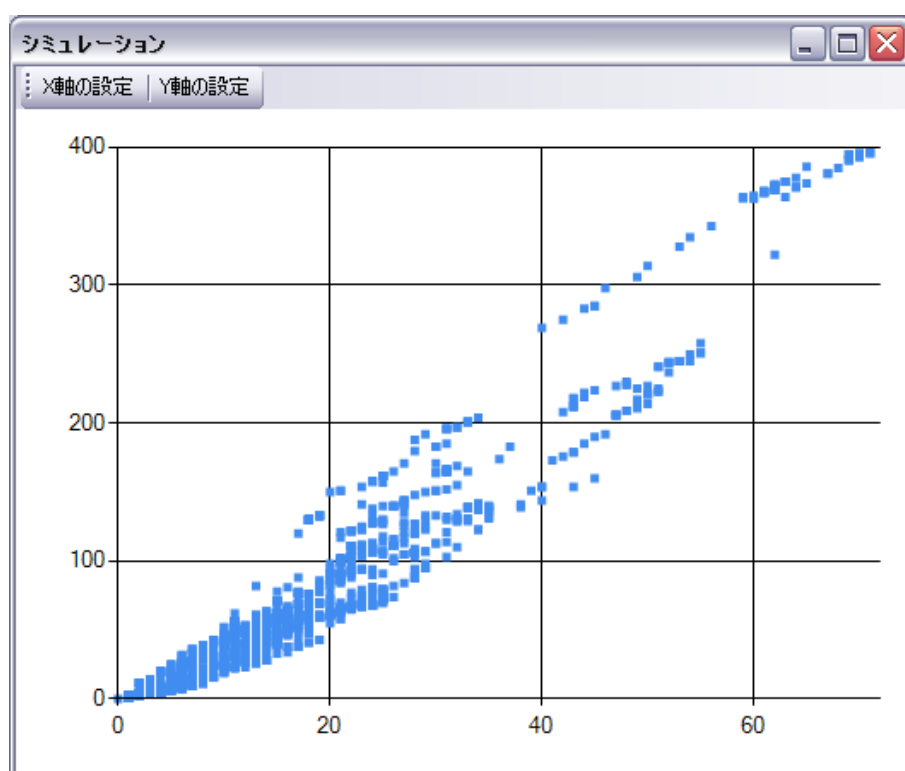


Fig. 3.10: “w” のコメント数と文字数の関係

3.2.5 音声による面白さの値の計算

音声による面白さとは一体何であるのか、何をもって音声を聞いた時に面白いと感じるのか。音を言語として理解し、その意味解釈を行いコンテキストを理解してそれが人間の機微に触れる内容である時に面白いと感じるのではないだろうか。つまり音声による面白さの値を計算するためには音声をテキスト化し、その意味を解釈する必要がある。音声のテキスト化については既に様々な製品が開発されており、Youtube は音声の自動翻訳機能 [19] を発表している。しかしこのテキスト化された内容から面白さの値を計算することは自然言語処理の技術を用いても難しい。

そこで本研究では音量が継続的に高いところがある程度面白い部分であるという仮定を置き面白さの値を計算することを試みた。

音声は WAV 形式 (PCM) の音源を用いる。しかし PCM 形式の音源は非常に処理負荷が高いため、各バイト毎の情報を利用するのは現実的ではない。そこで予め音源のサンプリング周波数とは別にサンプリング周波数を決める。そして再サンプリングした振幅を元にコメントと同様に畳込みを行う。しかし、畳込みの分解能をコメント数の畳込みの分解能と別にするると畳込みの合成ができないため、窓幅のみ変更が可能である。またステレオ音源では2つの波形があるが、普通の放送ではそれぞれの音源にはそれほど差がない。また両チャンネルを公平に扱うという考えのもとに2チャンネルの絶対値の和を用いる。

Fig.3.11 は音の畳み込みを窓幅を変えて行った結果である。ピークを散見することができるが、これではピークが多すぎて特徴的なピークを見つけることはできず、コメントの畳み込みのような場面ごとの相対的な評価が不可能である。よって本研究では音声による面白さの値の計算を断念したが、後述の音声によるクリッピングで音声データを利用する。

3.2.6 映像による面白さの値の計算

映像についても音声と同様であり、映像の意味を解釈し面白さを数値化するのは非常に難しい。そこで本研究ではオプティカルフロー [5]、[6] を用いた画像の変化度を求めその値が大きいところを抽出することを試みた。映像は AVI 形式の HUFFYUV でエンコードされた動画を用いる。ニコニコ動画の映像の基本サイズは縦 512、横 384 であり、このサイズの映像は処理に時間がかかりすぎる。そこで予め面白さの値、後述のクリッピング位置を決める際の計算においては元の動画を小さいサイズ (128 × 96) に変換しておく。本研究では鹿嶋らの研究と同様に LK 法を用いてオプティカルフローを計算する。ブロックマッチング法は計算量が多いため本研究では利用しない。

3.2.7 オプティカルフロー

オプティカルフローは映像のある点の輝度の移動を時空間微分することで抽出するアルゴリズムである。本研究で用いるのは勾配法と呼ばれる手法で以下の拘束方程式を元に算出される。拘束方程式は画像中の (x, y) 点における時刻 t における輝度を $E(x, y, t)$ とし、

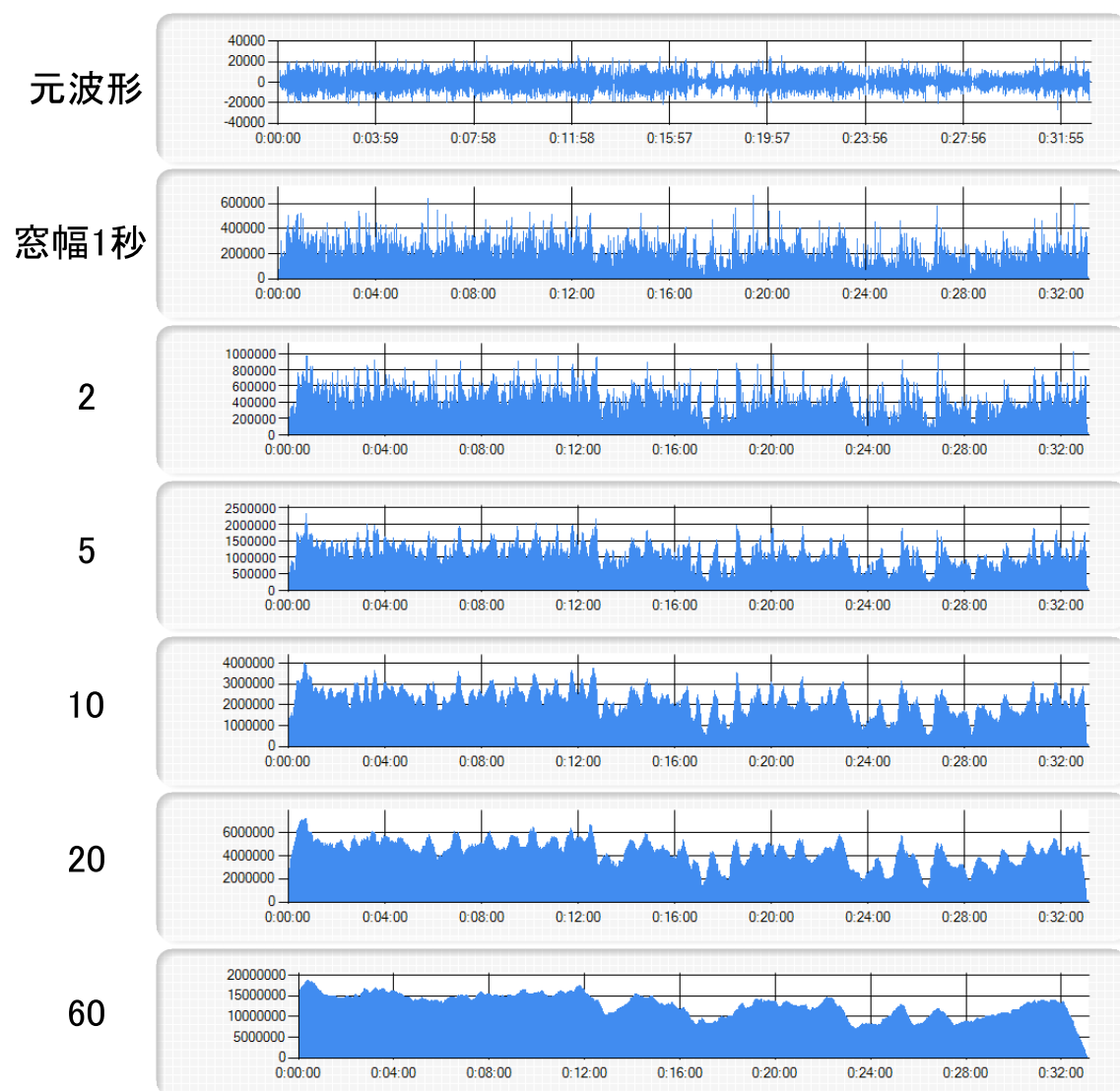


Fig. 3.11: 音声振幅の畳込み

窓幅1秒

2

5

10

20

60

120

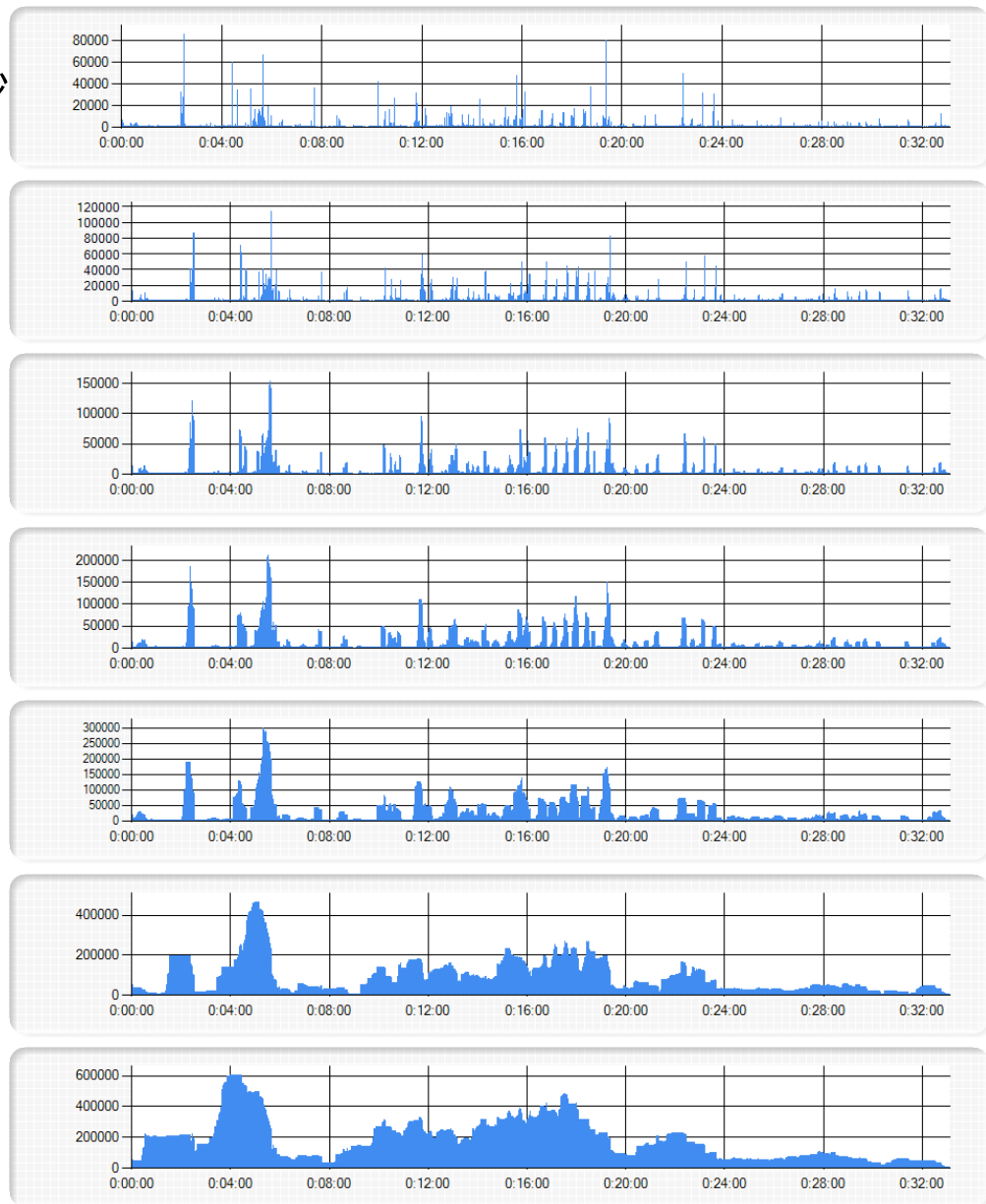


Fig. 3.12: オプティカルフローの畳込み

$\mu(x,y)$ を x 軸のフローベクトル、 $\nu(x,y)$ を y 軸のフローベクトルとすると以下のように示せる。

$$E_x\mu + E_y\nu + E_t = 0 \quad (3.5)$$

そしてフローベクトルが局所領域内で等しいとし、上記の式を最小二乗法により最適値となるものを選択することによって μ 、 ν を導き出すことができる。本研究ではこの領域を 15 ピクセル四方として計算を行った。Fig.3.12 はオプティカルフロー計算の結果を畳み込みで窓幅を変えて行った結果である。音声に比べると特徴的な部位が読みとれる。しかし映像の変化度のみを考慮したものであり、果たして映像がよく変化する動画がおもしろいかといえそれは違う。また映像の変化度は全体的にこういった内容であるのかにもよる。例えば Fig.3.12 で用いた動画で特徴的な山の部分は実は本筋とは関係ない動画の引用部分である。つまりこの畳み込みの値を利用すると結果的にどうでもいいシーンが一番評価されてしまい選ばれてしまうのである。本研究では以上の理由から音声同様に映像から面白さの値を得ることを断念し、クリッピングで利用する。

3.2.8 映像と音声による面白さの値について

映像と音声についての定義を述べたがこれらは真の意味で動画の面白さを表すものではない。音声、映像のおもしろさは本来はその“意味”を解釈する必要がある。

3.3 クリッピング

クリッピングとは動画の内容ごとに区切りをつけることである。このクリッピングはチャプターを設ける事によって、編集領域を決める際に内容がぶつ切りにならないようにするためのものである。単純なコメント量だけでは話の区切り、内容の節を取得することが不可能であるため映像、音声から特徴的な点を抽出しアーカイブ化に利用する。

3.3.1 音声によるクリッピング

音声のデータは畳み込みと同様に PCM のデータがあるサンプリング数の間隔で確保されている。音量がある閾値以下の点をクリッピングするが、この手法では完全にクリッピングを行うことが出来ない。なぜならば人の声だけが入っているとは限らず、BGM 等があると閾値以下になることが終始ないことがあるためである。そこでこの閾値をある程度緩く設定し、細かく音声途切れている部分のみをクリッピングする。そのため音声によるクリップ点が大量に作成されるが、コメントの畳み込みによる編集領域を元に制限を作り、クリッピングによるブレを防ぐ。

3.3.2 映像によるクリッピング

映像のクリッピングは畳込みで利用したオプティカルフローを用いて行う。本来オプティカルフローは映像の変化度を取っているため、場面の変化を抜き出すには適している。オプティカルフローの値の絶対値の和を取り、その変化量が閾値を超えた点をクリップ位置として取る。Fig.3.12の一番上のグラフは窓幅が一秒のためオプティカルフローの和そのものである。音声同様にクリップ点を緩く抽出し、クリップ点を多く設ける。

3.4 編集領域

3.4.1 閾値を用いた編集領域の決定法

前節の面白さの値を元に編集領域を決める。まず考えられるのが面白さの値が適当な閾値を超えた領域を編集領域とする手法である。

コメントによる畳み込みの値で閾値を超えた点を抽出する。そしてその点を元にさらに窓を用いて編集領域を決定する。窓の決定の仕方は大きく分けて二つあり、一つはその点の畳み込みに用いた窓を利用する方法。もう一つは窓を再定義する方法。前者は用いた窓内のコメントをすべて補足するという目的があるが、長い窓を用いると編集領域がそれに応じて延びてしまう。後者は窓を再定義するため柔軟な領域の定義が可能であるが、設定に対する論理的な根拠はない。そのため自分で各設定ごとに確認してその特性を掴む必要がある。

Fig.3.13は上記の方法で閾値を変化させた場合の編集領域を閾値ごとに表示したものである。

3.4.2 クリップ位置を用いた編集領域の決定法

閾値のみを用いた編集領域の計算ではぶつ切りになる可能性があるため、映像、音声から計算したクリッピング点を利用する。閾値を超えた点をまず抜き出し前述の編集領域を窓によって作成し、領域内のクリッピング点で一番外側の点までを新しい領域とする。

3.4.3 編集領域の探索

以上の手法を用いて編集領域の探索を行う。目的の時間と許容誤差を決定し、その範囲内の結果をリストとして返す。Fig.3.14はある面白さの値に対する閾値変化と編集領域の割合の変化を示したものである。縦軸の1はすべての領域が選ばれているということである。これは一例であるがこの変化特性は実験対象とした動画では一貫して観測することができた。Fig.3.14を見ると閾値が小さい場合、大きい場合では編集領域の変化が少なくその間では変化量が大きいことがわかるこれは閾値が中間地点付近でピークを持つ場合が多いためである。また変化量に差はあれどほぼ連続的に編集領域を確保できるため柔軟な圧縮が可能であることが読み取れる。つまり任意の圧縮率に対する編集領域を確保できるということである。

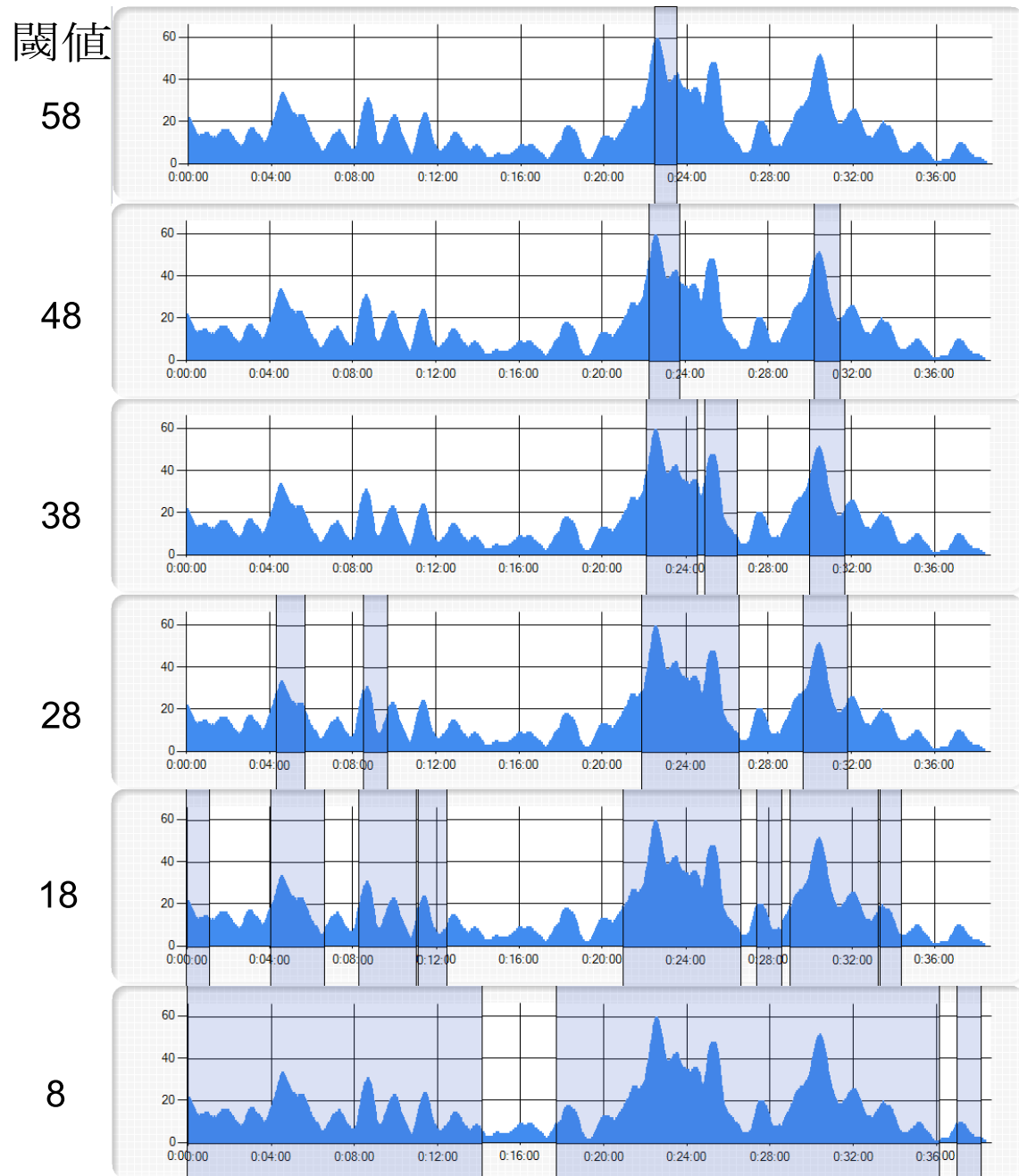


Fig. 3.13: 閾値の変化による編集領域割合の変化のグラフ

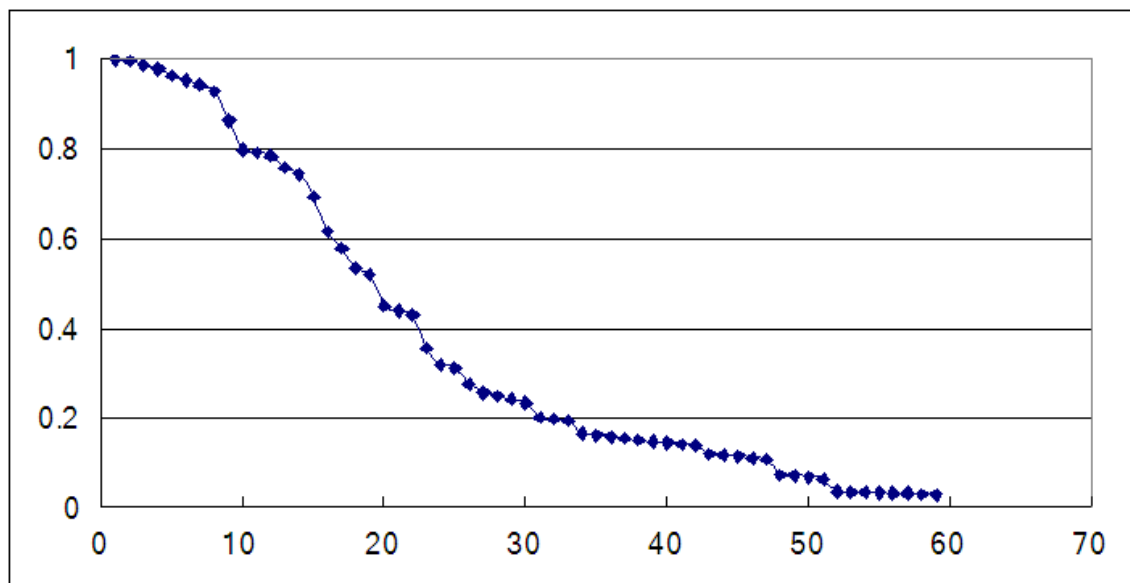


Fig. 3.14: 閾値の変化による編集領域割合の変化

3.4.4 アプリケーション

前節まで提案した手法を実装しアプリケーションを作成した。開発はMicrosoft VisualC#で行った。現在このアプリケーションはネットに公開中である⁴。(Fig.3.15)。アプリケーションの各部位の機能は以下のとおりである。

- 1の機能 主にコメント表示部分で、コメントの情報をセルで表示している。他にはコメントのXML、各クリッピング機能によるクリッピング結果、編集領域のテキストでの結果を表示している。
- 2の機能 動画の表示部分。現在Flash Playerが再生できる動画のみ再生可能。
- 3の機能 畳込み結果及び編集領域の表示部分。読み込んだ音声波形、オプティカルフローの絶対値和のグラフも表示される。

アプリケーションの詳細な設定方法については付録Bに記載した。

⁴<http://d.hatena.ne.jp/gomao14/>

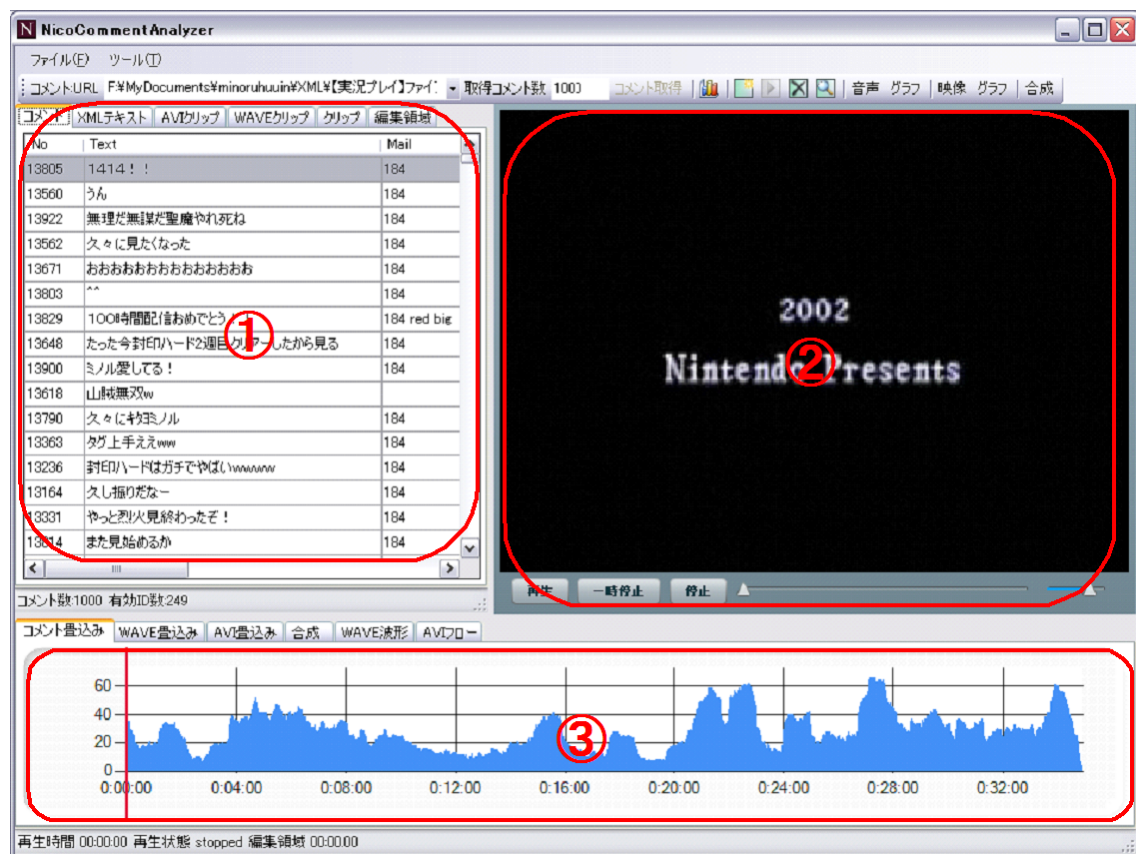


Fig. 3.15: アプリケーション

第4章

評価

4.1 実験

前章で説明した提案手法を用いて実際にアーカイブを作成した。現在ニコニコ生放送ではアーカイブが配信されているがこのアーカイブのデータを入手するのは困難であり、またアーカイブの視聴はプレミアム会員のみに限られているため現存のサービスの兼ね合いで実際のアーカイブを作成するのは難しい。そこでニコニコ動画に投稿されている動画からアーカイブを作成することにした。

4.1.1 実験の設定

畳込みの理想的な設定の論理的な根拠は導き出すことが現在不可能である。そこで一番最初の動画に関しては自分の視聴を踏まえて、必要だと思われる設定に目星をつけることにした。以後、二番目以降の動画に関しては一つの手法で抽出を行う。

畳み込みの設定は畳み込み窓幅を40秒、間隔を1秒、オフセットを1秒、編集領域の窓は窓幅を30秒、オフセットが0秒である。最終的に1/10の長さになるようにコメント数に比例した編集領域を割り当てた。

今回の目的は動画の面白い部分を抜き出すということが目的である。そこでコメントから視聴者が笑っていると思われる部分を抜き出すため前章で紹介した“w”を用いる。“w”のあるコメントを抜き出して、そのコメントの中から“w”の数を計測する。またコメント数は各動画につき数千あるが、今回は生放送との兼ね合いを取るためコメント数を500に制限した。

4.1.2 アーカイブの作成方法

作成したアプリケーションに前述の設定を施して編集領域を計算する。本研究ではアーカイブされている動画を再エンコードすることは考慮していない。これは既にフラッシュの技術で飛ばし再生¹が可能であり、フルアーカイブされているものを飛ばし再生によって編集領域を見ることを仮定している。また改正著作権法を考慮してダウンロードしたコンテンツの違法利用を防ぐため、また、ニコニコ動画が動画のダウンロードを推奨はしていないためダウンロード機能を搭載しておらず、自動的にアーカイブを作成する機能もない。そのためすべての編集領域を計算し、Video Studioを用いて手動でアーカイブを作成した。エンコードをふくめて作成にかかった時間は5時間程度だった。

今回アーカイブを作成したのは<http://www.nicovideo.jp/mylist/8147659>の実況プレイ動画である。単品の作品を除いて全74個の動画があるが全部のアーカイブを作成するのは間に合わなかったため、前18個(タイトル表記では1章から7章の範囲)のアーカイブを作成した²。元動画の長さは全450分、約7時間で、これを48分にまとめた。

¹任意の時間から動画を再生すること。

²<http://www.nicovideo.jp/watch/sm9617614>

4.1.3 アンケート

実際に作成した動画の最後にアンケートを載せて任意で答えてもらった。この質問は以下のとおりである。

- この動画が面白かったか
 1. 面白かった
 2. まあまあ面白かった
 3. 普通
 4. 微妙に面白くなかった
 5. 面白くなかった
- 今後この動画の続きを望むか。
 1. 作って欲しい
 2. あれば見る
 3. 好きにして欲しい
 4. 見ない
- シーンのつなぎ目で不自然な点はなかったか
 1. 特に気にならなかった
 2. 若干あった
 3. 不自然だった
- 動画を短くした方がいいか
 1. 短くしてほしい
 2. このままでいい
- あのシーンがなかったということはなかったか

4.2 結果

4.2.1 アンケート結果

アンケートの結果はFig.4.1、4.2、4.3、4.4のとおりである。この結果は2月8日現在のものである。

Fig.4.1「続きの動画が必要であるか」について、続きの動画についてはこのアンケートを集計する前に同様の手法で投稿してしまった³が、あれば見るという意見が多数であった。それほど期待はされていないものの、面白さが保証されているコンテンツのため視聴の意志が読みとることができる。ちなみに「好きにしろ」という選択肢は視聴者側から決定する意志を放棄する意図を汲むためのものである。ニコニコ動画に限らず動画共有サイトは投稿主が自発的に投稿することによってコンテンツが成立する空間であり、投稿主ではない一般視聴者は受動的であることが多い。

Fig.4.2の「動画は面白かったか」という質問に対してはまあまあ面白かったという意見が多数を占めた。今回の荒い抽出手法でもある程度の面白さを保証することが可能であるようだが、まだ満足といえるほどではないようである。

Fig.4.3の「動画のつながり目は不自然ではなかったか」という質問に対しては、6人ほど不自然さを感じているという意見をもらった。これに関しては作成した自分にも明らかにシーン間延びしていることが確認できた。そして次の質問のFig.4.4「動画の長さを短くしたほうがいいか」という質問はあらかじめこの反応を危惧してのことである。やはり結果としてアンケートに答えてもらった人の半数の視聴者から「短くしたほうがいい」という意見をもらった。また最後の「あのシーンがなかったということはなかったか」という質問に関しては特に上げられていなかったことから、シーンの抽出自体はうまくいっているもののシーンの合計時間を設定してしまったため必要ない部分まで引き取ってしまったことが読み取れる。

これを解決するためにはリスト内の各動画の重み付けを再定義する必要がある。今回のように合計時間をあらかじめ設定し再生数、コメント数、マイリスト⁴数、再生時間に応じて安易に編集時間を割り振るのは危険である。なぜならば各動画においてこれらの値は平等ではないからである。たとえば一番最初の動画はとりあえず見てみようという人が視聴するため圧倒的に数値は大きくなる。そして一般的に以後数値は右肩下がりになるが、あくまですべての動画が等間隔でなおかつ同じ条件で投稿されている場合に限る。しかし動画サイトの環境自体も日々変化するためこれらが理想系と一致することはまずありえない。こうした間延びを防ぐためにさらに圧縮するためには別の編集領域の計算手法が必要であると考えられる。

³<http://www.nicovideo.jp/watch/sm9629862>

⁴ニコニコ動画ブックマーク機能

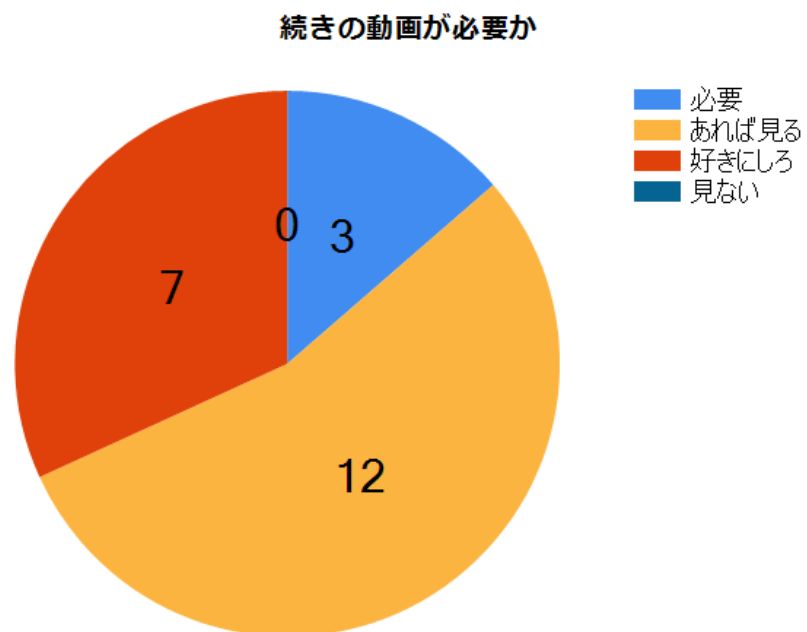


Fig. 4.1: 続きの動画が必要か

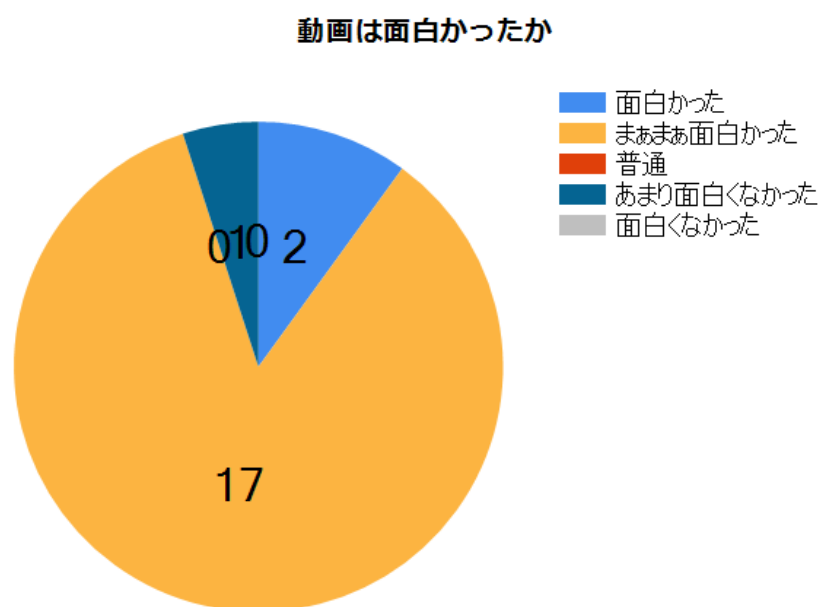


Fig. 4.2: 動画は面白かったか

動画の繋ぎは不自然ではなかったか

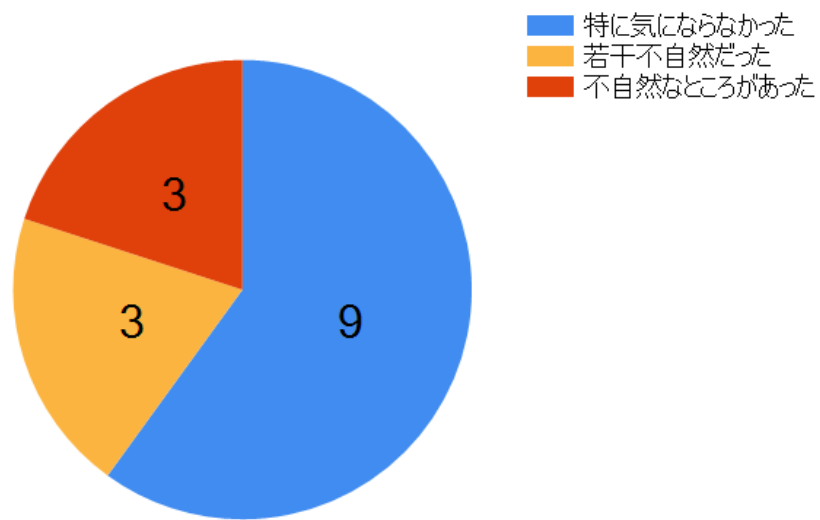


Fig. 4.3: 繋ぎは不自然だったか

動画の長さを短くした方がいいか

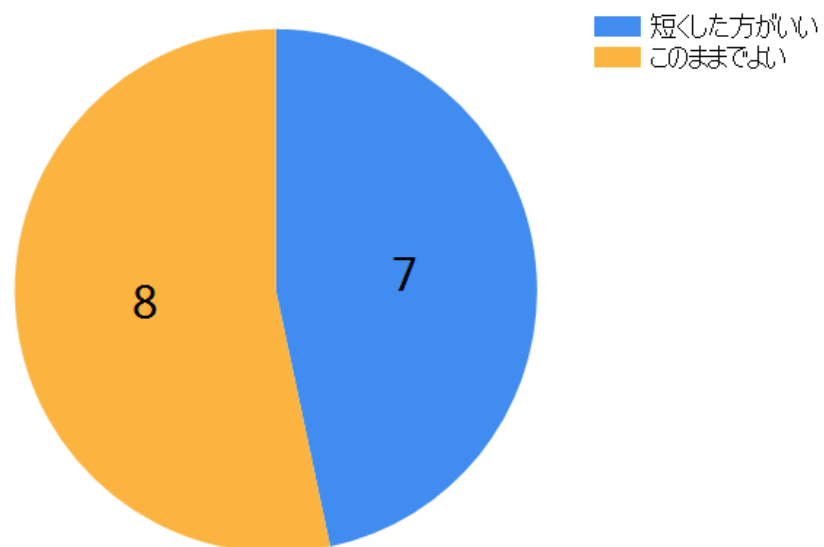


Fig. 4.4: 動画の長さは短くした方がいいか

第5章

考察

5.1 編集領域の決定法

前章のアンケート結果で述べたように現在の編集領域の抽出手法はさらに圧縮する上では適用できないことがある。まず閾値を用いる手法は一つの動画に関する圧縮を念頭においているため、相対的に面白くない動画から無理に編集領域を選んでしまうことがある。つまり定義している面白さの値は動画内で独立であり、絶対的な値ではない。そこで連続したシリーズを圧縮することを念頭においた手法を考える必要がある。

たとえば同じ面白さを持つ領域が二つある動画、一つある動画において各動画からコメント数を一定数取るとそれぞれの領域におけるコメント数は二つある場合よりも一つある場合の方が多くなってしまう。しかし両者において面白さの値が等しくなるような手法があればコメントを結合して計算することが可能である。そこで同じ面白さに対しては同じ人が同じ反応をすることを仮定すると、総コメント数に応じたサンプル数を取ることによってこれを解決することができると考えられる。そして得たコメントを結合し、全体に対して計算を行うことによって、必要のない編集領域は自然と除外されることであろう。

5.2 面白さの値

本研究において提案した面白さの値とは一体何であったか。実験を進めるに従いこの面白さの値が実際の動画の面白さを示すには不十分であると感じた。これは面白さの値が畳み込みで計算されていることと、単純なスカラー値で表現していることに起因していると考えられる。面白さの値をより厳密にするためには擬似人格のモデルが必要ではないかと考えた。観測した事象に対する人間の反応はそれまでの経験の積み重ねによって変化するものである。経験を保持できる人格のモデルによって編集後の動画の検討も容易になるのではないかと考えられる。“面白さ”という表現は動画の“良い”部分を映し出しているかのように見えるが、実際の動画にはそれ以外の意味も含まれており実際には中性的な存在であるはずであろう。

5.2.1 コメントによる面白さの値

本研究ではコメントをあくまで視聴者のレスポンスとして扱ってきたが、実際にはコメント自体もコンテンツである。そのためコメントの意味解釈も必要である。これは畳み込みによってコメントを積み上げて解釈する方法とは逆で、切り崩していく方向である。

5.2.2 映像と音声の面白さの値

コメントによる面白さの値の定義については解釈しやすいテキスト形式から計算されるが、音声、映像に関しては未だ意味の解釈が十分ではない。これは画像認識、音声認識分野における自動翻訳技術の発達によって解決できる可能性があるが、その場合はコメントの意味と動画の意味を対応づける必要があると考えられる。また式 3.1 では非常に簡潔

な和で面白さを表現したがこの表現は十分ではない。映像、音声、コメントは独立の情報ではなく、一般的な視聴者はすべての情報を得て反応をしている。これが無理であるために分解する形を取ったが、独立して解釈出来るのはコメント量が限界である可能性がある。3章で示したように情報の変化量からは面白さを解釈するのは困難であり、意味の解釈を試みた途端音声や、映像は独立的な存在ではなくなる。これを解決するには映像、音声、コメントをすべて同列に扱えるように変換する必要があるのではないだろうか。例えば情報をすべてテキストに還元し、意味と内容を並列に並べ音声、映像、コメントの境を無くすということである。テキストに還元された情報は扱いやすく本研究で提案した手法を利用できる可能性があると考えられる。

5.2.3 面白さとは

Fig.5.1 は本研究を通じて得た面白さの概念図である。一番最初に面白さの根源となるコンテンツがあり、それによって面白さが生まれ、そして最後に視聴者の反応があるという形である。至って当たり前の図であるが、この図から考えると式3.1の定義は少し矛盾していることがわかる。コメントの畳み込みから面白さをもとめる際はこの図において下から上に向かう方向であり、映像、音声、そしてコメントの意味解釈によって面白さを求める際は上から下に向かう方向である。つまり式3.1はこの双方向の物を和にしようとしていたのである。これは大きな問題である。なぜならば下から上方向に向かって求められた面白さはすでにその中に映像、音声による解釈を踏まえたものでなければならぬからである。これとは別個に映像、音声の解釈を含めようとすると二重に面白さを評価することになる。もし下から上方向の面白さと上から下方向への面白さが理想的に求まるのであれば、これらは和にしたりするものではなく、一致すべきものである。

逆にコメントが視聴者の反応を得る唯一の方法であれば、コメントだけで面白さの値を定義することが可能である。しかし人間の感情をテキストだけで表現するのはおそらく無理な話であろう。ただコメントはすべての情報を人間が噛み砕いて解釈した上でのアウトプットであるため、上から下方向へ直接的に面白さの意味を求めるよりも面白さの計算をしやすいということである。このことから面白さをさらに厳密に計算するためには視聴者のコメントとは別の反応を得る手法を考える必要がある。

5.3 ユーザと放送主のインタラクション

本研究ではニコニコ動画、ニコニコ生放送と二つのサービスで利用できる事を念頭においたが、実際は生放送サービスに準じたアーカイブ作成手法は投稿型にも利用できるはずである。そこで生放送をより強く意識した編集領域抽出を確立するためには一体何を考慮すればよいであろうか。

その生放送の特徴の一つに視聴者と放送主のインタラクションがあると考えられる。インタラクションは生放送の最大の魅力である。生放送では視聴者のコメントに対して放送主は即座に答えることができ、投稿型動画では受動的である視聴者がコンテンツを作り

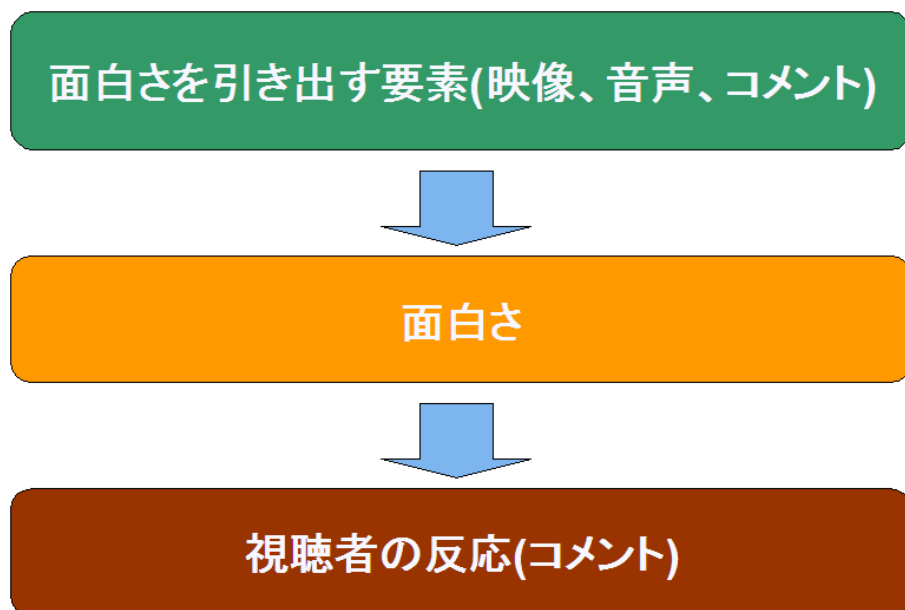


Fig. 5.1: 面白さとは

上げることができる。インタラクションを取り込んだ結果アーカイブが面白くなるとは限らないが、これはいままでになかった新しいエンターテインメントの提供の可能性である。

5.4 結論

動画には各時点において面白さの値があると仮定し、その値を元にして動画の編集領域を計算することを提案した。主にコメントの畳み込みによって面白さの値を計算しその値と音声と映像によるクリッピング点から編集領域を決定した。提案した手法を用いて実際にアーカイブを作成し、アンケートを得て評価を行った。そしてこの面白さの値によって動画の“面白い部分”を抽出できることを確認した。しかしこの面白さの値は未だ考察の余地があり、今後定義や呼称を見直す必要がある。

謝辞

本研究を進めるにあたり、多くの方々のご協力をいただきました。

担当の相田仁教授には数々のアイデアや有益なアドバイスをいただきました。2年という短い間でしたが大変お世話になりました。

また秘書の中山早百合女史には事務の面で研究生生活をサポートして下さっただけでなく、日頃お菓子を頂いたり日常的な面でお世話になりました。

技術専門職員の千葉新吾氏には研究室の生活面で大変お世話になりました。

そして同期の森川貴康君、藤木慎太郎君には楽しい時間をともに過ごせたことをとても感謝いたします。

またこの2年間の研究室生活において助け合った研究室メンバーの皆様にはとても感謝しております。

そしてネット上で僕の研究に役立つ数々の情報を発してくださった方々に感謝いたします。

最後にこれまでの成長を見届け勉学に打ち込める環境を与えてくれた家族を始めとする全ての方々に感謝いたします。
ありがとうございました。

2010年2月9日

参考文献

- [1] 磯貝佳輝, 斎藤義仰, 村山優子. “視聴者からのコメント情報を用いたダイジェスト動画作成支援システムの実装と評価”, DICOMO2009,(2009).
- [2] 青木秀憲, 宮下芳明. “ニコニコ動画における映像要約とサビ検出の試み”, 情処研報 2008-HCI-128/2008-MUS-75, Vol.2008, No.50, pp.37-42,(2008).
- [3] 伊藤聖修, 鈴木育男, 山本雅人, 古川正志, “ニコニコ動画におけるタグ共起ネットワークの特徴抽出”, 第 80 回人工知能学会知識ベースシステム研究会 (SIG-KBS), 1 月 15-16 日, NTT 武蔵野研究開発センター,(2008).
- [4] 小川一昭, 服部哲, 速水治夫, “視聴者からのコメント情報を用いたダイジェスト動画疑似生成の提案”, 情報処理学会研究報告, 2009(33), pp.145-150, 2009/3/18-19(2009).
- [5] C. Tomasi and T. Kanade. “Shape and motion from image streams under orthography -a factorization method.”, Proceedings of the National Academy of Sciences of the United States of America, 90(21),pp.9795-9802, November (1993).
- [6] Lucas B D and Kanade T, “An iterative image registration technique with an application to stereo vision.”, Proceedings of Imaging understanding workshop,pp 121?130,(1981).
- [7] 鹿嶋 雅之, 押川修士, 鶴田伸一, 佐藤公則, 渡邊 睦. “映像と音声特徴に基づいた映画のダイジェスト生成”, 平成 21 年度 (第 62 回) 電気関係学会九州支部連合大会,(2009).
- [8] 伊藤一成, 酒井康旭, 斎藤博昭. “音声と映像の一貫性を考慮したニュース動画の要約”, 日本データベース学会論文誌 DBSJ Letters, Vol.3 No.1, pp.5-8,(2004).
- [9] 吉高淳夫, 田中壮詩, 平嶋宗, “映画等を対象としたダイジェスト映像生成のための映像特徴に関する検討”, 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, Vol.2007, No.68(20070706), pp. 79-86,(2007).
- [10] 大塚裕子, 乾孝司, 奥村学. “意見分析エンジン-計算言語学と社会学の接点- 5 章社会の関心、意見を分析する実践 構文論と語用論を結ぶ分析および自動分類の試み ”, pp127-172,ISBN 978-4-339-02423-4,(2007).
- [11] 仁田義, 益岡隆志. “日本語のモダリティ”, ISBN 978-4874240359,(1989).
- [12] T.Fujiki, T.Nannno, Y.Suzuki, and M. Okumura. “Identification of bursts in a ducument stream.”, Proceedings of 1st International Workshop on Knowledge Discovery in Data Streams,(2004).

- [13] 金田哲, 永田昌明, 山西健司, “言語と心理の統計-言葉と行動の確率モデルによる分析”, ISBN 4-000-006850-4,(2003).
- [14] 石井恵, 井沢味奈子, 片岡良治. “ネット上の会話からの話題即時抽出技術の評価について (Web からの知識発見とネット応用)”, 情報処理学会研究報告. 自然言語処理研究会報告 2005(94), 33-40 ,20050929,(2005).
- [15] 栗原一貴, 伊藤乾, 五十嵐健夫. “編集と発表を電子ペンで統一的行うプレゼンテーションツールとその教育現場への応用”, コンピュータソフトウェア, Vol.23, No.4, pp.14-25,(2006).
- [16] Kurihara, K., Mochizuki, T., Oura, H., Tubakimoto, M., Nishimori, T., Nakahara, J., Yamauchi, Y., Watanabe, S. . “Borderless Canvas: Development of a Multi-display Discussion Software for Knowledge-emergent Presentations”, ED-MEDIA Conference Proceedings, pp.3676-3681(2009).
- [17] Kiyotaka Uchimoto, Masaki Murata, Qing Ma, Hiromi Ozaku, and Hitoshi Isahara. “Named Entity Extraction Based on A Maximum Entropy Model and Trans-formation Rules”, Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL), pp.326-335,(2000).
- [18] Oura, H., Mochizuki, T., Nishimori. “eJournalPlus: Development of a TabletPC Based Reading Support Software Toward Critical Reading”, Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008 (pp. 5684-5688),(2008).
- [19] Youtube, “自動翻訳機能 Auto-caps”, <http://www.itmedia.co.jp/news/articles/0911/20/news063.html>, <http://itpro.nikkeibp.co.jp/article/NEWS/20091120/340811/>, (2009).
- [20] ニコニコ生放送, “タイムシフト機能”, <http://blog.nicovideo.jp/niconews/2009/12/005930.html>, (2009).
- [21] NTT アイティ, “ダイジェスト映像生成システム「EasyDigest」”, <http://www.ntt-it.co.jp/press/2009/1111/091111bmj.html>, (2009).

発表文献

- [22] 大泉 拓, 相田 仁, “コメント付きユーザ配信型ライブ放送の「面白い部分」を抜き出したアーカイブ化”, 情報処理学会全国大会, Mar.2010, (2010).

付録

付録A ニコニコ動画のコメントXML

ニコニコ動画でのコメントは以下のようなフォーマットを持つ。

—— コメントのXML ——

```
<chat anonymity="1" date="1248662910" mail="184 ue green" no="7513" premium="1"
thread="1248436905" user_id="8j-O2XwNb_GCR1ZzuGMsI7p8dMo" vpos="82965 >お k
</chat >
```

anonymity IDの暗号化をするかどうか。暗号化されない場合は表示されない。

date コメントが投稿された時間のUNIXスタンプ。

mail コメントのコマンド¹が保持されるエリア。半角スペースで列挙できる。

no コメントの番号。

premium ユーザーがプレミアム会員²であるかどうか。プレミアム会員ではない場合は表示されない。

thread コメントのスレッド番号。同じ動画に対して違うコメント群が表示される場合がある。

user_id ユーザーのID。anonymityが1だと暗号化されている。

vpos 動画内でのコメント位置。1/100秒単位。

¹文字に対して装飾をする際に付加する。

²ニコニコ動画の有料会員で月500円の支払いが必要。専用回線の利用や動画投稿の優先権、生放送権利など様々な権利がある。

付録 B アプリケーションの設定方法

付録 B.1 畳み込みの設定

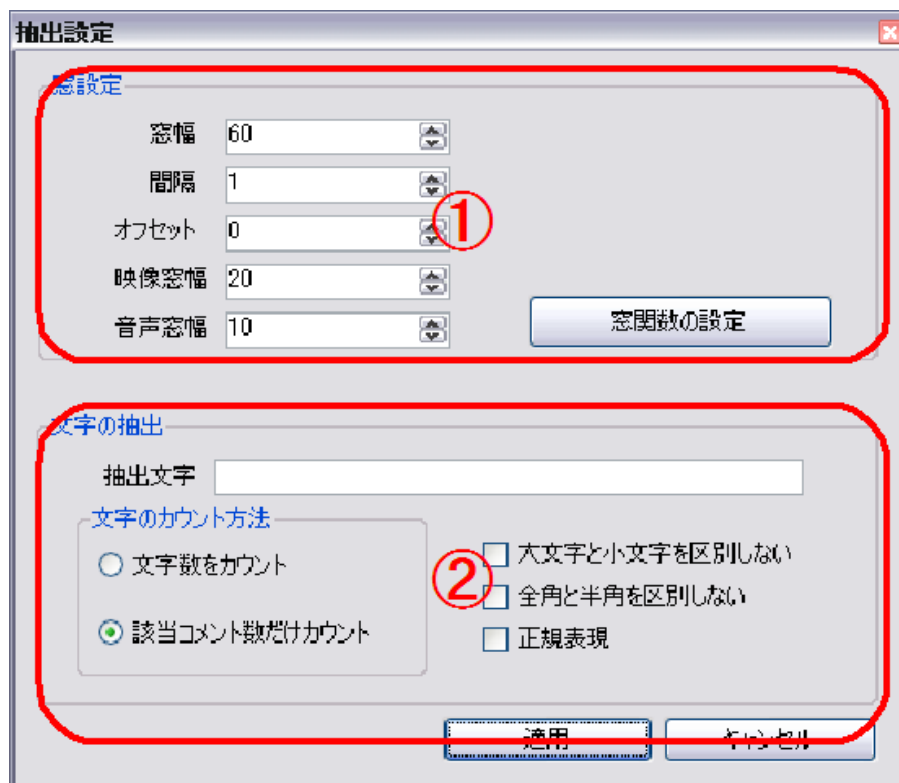


Fig. .2: 畳み込みの設定

付録 B.1.1 窓設定

窓幅 コメントを畳み込む際の窓幅。

間隔 畳み込みの数値を出力する間隔であり畳み込みの分解能。

オフセット ある時点における畳み込みの窓をどの時点からはじめるか。

映像窓幅 映像に関する計算を行う際の窓幅。分解能の兼ね合いにより間隔は共有である。

音声窓幅 音声に関する計算を行う際の窓幅。分解能の兼ね合いにより間隔は共有である。

付録 B.1.2 文字の抽出

抽出文字 特定の文字を抽出する際の文字の入力欄。空欄でコメント全体になる。

文字のカウント法 文字数をカウントするか、コメント数を数えるか。

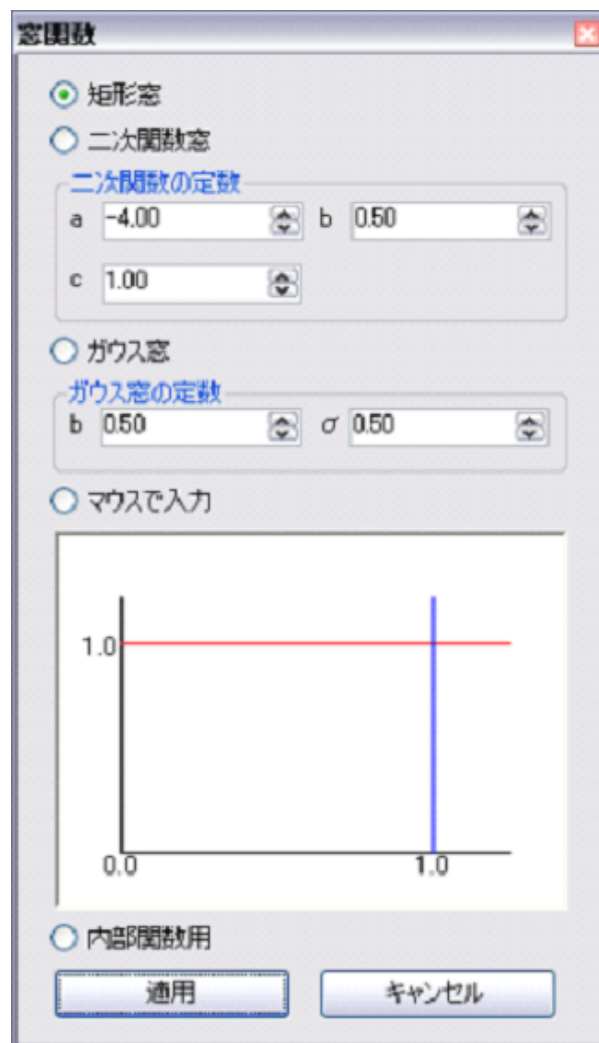


Fig. .3: 窓関数の設定

付録 B.1.3 畳み込みの窓関数の設定

矩形窓 矩形窓を利用。

二次関数窓 二次関数窓を利用。

ガウス窓 ガウス窓を利用。

マウスで入力 マウスで窓の外形を描くことが可能。

内部関数用 上記以外の関数を利用する際のデバッグ用。

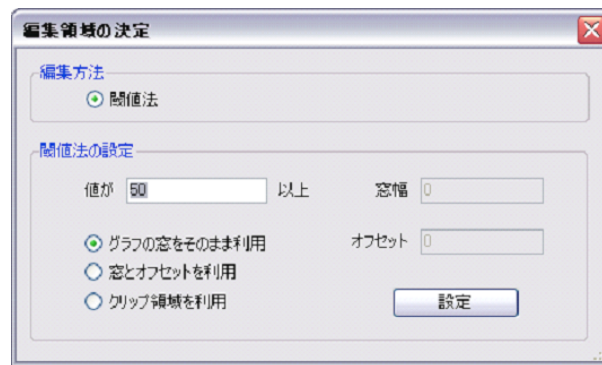


Fig. 4: 編集領域の設定

付録 B.2 編集領域の設定

グラフの窓をそのまま利用 畳み込みで使った窓の設定をそのまま利用する。

窓とオフセットを利用 右に入力された窓の設定を利用して編集領域を設定する。

クリップ領域を利用 クリップ点に挟まれた領域を編集領域に利用する。

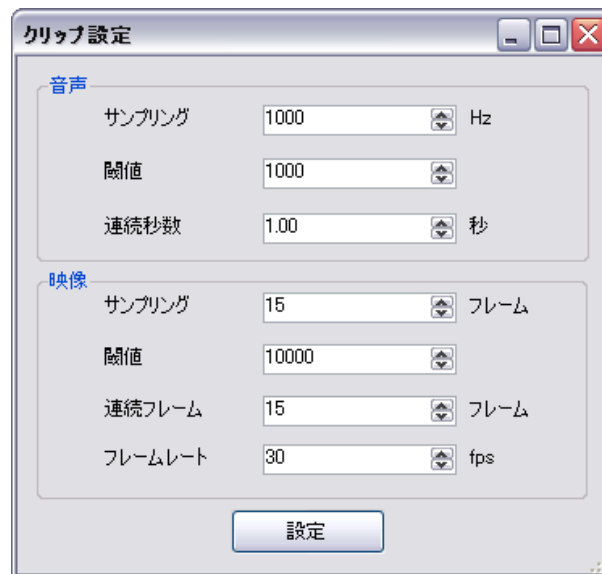


Fig. .5: クリッピングの設定

付録 B.3 クリッピングの設定

付録 B.3.1 音声

サンプリング 計算量を減らすためPCMのデータをサンプリングする間隔を設定。

閾値 クリッピングをするための閾値。

連続秒数 閾値を適用する範囲。

付録 B.3.2 映像

サンプリング 計算量を減らすため映像のフレームをサンプリングする間隔を設定。

閾値 クリッピングをするための閾値。

連続フレーム 閾値を適用する範囲。

フレームレート 読み込んだ映像のフレームレート。現在はフレームレートを読み込むことができないのでこれで代用する。