HORROR FILM GENRE TYPING AND SCENE LABELING VIA AUDIO ANALYSIS

Simon Moncrieff, Svetha Venkatesh
Department of Computer Science
Curtin University of Technology
GPO Box U1987, Perth, 6845, W. Australia

Chitra Dorai
IBM T. J. Watson Research Center
P.O. Box 704, Yorktown Heights
New York 10598, USA

ABSTRACT

We examine localised sound energy patterns, or events, that we associate with high level affect experienced with films. The study of sound energy events in conjunction with their intended affect enable the analysis of film at a higher conceptual level, such as genre. The various affect/emotional responses we investigate in this paper are brought about by well established patterns of sound energy dynamics employed in audio tracks of horror films. This allows the examination of the thematic content of the films in relation to horror elements. We analyse the frequency of sound energy and affect events at a film level as well as at a scene level, and propose measures indicative of the film genre and scene content.

Using 4 horror, and 2 non-horror movies as experimental data we establish a correlation between the sound energy event types and horrific thematic content within film, thus enabling an automated mechanism for genre typing and scene content labeling in film.

1. INTRODUCTION

Modern cinema has developed into a tightly intertwined visual and aural medium. The sound track of a film often has a direct say on the impact of the visual component of the film, and influences the way in which the visuals are interpreted [1]. In our recent research we have identified specific sound patterns that result from the manipulation of the sound energy dynamics, the change in sound energy intensity over time, in film audio tracks. The localised sound energy patterns, which we term sound energy events, relate to different high level affect, or emotional inflection of on-screen events portrayed. We refer to occurrences of sound energy events in film that directly correspond to the proposed emotional inflection for the particular event, affect events. The sound energy events that we have identified convey conventional meanings that are commonly associated with films of the horror genre. The manipulation of the sound energy represents a concrete and commonly used method to convey horror themes and to heighten the impact of on-screen events.

We have previously presented an algorithm for the detection and classification of these sound patterns and studied their association with various affect [2].

In this paper we investigate the underlying indexical nature of the affect events by examining the occurrences of the sound energy events within the film and the nature of the content they accompany. The high level semantic association between the sound energy and affect events which exists at the localised event level, can be extended to attribute a semantic correlation between affect events and the broader thematic content of the film. In this paper, we examine the usage of the sound energy events and associated affect at two levels of granularity of the medium – the film level and the scene level, and propose measures indicative of the genre and scene content.

We study occurrences of the affect events in 6 movies. Four of the movies, The Mummy (action, horror genre), Pitch Black (horror, science fiction), Scream (horror, mystery), and Aliens (science fiction, action, horror) contain multiple elements of horror. The remaining two movies, Twelve Monkeys (thriller, science fiction) and Titanic (drama, romance) that are analysed, are used to compare the use of the same sound energy dynamic patterns in non-horror films. Overall, 12.63 hours of film audio is used as test data.

Applications of this work lie in automatic, broad semantic classification of both film and scene content for many purposes including content labeling for consumption by distinct audience categories, and comparison of prevalent and successful use of audio cues in horror films for distillation of effective practices of the craft.

2. BACKGROUND

2.1. Related Work

Related research in the area of film audio analysis has addressed the detection of violent sounds in audio [3] and analysis of occurrences of particular sound effects in an attempt to determine a label for the thematic content of a film scene. Other related work includes the use of audio features for determining scene transitions within film [4], and the use of visual features to determine film scene transitions where the scene transitions are defined according to film grammar [5]. In [2] we introduced our current work concerning the examination of sound energy events in horror films and the event classification and detection algorithm.

2.2. Sound Energy Events

The sound energy events of interest are composed of specific patterns in the dynamical characteristics of the sound energy of film audio tracks, where the energy and its dynamics are manipulated through musical scores or sound effects, and often silence. The affect or feelings experienced pertaining to the four types of sound energy events studied are: (i) surprise or alarm (type 1); (ii) apprehension, or the emphasis of an event (type 2); (iii) surprise followed by sustained alarm (type 3); and (iv) apprehension building up to a climax (type 4). For details of sound energy characteristics and detection of events, refer to [2].
2.3. Terminology

The four types of sound energy events detected by our event detection and classification algorithm [2] are partitioned into two sets. **Affirming affect** events are those sound energy patterns detected by the algorithm that correspond to one of the four types of sound energy events as well as to one of the proposed affect under consideration. The remainder of the sound energy events are termed **non-affirming events**, and correspond to those sound energy events detected by the algorithm, which do not exhibit an affect being studied. The ground truth affect event set consists of all the affect events present in the film analysed.

3. FILM LEVEL: SOUND ENERGY EVENTS AND FILM GENRE

We first studied the presence of sound energy events within a film to determine whether a correlation exists between the occurrences of the events and the genre of the film.

For the four horror films that were examined, a large number of sound energy events were detected by the algorithm, both as non-affirming events and affirming affect events. There was also stronger support for sound energy event to affect event mappings within horror films, with a greater proportion of sound energy events accurately corresponding with their proposed affect. This suggests a link between the proposed affectivity and the use of sound energy events to communicate the affectivity within horror films. In particular, the number of affect events present in the horror films, when compared to those in the non-horror films studied, implies that a distinction can be drawn between the two genres based on the examination of the sound energy events.

Consequently, we hypothesise that we can detect films that are of the horror genre by the examination of the frequency of the sound energy events within the film. We further propose that the frequency of the sound energy events is linked to the degree to which horror themes are present within the film. Since sound energy events that are detected consist of both the non-affirming events and the affirming affect events, examining them allows us to determine a concrete link between the results produced by the sound energy event detection algorithm and the thematic content of the film.

We claim that the inclusion of non-affirming events in the analysis of the frequency of sound energy events occurring in the film, is valid. Errors will occur due to the presence of non-affirming events. However, we propose that the errors will have a minimal effect on the ultimate classification of the thematic content of the film when using the frequency of the sound energy events. The presence of the affirming affect events will influence the sound energy event frequency sufficiently to result in an accurate reflection of the horror content of the film. We examine the correlation of the frequency of the sound energy events to the genre of the films studied, along with the link between event frequency and the degree of horror content within the film.

3.1. Sound Energy Event Frequency Results

Table 1 displays the frequency characteristics of the sound energy events on a per hour basis for both the horror and non-horror films studied. The frequency of events are displayed as the average number of events occurring in the film per hour. The table also shows the duration of each film. The frequency analysis results of the sound energy events are broken down into three major groups within each sound energy event type: the ground truth affect events determined for the sample films, affirming affect events frequency – the frequency of the sound energy events detected by the algorithm that correspond to the proposed affect, and sound energy events detected – the frequency of the total number of sound energy events detected within the film, including both affect and non-affect events.

3.2. Sound Energy Event Frequency Analysis

The analysis for the results shown in Table 1 indicates that the occurrences of the events is correlated with the presence of horror themes within film (see column 4 across horror movies). Furthermore, in certain instances an association between the frequency of events and the degree to which horror is present in the film is apparent. The sound energy event type 1 exhibits the strongest link between the frequency of occurrences of the events, and both the presence of horror themes within film and the degree to which these themes are present.

In the case of sound energy event types 2 and 4, while there is a difference between the frequency values for the film types, this distinction is less pronounced in comparison to sound energy event type 1. The frequencies for sound energy events 2 and 4 are inadequate for comprehensively distinguishing between the genres of horror and action/horror, and action/horror and non-horror. However, there is a sufficient distinction between the frequency of the events in horror films and in non-horror films. Both of these factors are related to the more generic affect associated with sound energy event types 2 and 4 when compared to the affect of type 1 event. The results suggest that the method of analysing the frequency of sound energy events can be validly used as an indicator of film content relating to horror themes.

3.3. Combined Frequency and Genre Typing

We attempt to devise a single, readily computable metric to indicate the horror theme content within the film. To achieve this we combine the per hour frequency values for each individual sound energy event type within the film to calculate a total frequency index by summing the individual frequencies of each of the four event types.

<table>
<thead>
<tr>
<th>Film</th>
<th>Sound Energy Events</th>
<th>Ground Truth Affect Evts.</th>
<th>Affirming Affect Evts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch Black</td>
<td>37.20</td>
<td>28.05</td>
<td>23.17</td>
</tr>
<tr>
<td>Scream</td>
<td>35.54</td>
<td>26.48</td>
<td>21.67</td>
</tr>
<tr>
<td>The Mummy</td>
<td>21.54</td>
<td>20.45</td>
<td>14.37</td>
</tr>
<tr>
<td>Aliens</td>
<td>19.06</td>
<td>11.50</td>
<td>10.72</td>
</tr>
<tr>
<td>Titanic</td>
<td>15.00</td>
<td>6.34</td>
<td>5.66</td>
</tr>
<tr>
<td>12 Monkeys</td>
<td>13.00</td>
<td>6.50</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Table 2: Combined frequency characteristics for horror and non-horror films.

Table 2 displays the total frequency indices for each film examined, for ground truth affect events, affirming affect events, and sound energy events. The results for the total frequency index display similar properties to the frequency analysis for each constituent event type. However the index indicates a clearer distinction between general genre categories compared to individual frequency analysis of event types 2 and 4.
The values for the combined frequency index for the total sound energy events are of significance due to the apparent distinction between genres based on them. The films are split into 3 categories, both by frequency and genre: *Pitch Black* (37.2) and *Scream* (35.54) are pure horror, *Aliens* (19.06) and *The Mummy* (21.54) are action/horror, and *Titanic* (15.0) and *12 Monkeys* (13.0) are non-horror. This indicates that a general divide between the degrees of horror content of films is achievable based on the combined frequency of the total sound energy events detected.

### 4. SCENE LEVEL: SCENE CONTENT LABELING

We next examine the film at a finer scale. We investigate the occurrences of sound energy events at the scene level of film, drawing conclusions between sound energy event presence and the thematic content of scenes with respect to horror. We attempt to develop a method for the automatic detection of scenes containing strong horror thematic content within the film, given scene transitions in the film and time indices for the sound energy events occurring within the film.

#### 4.1. Horror Scene Determination

For scene extraction, we adopt the definition of a film scene as used in [5], one relating purely to visual aspects.

For our purposes, we extend the concept of a scene to determine a definition for a horror scene. In detecting horror scenes automatically we focus on the detection of scenes with a high degree of horror thematic content, rather than scenes that contain, for example, a single brief scary shot. Subsequently, we refer to a horror scene as a film scene that exhibits prolonged, continuous sections of horrific nature. For example, in *Scream*, the penultimate scene consists of a prolonged chase sequence between the killers and their victims. This constitutes a horror scene as a continuous danger towards the protagonist is portrayed. A further distinction is necessary in the case of action films that contain elements of horror. Where both action and horror themes are present within a scene, if action dominates horror, the scene is not labeled as a horror scene. For example, a scene that consists of a brief shock or scare that is subsequently by an action sequence is classified as an action scene.

As in the case of examining sound energy events at the film level, we hypothesise that occurrences of affect events are linked to the thematic content of the scene. More specifically, we propose that sound energy events are more concentrated within scenes in which horror themes are particularly prevalent. Consequently we examine the relation between the horror thematic content of the scene and the number of sound energy events contained within the scene.

#### 4.2. Ground Truth

There were two stages in determining the ground truth horror scenes for the six films examined.

- Initially the ground truth scene index is determined for the film in accordance to the guidelines detailed in section 4.1.
- Scenes that contain a strong degree of horror themes, in concurrence with the guide outlined in section 4.1, were determined and recorded as horror scenes. Scenes containing elements of horror, although not horror scenes, were also noted.

#### 4.3. Scene Content Analysis Method

We utilise the start and end times from the scene transition indices of the film and the time indices of the sound energy events detected to determine the number of sound energy events occurring within each scene of the film. Using the number of sound energy

---

Table 1. Event frequency (per hour) characteristics for horror and non-horror films.
events in the scene as a metric we determined a method for the detection and classification of horror scenes. The detection criterion adopted is as follows: if a scene contains three or more sound energy events, the scene is classified as a horror scene. The criterion was derived empirically to determine an optimum number of sound energy events for the classification of the scenes.

Figure 1 shows the graph of the number of false negative and false positive horror scenes plotted against the number of true positive horror scenes using a minimum of 2, 3, and 4 sound energy events to determine a horror scene. The optimum results were achieved using a minimum of 3 sound energy events. While there are valid horror scenes with one or two affirming affect events, the minimum of three sound energy events was selected to reduce the number of false positives returned by the horror scene analysis. Further, requiring a higher number of sound energy events also accounts for occurrences of non-affirming events which can invalidate the inference on horror content occurring in the scene.

![Figure 1. Determining the optimal classification of horror scenes.](image)

### 4.4. Results and Analysis

Table 3 displays the results on automatic determination of horror scenes. The overall precision for the detection of horror scenes in horror film is 92.9%, while the recall is 72.2%. While the size of the data set and the number of horror scenes precludes a more concrete analysis, this preliminary data suggests that the use of sound energy events to determine horror scenes is viable. However, while viable, the lower recall indicates that the method is less successful in detecting the full set of horror scenes within the films.

In the case of non-horror films it appears that this method of scene analysis cannot apply. For example, in *Titanic*, there are six scenes with 3 or more sound energy events, signaling horror scenes and given the length of the film, the frequency of the sound energy events is low yet the number of false positive horror scenes is high. This is due to the concentration of non-affirming events in scenes with low sound energy and scenes where the audio track exacerbates algorithmic errors leading to detection of non-affirming events. Consequently, due to the dichotomy between the low frequency of sound energy events and the large number of false positive horror scenes, we propose the use of the frequency analysis of the film as a pre-processing step. The frequency analysis for the entirety of the film is used to determine whether the film exhibits a low sound energy event frequency. This excludes non-horror films from any further scene analysis.

![Table 3. Results on the detection and classification of horror scenes.](image)

### 5. CONCLUSION

In analysing the frequency of sound energy events in film, we determine a correspondence between incidences of the events and the degree to which horror themes are present. The analysis of the frequency of the events is used as a feasibility study to determine the correspondence. Further work involving the use of more test data and various classification schemes is required to establish a more concrete link between thematic content and the use of the sound energy events in film.

At a film level, we establish a link between the usage of the affect events and the elements of horror as a genre in the film, and thus develop a broad classification of film content. Distinctions can also be drawn between the various degrees to which horror is present in the films examined based on the sound energy event frequency, thus aiding in genre typing. Analysis of further sample films is required to derive a metric relating the combined frequency of affect and sound energy events to the extent to which horror related themes occur in the film, thus aiding in genre typing.

Similarly, at a scene level, we attempt to determine a link between sound energy events and the thematic content of the scenes. The preliminary results suggest that the use of sound energy events to determine horror scenes in horror films is viable.

### 6. REFERENCES


