The time contour plot: graphical analysis of a film soundtrack

Sound and the Screen
University of West London, 20 November 2015

Nick Redfern,

School of Arts and Communication, Leeds Trinity University,
Brownberrie Lane, Horsforth, Leeds, LS18 5HD, UK.

n.redfern@leedstrinity.ac.uk
https://leedstrinity.academia.edu/NickRedfern

Abstract

Audio segmentation comprises a set of techniques for analysing the features of audio signals, including motion picture soundtracks. Parsing the structure of film audio allows us to identify scenes and change points, features in the audio envelope (attack, decay, sustain, release), the distribution of sound energy, and the presence of affective events in a soundtrack. Despite the availability of a wide range of software capable of audio analysis (e.g. Python, R, Sonic Visualiser, Audacity, Speech Filing System, Raven Lite, etc.) statistical analyses of cinematic style have yet to apply these methods to the formal analysis of film soundtracks systematically. In this paper I demonstrate the analysis of film sound using the time contour plot of generated by the **R** package **seewave**. The time contour plot is a graphical method for visualising the temporal structure of sound energy based on the short-time Fourier transform of a signal and allows analysts to identify interesting features in a film's soundtrack that warrant further examination and to communicate is a straightforward manner that is easy to comprehend. Applying this method to the soundtrack of *Behold the Noose* (Jamie Brooks, 2014), a short horror film, I show that the sound energy in the film increases non-linearly in creating of a state of heightened anxiety in the viewer.

Bio

Nick Redfern is Associate Senior Lecturer in Film Studies at Leeds Trinity University. He has published articles on the statistical analysis of film style in *Empirical Studies in the Arts*, the *Journal of Data Science*, the *International Journal of Communication*, and the *Journal of Japanese and Korean Cinema*.

Introduction

Nick Redfern

The time contour plot: graphical analysis of a film soundtrack

n.Redfern@leedstrinity.ac.uk

20 November 2015

Leeds Trinity University

At present there are no quantitative analyses of the aesthetics of film sound, but this is an area where the application of quantitative methods can contribute significantly to understanding style in the cinema.

There is a large literature on the audio segmentation of motion pictures that could provide a wide range of new methodologies and ways of thinking about film sound empirically (see Brezeale and Cook 2008 or Theodorou, Mporas, and Fakotakis 2014 for an overview). Audio segmentation comprises a set of techniques for analysing the features of audio signals, including motion picture soundtracks. Parsing the structure of film audio allows us to identify scenes and change points, features in the audio envelope (attack, decay, sustain, release), the distribution of sound energy, and the presence of affective events in a soundtrack. This approach has been used to detect narrative structure via audio pace (Moncrieff and Venkatesh 2007), exploring temporal coincidences between the visual and audio in montage editing (Zeppelzauer, Mitrović, and Breiteneder 2011), and to identify changes in the soundscapes of motion pictures (Orio 2013). These are typically the features of film soundtracks researchers are interested in but this work has been ignored by film scholars.

There is also a wide range of freely available software designed for this purpose that could be employed by film researchers, including Python, R, Sonic Visualiser, Audacity, Speech Filing System, and Raven Lite. The application of these methods and the availability of this software could potentially transform the analysis of film sound by allowing researcher to work directly on soundtracks in a detailed way but, again, this opportunity has been ignored by film researchers.

Introduction
Methods
Structure and evolution
Summary

Introduction

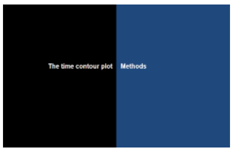
Key Aim: to demonstrate the use of the time contour plot of the normalised aggregated power envelope as a quantitative method for analysing film soundtracks

Example: *Behold the Noose* (Jamie Brooks, 2014)

Nick Redfern The time contour plot

In this paper I demonstrate the analysis of film sound using the time contour plot of the normalised aggregated power envelope generated by the **R** package **seewave**. The time contour plot is a graphical method for visualising the temporal structure of sound energy based on the short-time Fourier transform of a signal and allows analysts to identify interesting features in a film's soundtrack that warrant further examination and to communicate is a straightforward manner that is easy to comprehend. Applying this method to the soundtrack of *Behold the Noose* (Jamie Brooks, 2014), a short horror film, I show that the sound energy in the film increases non-linearly in creating of a state of heightened anxiety in the viewer.

Methods

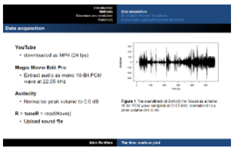


The time contour plot

Methods

In this paper I use the time contour of the normalised aggregated power envelope derived from the short-time Fourier transform of soundtrack. This method is implemented using the packages **tuneR** (v. 1.2.1; Ligges et al. 2013) and **seewave** (v. 1.7.6; Sueur, Aubin, and Simonis 2008) for the open source statistical software **R** (v. 3.0.1; R Core Team 2013).

The first stage in the analysis is data preparation. *Behold the Noose* was downloaded from YouTube as an mp4 file (24 fps) and loaded into a non-linear editing suite in order to export the film’s audio as a mono 16-bit wave file sampled at 22.05 kHz. (Higher sampling rates and/or stereo soundtracks can be used but this substantially increases the computational cost of the analysis). This audio file was then normalized to a peak volume of 0.0 dB using Audacity (v. 2.0.6) and re-exported as a mono 16-bit PCM wave file. This wave is then loaded into **R** using the readWave() function in **tuneR** (Figure 1).



Introduction
Methods
Structure and evolution
Summary

Data acquisition
Short-time Fourier Transform
Normalised aggregated power envelope

Data acquisition

YouTube
- downloaded as MP4 (24 fps)

Magix Movie Edit Pro
- Extract audio as mono 16-Bit PCM wave at 22.05 kHz

Audacity
- Normalise peak volume to 0.0 dB

R > tuneR > readWave()
- Upload sound file

Figure 1 The soundtrack of *Behold the Noose* as a mono 16-bit PCM wave sampled at 22.05 kHz, normalised to a peak volume of 0.0 dB.

Nick Redfern The time contour plot

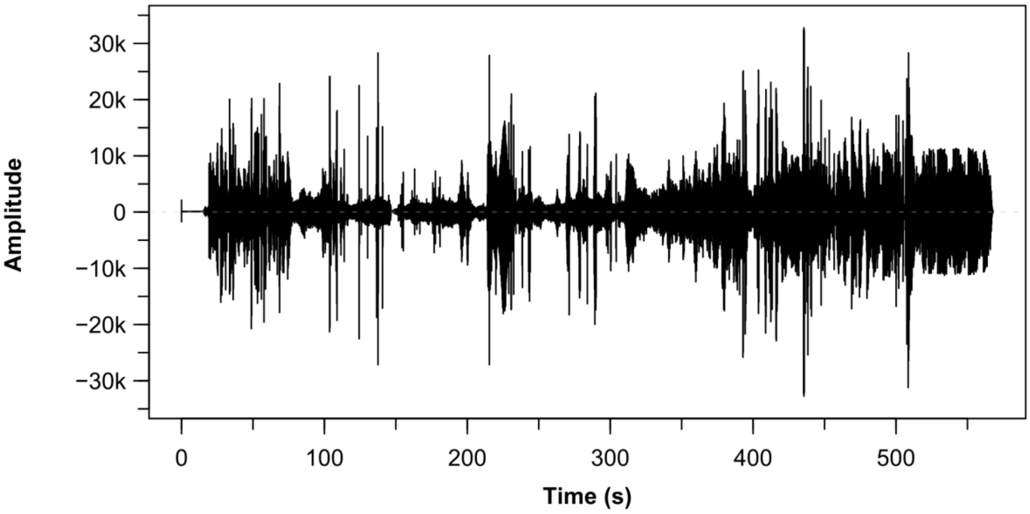
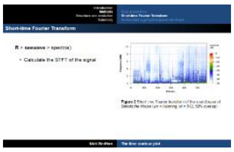


Figure 1 The soundtrack of *Behold the Noose* as a mono 16-bit PCM wave sampled at 22.05 kHz, normalised to a peak volume of 0.0 dB.

The next stage applies the short-time Fourier transform (STFT) to the wave to produce a 2D time-frequency representation of the signal called a spectrogram (see Goodwin 2008). The STFT divides the signal into a series of windows and calculates the Fourier transform for each window. The result is a Fourier transform of the signal localised in time dependent upon the shape (rectangle, Hanning, etc.), size (the number of samples within a window), and overlap of the window used. The spectrogram describes how the magnitudes of the individual frequencies comprising a signal vary over time, with the power spectral density indicated by colour. The **seewave** function spectro() calculates the STFT and spectrogram of a signal. Figure 2 shows the spectrogram of the soundtrack of *Behold the Noose* using Hanning windows with a length of 512 overlapped by 50 per cent.



Introduction
Methods
Structure and evolution
Summary

Data acquisition
Short-time Fourier Transform
Normalised aggregated power envelope

Short-time Fourier Transform

R > seewave > spectro()
- Calculate the STFT of the signal

Figure 2 Short-time Fourier transform of the sound wave of *Behold the Noose* (wn = Hanning, wl = 512, 50% overlap)

Nick Redfern The time contour plot

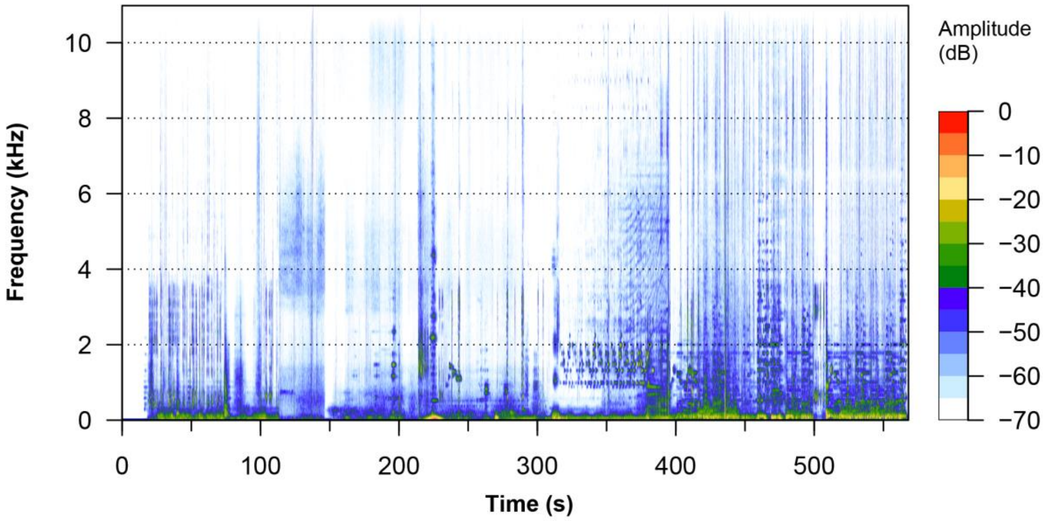
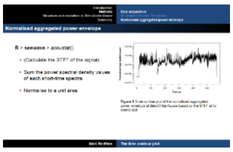


Figure 2 Short time Fourier transform of the sound wave of *Behold the Noose* (wn = Hanning, wl = 512, 50% overlap)

A time contour plot is generated by summing the power spectral density values in each of the short-time spectra in Figure 2 to produce an aggregate power envelope, which is then normalised to a unit area and treated as a probability mass function (Cortopassi 2006). The resulting plot shows how the energy of a signal evolves over time. It is not necessary to manually carry out the above calculations on the STFT as the **seewave** function acoustat() automatically produces the time contour. Figure 3 presents the time-contour plot of the normalised aggregated power envelope of *Behold the Noose*.


Introduction
Methods
Structure and evolution in *Behold the Noose*
Summary

Data acquisition
Short-time Fourier Transform
Normalised aggregated power envelope

Normalised aggregated power envelope

R > seewave > acoustat()

- (Calculate the STFT of the signal)
- Sum the power spectral density values of each short-time spectra
- Normalise to a unit area

Figure 3 Time contour plot of the normalised aggregated power envelope of *Behold the Noose* based on the STFT of its soundtrack

Nick Redfern The time contour plot

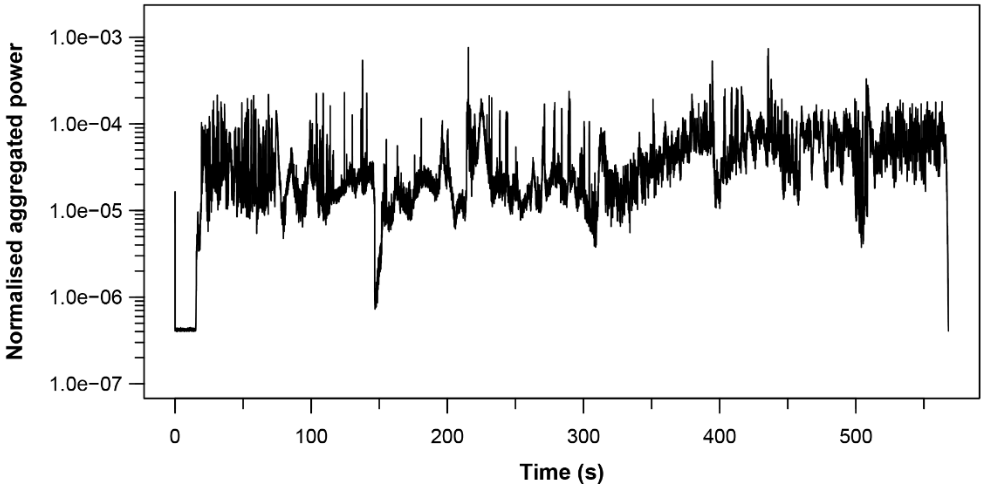
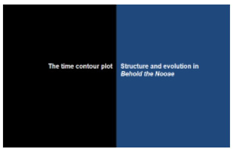


Figure 3 Time contour plot of the normalised aggregated power envelope of *Behold the Noose* based on the STFT of its soundtrack

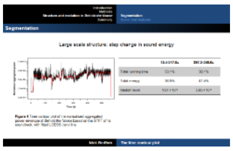
Structure and evolution of sound energy in *Behold the Noose*



The time contour plot

Structure and evolution in *Behold the Noose*

Analysis of the normalised aggregated power envelope allows us to describe the large scale temporal structure of *Behold the Noose* and to segment the audio and narrative structure of the film. It also allows us to understand the nature of the temporal evolution of sound energy for specific narrative features in the film.



Segmentation

Large scale structure: step change in sound energy

Visual inspection of the time contour plot immediately suggests the overall pattern of the evolution of the film's sound energy is a step function change from the slow build-up of tension in the second part of the film to the violence of the final segment linked by a transitional phase (see Figure 4). There is a clear change in the level of sound energy: the median value of the normalised aggregated power envelope between 15.4 and 317.0 seconds in Figure 4 is 1.97 x 10^{-5} and 5.85 x 10^{-5} between 397.3 and 568.0 seconds. The cumulative proportion of energy in the earlier part of the film is just over a third of the film's total (36.5%) even though at 301.6 seconds it accounts for over half the film's running time (53.1%), whereas the cumulative proportion of the later part of the film (which at 170.7s is 30.1% of the total running time) is just under half the total energy (47.4%).

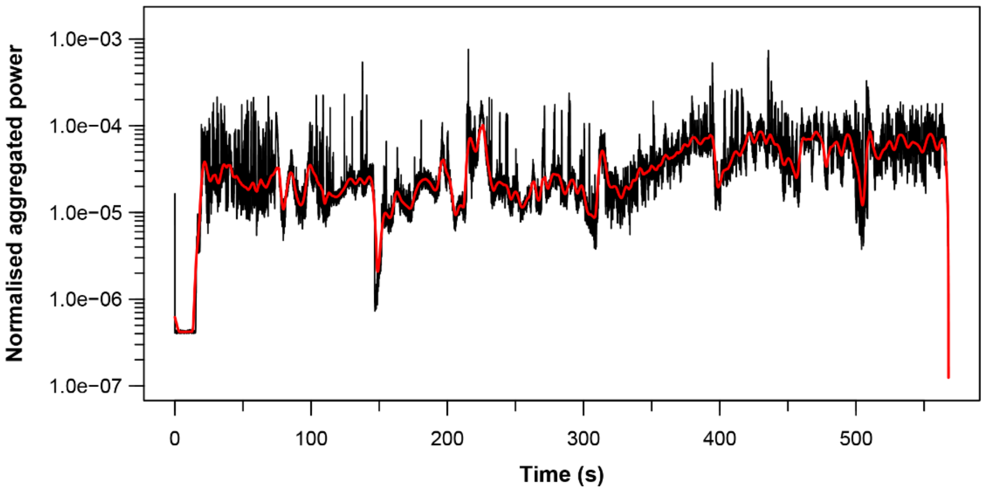
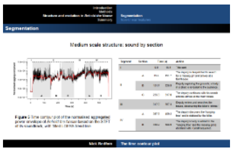


Figure 4 Time contour plot of the normalised aggregated power envelope of *Behold the Noose* based on the STFT of its soundtrack, with fitted LOESS trend line



Segmentation

Medium scale structure: sound by section

At the medium scale there are four main segments in the film and that segments II and IV can be further divided into three and two sections, respectively (see Figure 5). This structure is summarised in Table 1.

Segment I comprises a title attesting to the veracity of events in the movie and advising viewer discretion, but there are no sounds during this part of the film and the energy in this part of the film is very low.

The second segment of the film comprises three distinct sections. Section II.A establishes the film's premise as the Sheriff's deputy is called over the radio to assist in the search for a missing girl and drives to the farmhouse. On arriving, he exists the vehicle and retrieves his shotgun before heading off to search the premises. The sounds in this part of the film include the voice of the deputy and the dispatcher on the radio, the noises of the vehicle (the engine while driving and the parking brake on arrival), while the atmosphere of the sequence is created through a combination of music and crickets chirping in the background. Throughout this section the sound energy level remains even, though there are changes in the spread when no-one is talking (e.g. from 75.0s to 100.0s).

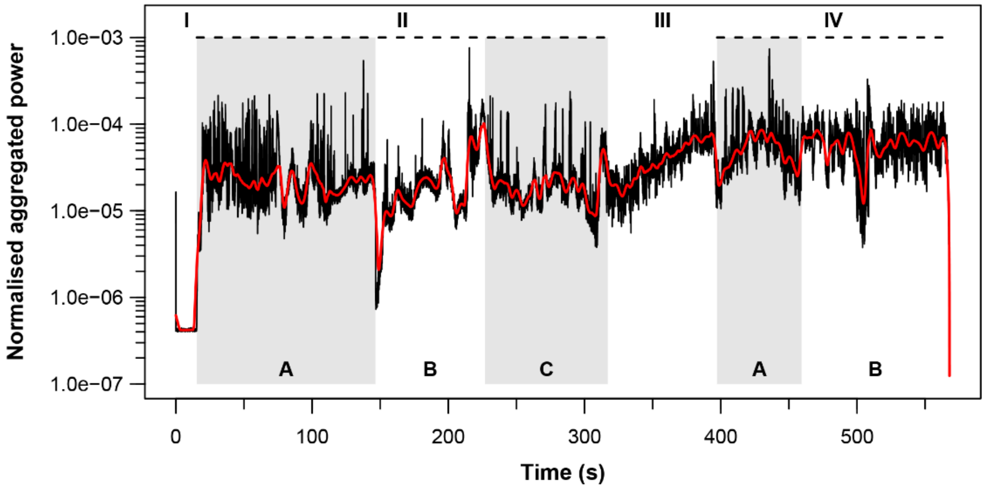


Figure 5 Segmentation of *Behold the Noose* based on the time contour plot of the normalised aggregated power envelope

Table 1 The structure of *Behold the Noose* based on segmentation of the on the time contour plot of the normalised aggregated power envelope

Segment	**Section**	**Time (s)**		**Action**
		Start	End	
I		0.0	15.4	Title card
II	A	15.4	151.7	The deputy is despatched to search for a missing girl and arrives at a farmhouse.
	B	151.7	226.9	Deputy exploring the grounds; a body in a shed is revealed to the audience.
	C	226.9	317.0	The deputy continues with his search until he arrives at the main house.
III		317.0	397.3	Deputy enters and searches the house, discovering the killer's 'shrine.'
IV	A	397.3	459.2	The deputy discovers the 'hanging tree' and is stabbed by the killer.
	B	459.2	568.0	The deputy's body is added to the 'hanging tree' and the missing girl is declared safe. Credit sequence.

Section II.B begins in silence as we see the deputy walking on a CCTV monitor, though there is no indication of who is watching. There is little dialogue in this part of the film (the deputy uses his radio twice), and the effects used comprise the deputy's footsteps and the chirping of crickets. Within this section there is a series of individual peaks, each associated with an increase in emotional intensity as the deputy enters the yard (163.0s), turns a corner into the unknown (183.5s), discovers the bloody sheet (196.5s), and, finally, at 216.0s the deputy opens a door to release a flock of cawing birds before the final 'true shock' of this sequence is revealed – a body hanging in the shed (the peak at 225.3s). These last two features produce a 'double peak' in sound energy.

The hanging body is shown to the audience but the deputy is unaware of its presence as we move into section III.C and he continues his search of the grounds. In this section the film returns to the relatively even sound energy level of section II.A with the voice of the deputy as he calls out and the crickets in the background. There is no music in this section to speak of, with the atmosphere created by a mixture of wind effects, throbbing noises, and ghost-like noises.

Segment III begins as the deputy enters the farmhouse. The sound energy in the early part of this sequence continues at the level of section II.C, but the sounds themselves have changed. As in section II.B there is little dialogue until the deputy radios for help (384.6s) and there are few effects aside from the buzzing of flies in the kitchen, which is eventually replaced by a similar non-diegetic noise that eventually distorts. The sound energy of this sequence increases from 327.9s as the deputy moves deeper into the house to 378.3s, when he discovers the killer's shrine to his victims – a scene that is first revealed to the audience with a shot of a skull in a jar. This increase in sound energy is related to the spatial experience of the film's 'terrible place:' as the deputy moves further into the house the energy of the soundtrack increases until peaking at the moment where he is furthest from safety and confronted with his terrible discovery. The peak level is sustained until 397.3s as the deputy surveys the photographs in the shrine and begins to panic, desperately radioing for help. The sound energy then rapidly decays to the level at which this segment began after a banging noise and laughter is heard off-screen and the deputy turns to exit the building. At 395.1s the song 'Witch Lynching' by Forever and Everest (2012) begins and dominates the remainder of the film's soundtrack.

The final segment of the film contains all the violence of the film and comprises two distinct audio sections. In section IV.A the deputy makes his way outside the farmhouse to discover numerous bodies hanging from a tree. The energy of the first part of this sequence (397.3s-421.5s) increases as the song 'Witch Lynching' takes over the soundtrack and the deputy calls out into the darkness, reaching a peak as the deputy gasps in shock at the sight of the hanging tree. The sound energy of this feature does not immediately decay and is sustained from 421.5s until the deputy is stabbed is by the killer and his shotgun fires. From 437.0s the energy of the sound track decays until 459.2s as the blade is driven deeper into the deputy's body and he finally falls to the floor. The volume of the song is reduced to bring the body horror of the deputy's goring to prominence and the soundtrack emphasises the sounds of the blade twisting in the deputy's body and his final, gasping breaths. Section IV.B begins with the rapid attack of the trumpets from the song and returns to the level of section IV.A, which is then sustained until the end of the film. The action in this part of the film sees the deputy's body dragged through the words and hanged before cutting to the deputy's vehicle as the dispatcher comes over the radio to announce the missing girl is safe and sound. The sound energy drops to a low at 505.0s as the dispatcher's voice fades and the volume of the music reduces before the sudden attack of the whistling refrain of 'Witch Lynching' returns as flames burst from the farmhouse's chimneys to illuminate the night sky. After this the credits roll over the remainder of the song.

Introduction
Methods
Summary

Non-linear features

Non-linear features

Non-linear increase in energy and emotion

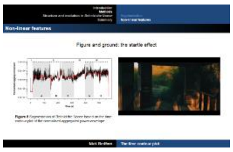
The time contour plot

The time contour plot in Figure 5 reveals a key feature of the sound mixing in sections II.B, III, and IV.A not previously identified in studies of sound in horror cinema: in each case the attack of the sound event increases slowly in the first part of the sequence before accelerating rapidly to a climax and follows a *non-linear crescendo* pattern. The decay of the sound event in section IV.A (437.0-459.2s) is also non-linear as the deputy succumbs to his fate. Smaller-scale sound events exist within these section-level events. For example, section II.B features some local events (the individual transient peaks) within a non-linear increase in sound energy that runs for the whole length of the sequence. In this section, each peak has a higher level of sound energy than the last and so we should not see these shocks as existing isolation but as part of a single dynamic structure producing a cumulative effect in the viewer. Because the sound energy does not fall back to the level prior to the peak these small-scale events contribute to the overall non-linear increase of the whole section.

A feature of horror film soundtracks noted by several scholars is the use of 'assaultive blasts that coincide with shock or revelation' (Lerner 2010: ix) or *stingers* to produce emotional effects in the audience. Stingers take the form of noises such as screams, orchestral music, or sound effects characterised by their suddenness, their stabbing shortness, and a sudden increase in volume to produce a startle effect (Hutchings 2004: 134-137; Baird 2000) that triggers a basic reflex of shock or surprise resulting from the collision of loud and soft sounds (Whittington 2014):

> The fact that these scenes are often equally effective underscores the significance of *contrast* between a relatively *loud* sound bursting into a relatively *quiet* scene. In gestalt theoretical terms: a figure that stands off perceptibly from a ground (Hanich 2010: 134; original emphasis).

Hanich (2010: 134) claims that the crescendo culminating in a short-sound stab is a less effective method of producing the desired audience response than a sudden, unexpected increase in volume, though he offers no empirical support for this claim (and does not specify the form of the crescendo). Analysis of the dynamics of sound energy in *Behold the Noose* shows this general description is unable to adequately account for the evolution of sound energy over time. The evolution in sound energy in section II.B occurs at both the local scale of the transient events and at the medium scale of the section. Consequently, the difference between ‘figure’ and ‘ground’ is not clear-cut: the peaks (or ‘figures’) associated with each event are part of the larger structure of the soundtrack evolving at a higher scale and contribute to the overall dynamic of the section as well as creating transitory moments of heightened tension. Because those small scale moments cannot be separated from the larger scale of the crescendo and so the claim that one method is more effective than another cannot be justified. When analysing horror film soundtracks it is therefore necessary to examine not only those moments of extreme horror but also their place within the larger structure of the soundtrack.


Introduction
Methods
Structure and evolution in *Behold the Noose*
Summary

Non-linear features

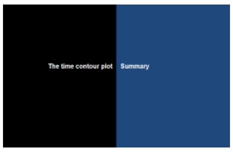
Figure and ground: the startle effect

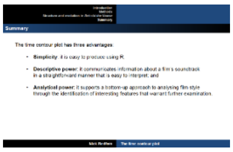
Figure 8 Segmentation of *Behold the Noose* based on the time contour plot of the normalised aggregated power envelope

Nick Redfern | The time contour plot

Conclusion

This paper demonstrates the use of quantitative methods for analysing the structure of sound in horror cinema using the short-time Fourier transform and the normalised aggregated power envelope. These methods have several features to recommend them for analysing the structure of sound in motion pictures. First, they are easy to compute using freely available software, though the computing power and the time required for analysing feature film soundtracks will be high. Second, they have descriptive power, communicating detailed information about a film’s soundtrack in a straightforward manner that is easy to interpret. Third, they have analytical power supporting a bottom-up approach to analysing film style that allow the researcher to identify interesting features in a soundtrack and then to look beneath them in order to understand what is going on here.


The time contour plot | Summary


Introduction
Methods
Structure and evolution in *Behold the Noose*
Summary

Summary

The time contour plot has three advantages:

- **Simplicity**: it is easy to produce using R;
- **Descriptive power**: it communicates information about a film’s soundtrack in a straightforward manner that is easy to interpret; and
- **Analytical power**: it supports a bottom-up approach to analysing film style through the identification of interesting features that warrant further examination.

Nick Redfern | The time contour plot

Applying these methods to a short horror film I identified a range of features to provide a detailed understanding of how sound functions in *Behold the Noose*. The time contour plot of the normalised aggregated power envelope allows us to segment the film and define its structure at different scales. At the micro scale individual moments of horror in the film stand out, while at the medium scale each section of the film can be easily distinguished from those that precede and follow it based on the time contour plot. A key feature revealed by this method is the non-linear sound mixing in the sequences in which the deputy searches the film’s ‘terrible place.’ It is a feature that plays a key role in how *Behold the Noose* builds tension to a climax to create an emotional response in the spectator. This is an original result (surprisingly) not previously identified in studies of horror film soundtracks. This feature is not identifiable from simply listening to the film’s soundtrack or from examining either the waveform or the short-time Fourier Transform of the soundtrack; but is immediately evident in the time contour plot, demonstrating the value of this method. Finally, at the macro scale the evolution of the sound energy is a step function change between the set-up of the narrative and early, transient scares and the emotional intensity and bloody violence of the deputy's murder.

References

Baird, R. (2000) The startle effect: implications for spectator cognition and media theory, *Film Quarterly* 53 (3): 12-24.

Beeman, K. (1998) Digital signal analysis, editing, and synthesis, in S.L. Hopp, M.J. Owren, and C.S. Evans (eds.) *Animal Acoustic Communication: Sound Analysis and Research Methods*. Berlin: Springer-Verlag: 59-103.

Behold the Noose (Jamie Brooks, USA, 2014, 9:28 min), https://www.youtube.com/watch?v=OhlDW9_Ehik, accessed 8 May 2015.

Brezeale, D., and Cook, D.J. (2008). Automatic video classification: a survey of the literature, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 38 (3): 416-430.

Cortopassi, K. A. (2006) Automated and robust measurement of signal features, http://www.birds.cornell.edu/brp/research/algorithm/automated-and-robust-measurement-of-signal-features, accessed 4 June 2015.

Forever and Everest (2012) Witch Lynching, on *Filthy Songs & Dirty Stories*, https://foreverandeverest.bandcamp.com/track/witch-lynching, accessed 20 July 2015.

Goodwin, M.M. (2008) The STFT, sinusoidal models, and speech modification, in J. Benesty, M.M. Sondhi, and Y. Huang (eds.) *Springer Handbook of Speech Processing*. Berlin: Springer: 229-258.

Hanich, J. (2010) *Cinematic Emotion in Horror Films and Thrillers: The Aesthetic Paradox of Pleasurable Fear*. London: Routledge.

Hutchings, P. (2004) *The Horror Film*. Abingdon: Routledge.

Kent, R.D. and Read, C. (2002) *The Acoustic Analysis of Speech*. Clifton Park, NY: Delmar.

Lerner, N. (2010) Preface: what about horror's ear?, in N. Lerner (ed.) *Music in the Horror Film: Listening to Fear*. London: Routledge: viii-xi.

Ligges, U., Krey, S., Mersmann, O., and Schnackenberg, S. (2013) tuneR: analysis of music, http://r-forge.r-project.org/projects/tuner/, accessed 8 May 2015.

Moncrieff, S., and Venkatesh, S. (2007) Film audio pace, *International Journal of Intelligent Systems Technologies and Applications* 3 (3-4): 296-308.

Orio, N. (2013) Soundscape analysis as a tool of movie segmentation, *Cinergie* 3: 157-163.

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, accessed 4 June 2015.

Sueur J., Aubin T., and Simonis C. (2008) seewave: a free modular tool for sound analysis and synthesis, *Bioacoustics* 18 (2): 213-226.

Introduction
Methods
Structure and evolution in *Behold the Noose*
Summary

Summary

Read the full paper:

Quantitative analysis of sound in a short horror film

@

https://leedstrinity.academia.edu/NickRedfern

Nick Redfern The time contour plot

Theodorou, T., Mporas, I., and Fakotakis, N. (2014) An overview of automatic audio segmentation, *International Journal of Information Technology and Computer Science* 6 (11): 1-9.

Whittington, W. (2014) Horror sound design, in H.M. Benshoff (ed.) *A Companion to the Horror Film*. Chichester: John Wiley & Sons: 168-185.

Zeppelzauer, M., Mitrović, D. Breiteneder, C. (2011) Cross-modal analysis of audio-visual film montage, in H. Wang, J. Li, G.N. Rouskas, X. Zhou (eds.) *Proceedings of 20th International Conference on Computer Communications and Networks*. Red Hook, NY: Curran Associates, Inc.: 1-6.