



Research & Development
White Paper

WHP 201

August 2012

**Affective Classification of
Large Scale Broadcast Archives**

Sam Davies *and* Denise Bland

BRITISH BROADCASTING CORPORATION

Affective Classification of Large Scale Broadcast Archives

Sam Davies and Denise Bland

Abstract

In this paper, we present an overview of our framework system for the affective classification of a large scale broadcast archive. Using a combination of video and audio processing we classify programmes according to their affective content, resulting in a mood vector for each programme. This is displayed on a two-dimensional graph, allowing users to select programmes based on mood. We also present an overview of our work on automatic event detection with the initial aim of identifying which sections of a sports match are of interest and ranks these by way of overall interest. This paper forms an overview of the British Broadcasting Corporation (BBC) Research and Development (R&D) department's work on automatically classifying TV programmes for entertainment re-use.

This document was originally published in proceedings of the SIGIR 2011 Workshop "Entertain Me": Supporting Complex Search Tasks. ACM Press, 2011, Beijing China.

Additional key words: Multimodal, feature extraction, semantic metadata, classification, retrieval, algorithms, management, experimentation.

White Papers are distributed freely on request.

Authorisation of the Chief Scientist or General Manager
is required for publication.

© BBC 2012. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/whp> for a copy of this document.

Affective Classification of Large Scale Broadcast Archives

Sam Davies and Denise Bland

1 INTRODUCTION

The amount of audiovisual material available to viewers in a digital format is growing rapidly as broadcast transmission capabilities increase and archived content is digitised. With this, it is important that users are able to find not only the media they want but also the segments of media they want. As such, metadata is required for each media asset to allow for inter and intra document searching. Within the British Broadcasting Corporation (BBC), all media assets that are likely to be reused have manually created metadata, contents of which range from brief synopses to detailed shot and topic listing. However, this is a time and resource expensive process. On average a detailed analysis of a 30 minute programme takes a professional archivist around 8 to 9 hours.

The BBC uses a system called LONdon CLASSification (LONCLASS), an in-house developed extension to the Dewey Decimal System for this classification. Programmes will also have an entry in the BBC's INFAX database, which contains all available metadata about a programme such as LONCLASS number, synopsis, and programme credits. Using this, factual aspects of a programme are classified and then easily found. Currently, BBC Information and Archives (I&A), the section of the BBC which is responsible for the archive, periodically release digitized collections of archived factual programmes themed around specific subjects. However, these are based on expanded INFAX entries. Within these collections, only those parts of programmes that are of relevance and interest are provided. As such, the synopses are all that is required for navigation and selection by viewers. This labour intensive process requires teams of professional archivists to search, tag and segment the archives by hand. As more content is digitised this manual segmentation and metadata generation will become less feasible meaning automation of this process will become more important. Currently only a small fraction of the archives are digitised but this is rapidly growing due to digitisation projects such as [1]. However the main purpose of manually generated metadata is for professional reuse. Frame accurate metadata is designed to allow users such producers and researchers to find stock shots, interviews or other precise sections. However there is currently limited provision for classifying programmes according to entertainment value – if a user doesn't want to find out about a particular event or person, if they just want to be entertained. Our Multimedia Classification project aims to automatically generate metadata that will allow for retrieval of content from broadcast archives when the user wants to be entertained.

Various UK broadcasters now offer 'catch-up' services, such as the BBC's iPlayer, with an industry wide move to integrate these with traditional television set top boxes. This ability to download and view vast amounts of content again presents a requirement for the ability to find and watch desired content. As viewers using these services have the ability to skip forward through the programme, identifying which parts of it are interesting would be of great benefit. This identification would be of even greater use in large scale sporting events, such as world cups or the Olympics, where large numbers of matches or competitions are broadcast either concurrently or in close time proximity. Event detection tools could also be of use in a production environment, allowing for quicker creation of highlights programmes by providing a candidate list of interesting or unusual events during an event.

2 MULTIMEDIA CLASSIFICATION

Our system comprises of three main sections; characteristics extraction, feature extraction and a final machine inference module. We take a multimodal approach analyzing both the audio and the video to create an affective vector for each programme. This is then displayed on a 2D graph with the affective adjective labels, Happy/Exciting and Serious/Lighthearted. These were not chosen as they are diametrically opposed, more that they were initially found to map to extracted features and characteristics. An overview of our system is shown in figure 1.

2.1 Characteristics and Feature Extraction

The system extracts characteristics from the audio and video signals using signal processing techniques described in [2]. These analyze different temporal and spectral aspects of the audio and video signals. These signals are then used to either identify objects in the audio or video or else in the machine inference modules on their own. Using statistical techniques as described in [2], features such as laughter, motion and shot cut frequency are identified. These are then used in the machine inference module to identify the affective content of the programme.

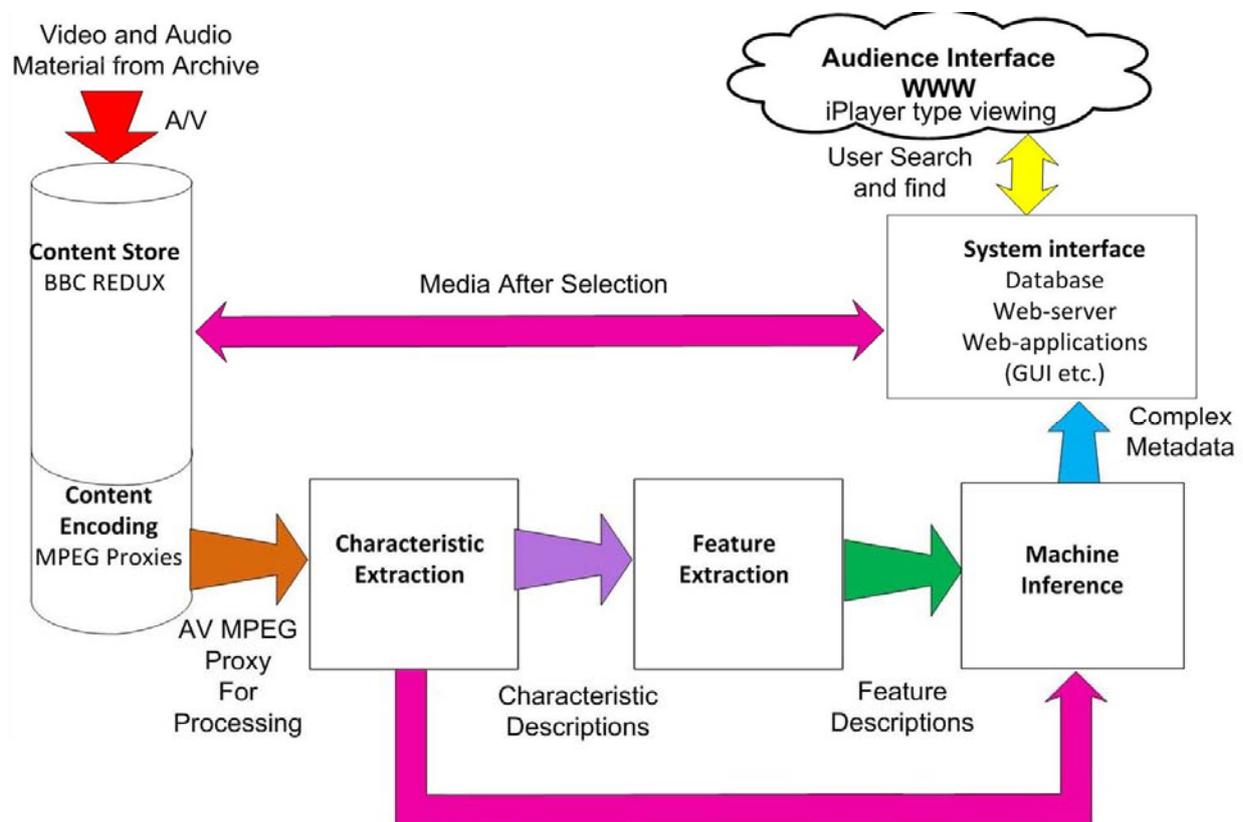


Figure 1. Overview of Multimedia Classification System.

2.2 Machine Inference

This module maps the features and characteristics extracted to the affective scales. Currently this is based on simple observed heuristics described in [2]. Current work is looking to incorporate machine learning methods such Support Vector Machines to increase the accuracy of the system.

3 INTERESTING EVENT TIMELINES

A further area of affective classification we are studying is that of interesting events; any event in a programme which a user may find interesting. In our initial study [3] we examined large scale sporting matches, creating timelines of interest within matches along with an overall 'interestingness' score for a match. Using signal processing of the audio only (to account for the

large number of radio only broadcasts made by the BBC of sporting matches), we initially segmented a programme into pitch/studio segments, then analyzed the pitch based segments for interesting events, looking for crowd excitation levels and referee whistles. Events were identified as peaks in these two sonic features. An example of this for one match is shown in figure 2. Ground truth data was taken from the BBC Sports Library, a professional service which identifies interesting events in some modern matches.

4 USER SEARCHING FOR CONTENT

Our current system presents users with content in one of two ways. Our multimedia classification system presents users with programmes arranged on a 2D graph. This shows programmes with similar overall affective content clustered together. This is shown in figure 3. When a user hovers their mouse over the programme marker, programme guide information is displayed.

Using this approach we aim to allow complex searches for content to be broken down into a simple graph.

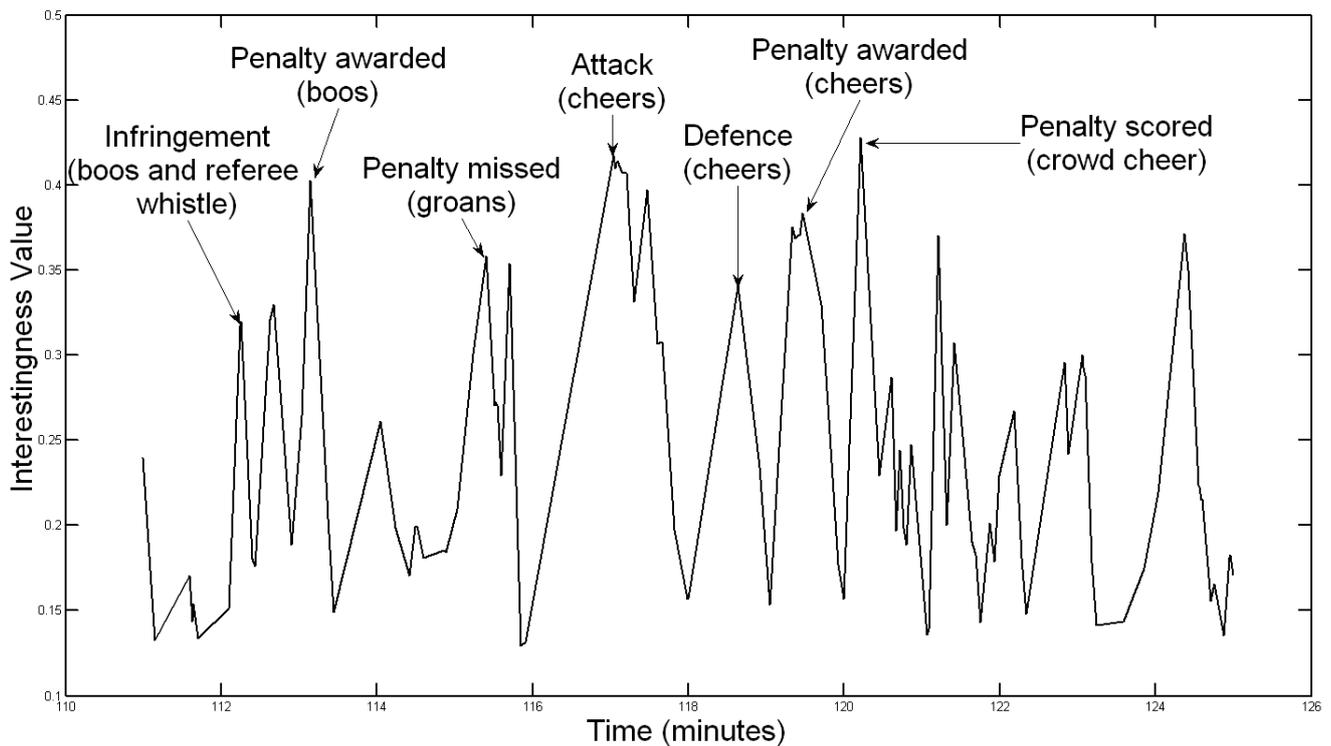


Figure 2. Event detection in sports broadcast.

Users may want to find content via a variety of methods; title, genre or subject or programme mood. Using this approach we allow users to combine these approaches. They can readily identify programme names from the display; genre and subject are contained within the programme guide and programmes with similar moods are clustered together.

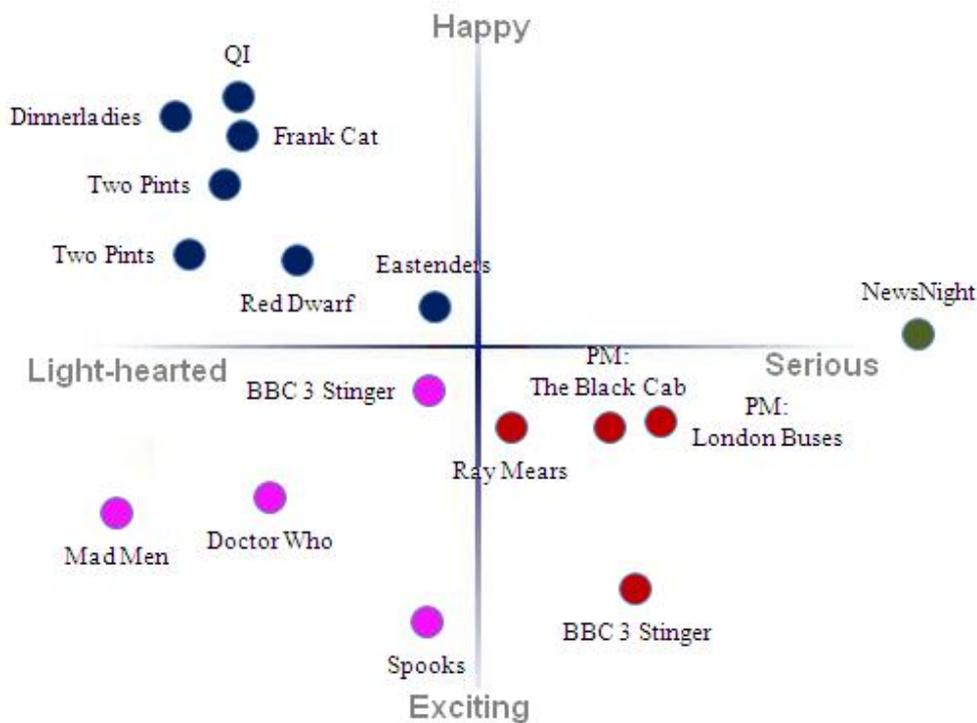


Figure 3. Overview of Multimedia Classification System.

The other approach is to present users with a time line of interesting events in a programme. Users could then select a programme they want using the system in figure 3, choosing only the sections that are of interest.

BBC R&D are investigating these multimodal searches to help user access the content in the BBC archive which they don't know exists. Current methods rely on the user having some idea about what type of content they like. We are trying to solve the issue of when a user doesn't really know what they want. The systems we are developing will allow users to search and browse based on traditional techniques such as keywords or name, but also on new techniques such as affect – or a combination of both.

5 REFERENCES

- [1] S. Cunningham and P. de Nier, "File-based Production: Making It Work In Practice," in *Proceedings of the International Broadcasting Convention*, Amsterdam, NL, 2007.
- [2] S. Davies, D. Bland, and R. Grafton, "A Framework for Automatic Mood Classification of TV Programmes," presented at the *5th International Conference on Semantic and Digital Media Technologies* Saarbrücken, Germany, 2010.
- [3] S. Davies and D. Bland, "Interestingness Detection in Sports Audio Broadcasts," presented at The Ninth International Conference on Machine Learning and Applications (ICMLA 2010), Washington, DC, USA, 2010.