

# ‘Sounds like trouble’

---

Peter W.J. van Hengel<sup>1</sup>, M. Huisman<sup>2</sup>, & J.E. Appel<sup>1</sup>

<sup>1</sup>Fraunhofer Institut Digitale Medientechnologie

Oldenburg, Germany

<sup>2</sup>Sound Intelligence BV

Groningen, the Netherlands

## Abstract

The newly founded project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology IDMT focuses on applications based on models of the human hearing. One of the domains for application of these models is the detection of incidents in the surveillance domain based on audio information. For this application a system designed to detect human verbal aggression was developed by Sound Intelligence. A case study with this system will be described here as an example of what can be achieved using audio technology in a supportive role for surveillance. Models of human sound processing incorporating aspects of neural processing are now further developed to indicate a surveillance operator possible incidents or unexpected events.

## Introduction

To improve the overall usefulness of camera-based surveillance systems it is important that situations with a high risk of injury and a relatively fast development, such as street-fights, collapses or accidents, are detected as quickly and as reliably as possible. Only then, appropriate action can be initiated. A system which can prioritize potentially dangerous situations autonomously and presents high priority events to a human observer for further analysis of the situation, would greatly reduce the chances of incidents being missed.

An evaluation of CCTV in one of the London boroughs in 2004 revealed that only about 30% of incidents such as criminal damage and emergency incidents happening in view of a surveillance camera, were detected by CCTV (Gill & Hemming, 2004). This rather low percentage is in contrast with the expectation of over half the people in residential areas believing that the police will respond quickly to incidents when CCTV cameras are installed (Spriggs et al., 2005).

With each operator handling about 80 cameras on average, the chance of looking at the right place at the right time is low. As one of the CCTV operators is quoted in (Gill & Hemming, 2004): “we are the eyes of the police but we don’t know where to look”.

In D. de Waard, J. Godthelp, F.L. Kooi, and K.A. Brookhuis (Eds.) (2009). *Human Factors, Security and Safety* (pp. 369 - 375). Maastricht, the Netherlands: Shaker Publishing.