**Developing an affect enhanced "Turrellian" RGB LED lamp designed to improve mood: towards multimodal affect-sensitive computing and devices.**

**EXECUTIVE SUMMARY**

**INTRODUCTION**

[Picture of final product]

**Context**

No aspect of mankind’s mental life has more pertinence to the quality and meaning of our very existence than emotions. Quite simply, emotions make our life worth living and sometimes worth ending. With the advent of computers, it became inevitable that systems and devices would be increasingly used to recognise, process, simulate and interpret human affects.

Coupled with the exponential increase in transistors on silicon integrated circuits (microchips), Moore’s Law has promulgated exuberant innovation in computing technology. Resulting in computing technologies that will increasingly embed themselves into the fabric of our everyday living spaces, project the human user into the foreground. Indeed, these user-centric and intelligent computer systems will seamlessly serve the needs of the user with software operating automatically in the background.

[Pictures of envisaged devices]

These advances in computing power, wearable computer devices and interaction modalities promises a new dawn for computer systems and applications. Indeed, since Rosalind Picard’s, notable paper and book on affective computing in 1995 – much research has been established towards developing affective computers, systems, interfaces and robots.

Additionally, in more recent times, prominent human computer interaction researchers Pantic et al. have encouraged the development of affect-sensitive multimodal human computer interaction (MHCI or HCI2) – specifically computing which can understand and respond to natural and autonomic modes of human communication such as emotion automatically via analysing a range of human modalities. These devices or interfaces should transcend the traditional keyboard and mouse, succinctly understanding and adapting to human affect, behaviour and signals.

In this vision of the future, often referred to as “ambient intelligence” humans will be surrounded by intelligent yet invisible multimodal computer devices that can anticipate all our daily needs. For instance, cars will automatically pull over if a driver becomes drowsy, our favourite music will be played by speakers when we come home weary after a long day at work and smart watches may prompt us with notifications to remind us to meditate when we are feeling stressed or agitated.

There are significant advantages to these increasingly empathetic and multimodal computers, interfaces and machines. Evidence from multimodal human computer interaction (MHCI) research finds that they are more human-like, usable, efficient and effective. Not only this, but these multimodal affect-sensitive computers and devices certainly have the capacity to significantly improve users’ day-to-day moods, emotions, routines and overall quality of life.

Spectrums of coloured lighting diffused by an internet of things (IoT) lamp has been chosen as the most appropriate natural medium to explore enhancing the mood of users. This is because different spectrums of lighting have been proven to have a non-invasive range of health and emotional benefits from improving concentration and sleep to even alleviating depression and migraines.

Additionally, there are an increasingly popular variety of IoT lighting devices designed to treat specific health disorders or on-command calibrate specific suitable environmental ambiences for their users. Therefore, there is undoubtedly value in adding more to this IoT lighting market by providing a device which acts automatically enriches users’ affective state, mood and experience.

[Pictures of rival IoT devices]

**James Turrell**

As the device-namesake James Turrell’s artwork has been a dominant influence in this research. Turrell emerged as one of the foremost artists associated with the Light and Space movement of the 1960s. His art encourages a state of reflexive vision that he calls “see yourself seeing” wherein we become aware of our own senses, emotions and of light as a tangible substance.

Particularly, Turrell’s work explores the immateriality of light – building on the sensorial and emotional experience of space and colour. His oeuvre, phenomenological *lightscapes* feature technically advanced LEDs, which are configured by computer programming. Collectively these innovative and photographic techniques enable light and its spectrum of colour to have a physical presence and fully immerse the space.

These expositions and spaces engage viewers – capturing the limits and wonders of human perception prompting an incredibly transcendental and emotional experience. As Turrell aptly notes, “my material is light, and it is responsive to your seeing”. This physical engagement with light and colour strongly inspires the aesthetic design of the Turrellian LED lamp.

In this sense, Turrell’s work particularly his expositions focusing on the Ganzfeld effect hones visitors’ senses by deliberately removing elements of perception and enabling colour to fully inhabit and immerse the space. The colour itself provides an emotional experience for viewers and the Turrellian LED lamp is keen to capture a similar emotional response for positive purposes through immersive coloured lighting.

Indeed, Museum of Fine Arts Houston curator De Lima Greene notes that, Turrell’s “pieces enact not just an emotional response to colour but a sharpening of all of your perceptual faculties. After you have spent some time looking at Turrell, you walk out and pay attention to how even an ordinary lightbulb casts light upon the wall”.

[Pictures of James Turrell and his works]

**Turrellian LED lamp**

This thesis describes the development and testing of a new IoT RGB LED lamp, the Turrellian lamp, which inspired by the phenomenological works of Turrell and research initially outlined above – utilises light, colour and affect-sensitive MHCI to improve user’s mood and quality of life.

Particularly, the Turrellian LED lamp offers an enhanced sensory experience to benefit users mental-wellness, cognition and emotion – few realise the importance of appropriate lighting and the subconscious effect of coloured lighting on mood. The lamp aims to subtly expose users to appropriate colours and hue profiles of lighting to enrich their present mood and cognition based on detailed research from colour psychologists Valdez et al.

Conventionally, individuals have accustomed to the importance of music for their cognition and quality of life – people everyday use radios, phones or even music streaming platforms to play their favourite songs and playlists. Surprisingly, in stark contrast, ambient environmental lighting has been broadly overlooked by society due to its imperceptible nature and perceived unimportance. However, like music, ambient environmental lighting holds significant importance for individual’s mood, health and general cognition.

With this realisation, more recently, there has been growing demand for smart IoT light devices, lamps and lighting systems. Generally, these devices fit into two key categories. In the first category, the light devices are designed to be manually calibrated by users to provide ambient lighting on-command. In the second category, the light devices are designed to provide “health” functionality to treat specific disorders and syndromes such as seasonal affective disorder (SAD), sleep disorders and even dementia.

[Pictures of sketches of the product, rival products and prototypes]

In contrast to these categories, the Turrellian LED is the first IoT lighting device to use insights from affect-sensitive computing to work synonymously with its users’ affective state (both conscious and subconscious) and offer mood enhancement through coloured lighting. Furthermore, the Turrellian LED has the capacity to operate automatically and autonomously. The goal of utilising insights from disparate fields such as ambience research, psychology and affective computing in the form of the Turrellian LED is threefold for users:

* To positively impact user’s cognition, affective state and mood through non-invasive coloured lighting.
* To encourage users to become committed to self-reflection and journaling practices – amplifying their daily routine with appropriate lighting and fostering emotional stability.
* To provide a low-demand and novel computer architecture for IoT devices, which is sensitive to users affect and offers multimodal human computer interaction.

Researchers within MHCI and affective computing have focused on the delivery of a wide variety of devices and applications such as more usable operating systems, driverless car design, healthcare, education, marketing or advertising research, video games, robots and even mirrors. However, as of yet no research paper has sought to utilise affective computing for building an ambient coloured LED lamp. This is despite much having been written and established regarding the influence of lighting and colour on users’ moods and emotions.

As a contribution in a new area of research, this thesis aims to support further work in this area by using open source and accessible prototyping platforms, providing an accessible GitHub repository for the codebase and by documenting developments and shortcomings to fully support further research or collaboration.

***Engineering design process and user experience testing***

Due to the pioneering nature of the LED lamp and the importance of users’ personal affective experience with the lamp. I decided to employ an engineering design process in which I would develop a beta ‘demo’ prototype for extensive experimentation and user-experience testing with candidates.

[Diagram of engineering design process]

This collaboration with eventual potential users would help provide useful data and research concerning the devices effectiveness and ease of use. Equally it would enable necessary adjustments and improvements to be successfully made to the interface and Turrellian LED product.

Testing on ? was conducted to provide feedback, experience, and expertise for the device. Due to the constraints of the Coronavirus pandemic and consequent issue of test subjects safely. Instead of engaging in a high number of test candidates, I adopted Sun Microsystems researcher, Nielsen’s model which encourages small usability tests for development.

[Nielsen’s model or allude to future chapters]

I worked with test subjects observing their relationship with the device, usability of the interface, conducted interviews and think-aloud studies for focusing on their responses to the Turrellian LED. A control lamp was employed with test participants for direct comparison to determine the effectiveness of the Turrellian LED.

**Map of project paper**

The thesis is clearly divided into four parts:

* **Literature review**: In this section I define affect sensitive MHCI, critically taxonomize recent contributions to the field. Before juxtaposing the research with recent relevant insights from affective science.
* **Design approach and prototyping:** Here I discuss the context for the Turrellian LED, colour psychology and the groundwork for successful delivery of a beta prototype.
* **User experience testing:**
* **Outcomes**

**Project aims and objectives**

The project core aims are twofold. Firstly, developing a prototype product and application which enriches the current IoT smart-lighting market. Secondly, to contribute to burgeoning academic research and discourse within affect sensitive MHCI via providing criticism of current approaches, contributing a novel solution and offering pathways for future research.

In order to achieve these aims the six objectives of the project are as follows:

1. Present a thorough definition and critical taxonomy of existing affect sensitive MHCI systems and literature.
2. Critique current multimodal approaches and offer insights from affective science – particularly embodied emotion and its contribution to affective computing design.
3. Explain the design approach and research foundations of the Turrellian LED and evaluate competitor devices.
4. Develop a fully functioning beta prototype – equipped with integrated hardware, physiological sensors, LEDs and a graphical user interface (GUI) controller.
5. User experience test the functioning beta prototype and collect initial results versus a control lamp.
6. Analyse results and offer clear pathways for future research – so as to continue successful development towards affect-sensitive multimodal computer devices.

**Affect-sensitive MHCI**

**Chapter introduction**

In this chapter, I firstly clearly define and signify affect sensitive MHCI and its specific position within the broader research area of affective computing. Afterwards, I secondly and specifically contextualise with a taxonomy of modalities and various approaches of prominent MHCI researchers. Lastly, I thirdly critically evaluate the current differing approaches with respect to developing the Turrellian LED’s affect-sensitive system architecture.

Rather than providing an exhaustive literature review of all past efforts to develop emotion sensitive computer systems – I focus on literature concerning automatic and multimodal analysis of spontaneous affective behaviour recorded in real-world settings. Indeed, this specific multimodal literature is most pertinent to developing devices and systems such as the Turrellian LED. However, for readers interested in exhaustive surveys of affective computing, please refer to:

**Definition**

Affective computing is the study and development of systems and devices that can recognise, interpret, process and simulate human affects. Originating with philosophical enquiries of the mind, this modern branch of computer science was properly defined under Picard’s 1995 paper. However, the focus of this research paper is affect-sensitive MHCI – a smaller subset of affective computing.

Affect-sensitive MHCI originated more recently under a paper written by prominent researchers Pantic and Rothankrantz in 2008. MHCI focuses on specifically next-generation HCI designs, which are capable of recognising users affective state – to become more human-like, usable and effective. Therefore affect-sensitive MHCI differs from broader affective computing because it is specifically focused on interpreting and processing the emotional state of users and thus improving man-machine interaction.

To ensure systems are multimodal, affect-sensitive MHCI designs monitor multiple non-communicative cues or modalities. For instance, facial expressions, body movements, vocal and physiological reactions to automatically determine the affective state of the user. This ability for computers to immediately empathise with their users replicates a cornerstone of human-to-human communication. Indeed, in face-to-face communication and interaction human beings socially process affective social signals with little conscious effort. Affect-sensitive MHCI thus envisages understanding their users affective state in a similar and automatic way.

**Taxonomy of approaches and significance of field**

As alluded to in the chapter introduction, there has been tremendous divergence in the approaches, modalities and methodologies employed by MHCI researchers designing affect-sensitive computer systems. Consequently, all human interactive modalities (sight, sound and touch) and non-verbal interactive signals (facial expressions, vocal expressions, body gestures and physiological reactions) have not been researched equally.

[Taxonomical diagram of approaches]

Currently most MHCI research has been preferentially invested into facial expression recognition and auditory speech analysis. These modalities are logically most desirable due to their non-invasive nature. Indeed, a camera or microphone can succinctly automatically document users vocal or facial changes and immediately deduce the underlying changes in a user’s affective state.

Additionally, the two modalities of facial and vocal expression most neatly mimic natural forms of human social and emotional communication, which depends on auditory and visual facial signals and channels. For example, we view mouth expressions to interpret if another party is smiling denoting happiness or the sound of guttural sobbing will usually preclude another party’s sadness.

Contrastingly, physiological reactions and bodily gestures require the use of intimate physical sensors for tactile processing of users changes in affective state. Often experiment participants have labelled these physical sensors as uncomfortable and mildly irritating. Resulting in less preference for physiological based architectures and gestural-based research and platforms.

Despite the different modalities used by MHCI for designing affect-sensitive computer systems and devices. There is broad consensus that a fusion of the above approaches will lead to the most sensitive and functional system. A system which utilises and processes all modalities automatically – enabling the machine to have a comprehensive apprehension of its user and their current affective state immediately.

[Diagram of fusion of approaches]

Providing machines with empathy and emotional intelligence is crucial for improving existing computer systems, devices and architectures. Furthermore, it has a wide variety of application in multiple fields. Machines capable of sensing and responding to users’ affective feedback are likely to be more persuasive, natural and trustworthy. Therefore, these systems have useful application within a plethora of fields for instance:

* Elderly care
* Retail purposes (e.g. Landsec)
* Marketing and advertising research (e.g. Real Eyes)
* Lie detector for security (e.g. Neurodata lab)
* Monitoring for inattention (e.g. Affectiva <https://www.digitaltrends.com/cool-tech/rise-of-emotion-tracking-tech/> )

[Pictures of respective field contributions]

Certainly, automatic MHCI designs unequivocally offer incredible usefulness and application across a variety of different fields. Now I will describe the latest research within three dominant and differing modalities: automatic facial emotion recognition (AFER), automatic auditory expression recognition (AAER)and lastly,

**Visual and facial MHCI approaches: Automatic facial emotion recognition (AFER)**

As alluded earlier, utilising facial recognition for automatic recognition of emotion (AFER) has become the most prominent methodology for designing affect-sensitive computer systems and architectures. To date, there have currently been numerous academic papers and start-up companies using AFER methodologies.

Historically, facial expression analysis to determine emotion was originally professed in 1862 by French researcher Duchenne who focused on the elector-simulation of individual facial muscles as responsible to produce facial expressions. This methodology for determining emotion via facial expression, was further enshrined by the tremendously influential work of Darwin and his book entitled, *The Expression of Emotions in Man and Animals*. Darwin’s work comprehensively accounts variation in facial expression and emotion.

More recently, this methodology centred on facial expression has been further extended by the prodigious work of psychologist Paul Ekman. Ekman’s research boldly suggested that the six basic emotions namely anger, fear, disgust, happiness, and surprise are universally transmitted through prototypical facial expressions. Furthermore, utilising the groundwork of anatomist Hjorstjo, psychologists Ekman and Friesen produced the Facial Action Coding System (FACS) to codify and taxonomize facial expressions into 32 Action Units (AUs) and 14 additional Action Descriptors (ADs).

Today, it is widely accepted that facial expressions serve as an essential nonverbal means for human beings to communicate emotions. Consequently, specific modern approaches have been developed by MHCI researchers designing affect-sensitive computer systems via visually measuring users’ facial expressions. Consequently, specific modern AFER approaches have been developed by MHCI researchers designing affect-sensitive computer systems via visually measuring users’ facial expressions.

Firstly, the message judgement approach, which aims to directly the meaning conveyed by a facial display such as being happy, angry or sad. Whilst secondly, the sign judgement approach, which contrastingly aims to specifically study the physical signal used to transmit the message instead – for instance, raised cheeks, nose wrinkles or jaw drop.

Both approaches are based off Ekman’s original psychological research. Message judgement computer vision systems categorise facial expressions into the six basic emotion categories. Contrastingly, sign judgement computer systems most commonly utilise FACS.

Both models are incredibly different and offer a range of advantages and disadvantages. Sign judgement systems offer a more accurate system for MHCI researchers because they are more comprehensive and objective – every facial expression is comprehensively described as a unique combination of AUs and ADs.

In contrast, message judgement approaches do not document all facial expressions and assume many to one correspondence (i.e. many different facial expressions correspond with one emotion), which directly contends with multiple studies from affective science. Despite this large shortcoming, MHCI researchers have interestingly preferred the less-specific message judgement approach. This is because a major impediment of sign judgement MHCI is that it is far too exhaustive.

Indeed, it takes several hundred hours of programming to become competent as an FACS and sign judgement analysis and each minute of footage takes roughly one hour to score for all available AUs and ADs – making it utterly impractical for automatic systems.

Fundamentally, both approaches are based off Ekman’s influential psychological research since the 1970s. Message judgement computer vision systems categorise facial expressions into the six basic emotion categories. Contrastingly, sign judgement computer systems most commonly utilise Ekman’s latest edition of FACS.

The two approaches are incredibly different and offer a range of advantages and disadvantages for AFER. Sign judgement systems offer a more accurate system for MHCI researchers because they are more comprehensive and objective – every facial expression is comprehensively described as a unique combination of different AUs and ADs. In contrast, message judgement approaches do not document all facial expressions and assume many to one correspondence (i.e. many different facial expressions correspond with one emotion), which directly contends with multiple studies from affective science.

Despite this large shortcoming, MHCI researchers have interestingly preferred the less-specific message judgement approach. This is because a major impediment of sign judgement MHCI is that it is far too exhaustive. Indeed, it takes several hundred hours of programming to become competent at FACS and sign judgement analysis and each minute of footage takes roughly one hour to score for all available AUs and ADs – making it utterly impractical for automatic systems.

Subsequently, MHCI studying AFER have focused on detecting the ‘basic’ six universal emotions. Although MHCI AFER researchers have made some tentative efforts to determine non-basic affective states, including fatigue pain, agreeing, concentrating, interest, frustration, thinking and unsureness. Indeed, these studies represent a growing trend towards developing systems which offer automatic analysis of spontaneous facial expression data – it is certainly a thriving research area.

Several studies have focused on automatic recognition of AUs rather than emotions from spontaneous facial displays. With the expectation to improve and enhance computer vision processes before developing a fully viable AFER computer system or device. However, the dominant conceptual framework followed by advanced AFER studies includes – facial registration, representations, dimensionality reduction and lastly recognition.

[Diagram from AA of Facial Affect]

**Face registration**

Face registration is the initial fundamental step in AFER – this process can be categorised into three categories. Firstly, (1) whole face registration is the most common method for computer systems. Rigid whole face registration maps an entire input face to a prototypical face. This has two advantages – the transformation can become less sensitive to registration errors and can cope with head pose variations better. Contrastingly, non-rigid registration enables registration of the face locally.

Secondly, (2) parts registration processes the face in terms of parts (e.g. eyes and mouth). Thirdly, (3) points registration focuses on the localisation of fiducial points on the face. All three face registration schemes used by MHCI AFER typically utilise techniques such as Active Appearance Models (AAM). An important decision to be made for registration is how to successfully deal with head-pose variations – a number of the studies suppress head-pose variation of subjects. Although arguably head-pose variation is an integral part of individuals affective behaviour.

**Spatial representations**

Spatial representations encode image sequences of the face frame-by-frame – MHCI researchers have encoded this in seven very different ways:

1. Shape representations – most frequently used, involves describing the face by simply concatenating the x and y coordinates of a number of fiducial points.
2. Low level histogram representations – extract local features, encode them to a transformed image and lastly pool the features of each region with local histograms.
3. Gabor representation – obtained by convolving the input image with a set of Gabor filters of various scales and orientations. Notably, a computationally costly process.
4. Bag-of-Words representation – BoW describes local neighbourhoods by extracting local features densely from fixed locations and then measuring the similarity of each of these features with a visual vocabulary dataset using locality constrained linear coding.
5. High-level data driven representations – Machine learning approach encoding features that are semantically interoperable. Deep learning techniques such as sparse coding and non-negative matrix factorisation are favoured.
6. Hierarchical representations – MHCI AFER systems encode an abundance of information in a low to high level manner to develop deep learning models.
7. Part-based representation – Process faces in terms of independently registered parts and thereby encode componential information.

Add a little discussion…

**Spatio-temporal representation**

Spatio-temporal representations morph a range of frames within a temporal window as a single entity – enabling modelling of temporal variation in order to represent subtle expressions more efficiently. This category can be subdivided into six different representations:

1. Geometric features – Analysis is performed of temporal variation and the corresponding muscular activity. For example, the recognition of AUs within temporal phases.
2. Low-level features from orthogonal planes – Extracting features from three orthogonal planes to determine AU recognition and basic emotion recognition.
3. Convolution with smooth filters – Involves convolution with small spatio-temporal features – incredibly accurate for determining subtle expressions.
4. Spatio-temporal Haar representations – Each dynamic Haar feature encodes the temporal variation in an image sequence with a pattern of binary values, where each binary output is obtained via thresholding the output of the Haar feature in the corresponding frame.
5. Free-form deformation representation – Registration technique that extracts features in the process of registration by computing the pixels spatial and temporal displacement.
6. Temporal Bag-of-Words representation – Specific to AU detection, represents an arbitrary subset of the given image sequence with single histogram which is computed.

Add a little discussion…

**Dimensionality reduction**

Dimensionality reduction is used to address several affect recognitions challenges such as illumination variation, registration areas and identity bias. Dimensionality reduction can be categorised into three classes – namely (a) pooling, (b) feature selection and (c) feature extraction methods.

Pooling is a sample-based discretization process. The objective is to down-sample an input representation reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. Feature selection aims at refining the facial representation by selecting a subset of its features and optionally weighting the selected features. This helps illuminate affective analysis by emphasising certain points on the face.

Lastly, feature extraction methods extract novel features from facial images. MHCI researchers will map input representation onto a lower dimensional space to discover a latent structure from the representation.

**Recognition**

Recognition within MHCI AFER studies has improved significantly – rather than just labelling a facial action with the corresponding emotion. Recent studies provide the intensity of the displayed emotion and corresponding facial action. Furthermore, for AU recognition, the output can be enhanced significantly by providing the temporal phase associated with the AU representation.

Several studies even better deal with spontaneous affect by recognising combinations of AUs rather than a singular AU in isolation. Indeed, spontaneously displayed AUs rarely appear in isolation. Most MHCI AFER techniques are reliant on machine learning techniques. As any machine learning application, the performance of an affect analyser is contingent on the quality and quantity of the training data coupled with the machine learning model.

For MHCI AFER systems, labelling data is a tremendously challenging process – indeed perceiving an emotion is quite subjective, therefore multiple annotators are regularly used. However, this presents the challenge of combing multiple annotations by different annotators. This has resulted in the emergence of another academic field focused on improving methodologies for AU annotation. For example, Zhang et al. have developed the most advanced interactive AU labelling system – focused on minimising collaborative human error.

Statistical modelling of annotated AU data is the next step. Certain studies have relied on generic models such as support vector machine (SVM) modelling. However recent studies have aimed to tailor specific statistical models for affect recognition. These specific models have focused on addressing multiple problems faced for affect recognition. These modelling issues include temporal variation of emotions, personalising affective displays, face configuration difficulties and correlated affective displays.

**Discussion of current AFER approaches**

[Table of recognition rates]

From this overview it is quite clear to see that current AFER techniques like many computer vision problems suffer from being incredibly analytical, mathematical and computationally taxing. Indeed, each of the processes, outlined above, involves a multitude of complex processes and mathematical image-processing solutions.

Evaluation is difficult to achieve as most AFER systems have been developed utilising completely different posed data sets. Making cross comparison and analysis of systems effectiveness at determining automatic affect especially difficult. Indeed, solutions suitable for posed settings are often utterly impractical for everyday life.

Additionally, MHCI AFER researcher’s insistence on utilising sophisticated statistical model’s recognition rates – makes it hard to determine if the system is indeed usable and effective in real world scenarios. Naturalistic data or a homogenised feedback mechanism – would be a far more robust means of determining the effectiveness of the AFER system.

Although more development is needed from MHCI researchers to ensure the development of a realistic facial expression analyser within a normal and unrestricted interaction environment. As the table shows, the recognition rates and performance of AFER has been improved through MHCI researchers utilising a combination of deep learning algorithms and machine learning techniques.

Situational and contextual dilemma?

**Auditory MHCI approaches: Automatic auditory expression recognition (AAER)**

Since the 1950s computer science researchers have conducted tremendous research on speech recognition. Consequently, there has been significant development in speech recognition i.e. successfully converting speech into a sequence of words – comprehendible for machine-based parsing.

However, multimodal and natural forms of verbal interaction between man and machine has not been realised. Despite MHCI researchers’ efforts, machines still lack the capacity to emotionally and empathetically verbally communicate with their users.

Contrastingly speech is a key expressive modality in human-to-human communication. Affective information is typically delivered in two forms – through explicit linguistic message and implicit message that reflects the tone of the message communicated.

Notably, affective scientists have yet to reliably and affectively deduce an optimal set of vocal cues coherent with specific emotion states or Ekman’s six universal emotional categories. In contrast human listeners are very adept at decoding emotional state from vocal expression and even non-linguistic vocalisations such as laughs, cries, sighs and yawns. However, MHCI researcher Cowie et al. have done a good job providing qualitive acoustic variations for certain emotions.

[Diagram of linguistic versus tone]

The linguistic content of speech carries emotional information. Some can be translated into specific emotional subsets utilising an appropriate dictionary or lexical parser. However, a sizable part affective information of the speech lies in the language dependent cultural and situational context. For example, the English language and culture is certainly steeped in higher amounts of irony versus other.

Sensitivity to situational and cultural context is certainly very challenging to successfully achieve. Nonetheless, some researchers have had success detecting specific emotional states whilst utilising an application. For example, Kwon et al. focused on detecting stress and Ang et al. tried to effectively detect user’s verbal annoyance and frustration.

MHCI researchers’ efforts have even expanded to automatic recognition of non-verbal vocalisations such as laughter or cries. In terms of data collection, there are many naturalistic settings to collect good audio data for experimentation – call centres and dialogue systems have both been effectively used.

Typically, AAER MHCI researchers and computer systems follow a basic processing pipeline divided into two categories: non-linguistic and linguistic analysis followed by eventual emotion classification.

**Non-linguistic analysis**

Vector features enable thorough analysis of the temporal evolution of the 25-50msec sound window or bite. Contrastingly, suprasegmental analysis for affective displays is calculated over the entire utterance duration.

Vector features are classified into two distinct classes – specifically, low level descriptors (LLDs) and functionals. Early studies particularly focused on pitch, duration and intensity. However, as analysis has improved, voice quality LLD features such as HNR, jitter or shimmer, such as spectral and cepstral measurements (MFCCs) have been far more extensively used.

Functionals are used to derive LLD statistics per utterance. The LLDs begin with dynamic features such as intonation, intensity, formants and spectral information. Next, systematic derivation of prosodic, articulatory and voice quality high level operation is performed by descriptive statistical analysis. From that point, the feature set is enhanced with automatic feature alterations, in order to find an optimal representation and a target classifier.

Although affective science has yet to conclusively, label all non-linguistic vocalisations as coherent with specific and discrete emotional states. Undoubtedly many human vocalisations are consistent with specific emotional categories. For example, laughter elicits amusement or nervousness contrastingly a yawn elicits boredom or tiredness.

Consequently, AAER MHCI researchers have made solid strides towards automatically classifying non-linguistic expressions such as laughter, cries, emotional outbursts and even coughs.

**Linguistic information**

Linguistic information schemas consist of ASR systems that identify specific words or phrases which can be correlated to an emotional state. Additionally, an updated dictionary is of paramount importance for the successful implementation of linguistics in emotion recognition.

Analysis from linguistics teaches the importance of prosodic features such as repetition and reformations in dialogue. For instance, repetition of the same phraseology may signify users intensifying anger or frustration. Furthermore, part-of-speech analysis finds that expressive nouns and adjectives are especially useful for decoding users affective state.

Based off these linguistic insights, AAER classification of emotional states has improved particularly in studies utilising salient emotion categories. Rather than classification into specific emotions, AAER MHCI researchers have used non-specific categories such as: negative, neutral and non-negative. Under this renewed categorisation, recognition rates have improved markedly.

Semantic labels have been introduced to develop recognisable word sequences for affective speech. In particular, Wu and Liang have used this as the basis for research with acoustic-prosodic information with some success for AAER affective recognition.

**Affective classification**

Much like earlier discussed AFER MHCI techniques. Machine and deep learning play a prominent role in AAER affective classification.

1. Gaussian Mixture Models (GMMs) – The most studied by MHCI AAER researchers – these offer low training and testing environments (i.e. accessibility) however optimisation has been noted to be difficult.
2. Hidden Markov Models (HMMs) – Another common technique, a stochastic process in which sounds are broken down into states – yet the optimum number of states has been difficult to measure.
3. Support Vector Machines (SVMs) – An increasingly popular process, SVMs offer optimality and generalisation. However, some MHCI researchers have struggled with successfully separating data.
4. Artificial Neural Networks (ANNs) – MHCI AAER researchers have been slightly ad-hoc – using a wide variety of ANN algorithms for AAER. Detailed performance comparisons are thus hard to determine.
5. Decision Trees – MHCI researchers have often used the C4.5 algorithm to varying degrees of success.
6. Random Forest (RF) – For AAER analysis has largely struggled with large datasets and resulted in high complexity.
7. Bag of Words (BoW) – A popular technique in natural language processing (NLP). Has been reasonably successfully used for affective recognition.

It is important to acknowledge that some MHCI AAER researchers have established a hybrid of classification machine learning algorithms to deduce affective state. This of course adds tremendous complexity and is highly computationally taxing. Although some MHCI researchers have shown improvement in affective recognition, for instance Schuller et al. improved recognition by 8% utilising GMMs and SVMs.

**Emotional speech databases**

Recently, there has been growth in the number of commercial emotional speech databases – based off the progress of public MHCI datasets and systems research covered in the earlier AAER sections. Most of the developed emotional speech databases are not available for public use. Consequently, there are little feedback systems or benchmarks for researchers to effectively collaboratively improve.

Additionally, due to legal and ethical issues, most of the databases use professional actors to test effectiveness of the database. Nonetheless, most of the databases claim to offer an effective means of determining Ekman’s six universal emotions from linguistic context.

[Table of effectiveness of databases]

Due to the earlier discussed privatised nature of the speech databases it is very difficult to cross-examine and analyse the databases: by thoroughly testing each database under the same experiment conditions. Therefore, caution must be applied to the exact effectiveness of each database.

**Discussion of AAER approaches**

Much like AFER, AAER similarly utilises many machine learning techniques to solve the issue of affective emotion recognition. Typically, high classification of databases and datasets has been achieved through utilising actors and consequently not spontaneous speech.

The classification range of 50% to above 90% for specific emotional subsets is obviously promising and encouraging – however, for the most part, the restrictions on accessing the computer systems and underlying database is immensely frustrating. Furthermore, certain pattern technologies such as multiple classifier systems (MCS) have yet to be effectively applied to AAER systems.

Due to the differing experimental setups – it is hard to achieve certainty which AAER system was independently most successful. Furthermore, this prevents the ability for effective academic feedback loops and benchmarks. Consequently, it is too difficult to directly apply these studies to developing a spontaneous multimodal affect-sensitive computer architecture such as the Turrellian LED.

**Physiological tactile affective signals**

Physiological signals refer to any data that is collected from the physical human body and its tactile systems. The signal signifies the underlying response when a human body is subjected to stimuli – physiological metrics are collected from the central and autonomic nervous system.

[Diagram of either nervous system]

Psychophysiology teaches that these signals occur unconsciously and thus offer an unbiased reflection of individuals affective state. Additionally, physiological signals are both detectable and inevitable because sympathetic nerves of the ANS get activated when a person is positively or negatively stimulated. Physiological signals have been particularly effective for placing participants on Lang’s valence model (see affective science).

Numerous physiological systems have been well researched within MHCI for coherently determining affective state, including: the cardiovascular system, electrodermal activity, respiratory system, muscular system, physical movement and brain activity. Studies utilising all of these physiological systems have demonstrated recognition ability for both discrete emotions and for determining negative valence or arousal. However, in line with the embodied emotion tradition, Lang’s valence model has been preferred for most MHCI physiological studies.

Physiological studies conducted by MHCI researchers typically involve a four-step process: (a) pre-processing the emotion elicitation stimuli, (b) feature extraction (c) feature reduction and lastly, (d) emotion classification.

**Pre-processing emotion elicitation stimuli**

Gathering a sufficient dataset is vital for the coherent functioning and development of a physiological emotion recognition system. Good physiological data is difficult for non-specialists to collect properly – the tactile physiological data must be elicited subconsciously and naturally from participants. However, MHCI have developed effective methods for eliciting natural emotions from participants:

1. Visual pictures (International Affective Picture System)
2. Audio music and sound clips
3. Movie and film footage
4. Personalised imagery

Raw physiological data often suffers from contamination with external influences. Consequently low-pass filters are used to smooth the raw signals before processing. For example, Kim et al. segmented physiological signals into samples of 160s and filtered each signal for the part least prone to movement artefacts. Contrastingly, Mandryk et al. developed a normalisation process producing a percentile variable between 0 and 100 for each physiological variable.

**Feature extraction**

Following pre-processing, statistical and featural information is extracted concerning the physiological signal. This statistical information can then be used for emotional content. Single physiological signals are rich in feature specific information and can thus provide many affect-relevant features. For example, 110 features were extracted by MHCI researchers utilising 4 physiological signals, namely, ECG, EMG, skin conductance and respiratory signal.

MHCI researchers have used multiple different mathematical models and methods for computing feature extraction. For example, Haag et al. computed the running mean and standard deviation of multiple physiological signals, namely ECG, BVP, EMG, SCG and respiration to help deduce tonic and phasic components regarding each analysed signal.

Physiological MHCI researchers have even utilised moving features for extracting feature-specific data. Moving and sliding features are recursively computed over the analysis window and offer approximations of the affective state and providing ample analysis of users constantly adapting affective state.

**Classification**

MHCI researchers employ feature reduction to remove any physiological signals that are not ideally corelated with the emotion profile. Such uncorrelated features typically negatively impact the reliable performance of affective classifiers. After extraction – a classifier must be trained to determine emotional states utilising the features presented. A wide variety of different algorithms have been successfully employed.

Machine and deep learning techniques have equally been well used including:

1. SVM – The most utilised technique, has provided a recognition rate exceeding 90% on several occasions. However, SVM does not work well for imbalanced datasets.
2. LDA – As a linear classifier the LDA decides class membership by projecting the feature values into a new subspace.
3. KNN – KNN is a lazy learning algorithm classifying method that is easy to understand and implement.
4. RF – For physiological MHCI high-dimension data the classification performance of RF and neural network algorithms is generally excellent.
5. CNN – An improvement on traditional neural networks. CNN offers weighted sharing and reduces network complexity.
6. DBN – DBN consists of simpler RBM models and gradually extracts deep features of the input data.
7. PNN – Feed forward neural network based on Bayesian strategy PNN offers simple structure and a fast learning ability.
8. LSTM – Utilises a sequence of long-term dependencies and contextual information. Produced mid to low recognition rates of 73.1% and 74.5%.

An overview of classification rates by different systems is noted in the table below. Successful classification rates range from 60% to 95% - with a sizable proportion over 90%.

[Table Overview]

**Discussion of physiological tactile approaches**

Machine learning techniques are once again utilised in the classification process for physiological affective MHCI techniques. However, in comparison to the earlier modalities of AAER and AFER, physiological (tactile) affective signals perhaps offer the easiest window for determining users affective state. Indeed, there are many correlations, which can be drawn from physiological-affective MHCI research, which are utilised for the basis of designing the affect-sensitive Turrellian LED:

1. HRV increases with positive emotion
2. Heart rate increases with high arousal
3. Skin temperature increases with positive emotion
4. GSR increases with high arousal

These correlations are incredibly useful for non-specifically determining participants current affective state. However, in contrast, for incredibly accurate affective reading – machine learning techniques must be robustly employed as each individual’s unique physiology. This is because individuals will uniquely display minorly different physiological changes for each affective state.

Consequently, a plethora of machine learning techniques for classification have been employed to reliably determine affective state across multiple different participants. Of course, adapting these systems outside of controlled laboratory settings will be challenging for MHCI researchers. Furthermore, participants during experimentation have found the physiological sensors uncomfortable and even irritating.

However, potentially this criticism is unable to have fully foreseen were unable to foresee the continued exponential growth of Moore’s Law and advent of more non-intrusive, comfortable sensors and wearables for accurate physiological data. Indeed, the demand for smart, ergonomic and non-intrusive wearable body sensors has grown in prominence.

From mobile devices, watches, smart running trainers, to even jewellery – often- AI driven “smart” physiological sensors designed to collect data on sleep, heartrate, steps, respiration, bodily posture and plenty more. User requirements and competitive markets will help to ensure these devices get smarter, less obtrusive and collect better physiological data on their users, which can then subsequently be fed to affect-driven computer devices such as the proposed Turrellian LED lamp.

**Chapter conclusions – fusion of approaches?**

As exposed by the taxonomy and thorough literature review, affect-sensitive MHCI still suffers from some shortcomings. Once these issues are resolved and considered – the development of empathetic devices begin to pervade our everyday lives.

This research area has experienced tremendous growth and improvements in recognition via the effective incorporation of computer vision techniques, machine and deep learning. However, this clear advantage is slightly detrimental because very few researches are competent in sophisticated techniques such as machine learning technology and emotion recognition.

Indeed, regularly PhD academics with these complex skills are hired commercially by large corporations as a part of a brain drain from academia. Therefore, developmental MHCI remains a niche subset of broader affective computing.

Another difficulty arising from utilising technical and diverse algorithmic techniques has resulted in a lack of appropriate feedback, constructive discussion and effective determination of clear benchmarks between researchers’ findings. Thus far MHCI researchers have approached the research area with very different methodologies, laboratory settings, participant numbers and participant demographics – resulting in a lack of standardisation and preventing reliable conclusions from being inferred between research methods.

Unlike AFER and AAER, physiological or tactile approaches enable me to infer four basic and testable correlations – which can form the basis of designing an affect-sensitive system for the Turrellian lamp. Although recognition rates may not be as high as sophisticated machine learning and computer vision approaches. This deliberate approach ensures the development of more accessible MHCI and affective computing. Equally this form of MHCI incorporates the increase in comfortable wearable sensors documenting our physiological modalities. Furthermore, as the next chapter notes physiological approaches within affective science have been most successful.

**Affective science**

**Chapter introduction**

The core ideas and concepts of affective computing and MHCI can be traced back to insights from affective science, neuroscience, psychology and philosophy. As noted in the preceding chapter, with the influential psychological-facial research from Ekman – insights from these disparate fields have enabled the development of all affect-sensitive computer systems and devices.

Crucially there is currently no scientific consensus defining an emotion. Subsequently in recent years, in particular, the definition of emotions has become vigorously researched and intensely debated. Yet to develop an affect-sensitive computer architecture we firstly need a clear theory and model of emotions - to effectively design a computer system sensitive to emotions.

In this chapter, I critically discuss insights from affective science most pertinent for developing the Turrellian LED system architecture and future affect=sensitive multimodal IoT devices. Notably, little academic literature has sought to deliberately synthesise affective science with developing multimodal IoT devices.

Many MHCI researchers have overlooked clearly classifying or defining emotions before designing their computer architecture and system. This oversight has consequently resulted in the design of overly complex systems for affective recognition and a failure to consider the significance or causes of participants modality changes.

Contrastingly, in this chapter, my research will firstly unify the fragmented discourse of affective science literature by providing a clear definition of emotions as bodily feelings or physiological changes. Afterwards, secondly, I will provide scientific evidence for this theory and its synthesis with the Turrellian LED system architecture. Lastly, I will introduce more recent research from affective science, most notably, enactive and scaffolded affective theory – which holds tremendous promise for the effectiveness of the Turrellian LED lamp design.

**Emotions as bodily feelings**

The most significant initial discovery within affective science came from, William James – who announced that emotions occur when the perception of an exciting fact causes a bodily change, and “our feelings of the … [bodily] changes as they occur is the emotion”. The same idea occurred to Carl Lange around the same time. This theory can be visually interpreted as:

[Diagram of Theory]

James-Lange offered an inversion of common sense with autonomic bodily changes preceding our emotional experiences. Therefore, central to this somatic feeling theory is that the body acts as a sounding board and is fundamentally integral to experiencing the infinite subtleties of the bodily states and changes coherent with emotional changes.

A good example of this theory in action includes the common experience of feeling internally pangs and tingles. For example, the common experience of in pressurised situation (i.e. prior to an exam or rollercoaster) it’s common to perceive a fluttery or churning feeling in the stomach.

These sensations are symptomatic of the brain-gut connection and have become colloquially termed as ‘nervous butterflies’. Furthermore, these feelings are fundamentally caused by the release of adrenaline and cortisol as part of a fight-flight autonomous response. Consequently, blood leaves the stomach to move to the muscles – the heart rate, breathing rate, blood pressure and alertness rises. Emotions are thus contingent on autonomous bodily changes, which individuals perceive as bodily feelings.

In support of this theory, James convincingly asks readers to abstract all physical bodily feelings from emotional experience and once all the bodily feelings are gone there is simply nothing left but “feelingless cognition”. Importantly, Lange’s research, which primarily focused on the vasomotor system – James’ research is highly inclusive of all patterned bodily changes for emotion.

Therefore, MHCI system designers can measure a vast array of physiological bodily changes to coherently determine the emotion being exhibited. For example, from James’ theory bodily changes and feelings in the viscera, facial expressions and instrumental actions can induce emotion. Crucially both James and Lange affirmed that different emotions create different bodily patterns or changes – providing each emotion with unique distinctness.

Subsequently, this Jamesian somatic bodily-feeling theory meshes nicely with the MHCI research discussed earlier. Indeed, legendary biologist Darwin observed multiple correlations between the body and emotional states. For example, Darwin’s list of identified fear symptoms include widely opened eyes and mouth, raised eyebrows, dilated nostrils, stiff posture, motionless, a racing heart, increased blood supply to the body, pallor of the skin and more.

Crucially, the influential work of Ekman fits very well with the somatic feeling theory as for each of Levenson, Ekman and Friesen’s six basic emotions – each of these emotions correspond to a unique bodily pattern and physiological state.

**Positive evidence for bodily feeling theory**

Despite early criticism, research and evidence for the theory has been increasingly successful over time. In 2002, a research paper on the autonomic nervous system concluded that the “theory was hard to disprove”.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Furthermore, consumers have readily purchased

Due to the sensorial works of artist James Turrell, the growing demand for ‘health’ orientated IoT light devices, significant findings from ambience research, neuroscience and psychology.

Furthermore, current market lighting devices are all manually calibrated by their users or medical practitioners to provide a comfortable ambience or treat a specific health condition and there is no concrete evidence of either categories effect on improving user’s mood. Consequently, at present there is no lighting device or application calibrated to improve overall mood, engender positive emotion, and work synonymously with its users’ affective state automatically. This is despite much having been written and established regarding the influence of lighting and coloured lighting on uses’ moods and emotion.

Disparate fields of study / Light as a medium for health and mood

Lighting, ambience research, artist James Turrell has conclusively found multiple health and emotional benefits coherent with

There are significant advantages to multimodal affect-sensitive computer devices and systems.

Devices capable of understanding users intimately – achieve mood improvement, be easier to use etc.

Multimodal human computer interaction (MHCI) researchers have processed different natural human modalities to extract user’s emotions with prominent techniques including facial expressions, vocal intonations, textual analysis and physiological reactions. Facial expression, vocal intonations and physiological reactions enable MHCI researchers to achieve automatic processing of user’s emotion.

Rather facial expressions utilising both machine learning and AI has become dominant and most central to providing automatic affective feedback recognition for computer systems. In contrast, body gestures or physiological signals, textual sentiment analysis and auditory analysis have to a greater extent represented a more secondary role in multimodal computing technologies affect-sensing research.

However, this research supports the contention that a broad fusion of approaches is most desirable to creating affect

***Overview of research paper***

Product release cycle

***Research output***

**Part 1 – LITERATURE REVIEW**

***Chapter – Affective computing***

***Chapter – Light, emotion and IoT light devices***

***Chapter – Embodied, scaffolded emotion and multimodal affect-sensitive computer architectures***

**Part 2 – PROTOTYPING, TESTING AND DEVELOPMENT**

***Chapter – Prototyping and hardware development***

***Chapter – Physiological affect-sensors***

***Chapter – Software and web application integration***

**Chapter Introduction**

In this section I exhibit initial sketches, prototyping and development of the project. The focus of this project was on aesthetic outcomes, interaction and early delivery of the lamp. Additionally, the lack of commercial devices utilising affect-sensitive platforms or LED coloured lighting to improve mood meant that a clear empirical and user-centric approach to research seemed most appropriate. Throughout alpha and beta development of the Turrellian RGB LED lamp – I incorporated an agile development cycle resulting in the development of several projects and incorporation of user-feedback for continual improvement whenever possible. Only here due to the consequences of the Coronavirus pandemic and pioneering nature of the project– I proceeded informally and opportunistically, utilising friends, family and colleagues whenever possible.

As discussed in the previous section, prototyping was carried out in MAX/MSP - a high-level music development environment, which allows fast development of musical and interaction ideas. To work in this way, I first had to develop a library which allowed simulation and control of Matsuoka’s Neural Oscillator (MNO) nodes in MAX/MSP.

**Part 3 – BETA PROTOTYPE AND UX TESTING**

**Part 4 – OUTCOMES AND FUTURE WORK**