An accelerometer based gestural capture system for performer based music composition

Jon Edwin Cobb

MSc by Research

University of York

Department of Electronics

December 2011

Abstract

Previous studies by research groups in the field of music technology have resulted in a variety of gestural capture systems that enable various levels of interaction and control in the generation of sound. The majority of approaches developed to date have focused on direct control of synthesis parameters. More recently there has been a move toward a higher level of abstraction in which performers can interact in the music composition process. This seems to be motivated by the desire to realise the full creative expression achievable through fusion of physical performance and music. Such a goal poses a number of technical challenges several of which have been explored in the current project.

The development of a small, low cost, low power accelerometer based gestural capture system is described. Using this device it is shown that significant subject to subject variation occurred (30% for a study group of n=36) when different test subjects executed the same movement. To compensate for this measurement uncertainty the development of a neural network based pattern recognition system is described which reduces variation to <5%.

Refinement of the design of the gestural capture system is described to enable it to be implemented as a wireless network of body-worn sensors to capture movement data simultaneously from all four limbs.

Theoretical consideration is then given to the scaling of the system to enable multiple performers to cooperate physically in the music composition process. This leads to the concept of a ‘performance orchestra’ where traditional musicians are replaced by instrumented dancers with direct control of accompanying sound.

List of contents

Chapter 1 - Introduction

Chapter 2 - Literature review and background theory

Chapter 3 - System Design, Experimental Methodology and Results

Chapter 4 - Conclusions

List of illustrations

Acknowledgements

The author expresses his appreciation and gratitude for the support, encouragement and enthusiasm of the project supervisor Professor David Howard.

Thanks are also due to the graduate administrator Camilla Danese for all the support and advice provided throughout this project.

Author’s declaration

This thesis contains the original work of the author except where otherwise indicated.

Chapter 1 - Introduction

This thesis documents a research project to develop and characterise a gestural capture and music composition system to allow interactive real-time motion-based control of computer music. The gestural capture system is worn by performers, typically an ensemble of dancers, who are conducted by a composer to simultaneously create movement and music. This chapter outlines the rationale for the project, defines the aim and objectives and presents the structure for the remainder of the dissertation.

1.1 Rationale

Generally dance is either free form, in which movements are decided by the performer, or choreographed in which case movements are orchestrated using a standard repertoire. In the majority of performance situations dance is a response to pre-composed music. Over the last decade, advances in technology have generated interest in exploring the inverse approach, to allow dance to be used as the basis for controlling sound generation and music composition. This has been motivated in part by the desire to utilise the rich expression that dance provides allowing it to be intimately linked to simultaneous production of music. The motivation for this coupling seems intuitively to arise from the sense of increased satisfaction that is experienced through other cohesive creative experiences, such as, operatic ballet.

The advantages offered by using body movement for gestural control of sound/music are complicated by the challenge of capturing the performer’s movements and of implementing mapping algorithms to translate this motion into sound. In recent years significant progress has been made to enable single performers to have intimate control over sound production. In these systems various external (e.g. video) or body-mounted sensors enable a performer to control a range of parameters, such as, tempo, amplitude, timbre, attack rate, and modulation. There remains, however, the basic difficulty of mapping pitch. Since the mapping of individual notes to body-mounted sensors is not usually a practical solution, a keyboard may still be used to assign pitches whose corresponding sound parameters are then controlled by a performer’s body movements. Alternatively, some investigators have achieved pitch control through the use of floor-mounted sensor mats which trigger different pitches through changes in the spatial position of the performer on the dance floor. Such a system provides a method that enables a dancer to move from simple sound control to music composition. However, this type of approach imposes constraints on the performer who must concentrate on being in the correct spatial position and this distraction can be detrimental to the musical composition. Moreover, some of the best conventional musical compositions result from the simultaneous contribution of multiple performers which is difficult to support in the ‘dance mat’ type approach, purely as a result of the problem of performers coordinating their efforts during real-time composition. This is overcome in electronic dance games by providing direction through visual conductor cues typically overlaid on video. Again the performer is forced to focus upon following the cues rather than on being fully engaged in the music composition process.

A proposed solution to this dilemma which is explored in the current project is the assignment of pitches and other musical parameters to a number of dancers who then collaborate to realise a musical composition. The coordination of this ensemble still requires either an independent conductor, or evolves ad hoc from the group. It its simplest form performers may just trigger pre-defined sounds. However, novelty arises through appropriate selection of motion sensors, allowing performers to apply a level of individual interpretation and artistic effect which reflects the interpretive freedom and control enjoyed by traditional musicians, and aids realism. In comparison to many of the systems reviewed in chapter 2, the current proposal appears initially relatively simplistic. However, a key advantage is the potential reduction in the complexity of the gestural mapping algorithms. As discussed in chapter 2, the complexity of the algorithms required to translate performer gestures into musical events has been a limiting factor on progress. The current proposal seeks to overcome this limitation by allowing a defined set of relatively simple gestures to be captured from a larger number of performers. In this way it is intended that the translation process for individual performers will be simplified. However there remain significant challenges to overcome in respect of recognition of specific performer movements and in automatically coordinating the contributions from many performers. For reasons of cost and practicality, the project has focused on developing and characterising the capture system for a single arbitrary performer using a scalable design that can be scaled up to an ensemble of performers.

The realisation of the concept of a ‘performance orchestra’ in which orchestral instruments are replaced by ‘dancers’ whose localised movements provide gestural control of sound, gives rise to a number of interesting technological challenges. The remainder of this report identifies these challenges and documents various practical and theoretical solutions towards the overall goal.

1.2 Aims and objectives

The principal aim of the project can be summarised as the development of a modular system to enable composition of music using a performing orchestra of instrumented dancers. Here the term dancer is not essential since it is intended that anyone capable of performing coordinated body actions could in principle be a part of the orchestra. However, it is expected that a performance would typically be a spectated event and therefore should provide a combination of both music and movement in order to enhance overall entertainment value. Consequently, performers are expected to be able to exercise reasonably well-defined movements, albeit without formal training. This is important as it implies that the range and types of movement supported by the system is tractably finite.

Despite the preceding constraint on performers of requiring a degree of precision in movement, it is known from common experience, that movement repeatability in a relatively novice performer is much lower than that of a trained dancer. This suggests that there will be significant uncertainty in the system's ability to automatically recognise specific gestures. Dealing with this uncertainty is a second primary aim of the proposed project.

To facilitate development of the system, other gestural capture technologies have been evaluated, including video, motion capture and ultrasonic scanning approaches. These approaches have been previously developed by other research groups and are known to have limitations in respect of cost, practicality and performance (refer for example to Dobrian and Bevilacqua (2003)). Consequently, it is envisaged that body-mounted sensors, such as, accelerometers and gyroscopes, would be utilised as the primary sensing mechanism.

Several key objectives were identified as essential to realising the project aims:

1. To evaluate suitable body-mounted sensing techniques for mounting on a performer's body to allow capture of body motions and gestures.
2. To develop a method of dealing with the uncertainty inherent in the acquired movement signals from performers to enable a reliable/repeatable mapping to sound events

3. To develop a scalable wireless based communication link between the performer(s) and the sound generation system.
4. To characterise the developed system and identify requirements for further enhancement.

1.3 Goals for personal development

In addition to the identified research goals the author sought to develop theoretical knowledge and practice with a range of technologies and apply these in a musical/audio context:

- To gain experience in interfacing and using accelerometers.
- To gain knowledge in the field of wireless sensor networks.
- To evaluate AI techniques relevant to the project.

1.4 Structure of the dissertation

The remaining chapters of this dissertation are organised as follows:

Chapter 2 provides a focused chronological literature review of recent progress in the field and introduces the relevant background theory that underpins the remaining chapters.

Chapter 3 discusses the design of the system, experimental methodology and the key results.

Chapter 4 provides a detailed project review, identifies areas for further work including potential refinements and enhancements and presents the project conclusions.

Chapter 2 - Literature review and background theory

The aim of this chapter is to provide a chronological literature review covering the most relevant recent developments. The focus is on key advances in relation to interactive dance systems for sound synthesis or composition of music. In this respect, systems implemented with integrated silicon sensors, and in particular accelerometers, are considered as these represent the most important developments, particularly in the area of gestural capture. Some consideration is also given to the gesture to sound mapping problem, to identify the key challenges and techniques being explored in the field. This chapter also provides essential background theory for the project which is of relevance in understanding material presented in later chapters.

2.1 Literature review

2.1.1 Capturing performance gestures using body-worn sensors

Despite the ubiquitous presence of computer music during the last two decades, few musicians consider that they are actually playing the computer as a musical instrument (Wessel and Wright 2004). This perception has arisen in part as a consequence of the relatively slow development or application of sensors to allow the same physical excitation as experienced with a traditional acoustic instrument. This challenge is referred to in the literature as gesture capture and in recent years has been significantly enhanced by the widespread availability of high performance motion sensor and motion capture technologies. A second common expectation of playing a musical instrument is that a single playing gesture corresponds to a single sound event such as generation of a pitch, chord or application of an effect like vibrato. Whilst a number of interactive music systems mimic this traditional one to one mapping, there has been significant interest in the research community in exploring high level mapping relationships between complex physical motions of the 'player' and corresponding generation of complex sounds or music structures. Moreover, with specific relevance to the current project, a number of research groups have extended the concept such that the instrumentalist/composer is also the physical performer, typically a dancer. This supports the concept of a symbiosis of creative dance and musical composition within an individual, or group of performers, that unifies creative expression.

Wanderley and Depalle (2004) reviewed the field of gestural control of sound synthesis and identified the following four key areas:

- Technologies for capturing gestures
- Methods of analysing a performer’s gestures
- Real-time sound synthesis methods
- Mapping from gestural variables to sound synthesis parameters

They identified these areas as comprising a highly specialised form of Human Computer Interaction (HCI) where multiple sound parameters, timing cues, rhythm structures and trained users combine into a complex real-time control system. Whilst the paper concentrates on control of sound synthesis it is relevant to the gestural composition system proposed in the current project as a number of similar challenges are evident. For example, in relation to the gestural capturing techniques the authors identified sensitivity, stability and repeatability as the most important sensor characteristics. Sensitivity is subdivided into spatial and temporal measures. Spatial sensitivity reflects the extent to which subtle movements by a performer can be captured and translated into musical parameters. Temporal sensitivity is considered to reflect the adequacy of the measurement system to capture anatomical movement and therefore provides a definition for the required sampling rate of the system. For humans the fastest motions are relatively slow, extending to a bandwidth below 100Hz. Stability is interpreted as being a measure of how well the system rejects small physical perturbations, sensor overshoot and saturation. The proposed use of modern accelerometers will alleviate many of the stability issues that reduced the performance of previous alternate solutions. These modern devices incorporate thresholds that can reject small movements; have adjustable coefficients for internal dampening; and have ranges well above the levels of acceleration normally experienced by humans, thus avoiding saturation. Furthermore, these modern devices support internal electrostatic calibration methods that enable compensation of systematic offsets arising from manufacturing tolerances or thermal variations.

Paradiso et al. (1998) developed an instrumented dance shoe that incorporated one plate of a capacitive sensor in the insole with a second plate located externally in the floor of the stage. The capacitively coupled sensing channel was used to determine the spatial location of the dancer on the stage via a transceiver also located on the shoe. Additionally, piezoelectric film sensors were incorporated in shoe, to measure foot pressures and a two axis accelerometer was used to detect pitch and roll. These signals were also transmitted wirelessly to a remote computer that performed mapping onto a simple three voice musical structure providing drum, bass and a melody. Through a training cycle the performer and software were 'tuned' to enable basic composition. The potential for interactive dance composition was evident with this approach however a number of limitations were identified: Only a single one shoe was supported which limited dance expression; the transceiver was relatively large covering approximately half of the rear side of the dance shoe; the transceiver required an external whip ariel extending to several inches above the heel which could interfere with the dancers motion; the need to train the performer to achieve successful mapping between movement and corresponding sound/music made the system relatively performer specific.

Wienberg et al (2002) developed a networked system for interdependent collaborative group composition of rhythms. Their wired system provided an interconnection of up to eight controllers with an allocation of one controller per musician and supported real time composition of complex rhythm patterns. Interestingly this system required a lead player to start a session by defining key parameters, such as tempo, which sets a schema for other musicians to follow. Beyond this basic rhythmic structure participants were able to contribute motifs in an ad-hoc fashion. This approach worked best when the participating performers had trained together with the system for some time. In which case, it became apparent that the system was providing an interactive form of musical social networking.

The work of Modler et al. (2003) details the use of artificial intelligence in the form of a Neural Network to deal with the uncertainty associated with the normal variation of hand gestures captured by video. The same uncertainty which is due to normal variation in human response occurs irrespective of the measurement technique and therefore will be an important issue in the development of the proposed system for the current project. Their justification for the use of video is based on the fact that body-worn sensors often require a focus on matching behaviour to the technology rather than creative expression. An interesting finding from the study by Modler et al. is that the artificial neural network demonstrated capability of recognising a gesture before it was complete. This predictive capability is important as it helps to reduce the latency between gestural action and triggered musical response. The system was also demonstrated to be robust in its capability of recognising specific hand gestures from a set of 14 motions. The success rate ranged upward from 80% dependent on the test conditions employed. The lower success rate of 80% was achieved using analysis of a live performer creating 30 random gestures. A key implication of these findings is that some form of intelligent processing is likely to be needed to deal with the problem of uncertainty in classification of movements that have inherent variability. For the proposed system it is expected that the detection rate will need to be significantly higher than 80% for live composition work to ensure the integrity of the music produced. This suggests that there is likely to be a need to tune each sensor node to the individual performer through a calibration or training phase. Also the system may need to be adaptive so that it can automatically compensate for 'gestural drift' that might be expected as the performer's muscle's warm up or become tired.

The study of Wessel and Wright (2004) is useful in providing an upper bound on the acceptable latency between gestural capture and sound/music generation of 10 milliseconds for real-time performance. They also argue that the latency should not vary by more than +/-1ms as psychoacoustic analysis has demonstrated that variations beyond this can be detected by the human ear. Furthermore the authors justify the use of the low latency Open Sound Control (OSC) protocol as a more robust and flexible solution for mapping real world continuous performance motion compared to the discrete event based MIDI protocol. In order to achieve acceptable levels of performance and to facilitate a scalable architecture the implementation required the use of field programmable gate arrays (FPGAs) whereas traditionally general microprocessors or digital signal processors have been employed.

Maki-Patola et al (2005) discuss the development and characterisation of four different gestural controllers for musical instruments. The instruments are 'playable' by a performer in a virtual environment. Of interest is their suggestion for the need to incorporate tactile feedback so that an instrument player has a physical sense of the virtual instruments response to being hit, plucked, strummed etc. In practice the realisation of this feedback seems to have been restricted to visual cues over the instruments physical response to excitation. This is probably a consequence of the technical difficulty of providing a realistic tactile feedback to a player since bio actuators still remain less advanced than sensing technologies.

Lynch et al. (2005) describe the development of a wireless inertial measurement system for an interactive dance environment. Their paper focuses on the development of the technology and in particular an inertial measurement unit and wireless communication system. The key challenge was to integrate these technologies into a unit that could be worn by a dancer at key sites over the body. Previous investigators had used off body measurement techniques, such as, ultrasound or video to capture motion and the required processing introduced unacceptable latencies when used to trigger interactive media. In contrast the sensor developed by Lynch et al. utilised two accelerometers, three gyroscopes and two magnetometers to realise a complete six degrees of freedom inertial measurement unit in a compact 25mm cube. The signals from these sensors were sampled into the digital domain and processed using a digital signal processor running a Kalman filter algorithm. This filter was necessary to compensate for the uncertainty in spatial position resulting from the drift error that is an inherent limitation of inertial measurement systems. Consequently, in order to achieve acceptable real time response the DSP was implemented as a custom solution on an FPGA. The communication system was based on a custom wireless protocol providing point-to-point transmission. The complexity of the measurement cube reflects technology that is already somewhat outdated. Today complete inertial measurement systems offering measurement over nine degrees of freedom are available as small single package solutions, although the cost is relatively high (around £100 at the time of writing). Similarly, small footprint flexible communication systems using standardised protocols are now widely available, for example, Zigbee modules.

Knapp and Perry (2006) discuss the design of a networked wireless music control system that allows many performers to collaborate in musical synthesis. The sensed input data to their system differs from that proposed for the current project in that it is based primarily upon capture of the performer's emotional states in real time. This was achieved by using biomedical transducers to obtain, brain wave activity (EEG), muscle activity (EMG) and heart activity (ECG). These signals were coupled into an existing medical wireless monitoring system produced by TeleMuse in the US. At the receiver, the signals were decoded and passed into a computer running the ChucK audio programming language. Algorithms were developed to control a virtual chorale vocal so that vocal related parameters could be modulated by the emotional state of the performer. Evaluation of the system was not reported in detail.

Aylward and Paradiso (2006) describe a 'multi-user sensor system for interactive dance'. This improves on the designs originally developed by the Media group at MIT by realising a more compact and unobtrusive wireless based system that can be worn on the limbs of dancers. Of particular interest in the context of the current study are the calculations presented to estimate the required bandwidth of the communication system. For example, an ensemble of five collaborating instrumented dancers is estimated to require a bandwidth sufficient to support a transmission rate of 400kbps of data. Whilst this is in the range of modern wireless data transmitters it is important to remember that the power consumption of these devices is proportional to the data transmission rate. For compactness, body-worn sensors are typically powered by lithium coin cells which provide fairly limited power/time capacities. Since a dance/music performance would conservatively require a sensor node operational time of around two hours it is important to reduce power consumption which is dominated by the power requirements of the transmitter. Thus minimising data rates is an essential design constraint. Therefore a key consideration of the proposed study will be to find a suitable balance between processing on the body-worn nodes and that done remotely in the computer performing the mapping process. To some extent this is a contradictory aim since higher performance local processing generally requires more power.

In contrast to some of the complex motion capture systems for dance described so far, the system of Feldmeier and Paradiso (2007) is relatively simple yet has been demonstrated to be scalable to several hundred participants. The system is aimed at non-performance dancers i.e. on a dance floor rather than a stage and enables control of the music being played through a public address system. The system comprises a tilt sensor constructed from the piezoelectric film, polyvinylidene fluoride (PVDF), which is applied to a comparator with an acceleration reference set at 2.5-G. The processed output of the comparator is connected to a radio frequency transmitter which sends a trigger signal to a remote base station receiver. The output of the receiver is decoded by a computer which runs mapping algorithms to translate the signal ensemble from the crowd into audio control signals. Thus the system acts like a polling system in which dancers vote for an increase/decrease in a parameter of the music being played. At any one time the currently selected parameter is displayed remotely using an optoelectronic display system. In the context of the current project, the relevance of this study is the confirmation it provides of the efficacy of controlling musical parameters through sensing signals from large interacting groups of dancers. Furthermore, their solution demonstrates the ability to achieve this at relatively low cost since the sensors were ‘giveaway’ items covered by the price of admission to the dance event. An important limitation of this system is its lack of support for real-time composition which is a key objective of the current project proposal.

2.1.2 Key issues for mapping gesture to music

Winkler (1995) provides a succinct introduction to some of the key strategies and approaches to the problem of interactively mapping gesture to computer music. Of particular interest is the recognition that the granularity of interaction between performer movement and corresponding control of sound/music 'lies on a continuum'. Thus a relatively coarse strategy is achieved by simply triggering pre-composed music, such as loops, in response to specific performance events, for example, foot tapping. In comparison a much finer level of control derived perhaps from subtle movement of the fingers can be used to control synthesis parameters or the playing nuances of virtual instruments based on physical models. For the current project, a control granularity between these extremes is envisaged, as dictated by measurement instrumentation (accelerometer based) that whilst sensitive, can only be practically mounted on major limbs of the performer rather than, for example, individual fingers. Whilst the degree of control is primarily dictated by the resolution of gestural movement the author also states that strategic decisions also have to be made regarding how input control signals derived from gestural capture are mapped onto musical composition, including, for example, parameters like harmony, tempo, register and dynamics.

Even though significant progress has been made with captured gesture to music mapping techniques, the actual implementation of the mapping algorithms has tended to be project specific and often heuristic. The required formulation of these approaches into a set of generic methods has so far not materialised. An alternative approach to defining a tractable mapping algorithm has been the use of artificial intelligence methods and in particular artificial neural networks (ANN). Cont et al. (2004) describe an ANN based mapping strategy that can be trained to respond to input motion by generating arbitrarily complex musical responses. The training can be expert based i.e. involving an experienced musical composer to interpret physical gestures from an instrumented performer or the performer can self-train the system by providing desired musical or sound responses. With the ANN approach small variations in input gesture can be effectively eliminated through averaging over the training phase. Consequently the classification, rejection and exception processing that must be incorporated in standard algorithmic approaches is no longer required thereby simplifying implementation of the mapping process. Whilst this approach is showing considerable promise in easing the mapping challenge the authors concede that its expansion to more complex input gestures is limited by the length of time required to perform the training cycles.

Many of the approaches considered so far have focused on the problem of mapping gestures that are quantified spatially or occur over a relatively short timescale. In contrast the work of Lee et al. (2007) is based on quantification of temporal variations in motion and gesture. In effect this is asking the question of what musically relevant parameters can be extracted from average motion of some defined epoch. Their system utilises accelerometers to capture performer motion data over timescales of typically a few to tens of seconds from which rhythmic values are determined. This provides useful information regarding a performer's intent for tempo and pulse. It is notable that the authors used frequency domain analysis to extract rhythmic values from the periodic motion of the performer. This contrasts with the more typical time domain analysis and is reflective of the sensitivity of accelerometers to movement artefacts that are manifest in the signal. Although the problem could be dealt with in the time domain, the required variations in rhythm suggests the need for an adaptive filter since the bandwidth of the artefact noise is variable. Such a complex solution can be avoided in the frequency domain where the peaks in the spectrum correspond to impulsive motions averaged over a FFT window of several seconds.

The recent work of Schacher (2010) is representative of the current state of the art in gesture to sound mapping for interactive dance. This work describes a maturation of the techniques developed over the preceding decade by various researchers and demonstrates the realisation of a system in which the dance performer can be truly considered as instrumentalist and conductor. To achieve this level of integration it is recognised that a number of sensing streams should be used to capture the full expression of the dancer. It is argued that in this way, the emotional content inherent in physical dance can be extracted, and this information then more easily mapped to corresponding emotional content in created music. This approach provides for a more natural and musically satisfying experience than was achievable in early mapping approaches where the generated music often lacked the subtle playing nuances that an instrumental musician would incorporate. To achieve this level of fine grained emotional mapping the author develops the concept of gesture as a 'meta layer of motion'. This means that sequences of captured primitive movements are recognisable as complete gestures that have emotional meaning and context. This information can then be mapped through a suitable translation process into music using for example an algorithmic composition technique. This multi-layered indirect approach to the mapping of physical variables onto music structure introduces a number of challenges. In particular for real time composition the need for data reduction is necessary to enable adequate data throughput. This can be achieved in part through elimination of redundant data as several of the sensing techniques capture overlapping information. Furthermore, data weighting is applied to prioritise contributions from different sensing channels. Despite these advanced techniques the author acknowledges that the system still requires some intervention from a human composer to guide the translation. The work is also of further interest in the context of the current study as it demonstrates successful use of a Zigbee based wireless transmitter (XBee) for an interactive dance application.

2.2 Background theory and design considerations

This section summarises the background theory for the projects and elucidates some of the key design decisions.

2.2.1 Accelerometer selection for performance gesture capture

The general concepts relating to the design of sensor systems for musical interfaces are comprehensively discussed in the literature and the reader is referred, for example, to the work of Miranda et al (2006). Here consideration focuses on the selection of sensors and in particular accelerometers for gestural capture.

Many of the previously published designs for interactive dance systems comprise multiple types of sensor. The motivation for this is that a true dance performance incorporates a full range of motions and the performer's position may also vary in three dimensional space. Therefore, many variables including position, velocity and angular rotation rates may need to be measured to fully capture performance gestures. At the present time the few single sensor solutions that can achieve all of these measurements are prohibitively expensive and typically two separate types of sensor, an accelerometer and a gyroscope are combined on a single PCB to implement an 'inertial measurement unit' (IMU), such a solution may have between three and six up degrees of freedom depending on the number of axis each sensor has. For the proposed system, it is intended that the performers forming the 'orchestra' will remain in a relatively fixed position and movement will be localised. Therefore precise positional information is not required. Consequently, gestural motion will be effectively restricted to approximately linear motion in three dimensions therefore measurement of angular rotation is not required. This reduces the problem to one of measuring accelerations in three dimensions. A key benefit of this simplification of the capture system is that it eliminates the drift associated with an IMU which arises when acceleration is integrated to derive velocity. This integration process introduces small errors that compound over time and typically a complex compensation scheme is required to minimise the error.

Accelerometers measure the rate of change of velocity so are sensitive to slowing up or slowing down motions. This has a direct parallel with the way in which some types of musical instrument are played, i.e. strummed, bowed or hit excitations exhibit acceleratory behaviour. This seems to partially explain the adoption of these sensors in the field of interactive dance systems.

Most accelerometers are sensitive to both static and dynamic forces. Static sensitivity is used in tilt sensing applications to measure orientation relative to the earth's gravity. Dynamic measurements are useful in determining transitions in motion. The units of measurement for an accelerometer are expressed in standard SI units as m/s^2, since the units for velocity are m/s, acceleration defines by how much velocity changes per second. On Earth the acceleration due to gravity, g, is 9.8 m/s^2 and therefore accelerometer measurement ranges are often expressed in terms of g. During normal unaided motion the human body typically experiences between 1g and 3g depending on the type of motion. Values over 4g have been measured in Irish jig dancing. The sensitivity of an accelerometer depends on the full-scale measurement range. A lower full-scale measurement range allows for a more sensitive detection of acceleration changes. In interactive dance/music applications, greater sensitivity provides for a finer control of sound parameters which is often desirable to give precise control over generated sound. Taken together, these factors suggest that an accelerometer with a full-scale range of around plus or minus 3g will be required.

Most modern accelerometers are tri-axial, producing independent output signals along three spatial directions, x, y and z. If we consider such a sensor mounted on the upper side of the forearm of a performer then a standard orientation might map horizontal arm motion to the x axis, vertical arm motion to the y-axis and arm motion away from and toward the body (i.e. push-pulling) to the z-axis. The output signals from each axis are simultaneously available and therefore intermediate arm positions can be resolved into components along each axis if necessary, thus allowing a full 3D determination of sensor acceleration. The response time of an accelerometer's output is expressed as a bandwidth and is typically in the range of 50-100Hz corresponding to minimum update rate of 20ms. In comparison 'fast' reaction times in humans are around 100ms. Thus a typical accelerometer is easily capable of tracking human motion.

Based on the preceding discussion, three accelerometers were identified as suitable for the current project: ADXL335 and ADXL345 (both Analog Devices) and the BMA180 (Bosch). Final selection of the ADXL335 was based upon the fact that this device provides direct analogue output signals whereas the other devices require a digital interface. The analogue interface allows for the output signals to be viewed directly on an oscilloscope allowing the device to be evaluated directly without the need for additional instrumentation. Furthermore, once the device was integrated with the control instrumentation the sampling rate was programmable via the microcontroller rather than limited to the internal sampling rate of the digital interface devices. The cost of this convenience is the potential for increased noise coupling into the analogue input channels. However, this can be resolved through inclusion of a ground plane on the printed circuit board (PCB) and by employing shielding within the sensor to interface connection wires. The final advantage of the ADXL335 is its very low current consumption of only 320 microamps which is ideal for battery operation, an essential implementation requirement for this project.

2.2.2 Selection of the processor for the gestural capture system

According to the data sheet the ADXL335 accelerometer has a nominal sensitivity of 300mV/g with a full scale measurement range of +/-3g. Therefore the output signal voltage range is given by (3g x 300mV/g) = 900mV. The bipolar measurement is offset on a DC bias of +1.5v so a unipolar analogue to digital convertor (ADC) can be used to sample the output. These figures are stated for an accelerometer supply voltage Vs of 3V. Assuming that this voltage is used as the standard supply voltage of the system, the voltage resolution per bit of an 8 bit analogue to digital convertor is $3/_{256} \approx 12mV$. Thus each converted bit of data has a weighting of (12mV/900mV) x 6g = 80mg.

Comparing this value to the force of gravity for the Earth, taken as 1g = 9.8 m/s^2 gives an acceleration resolution of 0.080 x 9.8 = 0.784 m/s^2 equivalent to a rate of change of velocity of approximately 80cm per second. Assuming a distance travelled during strumming or hitting a drum of 40cm allows for 2 such events per second this equates to a maximum effective playing tempo of 120BPM. For the purposes of the present study it was decided that this range would be acceptable for the development of the system. An important benefit of this restriction is that it allows for the use of an 8 bit ADC which is typical of the conversion bit depth available on many microcontrollers. It also allows for data values to be encoded easily on to the RS232 serial communication protocol which simplifies communication and increases the data transfer rate. However, there is clearly room for improvement in terms of accelerometer sensitivity.

It was originally intended to design a system in which all processing was performed in the sensor node so that the output from the system was the mapped musical event data e.g. MIDI events. This approach alleviates the need for significant remote processing of the data before it is sent to the sound/music generation system. A prototype using an ATMEGA168 processor (Atmel) was constructed and evaluated. The system operated at 10MHz and sampled the three input channels at 100Hz per channel consistent with the Nyquist criteria. Thus the total sampling rate was 300 samples per second. This data was then high-pass filtered using a digital filter to remove low frequency (0.5Hz) signal content. The origin of this low frequency interference was attributed to thermal variation of the accelerometer as the arm was moved. The air flow appeared to have a cooling effect which introduced low frequency signal artefact. This was proved empirically by using a freezer spray to stabilise the temperature which resulted in elimination of the low frequency artefact. The output of the filter was coupled to a discriminator used to threshold the sampled data. This was found to be necessary to eliminate movement jitter which appears to be an intrinsic aspect of human motion. The output of this stage was then used to 'address' a look up table which performed a simple mapping between the magnitude of acceleration and MIDI note velocity. Finally this data was assembled into a communication packet and transmitted. It became apparent that this level of processing could not be adequately supported by an 8 bit processor unless the processor clock speed was increased to its maximum speed of 20MHz. The lower clock speed was used initially to preserve power as the processor consumption scales linearly with clock speed. The 10MHz clock speed was selected initially to allow for the 2 hour target operation time from a 120mA/hour coin cell.

Given the limitations experienced with the 8 bit processor based system, a second prototype was developed using an ARM processor (NXP LPC1768) development board (Mbed). This 32 bit processor operates at 96MHz, has 16Kb of RAM and includes on board

DSP support allowing a convolution based digital filter to be executed very efficiently. However, two drawbacks were found with this approach, first the nominal power consumption was 150mA which effectively reduces the operational performance time to only one hour. Second, the Mbed development system is based on pre-defined library functions and offers little support for register level programming making it difficult to tune the system. As the author was not able to secure funding for a full high level development system for the ARM processor the refinement of the system would require assembly level programming which was considered infeasible given the complexity of the processing algorithms. For this reason a decision was taken to perform the processing remotely on a centralised computer and to develop the sensor node as a means of simply capturing and transmitting the raw accelerometer data. The final sensor node was therefore developed around an ATMEGA328 8 bit processor using the free AVR studio development system from Atmel. All code was written in the 'C' language.

2.2.3 Consideration for implementation of the wireless communication subsystem

The majority of systems considered in the literature review were found to use various low power digital radio modules. In most cases the performer modules were transmitter only and the sound/music processing system a receiver only. A pair of such modules type FM-RX2 and FM-TX2 (RF Solutions) were evaluated for the proposed application. These modules operate in the 418MHz UK public authorised band. The modules are capable of transmission/reception up to 300 metres which is adequate for the intended application. The digital data rate is also acceptable at up to 320Kbits/s. However, formatting of the digital data stream must be done externally in the application host processor which was found to impose a significant burden on processor performance. Furthermore, these devices all operate on the same frequency and therefore in a multi-sensor implementation some form of scheduling is necessary to avoid interference from simultaneous transmissions. The modules themselves are also physically large approximately 4cm by 2cm and thus not well suited to being worn by a performer (from the perspective of performance aesthetics).

As a consequence of the preceding limitations the author decided to investigate the feasibility of implementing the proposed system as a wireless sensor network based on Zigbee technology. This technology resolves most of the preceding problems in particular it is designed for networked applications which overcomes the issue of interference between modules. It is also used for low power communications with data rates up to 250Kbit/s. The device footprint of a typical implementation e.g. Xbee is approximately 2cm x 2cm including a built in ‘chip’ antenna which supports transmission distances up to 100m. The major advantage of this technology for the current application is the inclusion of a dedicated processor which implements the Zigbee protocol. This removes a significant processing burden from the host application processor allowing the latter to perform some of the sensor data processing locally. This allows for a reduction in the quantity of data transmitted which reduces power consumption of the sensing nodes. The Xbee modules are also transceivers which, allows a bidirectional transfer between the sensor nodes and remote host sound/music generation system (e.g. a PC). This bi-directionality is important as it allows the system to be adaptive in the sense that tuning information from the remote host can be sent to a sensor node for example to adjust filter bandwidth if the signal quality deteriorates.

Chapter 3 - System Design, Experimental Methodology and Results

3.1 Overview

This chapter details the iterative design of a performer motion-based composition system commencing with a derivation of the specification from which the requirements were identified. These requirements were then mapped into several exploratory designs which provided a basis for initial data capture and analysis. The hardware design was then refined into its final form. The development of a reliable signal processing algorithm is discussed as well as the characterisation of the various subsystems. The centralised processing of sensor data by a remote host to perform the mapping into sound/music is also presented.

3.1.1 Specification and requirements

The prototype system will comprise a body-worn network of four independent motion sensor nodes each with localised processing and communication subsystems. These will be termed gestural capture nodes. Each of these nodes will communicate with a body-worn master node which will then communicate with a remote centralised processing system that implements a mapping from the captured gestural data into sound/music.

It is a requirement that the system is scalable from a single sensor to many sensors on many performers. For practical design purposes this will equate to an upper limit of fifty performers each with a maximum of four sensing nodes. This means that up to 200 simultaneous sound generating events should be supportable. The communication system must provide adequate bandwidth to support this level of interaction.

It is a requirement that the system be capable of automatically recognising gestures which may exhibit significant variation either temporally for a given performer or from performer to performer. Tuning of a sensor node to a specific performer is not acceptable.

The final implementation of the sensor nodes should be in the form of battery operated units with wireless communication. The operational time for a sensor node should be two hours. Sensor nodes should be small enough to be mounted on the wrist or ankle and must not impede performer movement or be overtly visible to an audience.

3.1.2 Initial experimental studies

An early prototype system to enable capture of representative performer motion data was constructed and evaluated. This system comprised a proprietary three axis accelerometer (LIS 302DL ST Microelectronics) and a custom built embedded microcontroller interface, designed by the author. The interface design was based on an AVR-Mega168 microcontroller (Atmel) and was implemented using through hole technology for ease of construction and access to signals. Figure 1 shows the accelerometer mounted on the wrist and Figure 2 shows the interface board.

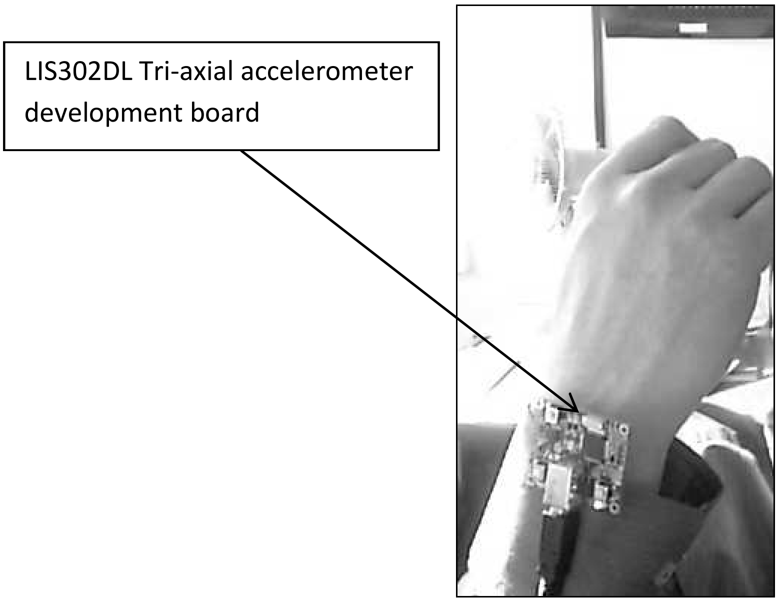


Figure 1 Three axis accelerometer type LIS302DL (ST Microelectronics) mounted on the wrist to capture motion of the forearm.

The prototype motion capture system was evaluated by performing simple upper limb motions along the XYZ axes. The output data from the accelerometer was connected via an I2C link to the interface board where a time stamp was added using a real-time clock (Dallas Semiconductor). Time stamped data was then be down loaded to a PC via a wired RS232 link at 9600 baud for off-line analysis. Figure 3 shows typical raw data which was plotted in Matlab.

The waveform in figure 3 shows the accelerometer response to motion of the arm along the Z axis, which for the orientation used in this experiment is a simple up/down raising and lowering of the arm. The graph shows an initial motion at Time=10s which was used to provide a visual synchronisation check against simultaneously captured video of the subject's motion. The graph also shows capture of a set of repetitive motions of the arm along the Z axis starting at Time=25s and continuing until Time=45s. Similar data was obtained for motions in the X and Y axis.

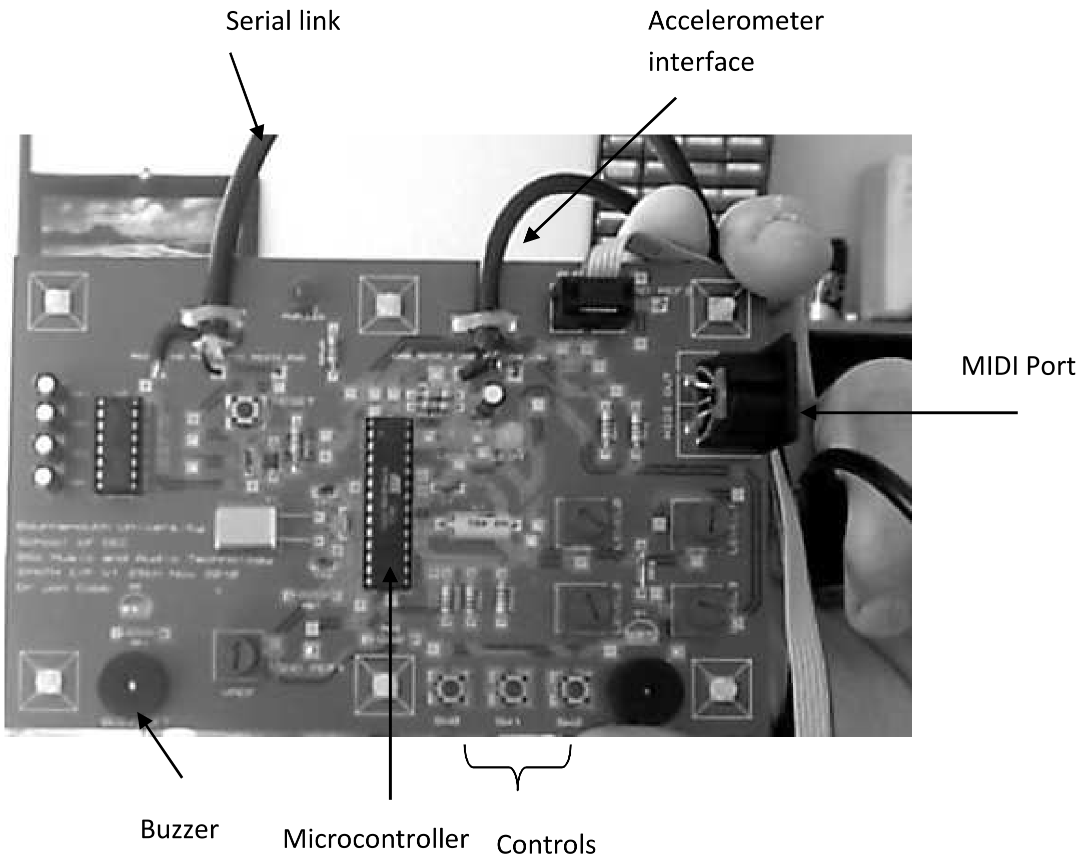


Figure 2 The custom built interface board for the accelerometer.

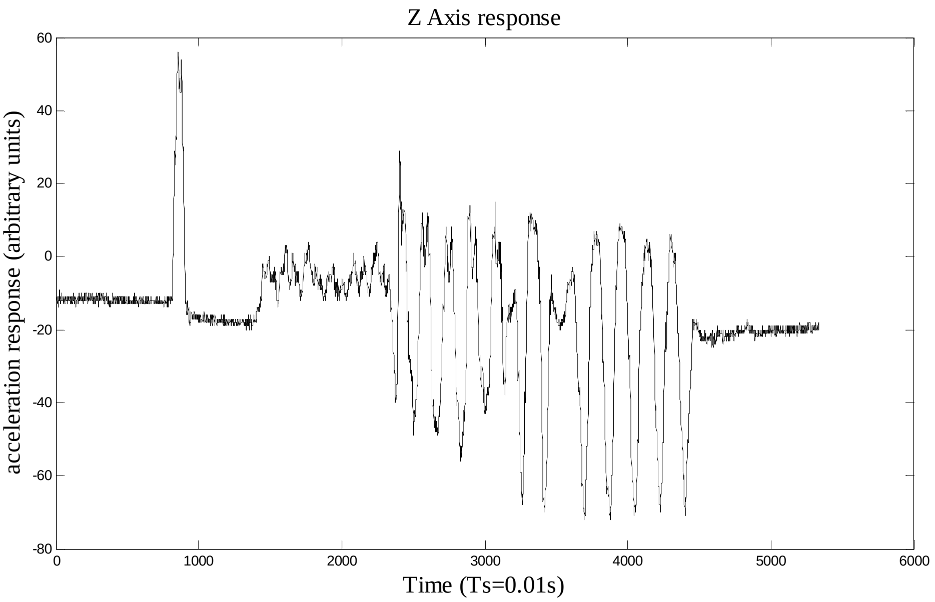


Figure 3 Typical signal acquired from one channel of the accelerometer.

The data shown in figure 3 appears reasonably well-defined however, signal processing was necessary to extract a clean signal. First, a low pass filter was applied to remove the high frequency noise evident in figure 3 e.g. in the 0s to 10s epoch. This filter also helped to reduce some of the movement artefact e.g. in the 15s to 25s time segment. This motion artefact arises due to the high sensitivity of the sensor which results in a notable response even to slight movements of a limb.

Standard Matlab algorithms (e.g. findpeaks(dataz);) were then applied to extract the maxima in the motion waveform. This procedure was complicated by two factors (i) Accelerometers exhibit DC drift which is evident in fig 3. (ii) There can be significant variation in the amplitudes of the peaks, for example, figure 3, shows a reduced peak at 31s and a double peak at 35s. These problems suggested that the basic approach i.e. threshold comparator followed by a peak detector resulted in missed events. This is significant since in the proposed application it is desirable to achieve 100% detection accuracy to ensure integrity/continuity of generated audio.

In order to solve this problem, various adaptive peak detection algorithms were developed and evaluated. The results demonstrated that if the algorithms were tuned to the dataset, event detection rates of 100% were achievable. However, this required substantial manual intervention which is not practicable for the intended final automated system. In the more general case the reliability of event detection was between 60% and 85% depending on the particular algorithm used. Furthermore, achieving the higher values in this range required a more computationally intensive algorithm which significantly extended the time required to process the data.

In parallel with the development of the event detection algorithm, other aspects of the system functionality were explored by manually identifying the event times from the filtered raw data. Four parameters were extracted: (i) The time of the peaks of the motion events. (ii) The magnitude of the motion events. (iii) The duration of the motion events. (iv) The rate of change of the signal. The challenge then became one of developing a technique for mapping these parameters into an algorithmic composition system that would automatically produce a musical score. The majority of algorithmic composition systems use a random number generator together with musical rule sets. In the current study the accelerometer magnitude data was used to seed a random number generator in Max/MSP (Cycling74) which then produced MIDI data to trigger a virtual synthesiser. The measured rate of change data was used to define note rhythm values.

The preceding study was useful in demonstrating the efficacy of gestural capture using an accelerometer. Consideration was then given to the extension of this concept into a music composition system in which a network of body-worn sensors could be used to directly trigger sounds or to control sonic parameters, such as, volume, after touch, pitch bend, sustain etc.

A diagram of the conceptual system to be worn by a performer is shown in Figure 4.

With reference to figure 4, sensors located on the body provide motion related signals in response to body motion. The event signals are relayed to a remote centralised processor which processes the signals and generates MIDI output data to drive a sound generator. The dynamics of the audio output can be made proportional to the magnitude of the motion, for example, a rapid change in acceleration could be used to produce a fast transient attack.

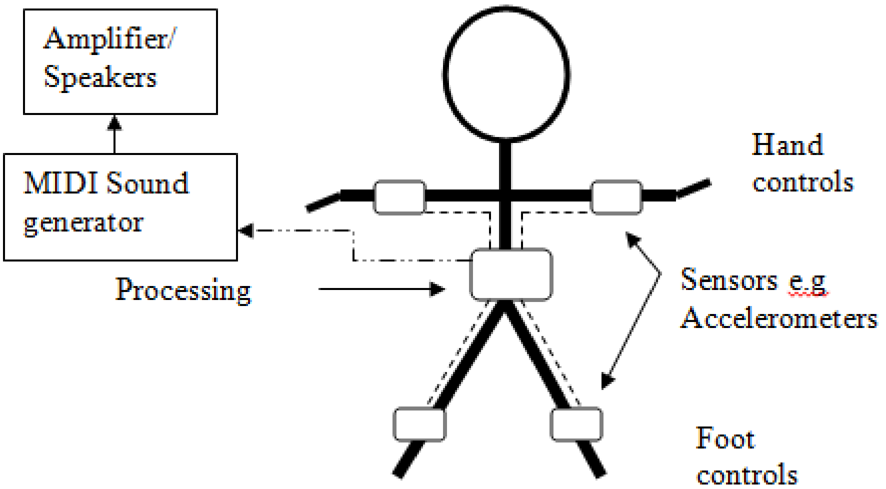


Figure 4 Conceptual architecture of the proposed system.

3.1.3 Design of initial bench prototype system

Figure 5 is a schematic diagram of the initial prototype designed for development and evaluation of the concept on the bench rather than as a performer mounted sensor subsystem. This design is a wired solution with a tri-axial analogue accelerometer type ADXL335 (Analogue Devices) wired to the interface controller board and connection to the remote host PC via a wired RS232 communication link.

Integrated circuit (IC) U1 is a microcontroller (Atmel ATMEGA168) which is programmed via the in-system programming interface connector AVR_ISP using C code developed in AVR studio. X1 is the crystal oscillator providing an accurate timing reference for the processor and divided internally to provide a real-time clock for time stamping of captured data. The standard reset circuit to the left is the principal method of 'rebooting' the system. To the upper left a set of three programmable switches are incorporated to allow user control of initiating data capture, selecting between MIDI and RS232 communications and to commence transmission of data over the serial communications link. Two programmable status LEDs (RLED0 and RLED1) were used to provide indication of the mode of operation, for example, RLED0 flashes green to indicate when data is being collected. The lower middle area of the schematic shows two buzzers which provide an aural indication to the performer. This functionality was included to assist subjects with training for standardised movements. For example, in order to get a 'standardised' strumming action, one buzzer generates a start and stop tone and the other an action tone. Using these sounds for guidance, subjects were able with minimal training, to achieve adequately matched movements by synchronising their action with the guidance sounds. This technique would of course not be appropriate in a live performance and so the problem of dealing with uncertainty in performer motion remains (resolution of this problem is considered in detail later).

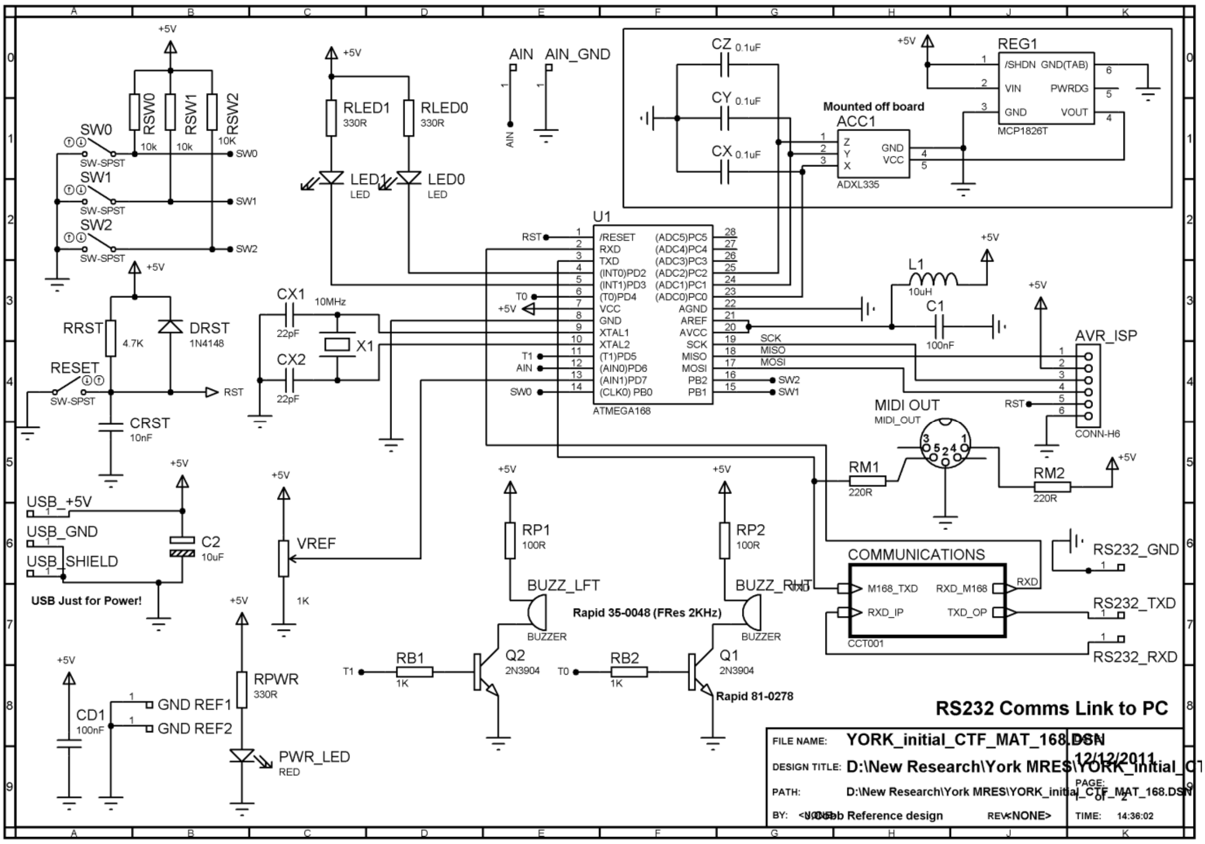


Figure 5 Schematic diagram of the initial wired prototype gestural capture system.

The lower right side of the schematic given in figure 5 is the communications subsystem. This supports two wired communications channels. The communications sub-block contains a RS232 level shifter type ST232 (ST Microelectronic) which provides a serial link to the PC. Data sent over this link is captured by a terminal emulator (Brays Terminal V1.3) from where it was buffered to a file for further processing and analysis using, the Matlab (Mathworks) and Max/MSP (Cycling74) applications. The second communications channel implements a MIDI output channel which was used to send MIDI event messages to a MIDI equipped tone generator. For the purposes of this study, a Yamaha TG300 tone generator was employed as the MIDI sound generation system.

The populated PCB hardware for this design has already been illustrated in the photograph of figure 2. However whereas figure 2 shows the interconnection to the LIS302 accelerometer via an I2C connection, the basic design shown in figure 5 was modified to provide an interface to an analogue accelerometer type ADXL335 (Analog Devices). This subsystem is shown in the boxed area to the upper right of the schematic. Each of the three output channels of the accelerometer is connected to the interface board by a shielded four-core instrumentation cable, having a grounded shield, a voltage supply to the accelerometer and three output signal connections one per axis of the accelerometer. The voltage supply to the accelerometer is derived from the main +5V supply using a regulator down to the required +3.3 Volts.

Capacitors CX,CY,CZ provide low pass output filtering from the accelerometer, and provide the dual purpose of setting the device bandwidth and implementing first order anti-aliasing filters on the input to the microcontrollers analogue sampling channels. In accordance with the ADXL335 datasheet the formula for calculating the required capacitance value is:

$$C = 1/(2\pi \times 32K \times f_{-3db}) \qquad [1]$$

Setting the bandwidth to 50Hz in accordance with the justifications given in chapter 2 then gives C=0.1uF.

3.1.4 Evaluation of the initial bench prototype system

Using the initial bench prototype system, test data was captured and algorithms developed to extract a reliable triggering signal for controlling MIDI events.

The X axis channel of the accelerometer was aligned in the vertical plane of arm motion to enable capture of strumming type gestures. Sampled data was sent over the serial line into a PC via Bray's terminal operating at 9600 Baud. ASCII and binary values were captured and imported into Excel. A graph of typical raw data from this type of measurement is shown in figure 6 after removal of the DC offset. Note the data from the accelerometer is an AC signal which is fed into the unipolar input convertor resulting in a large DC offset equivalent to a DC shift to the mid-point of the 8 bit data range of the sample convertor. Thus an initial processing step was to subtract the DC offset bias from the raw data. The DC value was determined as the average value of the sample data.

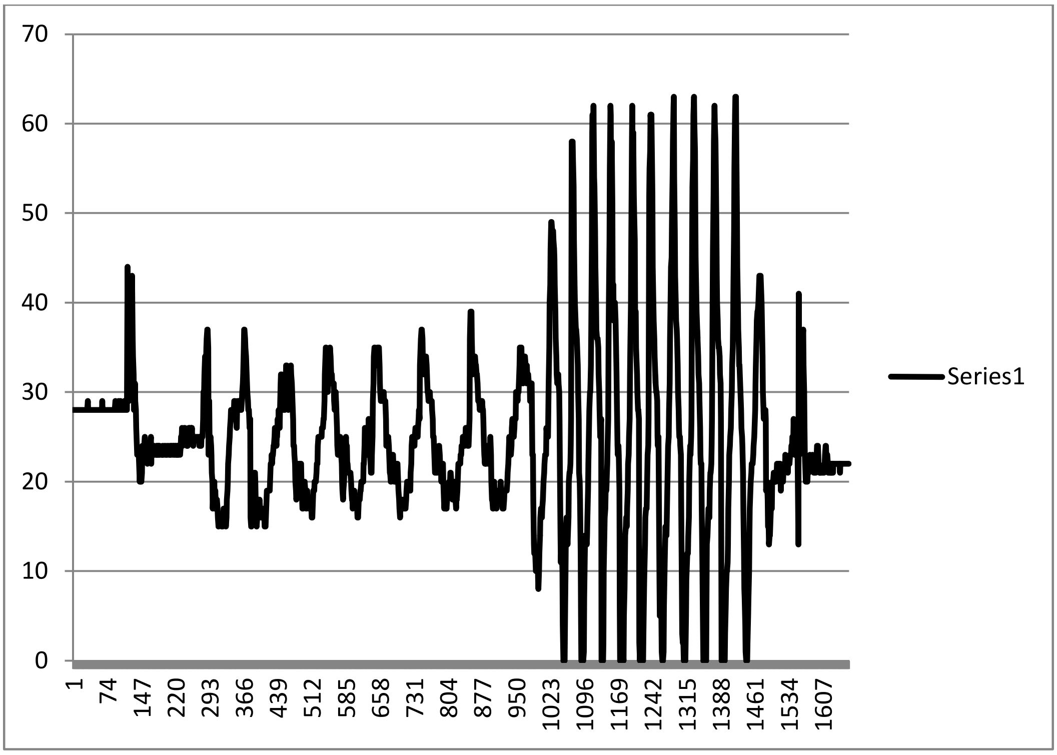


Figure 6 Typical raw data from an axis of the accelerometer following subtractive removal of the DC offset. This signal contains 20 cyclic strumming motions. The horizontal axis of this graph is in unit sample times (1/50Hz) = 20ms.

The signal displayed in figure 6 adequately captures both slow (initial epoch to 1000 samples) and fast arm movements (second epoch 1000 to 1500 samples) however there is significant noise due to interference coupling from the mains supply and also movement artefact (physical motion jitter) present in the signal. This made further time based analysis difficult. Consequently, the raw data for the signal shown in figure 6 was imported into Matlab as a column data variable. This comprises 1666x1 vector of type double. The filter design and analysis tool in Matlab (fdatool) was used to design a digital low pass filter (FIR) the transfer characteristic for which is given in figure 7. Using this filter, it was found empirically that typical motion data of interest in the current study could be adequately captured over a 10Hz bandwidth. Using a digital filter was preferable to changing the accelerometer output bandwidth in hardware as it allows for a wider bandwidth if necessary that can be changed under software control i.e. the bandwidth can, in principle, be made adaptive to parameters such as playing tempo.

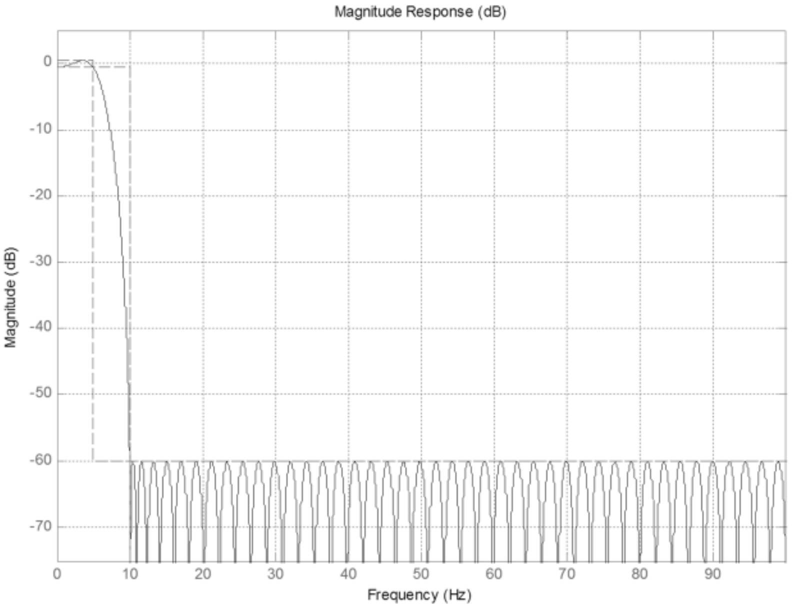


Figure 7 Transfer characteristic of a digital low pass filter designed in Matlab to remove electrical power line interference and physical movement jitter from the sampled accelerometer signals.

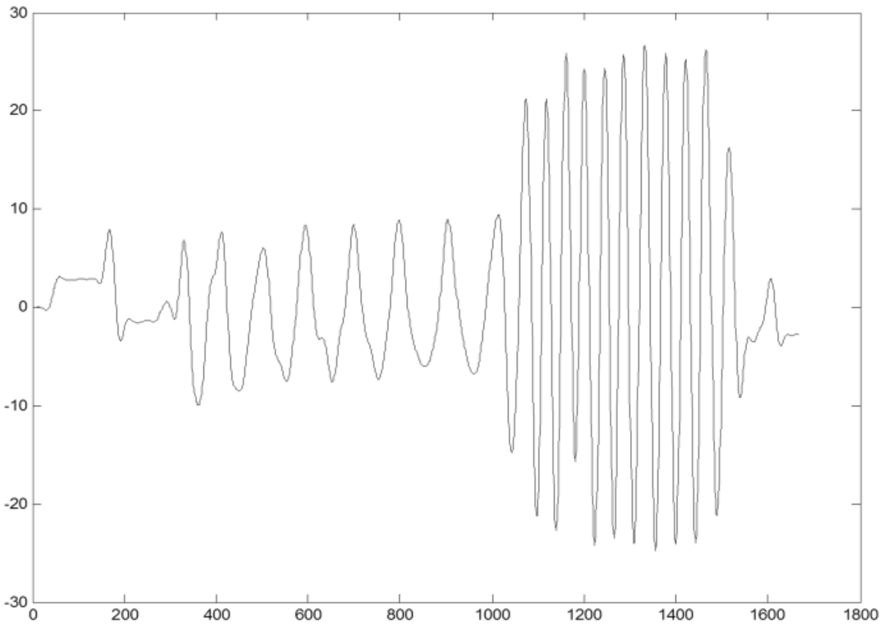


Figure 8 Digital filtered accelerometer output signal after low pass filtering in accordance with the transfer characteristic of figure 7. Here the horizontal axis is the measurement sample number and the vertical axis equates to ten times the equivalent g value so the +ve peak is at approximately 2.5g.

Figure 8 shows the resultant filtered output signal which no longer contains significant power interference/motion noise. As a consequence of this processing it was found that the algorithms used to detect specific threshold points in the signal were much more robust and reliable.

Note with reference to figure 8 the leading 300 samples and trailing 300 cycles are due to non-uniform arm movement during initialisation and completion of the experiment. The cyclic signal in the central portion of this graph shows eight cycles of moderate arm extension followed by 11 cycles of rapid arm extension. In this context (with the Z axis arranged perpendicular to the trunk) an arm extension is movement of the arm outward from the body and is followed by a retraction of the arm back towards the body (a ‘straight punch’ motion) with the arm at its normal elevation i.e. the lower arm at right angles to the upper arm at mid-chest elevation.

The next processing step was half wave rectification of the signal to isolate the positive half swings which were used to identify each unique trigger event. This is required since sound will only be triggered as the arm extends out from the body i.e. in the sense of hitting a drum. The rectification process was modelled in Matlab’s Simulink environment and as shown in figure 9 uses a simple thresholding process. The resultant rectified waveform is shown in Figure 10.

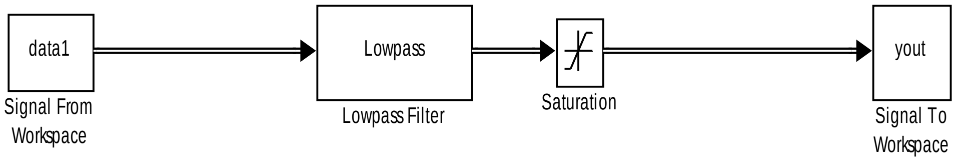


Figure 9 Development of processing model using Simulink – this model shows the low-pass filter and rectification steps.

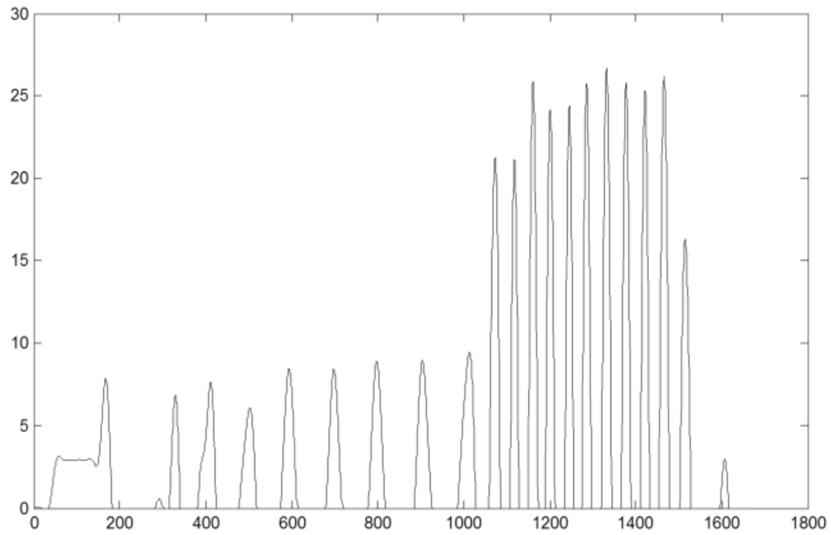
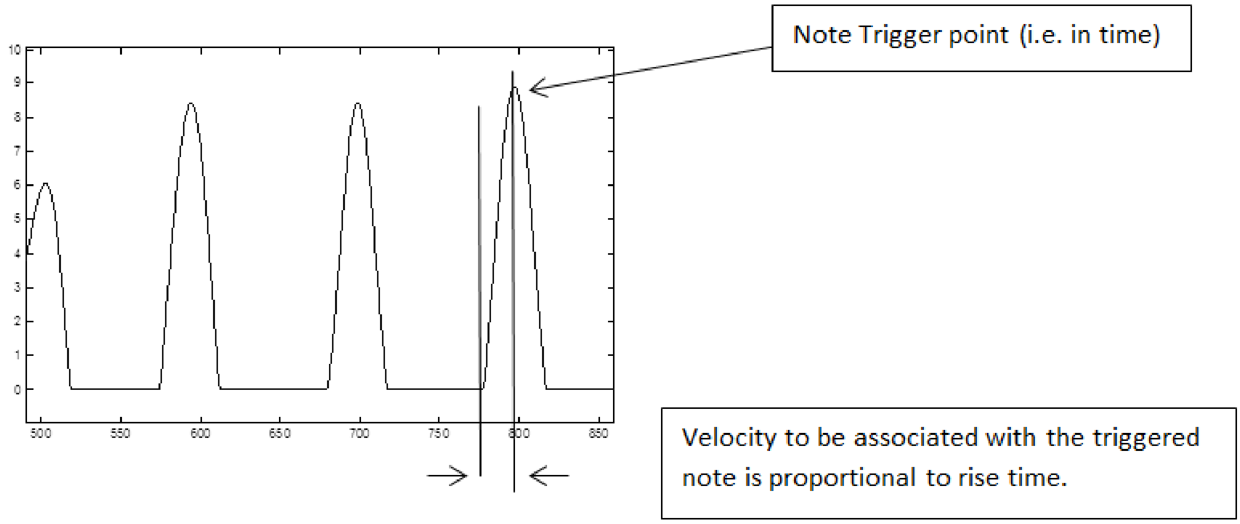


Figure 10 Low pass filtered and rectified data from one accelerometer channel

The processed data illustrated in figure 10 was used as input to develop the motion to sound mapping algorithm. The initial mapping algorithm was effectively a direct mapping of the motion event into a MIDI note trigger event and MIDI velocity value. The velocity value for the pre-selected note-on event was determined as the time taken from the start of the event to the peak. In motion terms this is the time taken from the start of movement to the final desired extension of the arm. Figure 11 shows how the MIDI velocity value is extracted from a motion event based on a sub-section of the data of figure 10.

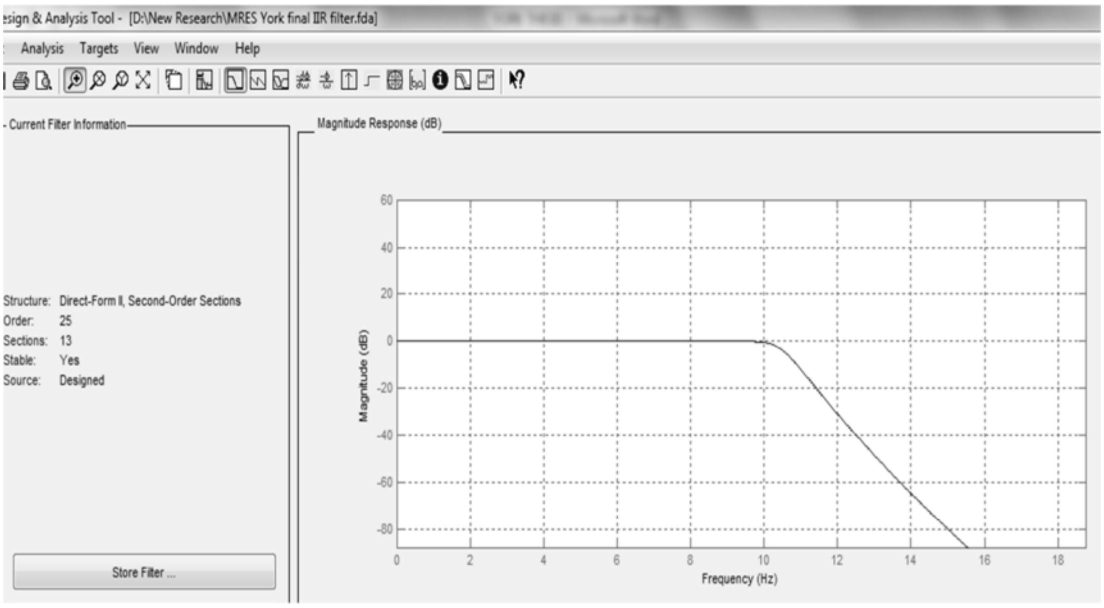


Note Trigger point (i.e. in time)

Velocity to be associated with the triggered note is proportional to rise time.

Figure 11 Extracting MIDI note velocity values from a gestural motion event via processed accelerometer data.

Using Simulink as illustrated in figure 9 proved useful for developing and evaluating various signal processing models. However the performance proved inadequate for real-time characterisation of the gestural capture system. Consequently, the signal capture, processing and mapping algorithms were translated for execution on the microcontroller interface unit previously presented in figure 5. The main challenge in translating the processing algorithm onto an embedded system was implementation of the digital filter. The filter used in the evaluation model was based on a Finite Impulse Response (FIR) implementation in order to ensure guaranteed stability. However, this filter had order 152 which after elimination of symmetrical coefficients still required 72 multiplication – accumulation operations (MACs) within the convolution algorithm. Although there is a single cycle multiplier in the microcontroller, most of the filter coefficients could not be held in the limited register set which affected performance. Subsequently, a revised design based on an Infinite Impulse response filter was designed for which the transfer characteristic is shown in Figure 12. This filter has less pass band ripple than the FIR design of figure 7 and can be implemented using far fewer resources i.e. 13 registers. Coefficients were rounded to 16 bit integer values to allow for a fixed point implementation. Overall this resulted in data throughput latency from system input to system output of 4 milliseconds compared to 24ms for the FIR based design. This latency was found to be adequate for slow to moderate equivalent tempos up to 120BPM.



Structure: Direct-Form II, Second-Order Sections
Order: 25
Sections: 13
Stable: Yes
Source: Designed

Figure 12 Final low pass filter design for system based on an efficient IIR implementation.

3.1.5 Bench calibration procedure

To calibrate the system response against known values of acceleration the tri-axial accelerometer was connected to the embedded interface unit of figure 2 via a shielded trailing umbilical wire of length 10 metres and relocated to the physics labs. Using standard ticker tape acceleration measuring equipment, it was possible to evaluate the system against known accelerations of calibrated masses. Each channel of the tri-axial accelerometer could be assessed by changing the plane of orientation of the sensor. Measurements of n=30 repeated cycles were made for different values of acceleration and the numerical average obtained in each case. The results showed that the system was accurate to within 3% of the reference acceleration value. It is important to note that these experiments were carried out at normal laboratory temperature (approximately 20 degrees C) with no separate temperature monitoring. This is a potentially a source of additional error as the accelerometer response is subject to a small amount of thermal drift through self-heating effects. For lab evaluation this drift is negligible however it would be more significant under the effect of stage lighting in a performance. Newer accelerometer sensors include an on board temperature sensor which can be used to compensate for this source of error. Modern designs also include internal calibration using an electrostatic force applied to the accelerometer reference mass. This would alleviate the need to develop an independent calibration rig for the system.

3.2 Evaluation of the sensor under real gestural motion

3.2.1 Test protocol

Following successful evaluation of the system under bench conditions the same set up was used to evaluate system performance from real subjects. A simple Velcro wristband was constructed to support the accelerometer and secure it to the participant's upper or lower limb. The wrist and ankles were found empirically to be the most suitable places for sensor attachment. For matched study purposes only right handed participants were recruited and the sensor was accordingly attached to the right wrist or ankle. When the sensor was mounted on the upper or lower limb the channel alignment mapping was X horizontal left-right arm motion, Y vertical up-down arm motion and Z inward-outward arm motion. With the sensor mounted on the ankle the X channel captures left – right leg swings, the Y axis captures lifting and lowering of the leg i.e. a stamping type motion and the Z axis captures inward – outward leg swing i.e. a kicking type motion.

Using the setup as described, representative data was obtained from volunteers recruited from the Creative Framework student cohort (BSc Music and Audio Technology) within the School of Design, Engineering and Computing at Bournemouth University. Ethical Approval was granted to enable up to twenty minutes of data to be obtained from a total of n=36 participants. The restriction on test duration was recommended as repeatability was being evaluated and potentially this would be affected by muscle fatigue. The duration of twenty minutes was set in accordance with normal clinical test practices. All participants were in good health and no exclusion criteria were defined.

Each participant was required to perform a set of pre-defined motions to simulate the following set of ten actions: Strumming; Bowing; Drum hit (vertical); Gong hit (horizontal); Hand clap; Punch; Leg kick; Leg Swing (horizontal); Leg lift (vertical); These motions were selected as typically generic forms of music/dance performance movement. White tape was used to set markers to delineate the range of motion for some of the tests. The buzzers on the instrumentation interface were also used to create aural prompts to guide the participant. Tests were performed in a standard instrumentation laboratory with benching used to provide support when necessary e.g. during leg kicks.

Each participant was allowed a five minute acclimatization period comprising three minutes of rest and two minutes of supervised practice to get accustomed to the various actions. This was followed by the actual test measurement phase in which each of the ten test movements was performed for thirty seconds. The measurement procedure was then repeated twice more so that each action was performed a total of three times. The full set of actions was completed in each cycle rather than the same action being repeated consecutively three times. The aim here was to reduce potential bias from fatigue and discomfort apparent in some actions, for example, hand clapping for one and half minutes. The replication of each test provided a basis for a simple measure of individual subject movement repeatability on a test by test basis.

A subset of five test subjects was used to assess reproducibility by repeating the measurement during the same lab time the following week. Through comparison of data it became apparent that accurate alignment of the accelerometer is critical for repeatable results. In subsequent tests this was done by carefully aligning a mark added to the wrist mounted sensor with the central line of the middle finger held out straight out. This was not possible on the lower limb where a make shift plumb line was used to align the ankle sensor with the vertical component of gravity.

Measured data was downloaded via an RS232 wired link to a laptop to overcome the problem of storing the data in the limited memory of the embedded microcontroller. Due to cost limitations only four sets of measurement sensor/instrumentation were constructed. One wrist and one ankle sensor were mounted per test subject. Whilst one test subject was being monitored the second sensor pair was affixed to the next participant who then commenced the acclimatization process. Data was transmitted as 8 bit binary with one start and one stop bit at 9600 Baud, no handshaking or error correction was employed. Data was buffered to RAM on the laptop and then stored as a comma separated value (.csv) file which was then available for analysis in Excel or Matlab.

3.2.2 Example test results from participant studies

When test data was plotted it became apparent that each time a test motion was performed there was a typical performer ‘settling’ period of up to five cycles before the motion became stable, this was apparent for the majority of tests. It was therefore decided to reject the first five cycles of each test, reducing the test set to 25 measured movements per test cycle. Figure 13 shows a typical result for a single strumming test which equates to a ‘strumming tempo’ of approximately 80 BPM. Plotting this data in the form of a histogram is useful as it shows important features such as clusters of reliable (+/-5ms) time and also the variance in motion timing that frequently occurred. For this example the mean value is 746ms with a symmetrical variation over a range of +/-33ms. This equates to a total variation of 8.8% across the data set for this test (25 measurements).

Figure 14 shows a similar analysis performed on a single data set for the X axis accelerometer data obtained during strumming. This shows similar features to the timing data in that some values of acceleration are matched to within 1mg however the spread about the mean of 997mg is +/- 10mg which equates to a variation of +/-2% across this data set.

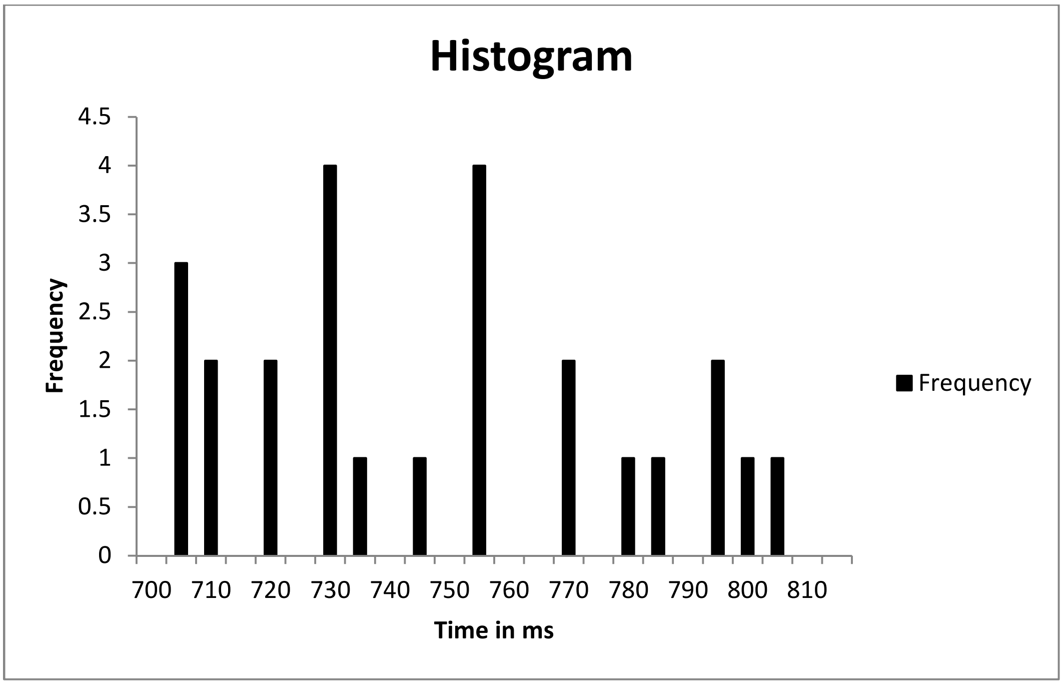


Figure 13 Variation in strumming times for a single test of 25 strumming cycles for a single participant.

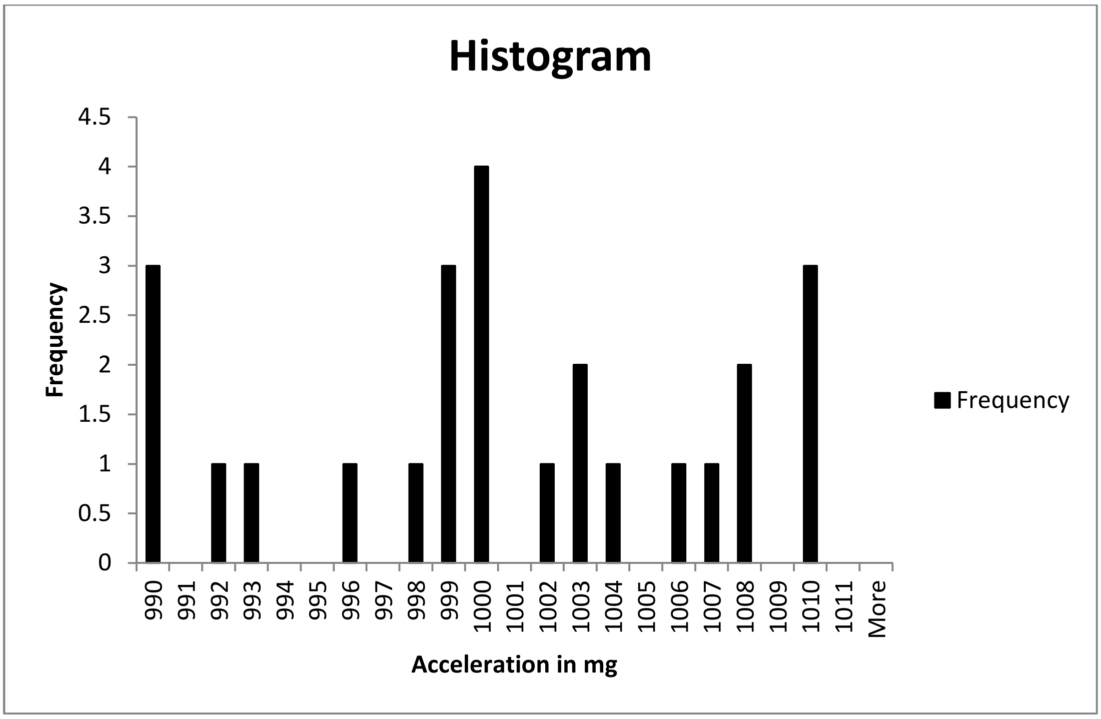


Figure 14 Variation in strumming acceleration values for a single test of 25 strumming cycles for a single participant.

When the full results from each participant were analysed in this way it was found that the overall spread of timing and acceleration data for an individual was within +/-10% of the mean for the individual. However, the variation in the means across all participants was +/- 28% of the group mean for timing values and +/- 16% for acceleration values.

The results of this investigation are consistent with the general perception that a person's ability to perform repetitive tasks is relatively good - however it takes time and practice to achieve a precisely repeatable response. Similarly more variation is expected from person to person as biological factors such as anatomical and physiological difference and physical experience are likely to affect measured values.

These finding introduced significant challenges for the development of the proposed system in which variation in response both within an individual and from performer to performer are required to be minimal. This is necessary to avoid the need to tune each gestural capture sensor unit to the individual. The requirement differs from most of the systems previously developed by other research groups in which variation is a key factor in achieving a versatile mapping to sound parameters, for example, it can support a wide range of oscillator frequencies in a synthesis engine.

To overcome the preceding critical problem various techniques were investigated to try and compensate for the normal variation observed in the performer motion responses. Initially an attempt was made to identify a minimal set of test procedures that would provide a small set of performer specific values to enable tuning of the output of the motion translation algorithms. This heuristic approach proved to be unsatisfactory and moreover, was impracticable since it would have required data for each performer to be obtained and stored in advance. Following discussion with the project supervisor (Professor David Howard, University of York) it was decided to investigate if artificial intelligence techniques could be used to reduce the uncertainty inherent in the measurement process.

3.2.3 Investigation of an artificial intelligence (AI) technique to reduce measurement uncertainty during gestural capture

Prior to the current project the author had little knowledge or experience of AI techniques and a number of texts which are listed in the bibliography were consulted. This resulted in purchase of the Neural Network toolbox for Matlab to provide a basis for practical exploration.

The first step in identifying a solution was to determine the precise nature of the type of AI problem involved. The basic difficulty was one of classifying the actual type of motion involved in order to be able to discriminate against the various types of motion which was required to ensure reliable translation into a corresponding sound/music event. If the music to be played by the ‘performing orchestra’ was known in advance then the solution would be trivial as the score could be provided to the translation system and pre-mapped to physical motions of the performer. The performer would then just be able to provide fine control over sound dynamics. This is effectively what is done in the majority of existing gestural composition systems. However, a key novel aspect of the proposed system is the ability for the ‘performing orchestra’ to be conducted in real-time with the conductor taking the dual role of composer. Furthermore, there is also the possibility of spontaneous un-conducted composition by the performers themselves. In both instances the musical content / structure is not known a priori.

The preceding problem of implementing a generalised (performer independent) gestural capture method became the main focus during the latter stage of the project. An automated classifier was required to enable different types of motion to be recognised and reliably translated into sound or music triggering events. The first stage of the classification process was a relatively trivial problem since six of the standard movements considered are generated by the upper limb and the remaining three by the lower limb. Since the data from each sensor could be uniquely labelled using an identification header in the transmitted data packet, it was only necessary to discriminate within the upper and lower types of limb motion. Furthermore, analysis of the typical waveforms produced under each form of motion revealed that classification of the three types of lower limb movement is also relatively straightforward since there is a direct one to one relationship between the physical motion and the accelerometer output on the corresponding axis. For example, a stamping of the leg produces a dominant signal on the Z axis which is easily discriminated from the X and Y output signals by a comparator with a coarse threshold. Conversely, several of the types of arm movement were difficult to distinguish as it was found that contributions on all three axis were often comparable as the arm tended to move more freely within a three dimensional space rather than a two dimensional plane. This introduced uncertainty into the discrimination process. To solve this classification problem for the arm it was decided to investigate the use of an artificial neural network (ANN). A conceptual model for the ANN that was envisaged is illustrated in figure 15.

With reference to figure 15, the classifier system was first trained by supplying a set of reference patterns acquired from performers. These patterns were manually mapped by the ‘composer’ onto desired sound or music event targets. During training validation mode input patterns were classified by the system and the system response was compared with the desired response, the error was then minimised by adjusting the weights. The desired result was a classification system that could respond to a wider class of input signal patterns than was used for training. Using this approach it was possible to create a generalised classifier without having to capture an exhaustive set of all possible input signals.

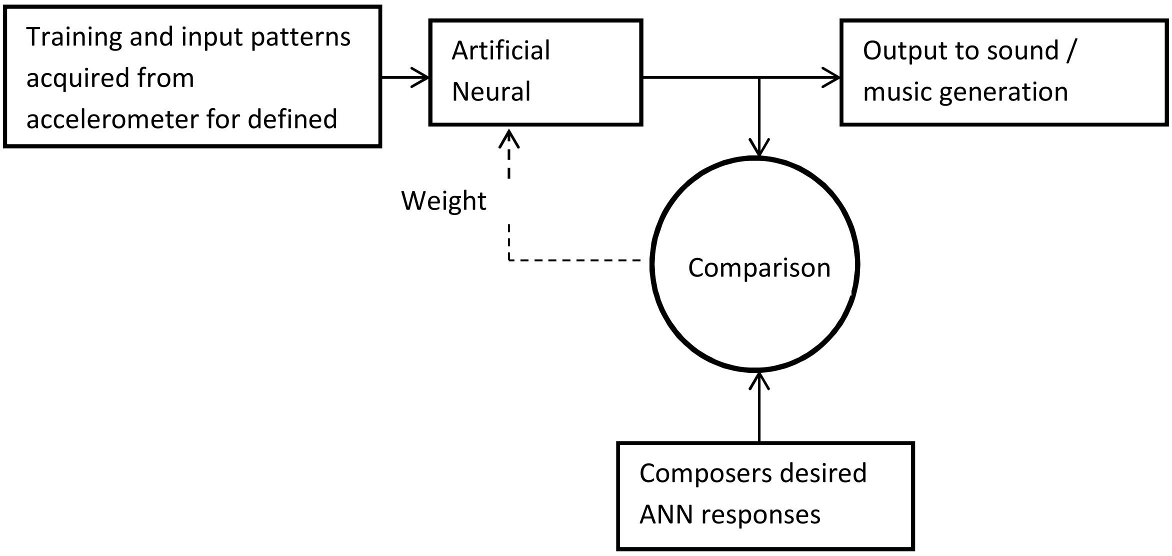

Training and input patterns acquired from accelerometer for defined
Artificial Neural
Output to sound / music generation
Weight
Comparison
Composers desired ANN responses

Figure 15 A conceptual model of the Artificial Neural Network approach used to classify a performers arm movements.

The data set used to develop, train and validate the ANN classifier was the data obtained in the evaluation of the gestural capture system as discussed earlier in this dissertation. The total available data set comprised a total of 750 motion events. Each of these events was characterised by six variables, the peak of the accelerometer response on each of the three output channels (X,Y,Z) and the corresponding event duration measured as the rise time from initiation of the event taken as 10% of the peak value. This threshold was used as the baseline is not always stable at zero due to the static component of the output (sensitivity to gravity). However, this static data is used as part of the valid signal data. The numeric values in the data set were normalised over a range of values from zero to unity to be compatible with the input range of the sigmoid activation function used in the implementation of the ANN.

Figure 16 illustrates the structure of the intended classifier in terms of the input data variables and output signals for an arm sensor. The input data variables were labelled as Xa, Xd, Ya, Yd, Za, Zd where a is the amplitude value and d the duration value.

Figure 17 Shows a section of the Excel database used to hold the data values. The value for each type of movement was recorded in a separate table. The table records a participant number and the test number for each repetition of each movement test. The table of figure 17 shows five data values for each of four participants. The amplitude values in these tables are based on an eight bit conversion over a 3V reference range giving a conversion factor of 12mV per bit. These values were then normalised over the required input range of 0 to 1. The timing values are also normalised.

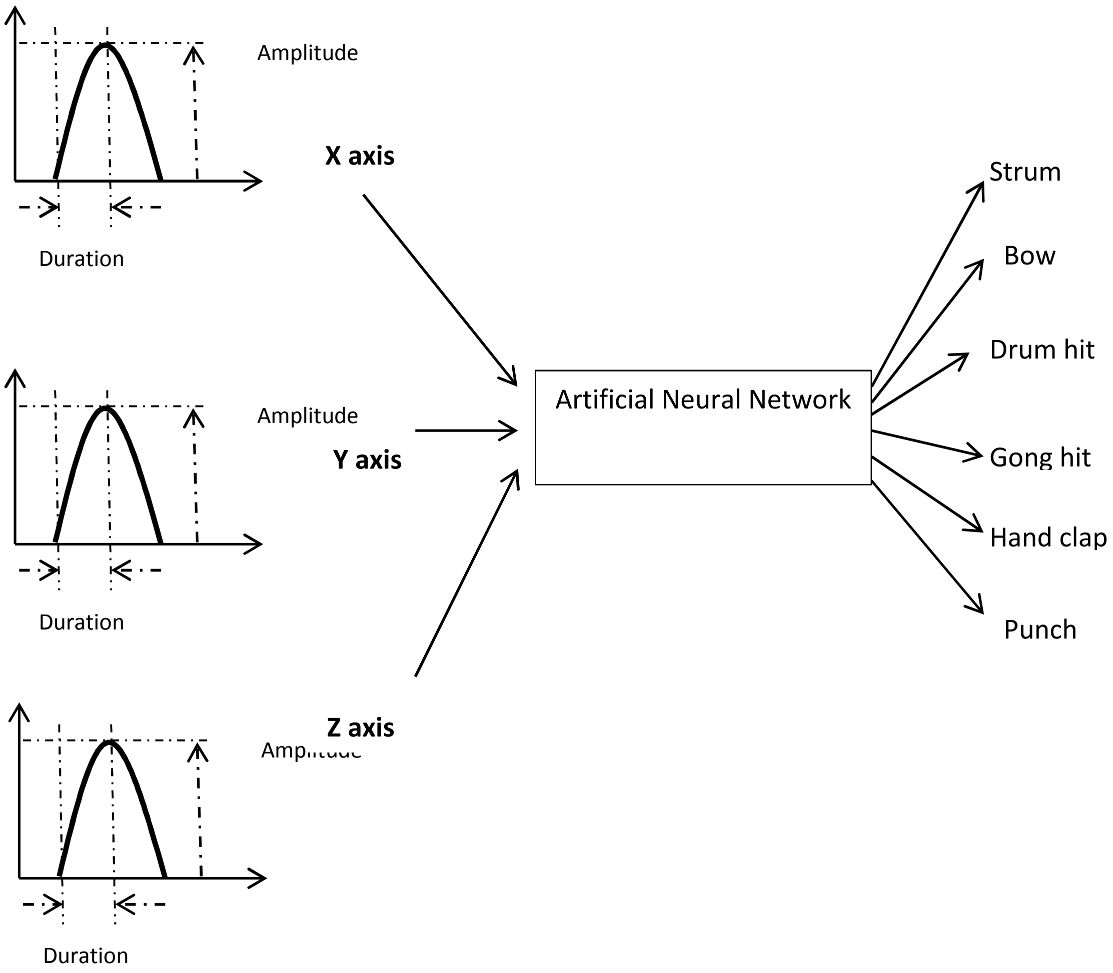
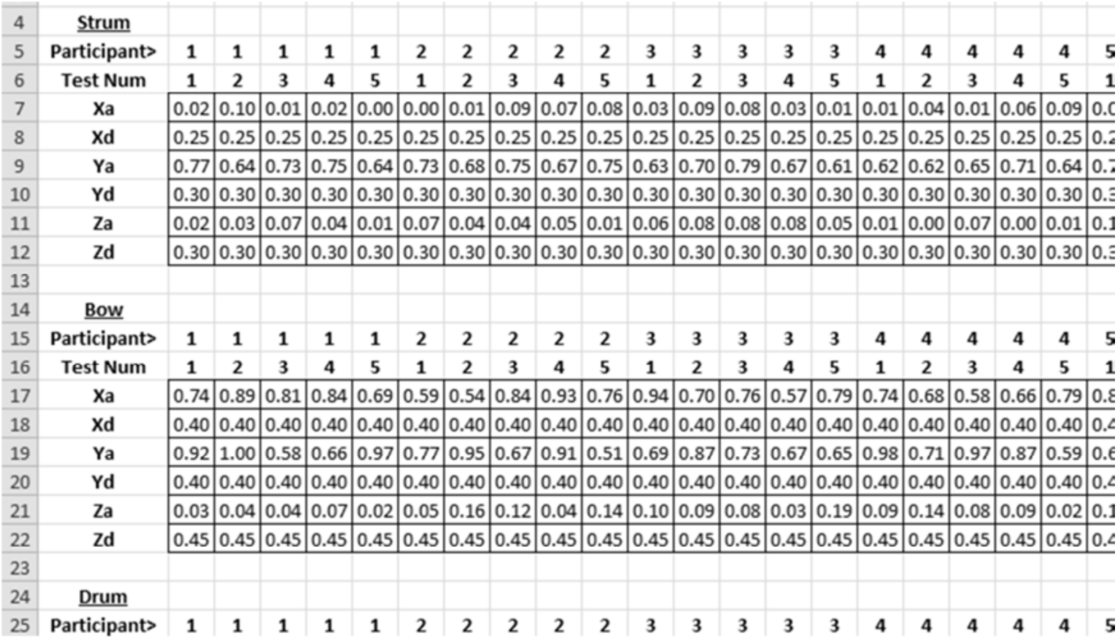


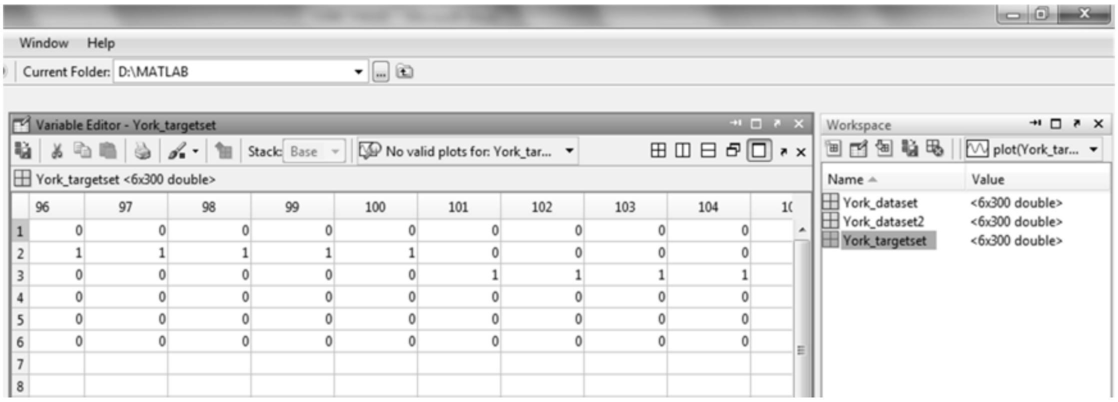
Figure 16 The ANN classifier takes six input variables (two per axis) and classifies this data into one of six outputs which can then be used to activate a pre-defined corresponding musical event.



4	Strum																					
5	Participant>	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5
6	Test Num	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1
7	Xa	0.02	0.10	0.01	0.02	0.00	0.00	0.01	0.09	0.07	0.08	0.03	0.09	0.08	0.03	0.01	0.01	0.04	0.01	0.06	0.09	0.0
8	Xd	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.2
9	Ya	0.77	0.64	0.73	0.75	0.64	0.73	0.68	0.75	0.67	0.75	0.63	0.70	0.79	0.67	0.61	0.62	0.62	0.65	0.71	0.64	0.7
10	Yd	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.3
11	Za	0.02	0.03	0.07	0.04	0.01	0.07	0.04	0.04	0.05	0.01	0.06	0.08	0.08	0.08	0.05	0.01	0.00	0.07	0.00	0.01	0.1
12	Zd	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.3
13																						
14	Bow																					
15	Participant>	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5
16	Test Num	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1
17	Xa	0.74	0.89	0.81	0.84	0.69	0.59	0.54	0.84	0.93	0.76	0.94	0.70	0.76	0.57	0.79	0.74	0.68	0.58	0.66	0.79	0.8
18	Xd	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.4
19	Ya	0.92	1.00	0.58	0.66	0.97	0.77	0.95	0.67	0.91	0.51	0.69	0.87	0.73	0.67	0.65	0.98	0.71	0.97	0.87	0.59	0.6
20	Yd	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.4
21	Za	0.03	0.04	0.04	0.07	0.02	0.05	0.16	0.12	0.04	0.14	0.10	0.09	0.08	0.03	0.19	0.09	0.14	0.08	0.09	0.02	0.1
22	Zd	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.45	0.4
23																						
24	Drum																					
25	Participant>	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5

Figure 17 Excel database of measurement values for input into the neural network

Figure 18 is a screenshot from Matlab after import of two datasets of measured values. For evaluation purposes different subsets of the full data set were created, each reflecting different values dependent on the pre-processing (thresholding/filter) values used in the signal processing system at capture time. Also shown in the main part of the window is the set of target values. These target values determine which set of sample variables should be mapped onto which output by the neural network. For the example shown, sample sets 96 through to 100 map onto output two which corresponds to a bowing motion whereas sets 101 through to 104 map onto output 3 which corresponds to a drum motion.


Window
Help
Variable Editor - York_targetset
Stack: Base
No valid plots for: York_tar...
Workspace
plot(York_tar...

Current Folder: D:\MATLAB

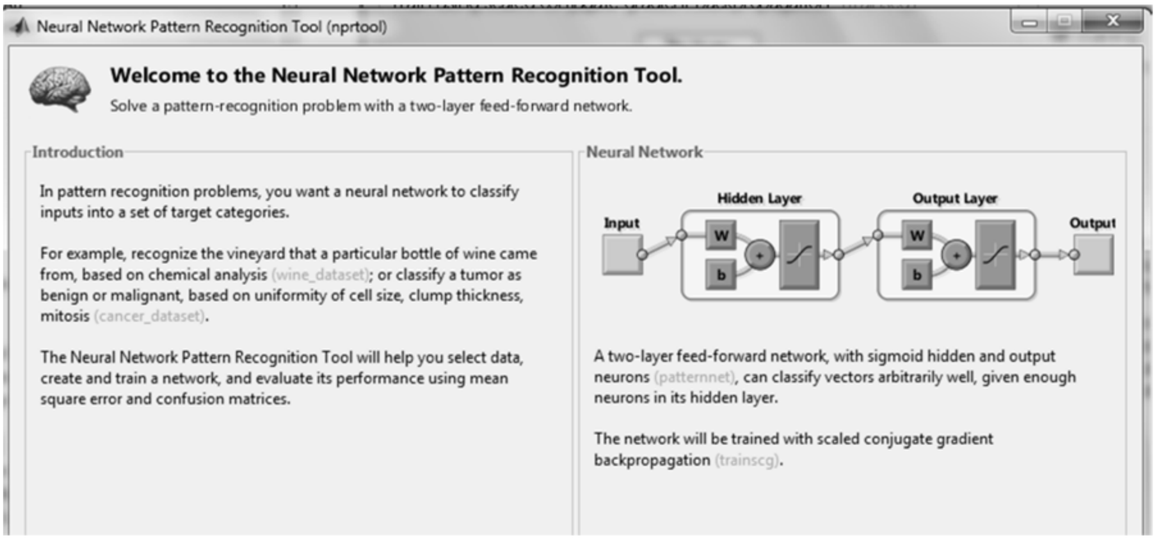
York_targetset <6x300 double>

	96	97	98	99	100	101	102	103	104	1(
1	0	0	0	0	0	0	0	0	0	
2	1	1	1	1	1	0	0	0	0	
3	0	0	0	0	0	1	1	1	1	
4	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	
7										
8										

Name	Value
York_dataset	<6x300 double>
York_dataset2	<6x300 double>
York_targetset	<6x300 double>

Figure 18 Screen shot of ANN input data set in Matlab after import from Excel.

Using these datasets the neural network was implemented in Matlab using the Neural Network Pattern Recognition Tool (nprtool). The generic architecture of the ANN is shown in figure 19.



Neural Network Pattern Recognition Tool (nprtool)

Welcome to the Neural Network Pattern Recognition Tool.

Solve a pattern-recognition problem with a two-layer feed-forward network.

Introduction

In pattern recognition problems, you want a neural network to classify inputs into a set of target categories.

For example, recognize the vineyard that a particular bottle of wine came from, based on chemical analysis (wine_dataset); or classify a tumor as benign or malignant, based on uniformity of cell size, clump thickness, mitosis (cancer_dataset).

The Neural Network Pattern Recognition Tool will help you select data, create and train a network, and evaluate its performance using mean square error and confusion matrices.

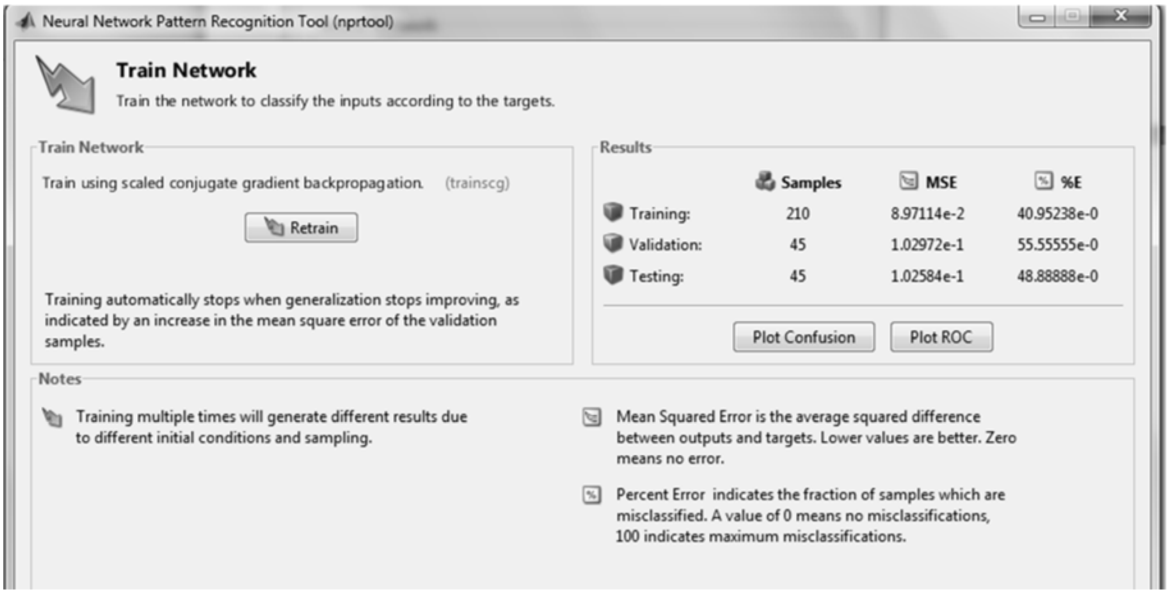
Neural Network

A two-layer feed-forward network, with sigmoid hidden and output neurons (patternnet), can classify vectors arbitrarily well, given enough neurons in its hidden layer.

The network will be trained with scaled conjugate gradient backpropagation (trainscg).

Figure 19 The architecture of the ANN is based on a two layer model with sigmoid activation functions on each neuron

Figure 20 Shows the result of creating an ANN classifier using a subset of 300 samples sets. Seventy percent of the set were used for training and fifteen percent each for testing and validation. In this example the number of input and output neurons was set to a figure of 10. Figure 21 shows a tabular summary of the classification success rate for the current example.



Neural Network Pattern Recognition Tool (nprtool)

Train Network

Train the network to classify the inputs according to the targets.

Train Network

Train using scaled conjugate gradient backpropagation. (trainscg)

Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples.

Results

	Samples	MSE	%E
Training:	210	8.97114e-2	40.95238e-0
Validation:	45	1.02972e-1	55.55555e-0
Testing:	45	1.02584e-1	48.88888e-0

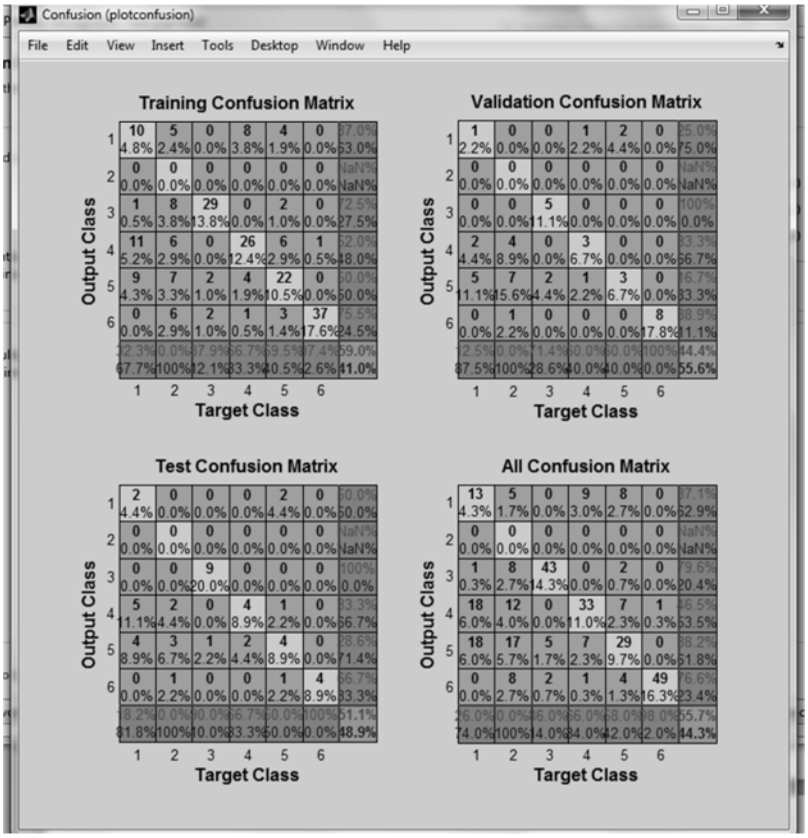
Notes

Training multiple times will generate different results due to different initial conditions and sampling.

Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

Percent Error indicates the fraction of samples which are misclassified. A value of 0 means no misclassifications, 100 indicates maximum misclassifications.

Figure 20 Post training results for an ANN using a database subset of 300 sample sets.



Training Confusion Matrix

Output Class	1	2	3	4	5	6	
1	10 4.8%	5 2.4%	0 0.0%	8 3.8%	4 1.9%	0 0.0%	37.0% 63.0%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
3	1 0.5%	8 3.8%	29 13.8%	0 0.0%	2 1.0%	0 0.0%	72.5% 27.5%
4	11 5.2%	6 2.9%	0 0.0%	26 12.4%	6 2.9%	1 0.5%	52.0% 48.0%
5	9 4.3%	7 3.3%	2 1.0%	4 1.9%	22 10.5%	0 0.0%	50.0% 50.0%
6	0 0.0%	6 2.9%	2 1.0%	1 0.5%	3 1.4%	37 17.6%	75.5% 24.5%
	32.3% 67.7%	0.0% 100%	87.9% 12.1%	66.7% 33.3%	59.5% 40.5%	97.4% 2.6%	59.0% 41.0%

Target Class

Validation Confusion Matrix

Output Class	1	2	3	4	5	6	
1	1 2.2%	0 0.0%	0 0.0%	1 2.2%	2 4.4%	0 0.0%	25.0% 75.0%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
3	0 0.0%	0 0.0%	5 11.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	2 4.4%	4 8.9%	0 0.0%	3 6.7%	0 0.0%	0 0.0%	33.3% 66.7%
5	5 11.1%	7 15.6%	2 4.4%	1 2.2%	3 6.7%	0 0.0%	16.7% 83.3%
6	0 0.0%	1 2.2%	0 0.0%	0 0.0%	0 0.0%	8 17.8%	88.9% 11.1%
	12.5% 87.5%	0.0% 100%	71.4% 28.6%	60.0% 40.0%	60.0% 40.0%	100% 0.0%	44.4% 55.6%

Target Class

Test Confusion Matrix

Output Class	1	2	3	4	5	6	
1	2 4.4%	0 0.0%	0 0.0%	0 0.0%	2 4.4%	0 0.0%	50.0% 50.0%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
3	0 0.0%	0 0.0%	9 20.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	5 11.1%	2 4.4%	0 0.0%	4 8.9%	1 2.2%	0 0.0%	33.3% 66.7%
5	4 8.9%	3 6.7%	1 2.2%	2 4.4%	4 8.9%	0 0.0%	28.6% 71.4%
6	0 0.0%	1 2.2%	0 0.0%	0 0.0%	1 2.2%	4 8.9%	66.7% 33.3%
	18.2% 81.8%	0.0% 100%	90.0% 10.0%	66.7% 33.3%	50.0% 50.0%	100% 0.0%	51.1% 48.9%

Target Class

All Confusion Matrix

Output Class	1	2	3	4	5	6	
1	13 4.3%	5 1.7%	0 0.0%	9 3.0%	8 2.7%	0 0.0%	37.1% 62.9%
2	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
3	1 0.3%	8 2.7%	43 14.3%	0 0.0%	2 0.7%	0 0.0%	79.6% 20.4%
4	18 6.0%	12 4.0%	0 0.0%	33 11.0%	7 2.3%	1 0.3%	46.5% 53.5%
5	18 6.0%	17 5.7%	5 1.7%	7 2.3%	29 9.7%	0 0.0%	38.2% 61.8%
6	0 0.0%	8 2.7%	2 0.7%	1 0.3%	4 1.3%	49 16.3%	76.6% 23.4%
	26.0% 74.0%	0.0% 100%	86.0% 14.0%	66.0% 34.0%	58.0% 42.0%	98.0% 2.0%	55.7% 44.3%

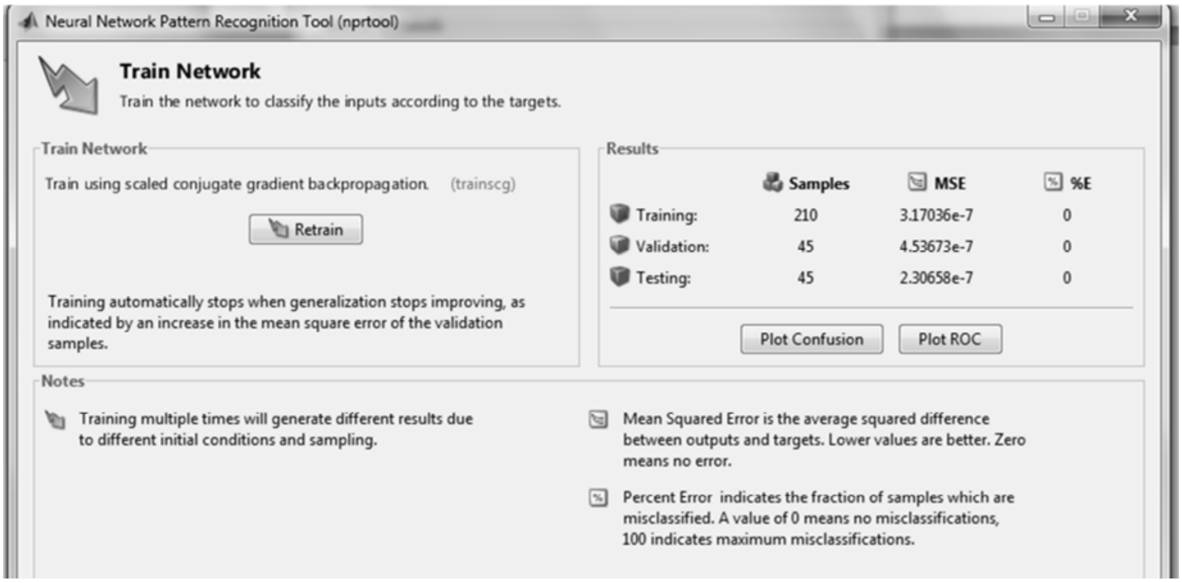
Target Class

Figure 21 Tabular representation of the classification test

Results of the initial tests for the ANN designs were not encouraging as summarised in figures 20 and 21. The classification error during testing is unacceptably high approaching 49%. Analysis of the confusion matrices showed a fairly uniform distribution of misclassification across the classes. A large number of such tests were performed on various subsets of the overall data and for differences in pre-processing parameters. The resultant error classification ranged from 24% to 56%.

In order to improve the performance of the ANN the raw data was analysed graphically and statistically to identify factors such as variance in the raw data, skew etc. As a result of these analyses it was apparent that the spread in the timing values was unacceptably large. An experiment was then conducted in which the buzzers on the interface unit were used to create an audible beat tempo and new data collected from five participants. This greatly reduced the spread in the measured timing values. Consequently, a revised full data set was collected from the new 2011 student cohort of n=32 students. Figures 22 and 23 show the results similar for those of figures 20 and 21 based on a subset of the revised dataset using a reference beat of 80 BPM. These results which represent the best obtained from many permutations of the sample set demonstrate that a classification success rate of

100% is achievable. The maximum classification error that occurred in the revised data set was 13%.



Train Network

Train the network to classify the inputs according to the targets.

Train Network

Train using scaled conjugate gradient backpropagation. (trainscg)

Training automatically stops when generalization stops improving, as indicated by an increase in the mean square error of the validation samples.

Results

	Samples	MSE	%E
Training:	210	3.17036e-7	0
Validation:	45	4.53673e-7	0
Testing:	45	2.30658e-7	0

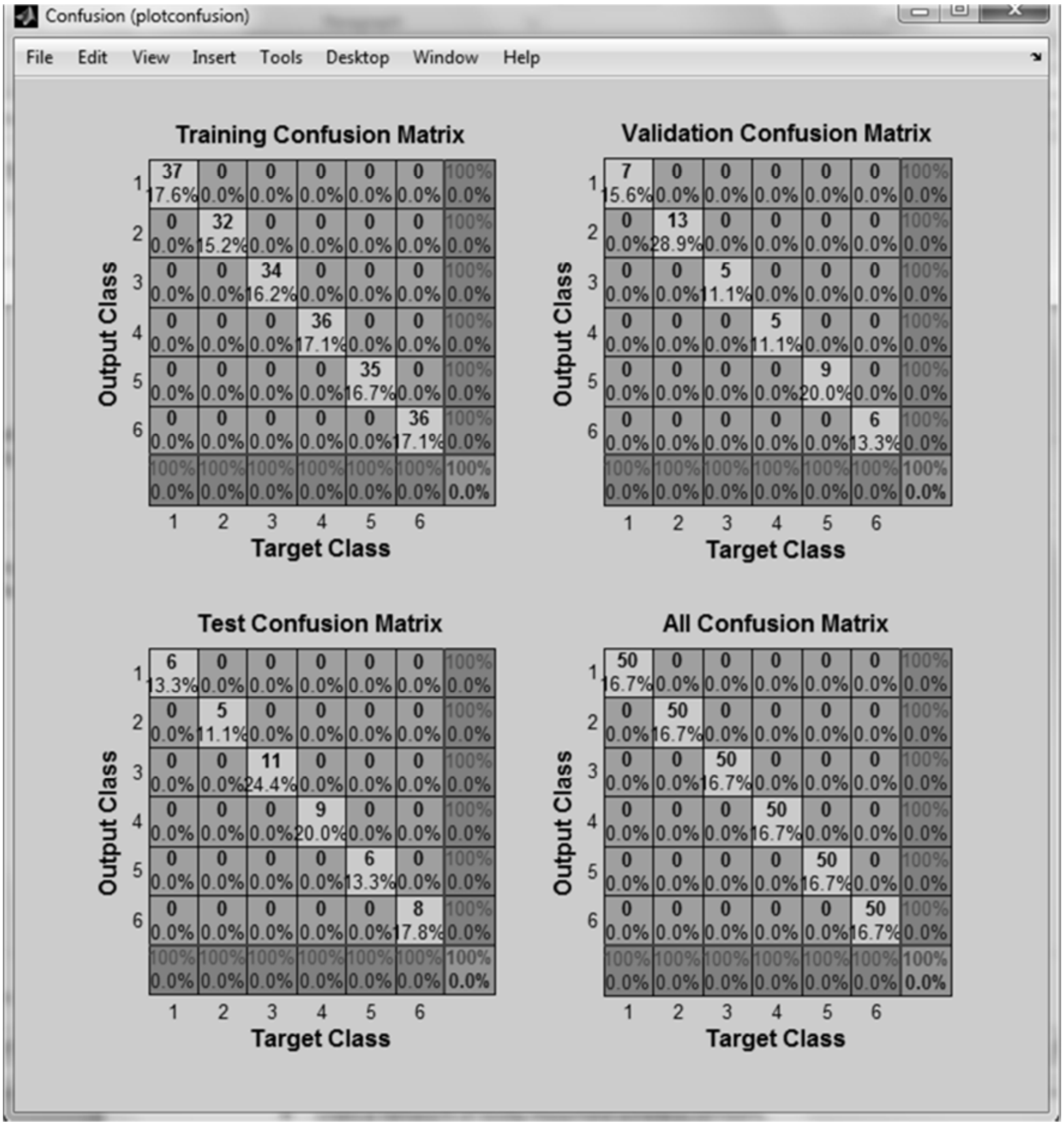
Notes

Training multiple times will generate different results due to different initial conditions and sampling.

Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

Percent Error indicates the fraction of samples which are misclassified. A value of 0 means no misclassifications, 100 indicates maximum misclassifications.

Figure 22 Best case ANN training result using a dataset measured using a reference beat tempo to aid participants.


Confusion (plotconfusion)
File Edit View Insert Tools Desktop Window Help

Training Confusion Matrix

Output Class \ Target Class	1	2	3	4	5	6	
1	37 17.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	32 15.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	34 16.2%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	36 17.1%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	35 16.7%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	36 17.1%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Validation Confusion Matrix

Output Class \ Target Class	1	2	3	4	5	6	
1	7 15.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	13 28.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	5 11.1%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	5 11.1%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	9 20.0%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 13.3%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Test Confusion Matrix

Output Class \ Target Class	1	2	3	4	5	6	
1	6 13.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	5 11.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	11 24.4%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	9 20.0%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 13.3%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	8 17.8%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

All Confusion Matrix

Output Class \ Target Class	1	2	3	4	5	6	
1	50 16.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	50 16.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	50 16.7%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	50 16.7%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

Figure 23 Confusion matrices for the best case ANN solution showing 100% classification rate.

At the present time work is on-going to convert the neural network designed in Matlab into C/C++ code so that the model can be executed in real time. It is encouraging that the complexity of the ANN is such that it could be realistically implemented in software to run locally on each local sensor processor. This would significantly reduce bandwidth requirement for the communication channel and help to reduce overall system latency.

3.3 Final hardware design for the gestural capture system

The development of the system discussed so far was based upon a sensor interface of dimensions 100mm x 160mm implemented using through-hole components. This section discusses the further development of the gestural capture hardware into a wireless sensor network small enough to be worn by a performer. A conceptual diagram of the final system has already been given in figure 4. The wireless system was designed based upon a star topology with four end devices surrounding a single coordinator node. This implements a local body network which itself becomes part of a larger star network with the principal coordinator located at the central PC which acts as the 'composer' node. The end devices on the performer are located at the wrist and ankles and are given corresponding Node identifiers: Left Wrist, Right Wrist, Left Leg, Right Leg. Configuration of the network is nearly fully automatic under control of the Zigbee protocol. At start up the individual performers' coordinator issues 16-bit addresses to the four end devices comprising it's local network. The central coordinator communicates with the performer coordinators using the unique PAN address. The Hayes compatible AT command set was used to set up and configure specific operational parameters of the XBee modules such as baud rate. The simplify design and synchronisation a simple round robin polling of each sensor node was envisaged rather than an event driven technique.

By replacing the wired RS232 link of figure 4 it was possible to demonstrate the implementation of the basic communication system. Phantom input signals were generated using potentiometers in place of the accelerometers to simplify testing. The potentiometers were connected across the supply to produce a varable 0 – 3 Volt signal at the wiper that was connected to the sampling input pins. Initially these signals were connected to the analogue input channels of the XBee modules which have their own on board processor. However, the onboard processor has limited IO and memory and was clearly inadequate to act as the processing host for the sensor sampling subsystem, classifier and communications system. The limited code memory available in the ATMEGA168 processor used in the first prototype system also became a problem and a decision was made to switch to the ATMEGA328 device which is pin compatible and offers twice the code space providing 32Kbytes. The remainder of the design was unchanged from that given in figure 5 except that surface mount devices (SMD) were used throughout. All passive components used a 0805 SMD foot print. This allowed for a final printed circuit board square geometry of 45mm x 45mm. The final size allowed mounting on a commercial wrist band available for the Apple iPod nano (Griffin). At the time of writing construction of the final hardware (wireless version) is delayed pending arrival of a new SMD pick and place machine with in the School of Design Engineering and Computing and Bournemouth University. This is due to be operational in the spring term of 2012. For this reason the remaining development time on the project has mainly focussed on exploring the theoretical issues and challenges in using the system for composition. To support this some further practical evaluation was done using the single node wired system with the gestural capture system tuned specifically to the author's movements so that it was not necessary to have an implementation of the ANN classifier. This allowed the necessary code to be contained within the 16Kbyte code space available on the ATMega168 microcontroller.

3.4 Sound control using a single sensor node

Figure 24 is a photograph of the equipment configuration used for development of composition software using a single sensor node. The ADXL335 accelerometer was connected to the sensor interface card described earlier. The interface provides a direct MIDI output connection to a Yamaha TG300 tone generator allowing simple mapping algorithms on the microcontroller to trigger musical events. To allow for a more complex composition system the interface also supports a wired RS232 serial link. The serial link was used to pass sensor data into the Max/MSP application (Cycling 74 version 5.1) and to receive data to configure the interface. The interface includes two buzzers seen in the foreground and three LEDs which can be used to conduct performer movement. As a whole this system provided a very flexible platform for development of system software and composition experiments. The M-Audio sound card is used to provide a low latency audio interface to Max/MSP.

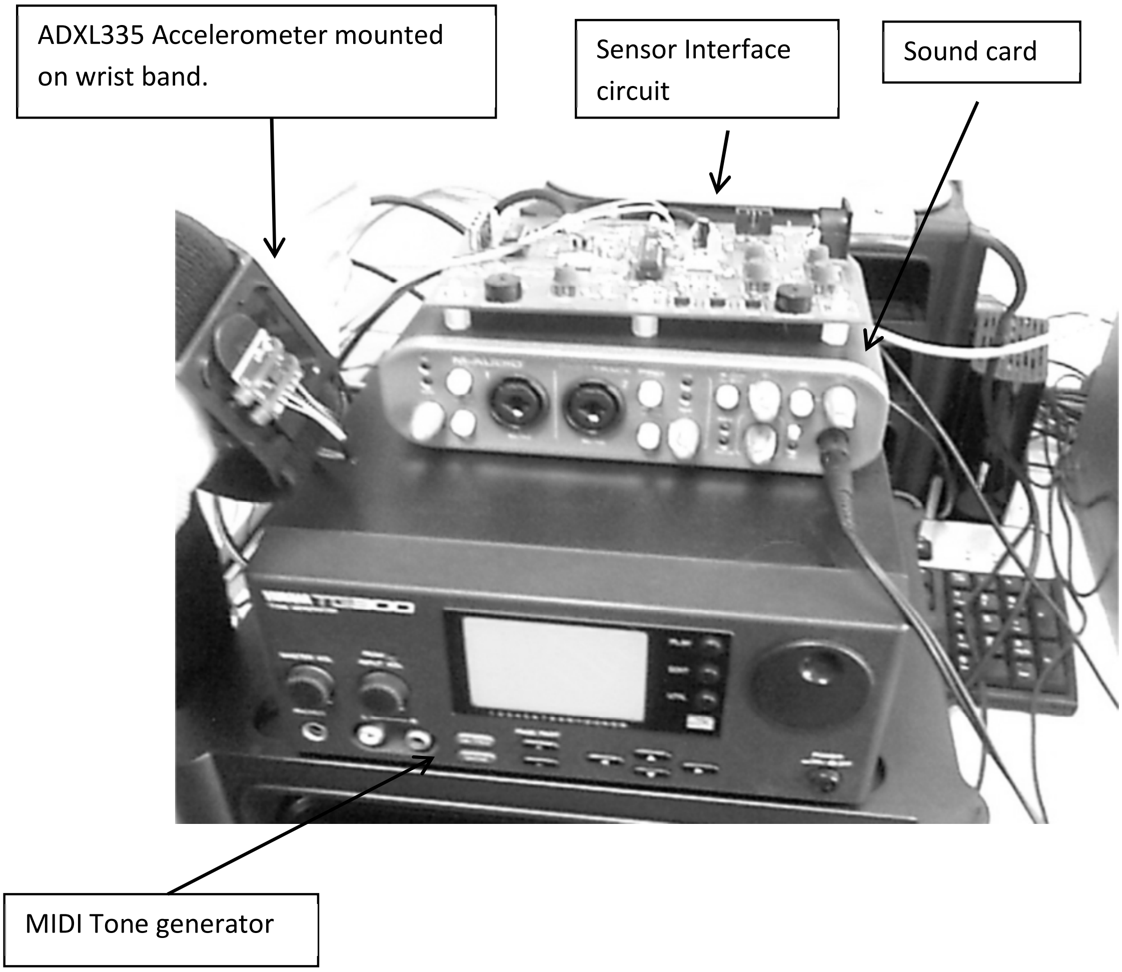


Figure 24 Photograph showing experimental setup for single sensor node evaluation

Using the experimental setup of figure 24 many basic experiments were performed to evaluate useful methods of controlling sound generation parameters based on characteristics of the accelerometer signals under different performer motion.

An obvious mapping of sensor output is magnitude of acceleration to MIDI note velocity. The main challenge was in efficiently detecting the peak of the acceleration signal and generating a corresponding note-on and velocity value with minimal latency. It was found that if the delay was greater than 20ms the loss of synchronisation between the visual event and corresponding sound generation became noticeable which is conceptually similar to a video lip sync error.

Another straightforward mapping is in stereo panning where the magnitude of the acceleration signal for the X axis was used to set the stereo position between a stereo pair of speakers. For this application the full bipolar output range of the signal was sampled so that a positive acceleration mapped to a right pan and a negative acceleration to a left pan. This worked particularly well for continuous panning in the form of a ‘ping-pong’ effect.

Other successful control mappings included pitch bend, modulation, glissando and echo time. However, it proved difficult to control many types of MIDI parameter without very controlled and repetitive arm movements. Similarly some types of time domain audio effect, such as tremolo (amplitude modulation), and vibrato (frequency modulation) were relatively easy to apply and provide an aurally satisfying response whereas many do not. These findings are in accord with many of the gestural capture systems demonstrated on line and there seems to be a fundamental limit to the generalisation of accelerometer based systems as universal controllers of sound parameters.

3.4.1 Composition using a single sensor node

For the reasons outlined in the preceding paragraph the main aim of the current project was to move away from sensor driven sound effects toward a more general form of sensor based composition. It proved to be possible to perform basic musical composition using a single sensor node. Figure 25 shows a Max/MSP patch developed to explore this concept. This patch is effectively a two dimensional sixteen bar step sequencer with note count along the horizontal axis and pitch along the vertical axis. The intersection points are LEDs which are sequentially illuminated in a cyclic scanning arrangement from top left to bottom right.

Along the top of the patch is a separate row of LEDs which act as visual triggers to enable the user to coordinate their motion with the sequencer action. It was found to be essential to set the timing signal to the trigger indicators at a slight leading phase to the sequencer timing signal to enable the user to return their arm to the détente position before triggering an event. The optimum timing relationship is tempo dependent and was determined empirically by adjustment of the trigger level control. A reasonable period of familiarisation with the technique is still required to enable the user to use the patch competently. Furthermore, at tempos above 100BPM it proved physically difficult to coordinate arm movement with the sequencer timing.

In operation the user can use the visual display of note, note position and trigger to set a selected note with a note velocity set proportional to the magnitude of acceleration on a selected output from the sensor. This was found to work well in practice. It proved difficult in general however for a user to successfully control other sound parameters mapped to the other accelerometer output channels at the same time. Unless the tempo is set below 40BPM there is inadequate time for a user to adequately conceptualise and execute a three dimensional movement within the time available before the next note becomes active.

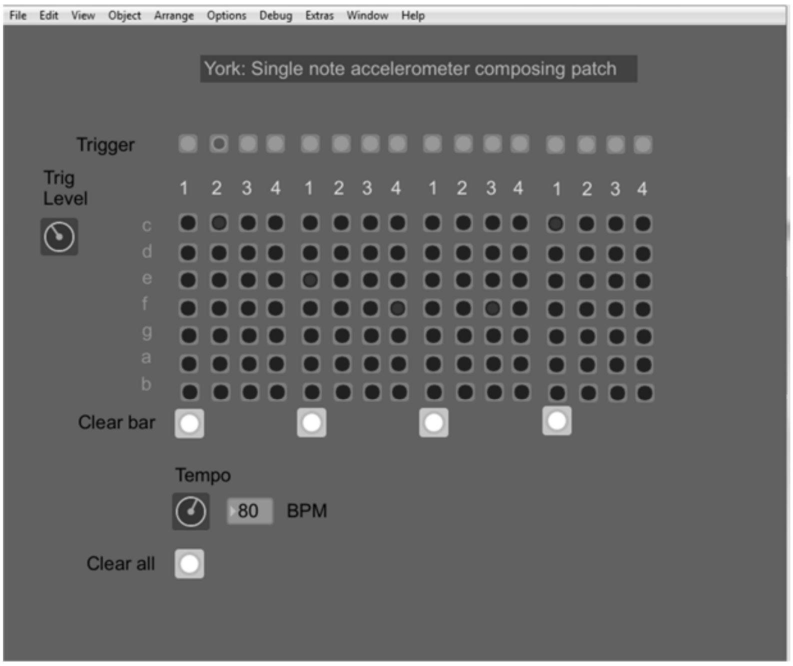

File Edit View Object Arrange Options Debug Extras Window Help
York: Single note accelerometer composing patch
Trigger
Trig Level
1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
c
d
e
f
g
a
b
Clear bar
Tempo
80 BPM
Clear all

Figure 25 A Max/MSP patch developed to explore the concept of single sensor node composition.

3.5 Scaling up for multiperformer composition

A key aim of this project was to overcome the aforementioned problem of overloading a performer with the burden of creating complex movements in order to achieve good control of musical parameters. Thus figure 26 shows one proposed solution of how in this case three performers could be coordinated using visual cues that are continuously updated. The information in the diagram would be displayed to the performers to indicate the next required movement. This is the the same type of cueing information as used in dance type games on consoles. However, here the music is being generated by the triggering information from the sensors on each limb. Furthermore, the triggering is not simply off/on there is a degree of performer control over sound parameters depending on the dynamics with which the movement is performed.

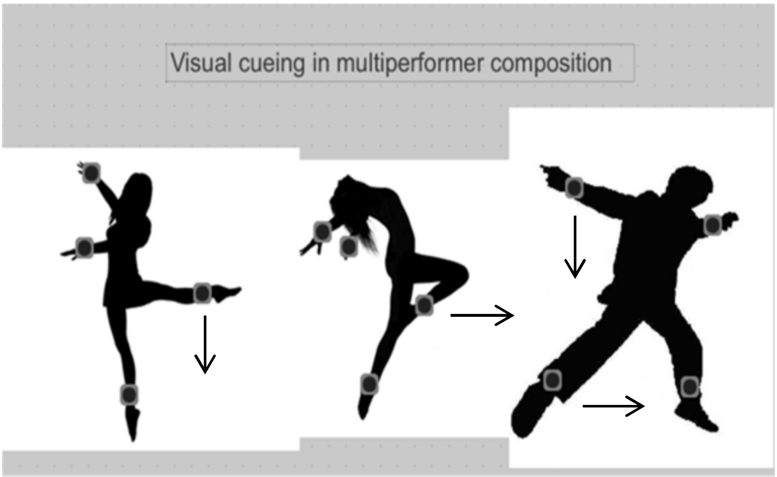


Figure 26 Concept of controlled multi-performer composition

With reference to figure 26 we can imagine a sound mapping in which the sensing nodes on the two female performers each map to a C major key e.g. C3 to C4 with a piano timbre and the four nodes on the male performer mapping to percussion e.g. bass, snare, tom, and cymbal. As well as the triggering of notes and percussion the control of the dynamics for generating these sounds is fully under the performer’s control. This is one of the key novel concepts of the current project. Thus performers control only those features of composition that have a clear correspondence with dance movements i.e. dynamics whereas the performer does not need to be distracted by the musical composition in terms of pitches since this information has been pre-assigned to each sensor node.

The concept shown in figure 26 is easily adapted to ad-hoc performer and conducted compositions. With performer based composition the dancers themselves would have full control both over movement and music. However, it is an interesting question as to whether, in this situation, a group would generally create a self-coordinated response or tend to chaos. For this reason the live conducted composition is expected to be the more practicable approach. Using this technique a ‘performance conductor’ could compose music in real-time by signalling to specific performers to make specific movements which would then trigger musical notes. Furthermore, the conductor could have coarse control over some of the dynamics of each sound. Such a technique offers intriguing possibilities however it also raises further challenges. For example, how would a chord be played? A triad could be mapped to three sensor nodes on a single performer, however whilst two notes could be played using mirrored arm movement the dynamics of the note mapped to the leg would be physically difficult to control in the same way. The alternative here would be to map each note of the triad to three different performers however the signalling by the conductor to the three performers and the successful execution of a coordinated response could prove difficult. These problems would scale up as more ‘instruments’ and musical passages were added to the composition. Clearly some combination of visual cues, rehearsal and choreography would still be necessary to ensure the overall result was acceptable.

Chapter 4 - Conclusions

4.1 Introduction

A critical appraisal of the research and development conducted during this project is presented. Key results are discussed and recommendations made for further refinement and enhancement of the system are proposed. The contribution of the work to the field is appraised.

4.2 Project review

The work presented in this dissertation reflects a hybrid design and research project that has evolved since its conception along a number of interrelated paths. The original idea that the project would simply involve building hardware to a defined specification and conducting an evaluation study to characterise its performance proved to be unrealistic. Review of the literature revealed that there were considerable challenges remaining in terms of reliable motion capture and in translating acquired movement data into sound and music. However, many of the published techniques were found to have progressed to an advanced level for single person use and there was little scope for novel contribution in this respect. Furthermore, many of the limitations identified in the literature in relation to sensor-based motion capture have effectively been addressed through other commercially driven approaches including the Nintendo Wii controller and more recently video based systems, such as, the Kinect interface for the Xbox 360. These technologies are currently effective in supporting up to four simultaneous participants in interactive games of which current relevant examples include Just Dance, Dance Dance Revolution, Dance Central and Dance Star Party. These games provide automated choreography and use indirect motion sensing to assess user performance. Although software development kits are now becoming available to allow customised use of the sensing interfaces and graphics/sound engines these systems are not inherently scalable to an arbitrary number of participants. It is feasible that the sensing technology could be replicated and with sufficient computing power many users supported. However costs are currently prohibitive, at the time of writing the Kinect video sensor is around £100 and the Wii motion controller around £30. In contrast an accelerometer sensor and microcontroller can be purchased for a few pounds and allow distributed systems to be realised that can support many simultaneous interacting users. This was the basis for the novel aspects of the current project to try and address some of the challenges such a distributed multi-user system presents in the context of interactive music composition. The realisation of a system to explore these challenges and explore solutions was the principal aim of the project.

Once the aim had been established, the objectives defined in chapter 1 were identified. The first of these objectives, to develop a body-worn sensor to enable gestural capture, was initially thought to require a straightforward upgrade of existing published designs using newer sensors to give a performance edge. However, most existing designs use multiple sensors including accelerometers, gyroscopes and magnetometers to realise inertial measurement units in several degrees of freedom. Typically this approach is not used for fully capturing a dancer's position and dynamic motion rather it tends to be used to provide a large number of motion related signals that can then be mapped onto synthesis and audio effects parameters to provide fine granularity over sound. Whilst this has resulted in some powerfully expressive mechanisms for sound production it does not adequately engender composition of music.

Consequently, the decision was taken to implement a much simpler, lower cost gestural capture interface that would facilitate music composition. In principle this could be done using only switches to allow performers to trigger note production and indeed such systems already exist in the form of sensing floor mats where spatial position is mapped to specific pitches. However, this is a coarse approach that does not permit a performer any level of expressive control over the way in which a triggered note is sounded. One solution to this problem that has been explored both by researchers and in some commercial products has been the replacement of simple on/off switches with pressure sensitive transducers notably the Force Sensing Resistor (FSR). However, it is apparent that this type of approach requires conscious effort and focus by the performer to interact with the sensor to achieve the desired result. In contrast the use of an accelerometer does not require significant additional effort on the part of the performer beyond the natural movements that are part of the physical performance. Overall this means that the focus of effort by a dancer remains in articulating expressive body movements whilst still allowing a degree of fine control over musical expression. Based on these arguments the design presented here uses a single accelerometer per sensing node.

It was fortuitous, on reflection, that the gestural capture unit developed for this project was based around a tri-axial accelerometer from the outset. In principle it would have been possible to implement a solution using a single axis device however, the additional information obtained from a three channel sensor proved to be essential to achieving a generic sensor that could be used on any performer with a minimum of tuning/calibration to the individual. This requirement was driven by practicality in that a lengthy period of configuration of equipment prior to a performance was deemed undesirable.

Thus an important aspect of this project that materialised from evaluation of the initial prototype devices on subjects was to deal with the uncertainties arising from variation in the signals acquired when the same movements were performed by different people. In vitro characterisation of the calibrated sensors proved that the problem was not due to instrumentation tolerance and it was therefore attributed to the intrinsic anatomical and physiological differences of the subjects.

To deal with the uncertainties posed by variation in repeatability of subject performance and subject to subject variation some statistical analysis was applied to measurement data to understand the nature of the variation. As the histograms given in chapter three illustrate, the frequency profile of the data does not follow a standard distribution and therefore, quantitative statistical inference methods (predicting responses based on anatomical measurement, for example) could not be applied to solve the problem. Consequently, considerable time was expended in trying to develop heuristic solutions by generalising against specific variables observed in the data, specifically peak amplitudes and times to peak.

The most successful approach developed in this manner was a simple combination of linear filtering to remove movement artefact and triggering with an adaptive threshold. Whilst partially successful signal processing required subject specific parameter tuning and it proved impossible to make the sensor response subject independent. Therefore, the decision was made to investigate the use of artificial intelligence as a method of subject independent signal classification.

A number of AI techniques were considered to assess their suitability. An expert system approach was judged infeasible as the database of expert rules would have required an extensive detailed evaluation of signals captured from a large number of subjects. Additionally, such a rule set which is effectively a signal look up table to enable comparative classification would have required substantial memory storage either locally or in the centralised computer. Even if it were possible to create such a database, query response time would be expected to limit performance. Similarly the quantity of prior information necessary to implement a solution based upon the naïve Bayes classifier technique was deemed prohibitive. Fuzzy logic and fuzzy sets were also briefly considered however with the limited number and type of signals being produced by the accelerometers defining suitable behavioural descriptors proved challenging. Once again the requirement for an analyst to assess a wide range of signal parameters and assign appropriate values was prohibitive. The only remaining viable AI technique for this application was an Artificial Neural Network which enables general classification based on a tractable subset of data.

The ANN design discussed in chapter three implements classification based on pattern recognition. Various attempts were made to identify an optimally configured data set and it was found empirically that a set of around 300 samples (each of six measurement variables) was adequate to achieve consistent classification performance with an error rate of better than 5%. It is notable however that reducing the number of variables to three, so that only the peak amplitude values are used resulted in an increase in error rate of between 40 and 50% (the actual figure varies each time the ANN is retrained as different data subsets are randomly selected for training and validation). Thus a key finding of this study is that both amplitude and duration data are essential measurement variables to enable reliable classification.

The third key objective of the project was to investigate and develop a wireless sensor based implementation of the gestural capture system. A key design decision was with regard to the balance of processing between the local sensor node processor and the processing performed remotely on the central computer supporting the composition system. The importance of this stems from the bandwidth limitations of low power radio modules. If the processing is done local to the sensor then the volume of data transmitted can be minimised. This however places a significant processing burden on the local processor. It was found that the operational lifetime of the battery (coin cell) operated sensor node was slightly less than one hour when executing the basic processing and mapping algorithms at a processor clock speed of 10MHz. With the additional implementation of the neural network classifier processor power consumption is likely to rise. Furthermore, it was found necessary to increase the processor clock rate to 20MHz to achieve acceptable data processing and throughput rates. Conversely, if the processing is off-loaded to the remote processor (e.g. a PC) the key challenge identified was in achieving adequate data throughput across the wireless link. The data throughput of typical conventional low power radio modules was found to be typically around 40Kbit/s or

5Kbytes. At the design specification sampling rate of 50Hz per accelerometer channel, the message transmission rate is 5Kbytes/150samples ≈ 30 measurements per second. This raises the question as to whether enough measurements could be transmitted to allow the composition engine on the central processor to perform required processing and look ahead scheduling of real time musical events.

To this end Zigbee was identified as a suitable protocol and the XBee implementation of this protocol supports transmission rates up to 200Kbit/s allowing a five-fold increase in data throughput. The technology is also physically small and the photograph of figure 27 shows a mock-up of the proposed wireless implementation with the XBee module located above the sensor / local processor board. These units are mounted on a wrist band and a one pound coin illustrates the scale. Clearly it is feasible to build a system that can be worn discretely by a performer and this overcomes the issue of the large physical size limitation of some of the systems reviewed in chapter two. In figure 27 the wires are connections to an external dual rail power supply however in the final implementation it is intended to use a pair of three volt coin cells also mounted on the wrist band. The weight of the entire system is estimated at around 30g which is negligible in terms of impact on the performer's movements.

A further benefit of using Zigbee modules which includes a dedicated radio processor was that the processing load on the sensor processor is reduced as it simply sends packets over a local serial link to the Zigbee processor. This improves computational efficiency and helps to reduce overall latency.

For evaluation purposes the sensing system worn by a performer was constructed with four tri-axial accelerometers connected by wire to a central microcontroller (ATMega168) worn on the subjects waist (refer figure 4). With this approach only two output channels were sampled per sensor as the processor only has eight on board ADC channels. The two data wires and power lines were run along the trunk and limbs to the ankle and wrist sensors. Wiring was fixed using surgical tape. This approach was acceptable for investigative purposes however it was evidently too cumbersome and unreliable for a real performance. Furthermore, locating the instrumentation on a person is time consuming especially if the wiring has to be cut to length to suit anatomy. Therefore, the decision was made early in the project to implement a wireless network on the actual performer. This enabled each limb sensing node to communicate wirelessly with a waist mounted coordinator node that implemented the communications link to the remote composition system.

The implementation of the body-worn wireless sensor network was based upon the radial star network topology. Each sensing node has a unique address and communicates with the coordinator on a point to point basis. This relatively simple arrangement allowed series one XBee modules to be used which helped to reduce cost. Series two XBee modules support more complex networking topologies which in principle would allow a more robust networking implementation. For example, if the coordinator unit on a performer was faulty then that performer's sensor nodes could in principal re-route through the coordinator of another performer. This level of robustness and fault tolerance may be important for the current application as loss of signal integrity from a performer could manifest as audio dropout. Further evaluation of the most appropriate networking topology was beyond the scope of the current project.

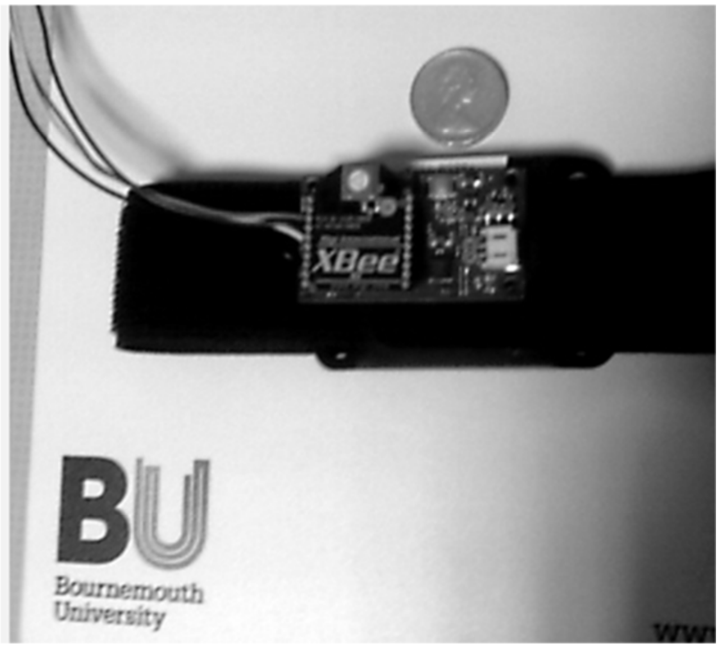


Figure 27 Photograph showing the intended final implementation of the sensor node as a body-worn wireless sensor.

The final project objective of implementing and evaluating a full multi-performer composition system based on this technology has not yet been realised. This is in part due to cost considerations and moreover the time required in overcoming the technical challenges of implementing the system which have been discussed in detail throughout this report. Some preliminary studies have been made based on a single performer using a sensor node tuned to the individual (in lieu of mapping the neural network design into C code to run on the microcontroller). Simple composition patches have been implemented in Max/MSP and used to demonstrate the efficacy of the concept. Furthermore, consideration has been given to the potential issues in scaling up to enable multi-performer composition including automated conduction, ad hoc performer based and live real time conductor based strategies.

4.3 Recommendations for further work

This report marks a timely summary of the development to date of a gestural capture based music composition system. Work is on-going to construct the final wireless version of the body-worn sensing nodes which will be implemented using surface mount technology and Zigbee RF modules. Here some of the key remaining challenges are summarised and briefly considered.

The analogue accelerometer used successfully in this study does not offer the performance and capabilities of newer digital devices. In particular accelerometers such as the BMA180 (Bosch) and ADXL345 (Analog Devices) include electrostatic calibration which overcomes the need for an external calibration platform. Furthermore, these devices include temperature sensors to compensate for thermal drift. This is relevant in the current application due to the heat from stage lighting and warrants further investigation.

Implementation of the neural network in C code to run on either the local sensor processor or the remote composition processor is a priority as this will allow multi-performer composition techniques to be explored in depth without the need to tune each sensor to the individual. It is expected that execution of the ANN based pattern recognition algorithm on the sensor nodes will be preferable as this will minimise the quantity of data that needs to be transmitted to the composition system. This level of localised processing may require a move toward a more powerful processor to achieve satisfactory performance.

At the present time there are many unknowns with respect to the communication system in particularly the potential increases in latency caused by RF interference/packet collisions are a concern. The Zigbee protocol requires transmitters to try retransmitting after a random short delay time following a failed transmission. This could potentially cause the composition engine to stall and strategies need to be investigated to deal with this scenario intelligently.

List of references

Aylward, R. and Paradiso, J. 2006. Sensemble: A Wireless, Compact, Multi-User Sensor System for Interactive Dance. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France. pp134-139

Cont, A., Coduys, T., and Henry, C. 2004. Real-time Gesture Mapping in Pd Environment using Neural Networks. Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME04), Hamamatsu, Japan. pp39-42

Dobrian, C. and Bevilacqua, F. 2003. Gestural Control of Music Using the Vicon 8 Motion Capture system. Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada p161-163

Feldmeier, M. and Paradiso, J. 2007. An Interactive Music Environment for Large Groups with Giveaway Wireless Motion Sensors. Computer Music Journal. pp50-67.

Knapp, R. and Cook, P. 2006. Creating a Network of Integral Music Controllers. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France. pp 124-128.

Lee, E., Enke, U., Borchers, J. and Jong, L. 2007. Towards Rhythmic Analysis of Human Motion using Acceleration-Onset Times. Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07), New York, NY, USA. pp 136-141

Lynch, A., Majeed, B., O'Flynn, B., Barton, J., Murphy, F., Delaney, K., and O'Mathuna, S. 2005. A wireless inertial measurement system (WIMS) for an interactive dance environment. Journal of Physics: Conference Series 15 . Institute of Physics Publishing. pp 95–100

Maki-Patola, T., Laitinen, J., Kanerva, A. and Takala, T. 2005. Experiments with Virtual Reality Instruments. Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada. pp 11-16

Modler, P., Myatt, T. and Saup, M. 2003. An Experimental Set of Hand Gestures for Expressive Control of Musical Parameters in Realtime. Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada. pp 03-146 to 03-150.

Paradiso, J., Hu, E. and Hsiao, K. 1998. Instrumented Footwear for Interactive Dance. Version 1.1, Presented at the XII Colloquium on Musical Informatics, Gorizia, Italy, September 24-26, 1998. pp 1-4.

Schacher, J. 2010. Motion To Gesture To Sound: Mapping For Interactive Dance. Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), Sydney, Australia. pp 250-254.

Wanderley, M. and Depalle, P. Gestural Control of Sound Synthesis. PROCEEDINGS OF THE IEEE, VOL. 92, NO. 4, APRIL 2004. p632-644

Weinberg, G., Aimi, R. and Jennings, K. The Beatbug Network – A Rhythmic System for Interdependent Group Collaboration. Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME-02), Dublin, Ireland, May 24-26, 2002 pp 02-01 to 02-06.

Wessel, D. and Wright, M. 2004. Problems and Prospects for Intimate Musical Control of Computers. Computer Music Journal: Vol. 26, no. 3. Cambridge, MA: MIT Press: 11-22.

Winkler, T. 1995. Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music. Published in Proceedings of the 1995 International Computer Music Conference.

Bibliography

Miranda, E. and Wanderley, M. 2006. New Digital Musical Instruments: Control and Interaction Beyond the Keyboard. ISBN 0-89579-585-X

Birnie, P. and Fairall, J. 1999. An introduction to Low Power Radio. ISBN 0-9537231-0-0

Faludi, R. 2011. Building Wireless Sensor Networks. ISBN 978-0-596-80773-3

Barnett, R. and Cox, S. 2003. Embedded C Programming and the Atmel AVR. ISBN 1-4018-1206-6

Akl, A. 2010. A novel accelerometer based gesture recognition system. MSc Thesis Dept of Electrical and Computer Engineering, Univ. Toronto.

Gray, J. 1989. Dance Technology. Current Applications and Future Trends. ISBN 0-88314-429-8

Haykin, S. 2011. Neural Networks and Learning Machines 3rd Ed. ISBN 0-13-129376-1

Negnevitsky, M. 2005. Artificial Intelligence a guide to intelligent systems. 2nd Ed. ISBN 0-321-20466-2

Ifeachor, E and Jervis, B. 2002. Digital Signal Processing – A practical approach. 2nd Ed. ISBN-0-201-59619-9