Animatronics: The Development Of A Facial Action Sensing System To Enhance Performance Control

Adrian Woolard B.Eng.(Hons)

A thesis submitted for the degree of Doctor Of Philosophy

in the
Department of Electrical And Electronic Engineering
at the
University of Newcastle upon Tyne

July 1994

Abstract

This thesis presents the initial exploratory research into an original and novel technique to enhance performance control in animatronics. An animatronic system is defined as a 3-D electro-mechanically driven facial model that can move in certain ways, when controlled by a human performer to create the "illusion of life" for a viewer. The vital elements in this form of performance are the synchronisation of lip movements to an acoustic speech signal and the animation of emotive expressions.

A novel optical sensing technique is proposed based on the hypothesis that the input of distinctive articulatory or emotive movements from the performer's face would provide a more 'natural' form of control. The principle that the movement of a minimal set of points at key positions on the face can produce sufficient control information to describe the overall action is proposed to achieve this hypothesis.

A comprehensive investigation into human communication, including visual speech perception and non-verbal facial expression, to define the optimum set of key points is described. Conclusions are also drawn on the primary facial actions required for successful lip synchronisation.

Both the theoretical and practical aspects of the realisation of a prototype system are described. A methodology is presented for the assessment of the sensing system and the overall objectives based on the design and construction of an animatronic face, of the same dimensions as the researcher's, to produce animation of the desired actions with similar displacements. Objective analysis is achieved through the comparison of measurements by the sensor system from the performer's key point movements and those of the animatronic model. Perceptual data is generated through the visual analysis of the animated facial movement. The results and analysis of the investigations are presented in the thesis.

The thesis discusses results obtained which indicate that, given certain valid assumptions, the sensor system is capable of consistent facial motion detection. It can provide sufficient control for the animatronic model to produce a limited set of facial actions in a realistic manner. Results indicate the possibilities for improved lip synchronisation and, hence, "overall character" performance.

This thesis is dedicated to Helen and my parents for their unstinting love during the highs and lows of the last few years.

Acknowledgements

The researcher would like to thank, in true Hollywood style, a great number of people behind the scenes who have all helped in some way to realise this final piece of work.

This thesis is dedicated to Helen, who has been a never-ending source of inspiration and confidence. Thank you for all the love and support during even the darkest days. Much love and gratitude goes to my parents for their unerring belief in their son every Sunday afternoon.

At Newcastle, my thanks go firstly to my supervisor, Oliver Hinton, for his enthusiasm in supporting this project and for coping with a "gurning" student during research meetings. Big thanks also to Doctor Iggy for his proof reading, advice, pints of beer and impressive d.j. skills and to my colleagues, Nick and Khu, for their awesome help during the last few weeks of madness.

On the technical side, big thanks to Jim Harrison for all the help with the photographs and video. I told him he'd make it above the key grip in the credits. Thanks to Eddie, Steve, Alan and the other technicians for all their help and poor jokes during the construction of the animatronic monster. Similarly, thanks to the technicians in the electronics workshop and also to the office staff for turning the other way when I needed print outs and photocopies. Thanks to Roy Booth for the design and construction of the Data Acquisition board and to Steve Donald for advice on design of the replica.

Big cheers to all my friends and flatmates who had to put up with the trials and tribulations from their deranged friend. To the other members of Hug, thanks for the fun and memories over the last few years. To my fellow researchers; Khu, Nick, Dave, Jalil, Farid, Jay, Mary-Jo, Parivash and Sanjev, my warmest thanks for all their friendship, help and support during even the longest days and nights. The microwave is all yours and remember "it's not just for Christmas, it's for life!".

At the Jim Henson Creature Shop, my deepest thanks to, firstly, Dave Housman and John Stephenson for their support and belief that I was a valid investment. To all the others; Quentin, Pete, Verner, Terry, Dave, etc., many thanks for the advice and good times during the Turtle days.

This research was sponsored by Jim Henson's Creature Shop, London. Many thanks for their help and encouragement.

Table Of Contents

Table Of Figures

Table Of Tables

Table Of Video Sequences

Table Of Contents

Table Of Figures

Table Of Tables

Table Of Video Sequences

Sequence	Title	Time	Brief description
V.1	Introduction	0 min.	Thesis title courtesy of the replica
V.2	Drive System	30 sec	Description of the actual mechanical construction of the replica
V.3	Facial Action Permutations	4 min. 30 sec	All drives in the replica face driven to produce all the possible permutations of displacement
V.4	Key Point Actions	12 min. 30 sec	Examples of the reduced key point actions for ramp input
V.5	Experimental Data	13 min.	Animation of the recorded data used in the analysis section of Chapter 6
V.6	Experimental Data With Different Conditioning	16 min. 50 sec	Examples of recorded data animated with different types of conditioning
V.7	Primary Actions With Acoustic Signal	17 min. 40 sec	Examples of replica animation in synchronization with acoustic signal
V.8	Primary Actions In Comparison With Live Face	20 min. 50 sec	Examples of the replica animation with live face on screen providing real time control
V.9	Lip Synchronization Example	25 min. 40 sec	Lip synchronization of chosen sentence
V.10	Final Lip Synchronization Examples	27 min. 30 sec	Examples of lip synchronization, including the alphabet, numbers, thesis title, and other performance sequences
	End	36 min.	

PUBLICATION

One conference paper has emerged from my postgraduate research so far.

Title : "A Low Cost Animatronic Facial Action Recognition System".

Conference : I.A.R.P. 2ND Workshop on Sensor Fusion and
Environmental Modelling, Oxford, September 1991.

Animatronics: The Development Of A Facial Action Sensing System To Enhance Performance Control

"The function of the imagination is not to make strange things settled

so much as to make settled things strange"

G. Chesterton 1901

Chapter 1

Introduction

Chapter 1

Introduction

1.1 Context, Background And Motivations Of Research

Darwin was of the opinion that the face is the most powerful area on the human body for communication **[Ekman73]**. This opinion acknowledges the fascination displayed by every age and culture throughout history. The sculptors of Ancient Greece, portrait artists such as Leonardo Da Vinci, anatomists such as Duchené in the 19th century and present day facial plastic surgeons have all attempted to produce greater understanding or better representations of the physical structure and complex actions of the face **[Faigin90]**.

The face commands our attention when people communicate due to the concentration of the main sensory and articulatory systems; vision, hearing, sound production, smell and taste. Communication is broadly defined as the method of transmitting and receiving information between two or more individuals. In humans the primary form of communication is by use of spoken languages where the majority of information is perceived audibly **[Massaro87]**. However other important messages are understood visually. Emotional signals, such as facial expressions, gaze and body language, and the actions associated with the production of speech are all perceived through the visual channel. In particular these articulatory gestures can offer important clues to

understanding what has been said when the acoustic signal is degraded or if the listener is hearing impaired.

An important area for communication is that of entertainment where the individual must produce a performance for an audience. An actor's performance is loosely defined as the ability to realistically convey the words and emotions of a mythical character **[Hilton87]**. A different yet popular form of performance is the art of puppetry. In puppetry, the performer acts out the role of a character through the movements of the puppet, an inanimate object, creating the "illusion of life" **[Engler73]**. The performer is the term used to describe the actor, puppeteer, operator or controller that produces the character performance. The actions produced are animations of perceived life-like motion, expressions, gestures and emotions that convince the viewer that it is a living creature. The success of this illusion is primarily dependent on the performer's ability to control their physical actions in relation to the pre-defined script, i.e. the message. As technology has developed, so variations on puppetry have emerged and this body of research is concerned with one such descendent, that of animatronics.

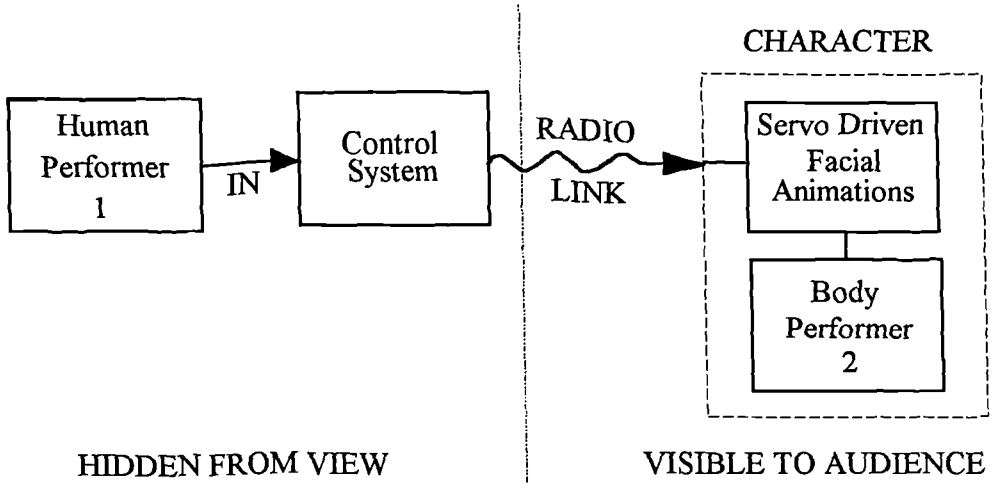
"Audio-Animatronics", to use its original title, was conceived by Walt Disney in 1956 to describe the three dimensional animation of life-like figures in his Disneyworld Amusement Park **[Thomas81b]**. He rather grandly defined it as "a combination of all the arts; the 3-D realism of fine sculpture, the vitality of a great painting, the drama and personal rapport of the theatre and the artistic versatility and consistency of the motion picture" **[Thomas81b]**. A more accurate definition is that "animatronics is the art of animating a life-like figure of a person, animal or creature by electro-mechanical means" **[Chambers93]**. Whilst animatronics continues to be utilised in amusement parks it has diversified to other entertainment mediums such as film and television. Here, due to the intimate nature of the camera, there is a desire to create fantasy characters that can produce a more subtle form of motion and expression than the gross actions necessary for amusement park entertainment. Examples of successful film fantasy characters are "E.T." (1982), "The Dark Crystal" (1983), "Gremlins" (1984) and "Teenage Mutant Ninja Turtles" (1990). An example is shown in Figure 1.1.

Figure 1.1 Photograph Of A Teenage Mutant Ninja Turtle (1990).

© Jim Henson Productions 1990

The realisation of this life-like animation is dependent on a number of elements: the appearance of the figure; the physical production of its actions; and the techniques employed to control these actions. Whilst acknowledging the importance of the artistic input in the design and production of animatronic characters, this research focuses purely on the engineering aspects of realising such complex motion control systems.

The descriptions of animatronic techniques, unless stated to the contrary, are based on the research sponsors, The Jim Henson Creature Shop. This is due to the lack of published material and a reluctance within the industry to reveal its techniques. Each animatronic character can be considered as two distinct elements; the body and the face, where the final performance is the result of the combined effects of both elements. Figure 1.2 shows the block diagram of the typical system covered in this research. The body movements, including the global head actions, are performed by a human performer inside the character's costume and the facial actions are created using servo-driven mechanisms controlled by a separate performer hidden from view.



IN - human performance input

RADIO LINK - transmitted drive control signals to allow character freedom to move

Figure 1.2 Block Diagram Of A Typical Animatronic System

The specific interest of this research is in the animation of facial expressions and speech production to improve the visual perception of animatronic film characters. The task of animating the complex lip movements and more importantly of synchronising these movements to a pre-defined soundtrack represents a major problem involved in successful animation. A good example of the significance of lip synchronisation is the anecdotal evidence of watching a poorly dubbed movie. Here the confusion between the perceived audio signal of the soundtrack and the different visual signal quickly becomes apparent and ultimately has a detrimental effect on the performance. The critical elements in successful lip synchronisation are, firstly, the generation of the correct lip shapes and, secondly, the production of the correct timing, rate and rhythm in order to sequence groups of lip shapes together to form sentences. Continuous speech also produces other vital speech factors such as co-articulation, intonation and expressive blending that have to be animated in the final output.

The motivation for this interest in lip synchronisation originated from the comments of hearing impaired individuals who indicated that only minimal information was

perceived through the visual speech actions of an animatronic character **[Housman90]**. The main reasons for this are either:

1. The physical techniques to produce the animated actions are inadequate.

2. The method for controlling the performance fails to provide sufficient information for correct action production.

As interest is primarily in the electronic aspects of animatronics, this project concentrates on the possible improvement of performance control techniques to enhance the animation of facial actions.

The role of the human performer is essential in the production of facial animation. Judgements are made firstly on the tasks required by the script, and secondly, on the input signals necessary to convert the script into an animated sequence of facial actions. Though the simulation of those behavioural decisions has been attempted, **[Reynolds82]**, it is vital to enhance the human decision making and maximise the control of the output animation. Performance control systems are therefore designed to implement two tasks: firstly, to allow the physical input of control signals by a human performer; and secondly, to allow the creation of rules to map between the control inputs and the drive outputs to produce the desired animation.

The current approaches that exist in animatronic control originate from the techniques of hand puppetry. Previous systems have used either direct hand control or numerous performers manipulating direct cable controls to produce the overall facial action. The present system, as seen in Figure 1.2, consists of a single performer inputting signals through multiple arm, hand and finger joystick analogue controls. Each control input can be mapped to any number of output drives by using a purpose built processor system. This enables the performer to produce the complex facial expressions from combinations of manual movements. This relationship between input hand movement and output facial expressions is not an inherent one. Superior hand-to-eye co-ordination and manual dexterity along with unnatural mental operations have to be developed by the performer in order to produce accurate lip synchronisation. These are specialised performance skills that take a significant amount of time to develop, if ever.

It is clear that an alternative approach which can manipulate the natural signals created by actual speech production to derive the control input, has the potential to reduce the need for these specialised skills and also to improve the animation of lip synchronisation.

Within the comparable field of computer generated facial animation, similar control problems exist even though the final application may be different, for example in communications systems. The most common approach is to extract the control signals from the analysis of the actual acoustic soundtrack **[Pearce86]**, **[Choi90]**, **[Lewis91]**, **[Morishima91b]**. The success of this approach is dependent upon the clarity of the input speech signal and on the accuracy of the analysis rules developed to identify the individual speech phonemes and hence the visual equivalents, known as visemes. Information on timing, rate and rhythm must also be derived. Similar techniques are adapted in text driven systems **[Morishima91b]**. The consequence of these analysis techniques is the segmentation of the continuous speech signal that can result in stilted or unnatural animations. This is due to the loss of information concerning the co-articulation between phonemes and the important expressive signals. To overcome such problems requires a large amount of processing, and these can still result in an animation with little of the variety associated with human speech.

Given that the overall objective is the production of visual speech animation, an obvious approach would appear to be the extraction of control signals from actual human facial movements. Image recognition methods have been developed to derive data on lip shapes and actions, though not necessarily for performance control **[Storey88]**, **[Petajan88a]**, **[Morishima92]**. Brooke, and later Waters, have successfully extracted information from the physical changes of the face. Through the analysis of a sequence of images, where white dots highlight specific key articulatory points, animation control data was extracted **[Brooke83]**, **[Waters89]**. This method, though severely limited by the large amount of off-line image processing required, avoids the errors resulting from the segmentation of the speech signal. This allows the capture of expressive and co-articulatory information from a limited set of facial points rather than the overall image.

In summary it is proposed that the development of a real time technique to accurately sense the motion of the performer's face would enhance the control of an animatronic character's facial performance. The solution is based on the principle that the

movement of an optimum set of measurable points at key positions on the facial surface can produce sufficient control data to describe the overall visual action. The movements on the face are a direct result of actual speech production and facial expression and hence provide information on the effects of continuous speech co-articulation, timing, rate and rhythm and of expressive blends. This solution has the advantage of allowing the performer a more innate form of control and in principle should reduce the need for highly specialised skills.

1.2 Summary Of Objectives

In order to produce the final system, the first objective is to develop a suitable scientific framework through the correlation of a wide range of disciplines. Until now, the field of animatronics has been largely based on practical and subjective approaches, with little literature available. Hence there is a need for a broad review, from which can be determined firstly a suitable technique to achieve facial action sensing and, secondly, the optimum set of key points on the face that can provide a sufficient description for a minimal workable model of visible articulatory gestures.

The primary objective is the practical realisation of an electronic technique to directly sense the physical changes at key facial positions. This system should interface with the existing Performance Control System from the research sponsors "The Jim Henson Creature Shop" to have any practical application. This creates the following design criteria; real time continuous operation, low cost, non-restrictive to the performer and the capacity to allow both articulatory and expressive facial action inputs. To assess the ability of the control technique, a comprehensive method of solution has to be devised to solve the problem of generating objective measurements given the highly subjective nature of visual action recognition.

To this end, it was proposed to design and construct an animatronic replica face, of identical dimensions to those of the live face, which would displace the same set of key points. This approach simplifies the mapping rules between control signal and drive action, and allows both objective and subjective analyses to be made.

The principle is that overall control actions are sensed from the live face. The signals are in the form of individual signals from the key points. These can then be mapped to

their equivalent driving points on the replica face and the subsequent recombination of the individual displacements should result in the final output actions being the same perceptually. Objective analysis is possible through measurements, by the sensor system, of both sets of key point actions. Subjective data can be generated through the visual analysis of the animated facial movements, by a set of independent viewers.

1.3 Description Of The Thesis Layout

The thesis is structured in the following way. Chapter Two is concerned with the engineering aspects of the field of animatronics. Descriptions are given of its multi-disciplinary origins and examples of past work and, where literature has been available, the techniques involved. Each element of the present system is detailed and a review of the existing performance control techniques used in animatronics and computer generated facial animation is presented.

Chapter Three investigates the human facial communication system to determine primary facial actions suitable for recognition and animation (emulation). The principles of facial physiology, speech production, lip-reading and visual articulatory gestures, facial expressions and emotional signals are all examined in detail.

Chapter Four correlates the information from the previous chapters and describes in detail the proposed method of solution. The theoretical principles of sensing the actions of key points, the generation of control parameters, the mapping relationships between control and drive and the construction of an animatronic replica head to facilitate system assessments are all explained. The methodology of analysis and the overall system design are discussed and a photogrammetric investigation into actual human facial expression is described.

Chapter Five describes the practical realisation of the individual elements of the proposed system. The electronic design of the optical sensing technique and a detailed examination of the sensing system in isolated conditions are presented. A method for data acquisition and playback and the mechanical design and construction of the animatronic 3-D face are described.

Chapter Six presents the results and analysis of the final system. The individual theories and techniques are investigated through the analysis of objective and subjective experimental data.

The final chapter discusses the overall conclusions from the research. The proof of the project objectives and the overall principles are considered. Theoretical and practical improvements are presented together with the further work to be undertaken and the overall contribution of the researcher.

Chapter 2

Review Of Animatronics And Present Performance Control Systems

Chapter 2

Review Of Animatronics And Present Performance Control Systems

2.1 Introduction

Animatronics is "the art of animating a life-like figure of a person or creature for entertainment by artificial means, principally electro-mechanical" **[Chambers93]**. **[Parke82]** distinguishes animation from simulation on the basis that "simulation is intended as an exact model whilst the goal of animation is to communicate an idea, story or message". The process of animation allows for artistic licence in the creation of actions that produce the "illusion of life" for the viewer. This "illusion" is a broad term that suggests the viewer's perceptions of a character's animated actions are sufficiently similar to their idea or experience of human movements and expressions, that they accept the actions as the product of a living form.

As stated in Chapter 1, this research concentrates on the role of the face in animated performances. The successful animation of lip movements and expressions can convey important messages about the character to the viewer. Although most people cannot lip read, i.e. identify the meaning of speech from the visible actions alone, viewer's have a passive notion of the correct mouth movements during speech **[Lewis91]**. Bad lip synchronisation is quickly recognised by viewers but there is no adequate definition for the production of 'good' or realistic lip synchronisation. For

example, is it the production of visually recognisable lip shapes or is it the accuracy of their timing with the soundtrack ?

The task of controlling the animation of these movements represents a major problem in the creation of realistic animatronic characters. Various techniques have been developed in animatronics and in the comparable field of computer generated facial animation to control lip synchronisation with varying degrees of success. All of the techniques involve human performers at some stage in their operation, principally in the role of decision makers. The objective of this chapter is to consider the techniques available that enhance the performer's control of the output actions thus creating improved animated facial performances.

Section 2.2 presents an overview of the origins of animatronics, indicating the relationships between the various disciplines that have led to the present systems. It also seeks to clarify the different terms used in special effects. Section 2.3 describes the present animatronic system developed by Hensons to control facial performances. Section 2.4 evaluates other performance control techniques used to produce lip synchronisation in computer generated animation.

2.2 The Origins of Animatronics and Its Related Fields

Before considering present animatronic systems and, specifically performance control techniques, there is a need to examine the origins of animatronics. Figure 2.1 shows the developments and inter-relationships of the various fields.

The fields that have direct or major relevance to animatronics are discussed in the following sub-sections and where possible they are considered in a chronological order. However certain fields developed, and still continue to develop, over the same period as others.

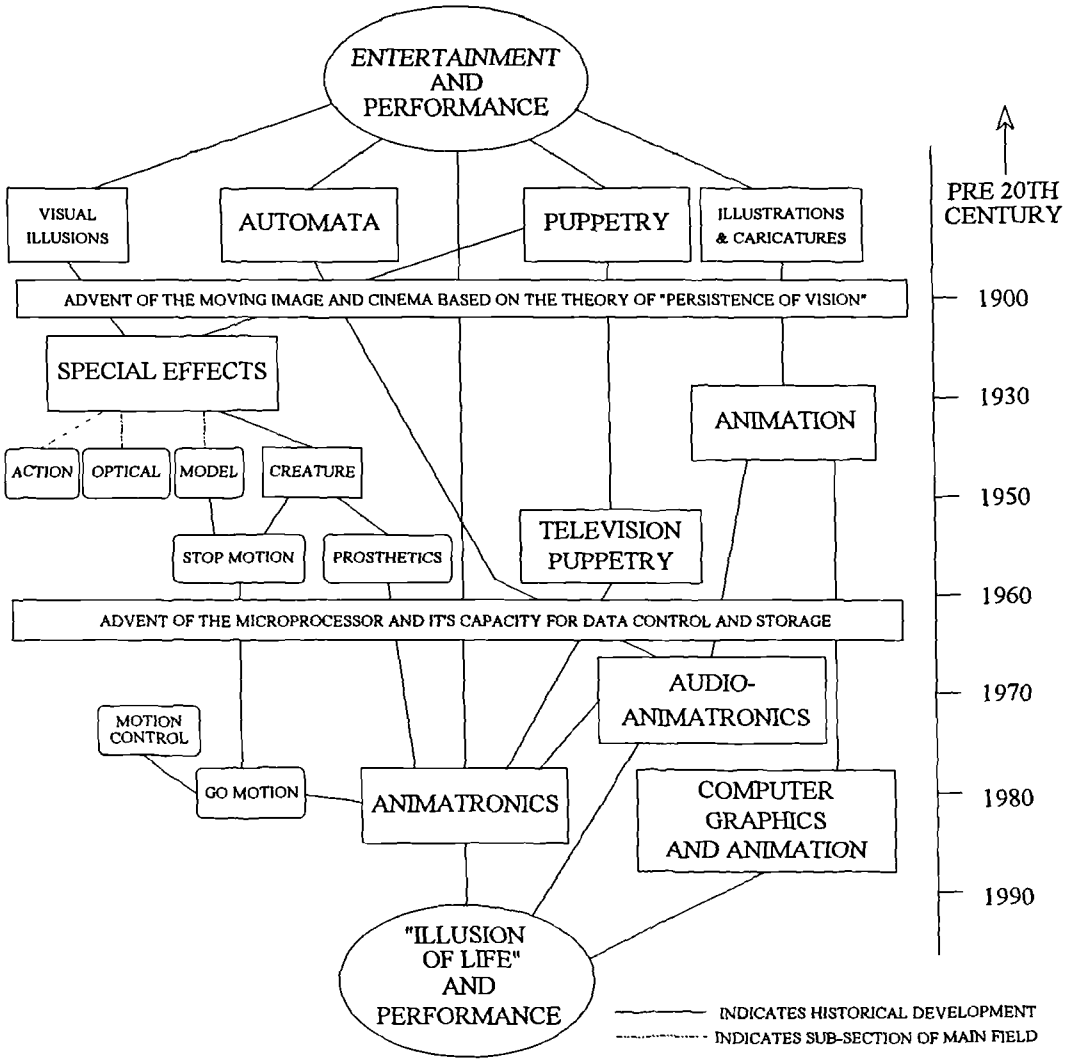


Figure 2.1 Chronological Diagram Of Origins Of Animatronics

2.2.1 Automata

An automaton is defined as "a machine that has the form of an organised being and contained within itself is a mechanism capable of creating movement and hence simulating life" **[Bailly87]**. This manifested itself as clocks in the 14th Century Europe and later as mechanical toys such as 'Tableau Mecaniques or Animes' and musical boxes in France circa 19th Century. The main drives were clockwork, and later electricity, which powered a variety of mechanisms including cams, followers or springs. These in turn produced the final actions through systems of pull rods, levers, strings and pulleys **[Bailly87]**. More recent examples of automata are the kinetic art of Jean Tinguley **[Hulton75]** and Jim Whiting.

2.2.2 Puppetry

[Engler73] defined puppetry as "a performing art where the puppet is an object manipulated by a puppeteer to encourage an audience to accept the living existence of an otherwise inanimate object". Puppetry was originated as an alternative means of dramatic expression to theatre and has existed for many centuries in many forms; marionette, glove, rod, shadow, paper, automata or Bunraki.

Of particular relevance to this research is the glove, or hand, puppet and specifically the mouth puppet where the actions are controlled directly by the puppeteer's hand inside the puppet's head. The illusion of speech is created by synchronising the opening and closing of the fingers and thumb, and hence the upper and lower jaw, to a spoken soundtrack **[Engler73]**.

2.2.3 The Moving Image and Cinema

The advent of the moving image and film at the end of the 19th Century gave performers a broader audience and also introduced new and more impressive methods to 'trick' the audience. It is based on the phenomenon of the persistence of vision that was first recorded by Ptolemy in AD 147 and is defined as "the eye's ability to retain an image for a brief moment after the object itself has been removed" **[Rovin77]**. Film achieves the illusion of movement by presenting a sequence of still images at a sufficient rate, 24 frames per second, to manipulate this phenomenon.

2.2.4 Special Effects

Film special effects originated from stage illusions, puppetry shows and trick photography. They are used in motion pictures when the scenes required are impractical, dangerous or even impossible to film in a normal manner **[Smith86]**. The overall term "special effects" covers a number of specific fields; action, optical, miniature model and creature, of which only creature effects are of relevance to animatronics.

2.2.5 Creature Special Effects

Creature effects is the broad term used to define the animation of fantasy creatures or the simulation of real actors or animals in daring and spectacular stunts **[Smith86]**. The three main areas are stop motion, go motion and prosthetics. It should be noted that wherever possible body performers are incorporated into the design of a character to provide the gross body movements. At its most primitive this can be seen as the walking monsters in the Godzilla series (1954 onwards) **[Rovin77]**.

2.2.5.1 Stop Motion

The stop motion technique is defined as the photographic process in which 3-D figures are exposed onto film on a frame by frame basis. Between each exposure, the figures are moved very small increments so that when the sequence of film is replayed, the illusion of movement of the subject is created **[Eastman83]**, **[Smith86]**. The figure is constructed with internal articulated joints or armatures to allow these increments whilst retaining the overall shape and rigidity. Major examples of this technique include "King Kong" (1933), the work of Ray Harryhausen (1953-1973) and more recently the work of Nick Park on "Creature Comforts" (1992).

This method is limited in a number of ways. Firstly, it is very time consuming to produce the animation. Secondly, unrealistic 'stilted' actions can occur if excessive changes are made between frames. Finally, in standard filming a 'blur' is created between the frames which 'aids' the visual perception of motion. This, however, does not exist when film is exposed one frame at a time and as a result the final sequence is perceived as false motion **[Hutchinson87]**.

2.2.5.2 Go Motion

'Go motion' effects were created by Industrial Light and Magic, a special effects' company, to overcome the unrealistic actions of stop-motion **[Smith86]**. It is based on principle of Japanese rod puppetry and makes use of motion control systems. Motion control was conceived as a tool for cameramen where complex camera movements had to be repeated. By driving the camera rig along a track the motor control signals can be recorded, allowing for exact repetition of movements on playback. Each joint or body part of a character is connected to a rod which in turn is connected to a precision motor. The actions are then individually programmed in via manipulation of a joystick to produce the overall action, using the ability to record and replay the exact position of each rod at any given time. Film can be now recorded as a sequence removing problems of the missing 'blur' in stop-motion. The limitations of this method are the precision of the drives, the method of control and the concealment of the rods.

2.2.5.3 Prosthetics

The use of prosthetics is another important technique in the creation of fantasy characters. They are defined as "make-up effects to create the transition from the human face and form to what might be termed a puppet" **[Kehoe85]**. Examples of make-up effects are "*The Wizard of Oz*" (1939), "*The Planet of the Apes*" (1968), "*The Elephant Man*" (1980) or "*Hellraiser*" (1987). Examples of make-up and mechanical effects include the apes in "*2001: A Space Odyssey*" (1968) which gave the body performer control over the jaw and lip movements **[Rovin77]**, the human to werewolf scenes by Rick Baker for "*The American Werewolf In London*" (1981) **[Kehoe85]**, **[Cinefex16]** and the man-machine effects on "*Terminator II*" (1991) by Stan Winston **[Cinefex47]**.

2.2.6 Film Animation

In a similar way to stop motion, traditional animators created characters from sequences of static images, each one slightly different, which, when replayed, produced the illusion of movement. The most successful cell animation was produced by the Disney Studios **[Thomas81a]**.

Of particular interest, are their techniques for synchronising the lip movements. **[Levitan79]** and **[Thomas81a]** described the following basic ideas, which also have a relevance to puppetry. Animation avoided the production of too much detail in any one drawing. Due to the limited number of frames, not every phoneme was animated in a sentence. The effect of doing so produced confusion in the perceived facial actions. The soundtrack was analysed where key frames occur and noted on an exposure sheet. The correct 'in-between' frames were produced. This phonetic breakdown established a limited set of lip shapes. The following categories were employed: open vowels (/a/, /e/, /i/); bilabial consonants (/b/, /p/, /m/); and oval mouth shapes (/u/, /o/, /w/).

These pragmatic principles produced quite acceptable lip synchronisation despite the fact that there were no hard rules, as **[Thomas81a]** explained in his definitive book on Disney animation. The animation of motion and expression, the creation of believable characters and their performance are all directly comparable to puppetry and animatronics.

2.2.7 Television Puppetry

The development of television led to the creation of new forms of puppetry. The work of Jim Henson on "Sesame Street" and "The Muppets" is a prime example of this **[Finch82]**. Originating from glove puppetry, the performance was achieved by direct hand and arm control. The right hand, inside the body of the puppet, controlled the head, eyes and mouth whilst the left hand manipulated thin rods connected to the limbs of the puppet. Lip movements were achieved by synchronising hand actions, which operated primarily the jaw, to the spoken soundtrack.

A more recent example of puppetry on television is Spitting Image. Here, facial expression is achieved through the caricature of the individual's features rather than their actual motion.

2.2.8 Audio-Animatronics

As stated in Chapter 1, 'audio-animatronics' was conceived by Walt Disney in the 1950's. The original idea was derived from mechanical automata with the desire to produce the animation of 3-D characters **[Thomas81b]**. By adapting the principles of cell animation to control the mechanised characters, Disney created a new form of entertainment within his amusement parks. These systems used a hydraulic or pneumatic drive approach, with electronic control, to produce the output movement. The control signals were created pragmatically by animators to synchronise the actions to music, dialogue or other effects, and then stored on computers for repeated playback.

The constraints placed on this type of system are significantly different from the later animatronic systems developed for film and television use. These include increased overall concealment from the audience, greater use of gross actions with the subtleties of movement lost due to visibility limitations, high drive performance often at expense of action speed, greater durability of drives and surface materials with minimal technical support and the ability to perform independent of human control, by use of programmable action sequences **[Hitchcox90a]**, **[Korane90]**, **[Bailey82]**.

Other than amusement parks, the use of the same principle of repeatable actions to create a performance has been applied to museum displays **[Hitchcox89]** and also to conference display exhibits. An example of this kind is the work by Jim Hennequin and Spitting Image. Again the facial expression is achieved primarily by caricature and jaw movements are controlled by recognising the start and end of each word from the pre-recorded soundtrack **[Minson89]**.

2.2.9 Computer Graphics and Animation

Computer animation was first conceived to assist traditional animation in executing the tedious and time consuming tasks such as image colouring. As computers have developed, research has attempted to generate 3-D models of faces and, specifically, human representations. The potential applications for these models are entertainment, telecommunications and scientific simulations **[Tost88]**. The modelling of the human face is based on three approaches; key frame, parameterization and muscle models.

Key frame models are based on the same principle as traditional animation. The desired image is fully specified at a certain moment in time and then another image is specified some frames later. An algorithm then generates the necessary frames in-between. This is limited by the large amount of data required as every point in each frame must be defined.

[Parke82] developed the parameterized model. In this model, an appropriate set of parameters are defined to describe, firstly, the conformation of the face and, secondly, it's expressions. The face is created as a polygon mesh where each polygon represents a surface texture or colour that is produced at the end through a process known as rendering **[Swain92]**. The shape of the topology is defined by the conformation parameters that are either defined by trial and error or from the measurement of a real face. The animation is produced by altering the values of the expression parameters. Each parameter alters the 3-D positions of pre-defined polygon vertices. Hence by creating sequences of parameter changes, the polygon mesh will produce different visual shapes.

In a development to Parke's surface model, **[Platt81]** produced a model based on the actual underlying muscle structure and its connections with the skull and skin. **[Waters87a]** and **[Guenter89]** produced adaptations of Platt's model based on the Ekman's facial action units (c.f. Section 3.4). The output surface image is deformed by defining changes in the parameters that specify individual muscles.

The term 'motion control' describes the method of manipulating facial models to produce the animated sequences of expressions. Guiding control allows the animator to define the set of key frames, in terms of the limited set of parameters required to produce it, and then the intermediate frames are generated by interpolation. Program

control describes the specification of motion in algorithmic terms and task control uses pre-defined high level commands to create complex movements **[Tost88]**. The actual techniques to 'drive' these types of motion control are principally program to image, acoustic to image and image to image. Program driven animation can generate all three levels of control by allowing a programmer to input pre-defined commands to produce time varying sequences of images **[Magnenat Thalmann89a]**. Acoustic driven animation produces parameter changes through analysis and conversion of an acoustic input signal **[Lewis91]**. Image driven animation extracts parameter values from sequences of facial images using either specific key points or overall image analysis **[Brooke83]**, **[Terzopoulos90]**. These control techniques are examined in detail in Section 2.4.

In entertainment, where realistic facial motion is required, the actor's face is captured through a technique known as rotoscoping. Live action footage of the actor, performing desired motions, is recorded and the frames provide the guide for the resulting animation **[Robertson88]**. The actor produces different expressions that are captured and used as key frames in the final animation and the motion is produced by software interpolation between these frames. These techniques have been used in "Terminator II" **[Cinefex47]** and "The Lawnmower Man" **[Cinefex50]**.

2.2.10 Present Animatronics

Present animatronic systems are the result of techniques developed in hand puppetry and motion control. Its application is primarily in film and television which produces a different set of design objectives to those of audio-animatronics. Facial animatronics has to produce a greater subtlety of motion and expression due to the proximity of the camera, and hence the audience. The gross actions of audio-animatronics produce unrealistic performances when viewed from close distance. The major advantage of animatronics is the ability to physically appear on the film set and interact in real time with other characters and performers. This reduces the need for the precisely defined script directions required in the production of computer animation.

Animatronic characters can be defined as two distinct types; animated fantasy characters and simulated real animals or people. The latter are produced when the scenes cannot be created in normal circumstances **[Smith86]**. Examples of this

include the close-ups of animals in "The Bear" and "Gorillas In The Mist" **[Cinefex46]** or the sharks in the "Jaws" series **[Hitchcox90b]**. Human faces tend to occur only in dangerous or impossible situations, for example in "Terminator II" **[Cinefex47]** or "Aliens^3" **[Cinefex50]**. Examples of animated fantasy characters include "E.T." **[Smith86]**, "Gremlins II" **[Cinefex46]**, "The Dark Crystal", "Labyrinth", "Teenage Mutant Ninja Turtles I & II" and "The Dinosaurs" for television. Present animatronic systems are discussed in detail in the following Section 2.3.

2.3 Present Performance Systems in Animatronics

Due to the lack of published material and reluctance within the industry to reveal its methods, the descriptions of animatronic techniques, unless stated to the contrary, are based on the research sponsors, The Jim Henson Creature Shop.

Having discussed the origins of animatronics, in Section 2.2, it was necessary to examine the actual processes involved in the production of a "successful" facial performance. The goal of an animatronic system is the creation of animated actions that communicate messages, defined in a script, to the audience in such a way that they believe it to be the actions of a living creature. The process of creating the final performance is the combined result of a number of distinct elements and the success of the system is dependent upon all of them. These elements are as follows:

1. the task defined by the script;
2. the decisions made by the performer regarding the facial actions required to communicate the message of the script;
3. the technique to input these decisions into the system and their actual production;
4. the defined mapping relationship between the input control signals from the performer and the output drive signals;

5. the design of the drive mechanisms to produce the required output actions; and

6. the final overall visual appearance of the character.

These are shown in the schematic representation in Figure 2.2. This system can be considered as a "single shot", open loop control system that is task oriented. It is single shot in terms of the fact that only one performance is considered as the final output. However, due to the nature of the work in film, rehearsal allows the performer the ability to improve their control actions through the visual feedback of the output actions.

2.3.1 The Task And Human Performance Control

The system is primarily task orientated in the form of a written script. The script provides an overall description of the message to be communicated in terms of the dialogue to be spoken or in the form of a pre-recorded soundtrack. It will also describe the non-verbal, or emotional, content of the performance as well as possible interactions with other characters. Look ahead to Sections 3.3 and 3.4 for discussion on verbal and non-verbal communication.

The role of the human performer within the performance system is essential. The decisions taken are complex processes as they represent the conversion of a descriptive written task into specific input control signals. These signals are designed to reproduce the task in form of output motion. Firstly, the performer decodes the overall task into more specific terms, and judgements are made on what the final actions should be to synchronise with the dialogue and to convey the correct emotional meaning. Secondly, the performer produces signals to convert the performance decisions into sequences of animated actions. These actions appear simultaneously due to the combined blends of expression and dialogue.

The techniques for the production and capture of these input signals represents the primary interest of the research and specifically the possible automation of this conversion between the mental performance decisions and the physical control actions.

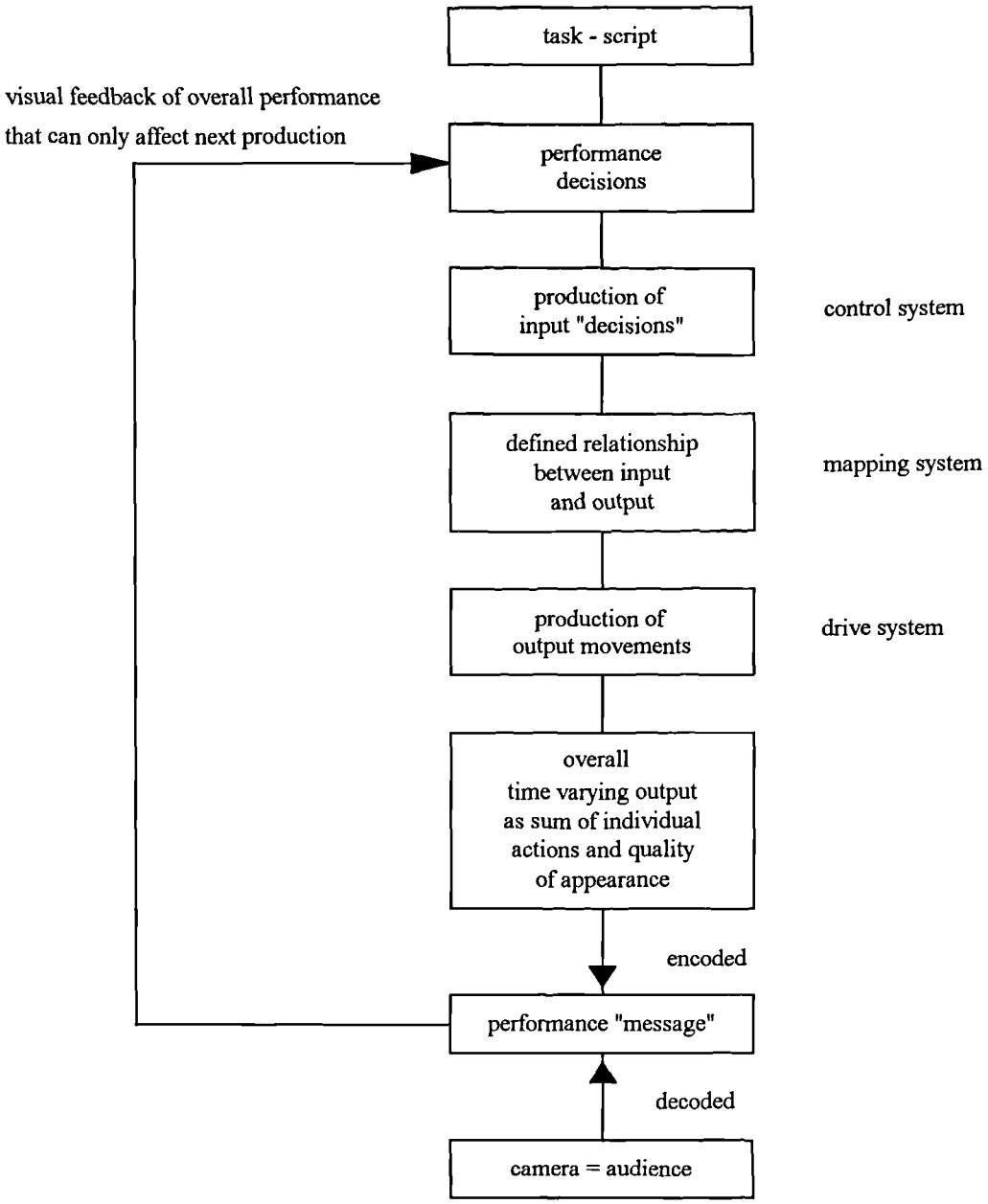


Figure 2.2 Schematic Representation Of Animatronic Performance Systems

2.3.2 Previous Animatronic Control Systems

Present animatronic systems are the result of techniques developed in hand puppetry. The most straightforward production of animated actions was by the direct manipulation, through the performer's hand, of the inside of the character's face. The hand and fingers represented the control input and the drive output in the overall system as shown in the block diagram of Figure 2.3.

Direct cable action was developed as a result of the physical requirements in a character's design. In this type of system, wire cables were attached to the inside of the face and then directly linked to individual hand controls. Each hand control, or lever, operated through a single degree of freedom and allowed one physical action to be produced on the face. The degree of freedom, (D.O.F.), defines the number of axes of motion for each input element. The overall system, therefore, required multiple performers to produce the overall expressions. This is shown in Figure 2.4. This technique for animation was also limited by the physical linkage between control and face. For this reason, radio controlled servo mechanisms were adapted to produce the drive actions, via mechanical linkages, in the character's face with control still from individual performers as shown in Figure 2.5. Both of these methods were limited by the problems of co-ordinating the actions and decisions of multiple performers to ensure that the correct overall movement and timing were achieved.

2.3.3 Present Henson Animatronic System

The present system, developed by the sponsors, overcomes these limitations by using a multiple input, hand control, system coupled with a processor system. The processor defined as the Henson Performance Control System (HPC System) allows the performer to define the mapping relationships between any individual control and any number of output drives by creating 'soft' wire connections between them. These maps can then be stored and edited easily and provide the performer with the ability to control all the resultant actions. Figure 2.6 shows a representation of a present animatronic system. It can be considered as three distinct sections; the acquisition of performance control signals, the production of output actions based on drive signals and the mapping between control input and drive output signals.

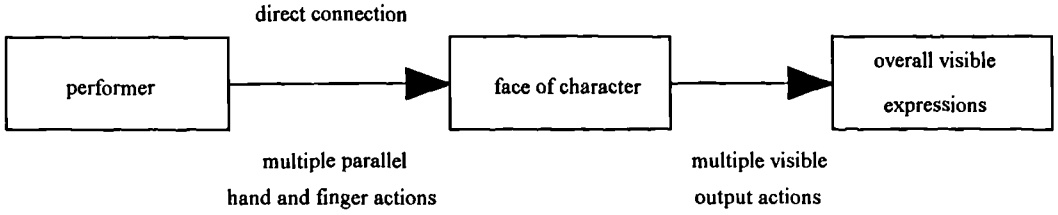


Figure 2.3 Block Diagram Of Direct Hand Control Animatronic System

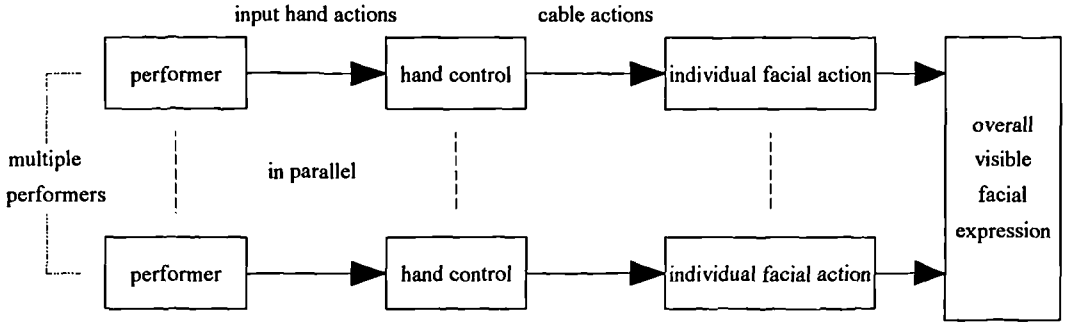


Figure 2.4 Block Diagram of Multiple Operator Direct Cable Animatronic System

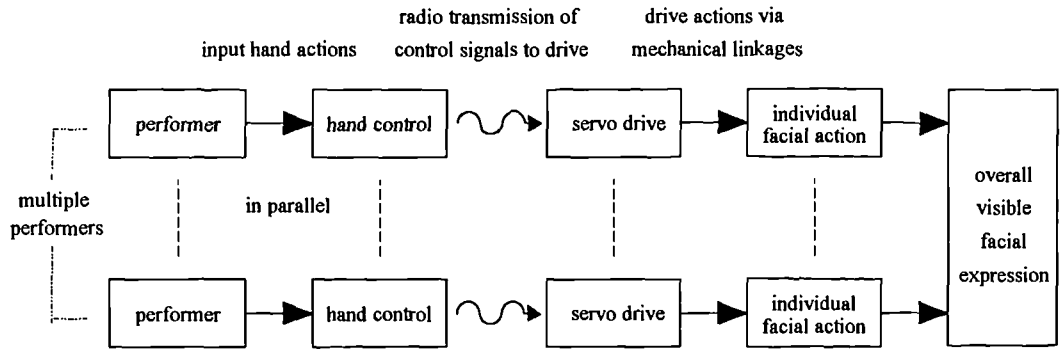


Figure 2.5 Block Diagram of Multiple Operator Radio Control Animatronic System

The overall system design is based on a principle of neutrality, where the drive system will remain at rest whilst no control signals are inputted to the system The resultant effect of this is that the overall face is said to be in a neutral setting or that it is 'expressionless'. Therefore any control input by the performer will result in a distinct facial change from the neutral face.

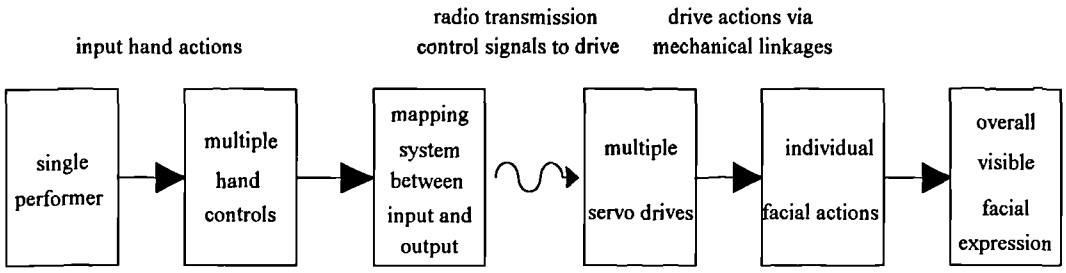


Figure 2.6 Block Diagram of Single Performer Remote Control Animatronic System

2.3.3.1 The Control Input System

The present technique for control input consists of two joystick controllers as shown in the representation in Figure 2.7 and the photograph of Figure 2.8. A "waldo" is any mechanical agent controlled directly by a human limb. Their design is based on the kinematics of the human upper limb to allow the maximum amount of input from finger, hand and arm actions. These input actions are measured either through analogue levers, with a single degree of freedom, or joysticks, which have two degrees of freedom. The displacement of each control is measured by a linear scale potentiometer at the joint of rotation producing continuous, analogue signals.

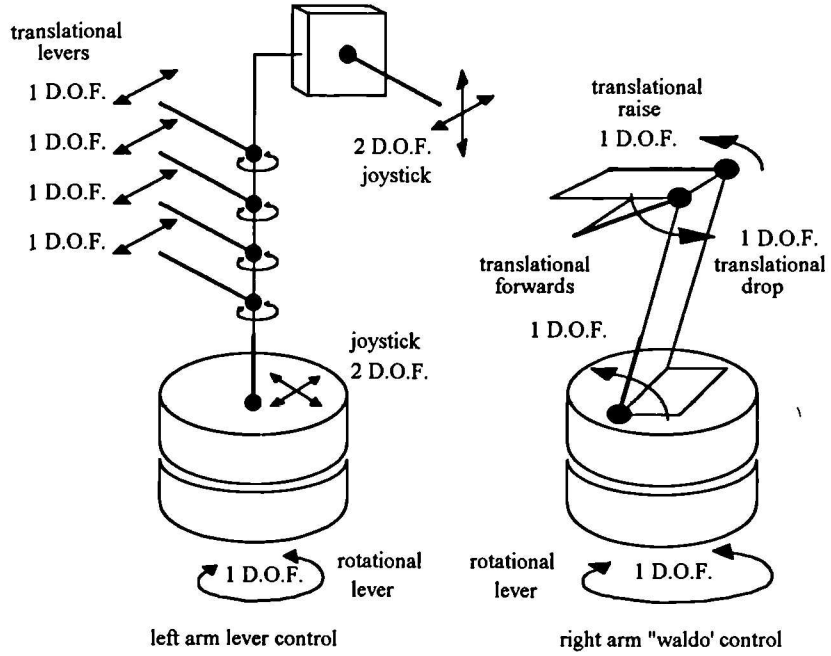


Figure 2.7 Basic Representation Of Multiple Input Hand Control

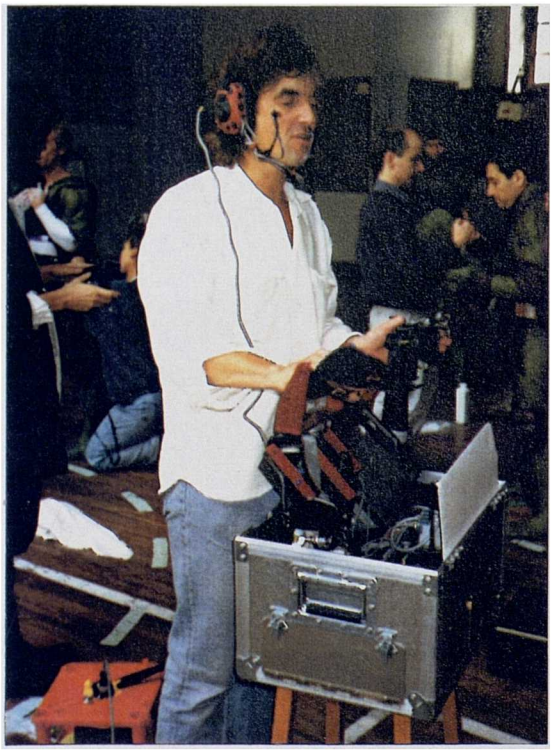


Figure 2.8 Photograph Of Henson Multiple Hand Control Input System

2.3.3.2 The Drive Output System

The drive system manipulates the 3-D animatronic face to realise the desired output expressions and performance. The main objective of the system is to produce actions, through the deformation of the model's skin, which resemble life-like facial expressions. These displacements should be of sufficient magnitude to be visually distinct and should be produced at such a rate to produce the smooth motion associated with actual facial actions. They must also be able to combine with other actions to produce perceptually correct overall expressions. The drive system can be considered as the combination of the following elements; the drive mechanism, the mechanical linkages between the drive and skin, and the face itself.

The drive mechanism can use a variety of actuator types, such as hydraulic or pneumatic, but this research concentrates only on the use of servo motors. The servo drive is a direct current (d.c.) motor whose output shaft position is defined by an input signal in the form of a varying pulse width, known as pulse width modulation. Closed loop feedback is incorporated to damp the motor action to prevent any overshoot. The pulse width signal is definable within the HPC system as a digital integer value, or motor offset, between –127 and +128, where zero always represents the neutral position of the motor. Previous research by Hensons has shown that the servos used throughout this research were entirely satisfactory and no development work was undertaken within this project.

The face is constructed from latex rubber which is sufficiently flexible to simulate actual skin. Similarly, the materials used are classed as the most effective for current applications and no research was undertaken on their performance.

The mechanical linkages represent the method for translating the positional changes of the drive mechanism into the physical displacements of the facial skin. These physical displacements are defined by animatronic designers based on their interpretation of the tasks required by the script. The design of the system, therefore, has to overcome a number of different constraints and is recognised as a highly specialised engineering problem. The process of design has to consider the following criteria:

1. the definition of the type of overall expressions and their component actions required by the script;

2. the separation of these expressions into distinct mechanisms to produce the individual actions which may also interact with each other;

3. the definition of the magnitude and direction for each action in 3-D; and

4. the selection of the type of drive and mechanical linkage required to produce the individual actions at sufficient rate and with smooth quality of movement.

Once the system is constructed, the designer will define positional limits for each individual drive, in terms of motor offset values, to prevent physical damage to the drive, linkage or skin. The performer can then control the drive, and hence the visible displacements, between these parameters. Alternatively a set of smaller parameters, or reference limits, can be defined depending upon the actual size of movement required. Each reference represents a final visible displacement of the facial skin.

2.3.3.3 The Control (Mapping) System

As stated earlier, the control, or mapping, system used throughout this research is contained within the HPC system. It is designed to allow the performer to connect any number of control inputs to any permutation of output drives. To achieve this, a map is created by the performer which links a limited set of known input reference parameters to an equivalent number of known drive references. Between these limits, the characteristic of the map is derived by linear interpolation. For full details of the mapping theory refer ahead to section 4.3.

2.3.4 Performance Control Of Lip Synchronisation

Lip synchronisation using the present system is dependent upon the pragmatic decisions of the performer, which adapt similar techniques to those discussed in cell animation in section 2.2.6.

Given the lines of dialogue in the script, or a recording of the speech, and the description of the overall emotional content, the performer must execute a number of complex functions to produce the correct lip actions. The first step is to break the overall speech down into the primary visual speech segments and consider the actions required to blend between them to convey continuous speech. The next step is to identify the best method of constructing distinct lip shapes from the drive actions available. This can be achieved by creating overall speech expressions from multiple drive actions which are mapped to a single control. Alternatively individual actions are mapped to single controls and the lip movements are produced by multiple physical inputs. For the final performance, all of the hand control actions have to be produced to create the overall animation of speech and expression.

These types of hand control systems along with more pragmatic performance approaches are commonly used throughout the industry. Techniques have been developed to record and playback specific lip actions to allow the performer to concentrate on the expressive signals.

"Total Recall" (1990) contained a number of examples of lip synchronisation. The 'Johnnycab' head was constructed with servo driven mechanisms and the actions were controlled by a computer with a sequence of key frames devised, pragmatically by a performer. The second example is a mutant head protruding from an actor's stomach. The head was a combination of direct cable operation of facial and head actions by 15 performers and servo driven cable lip actions. Each mouth shape was created one step at a time to synchronise to the sound track **[Cinefex43]**. "Gremlins II" (1990) contained one specific character that was required to lip synchronise to an existing soundtrack. The servo driven mechanisms were driven using a computer system that recorded the performer defined inputs at half speed and played back at normal rate, the same principle as motion control, with the other facial expressions added through real time joystick control **[Cinefex46]**. "Aliens III" **[Cinefex50]** and "Death Becomes Her" **[Cinefex52]** both contained human faces which had lip actions synchronised in the same way.

These techniques allow the performer significant control over the final animated output. However the creation of accurate lip synchronisation, which is an essential part of the overall performance, is the product of a number of complicated mental and physical processes. The complexity of this technique would suggest that the

development of a system to automatically derive these control signals would greatly enhance the performance by allowing the performer to concentrate on the development of the more esoteric ideas of character performance.

2.4 Other Methods Of Performance Control and Facial Action Recognition

In considering the development of performance control methods for animatronics, it is necessary to review other existing techniques used in the computer generated facial animation. The present methods to control the performances are either by the pragmatic decisions of the programmer, in the form of program driven animation, or through the automatic approaches of analysis of the speech signal, both acoustically and visually.

2.4.1 Program Driven Facial Animation

Program driven animation describes the technique for producing the control signals in the form of commands, via a computer, which will create the desired sequences of output actions. Using pre-defined descriptive terms, the programmer controls the type and intensity of actions required with respect to time based on their pragmatic assessments of the overall script. This technique originated from the methods of the traditional cell animators who prepared exposure sheets to plan the sequences of animated actions.

[Pearce86] developed a facial animation technique based on the specification of the phonetic script by the animator. By defining and then applying keywords, from a hierarchical script, sequences of movements were created which were animated using Parke's mesh model **[Parke82]**. Synchronised speech was achieved by using the same encoding of words to drive both the visual and audio synthesised outputs. The use of synthesised speech is limited in that it is difficult to achieve natural rhythm and articulation. Each keyword is represented by a key frame and the final animation produced by interpolating between these frames. Each function was defined as follows; the part of face to be moved, the type of movement, the initial and final

frames for action, the parameter value of the action at the final frame and the type of interpolation. For example, {jaw open, start frame number 12, stop frame number 25, parameter value 0.8 (80% of full range), linear}.

[Magnenat-Thalmann89a], **[Kalra91]** developed a hierarchical technique to control the facial animation. This has been presented in their film "Rendez-vous à Montreal" in 1988 with synthetic representations of Marilyn Monroe and Humphrey Bogart. The face is represented by a set of facial parameters based on the abstraction of facial muscle actions rather than the muscles themselves.

Actions of the abstract muscles can be controlled on three levels; parameter, expression and script. At the lowest level; parameter, individual muscle actions are grouped as simple overall actions. Examples of these are lip_compress, lip_protrude and left_zygomatic. The animation of the lips is then produced, by defining in a program, a time varying sequence of these parameters. At the expression level, more complex entities are constructed from combinations of parameters. Once the expression is created then the parameter level can be ignored. The intensity of the expressions can be modulated between 0 and 100% and they can be blended with other expressions through the use of multiple, parallel sequences, or 'tracks'. At the script level, a sequence of facial expressions is defined for a certain period of time. The values of the muscle parameters at specific key frames are defined by the programmer and then the final animation is created by interpolating between them.

A similar technique was proposed by **[Patel91]** with the levels defined by muscle, action unit and expression. The user defined the start and stop frames, the type of interpolation and the level of intensity. **[Choi91b]** also developed a hierarchical technique based on Ekman's Facial Action Coding System (described in detail in Section 3.4), with the parameters of each feature point at the base level and the emotional descriptors at the highest.

These techniques for animation controlled by programmed codes have a number of drawbacks. The production of animated sequences is one of trial and error based purely on the pragmatic decisions of the programmer. These methods are not real time and are not possible to automate and are therefore time consuming and inflexible to changes in the overall script. It is also limited by the terms used to describe the actions.

Synchronised lip movement is based on the programmer's correct assessment and creation of each speech segment, their correct timing and intensity and the ability to program the interactions that occur in continuous speech. Further problems exist in generating the natural rate and rhythm of facial actions and in producing the correct blending between expressions. The changes from one expression or segment to the next are not necessarily linear and different individual actions may change at different rates.

2.4.2 Acoustic Driven Facial Animation

The movement of the lips during speech is directly related to the physical production of speech (described in detail in Sections 3.2 and 3.3). For this reason, a number of techniques have attempted to control the actions of the lips automatically from an acoustic soundtrack, allowing the extraction of vital information on the correct rate and rhythm. This type of animation has a number of applications apart from entertainment, specifically in broad-band visual communication systems, **[Morishima92]**, and in the development of tools to help the audibly impaired **[Vila90]**, **[Girard88]**.

The most naïve approach for automatic lip synchronisation is to recognise the start and finish of each word and open and close the jaw accordingly with the degree of opening proportional to the loudness of the sound. It is obvious that this does not reflect natural articulation (c.f. the description of speech production in section 3.3), and the use of this method results in unnatural animation **[Lewis91]**.

The following methods are based on automatic speech recognition systems. Most of the present systems identify words through the transformation of the speech signal into a simplified representation. Such systems are only capable of identifying isolated words due to a number of problems as highlighted in **[Lewis91]**.

[Lewis87] and **[Lewis91]** proposed a technique to automatically identify mouth shapes, or visemes, corresponding to a given speech segment, or phoneme by analysing digitised speech using "linear prediction" rules. Their approach was to obtain a representation of the speech as a timed sequence of phonemes and then establish a correspondence between phoneme and mouth position in order to drive

Parke's face model. The defined set of mouth shapes obtained should therefore be "a compromise between robust identification and sufficient visual information for animation" **[Parke82]**.

[Hill88] used a technique based on the rules applied by speech synthesisers, such as the speak and spell toy machines, to convert text inputs into acoustic outputs. Each mouth shape was defined as a set of parameters in a library of all the possible visual segments. An utterance was then entered as text and converted into a sequence of these segments. The output algorithms then apply pre-defined rules, on the possible interactions, intonation and rhythm, to create a continuous set of parameter values. This was then applied to the control points on Parke's face model **[Parke82]** to produce the lip actions along with the synthesised speech.

[Morishima91a] developed a real time system to realise intelligent communication systems using two types of speech to image motion schemes The first approach was to analyse the input acoustic signal by vector quantization using code books. Each code word corresponded to a distinct acoustic group, known as a phoneme, and it's comparable visible lip shape which was recorded and stored during training procedures. The second approach used neural networks to train the mapping between speech and image parameters. The final animation was then produced directly from speech input in real time using parallel processing systems. **[Morishima91b]** also produced a text to image conversion scheme using similar techniques.

These methods for deriving control parameters from the acoustic signal have a number of limitations. Acoustic recognition systems segment the continuous acoustic signal into discrete units that will hopefully correspond to phonemes. Even if segmentation is correct, given possible noisy conditions, the identification process is complicated by the variability with which a phoneme can be spoken. The description of a speech wave is significantly reduced when it is converted into a symbolic representation. At every successive encoding, additional information from the original signal is lost. A significant amount of processing has to be applied to develop sufficient rules to compensate for the problems caused by the segmentation. These rules have to deal with the complex co-articulation problems that not only exist between consecutive phonemes but as many as five segments ahead **[Pelachaud92]**. Acoustic recognition systems also fail to identify the effects that expressive blending has on the visible speech production which produce the variety in performance.

2.4.3 Image Driven Facial Animation And Automatic Visual Speech Recognition Techniques

Given that the overall objective is the animation of facial expressions and more specifically the lip actions associated with speech, a number of projects have attempted to extract control signals directly from images of speakers. A number of research projects have also developed automatic visual speech recognition systems. In these studies the goal is not the extraction of animation control parameters but the recognition of the phonetic qualities associated with each visual image.

The measurement of articulatory movements presents a number of problems. These movements are small and rarely, if ever, exceed 25 mm in amplitude. For this reason, the techniques for recording measurements must be sensitive relative to the overall size of the head and it is important to separate these articulatory actions from the effects of the global head movements of the talker **[Brooke83]**.

The methods for producing animation from facial images can be considered from two distinct approaches: real image key framing and parameter extraction. Real image key framing is the method of producing animated sequences using actual recorded images of expression with interpolation between the defined key frames. Parameter extraction techniques analyse image frames to derive information on the changes produced by facial actions.

In entertainment, where realistic facial motion is required, the actor's face is captured through the technique of rotoscoping. The actor produces different expressions which are captured and then used as key frames in the final animation. The most accurate method is to rotate a low power laser through 360° collecting digital information on the shape of the actor's head **[Robertson88]**. Other methods include the use of two cameras to capture still images of expression which then have grids manually applied before being digitised as key frames **[Swain92]**. Both methods have a number of drawbacks. Firstly, it is not a real time operation as the capture of static facial expressions can take up to 15 seconds to be recorded. Secondly, the animation is defined by the programmer through time consuming matching of the correct key frames to the soundtrack. Finally, significant processing and data storage are required to generate the final animation.

[Petajan88a] produced real time measurements of the oral shape using a miniaturised head mounted television camera, to remove the global head action effects. The input image frames were digitised, clipped and thresholded to produce binary images that were then smoothed to produce compact code books of images. **[Petajan88b]** used the code books to translate incoming images of spoken numbers into corresponding symbols. These symbol strings were then compared to stored sequences representing different words in the vocabulary. Similarly, **[Brooke89]** used the same apparatus to investigate vowel recognition. This process is computationally intensive and requires efficient image encoding to perform a reasonable number of comparisons, whilst the early encoding and categorisation of the continuous speech signal results in the loss of relevant speech information.

The extraction of facial parameter changes from the analysis of time varying images has been attempted in a number of ways. **[Terzopoulos90]** developed an image processing technique to estimate the control muscle parameters for Water's muscle model **[Waters87b]**. The method tracked deformable contours, analogous to known muscle contractions, on the speaker's face and then related the changes directly to the facial model. For example, the nasiolabial ridge (c.f. Figure 3.4), was highlighted by black make up and then tracked through series of images. The measured changes in shape and position were mapped to the zygomatic muscle parameters on their model. This method is limited by the significant amount of image processing algorithms necessary, both to identify the image contour from the rest of the facial features, and, secondly, to determine the actual change in shape and direction of the contour and finally to map these changes to the driving system of the model.

[Williams90] proposed a technique to directly acquire the expressions, in the form of textures, through video tracking of a real face. Reflective spots were positioned on the face and their initial co-ordinates were registered by operator. The system automatically tracked each point to generate positional changes relative to the initial frame which were then applied to comparable points on the image model. Again, the results were limited by the large amount of processing required to extract the parameters from each image but the method is at least based on the principle of extracting a live performer's actions to control the animation.

[Brooke83] produced a system to record, measure and analyse the visible articulatory movements of a speaker's face performing vowel-consonant-vowel (VCV) syllables

from video images. The face was marked at key articulatory reference points with white gummed paper to reduce the amount of image processing required. By use of a mirror, the 12 articulatory points were visible in all orientations within each video frame. A similar experimental set-up is described in Section 4.3. To overcome global head actions, Brooke developed a 3-D reference axis to allow free and natural actions. By referring to these references, all point movements were measured relative to a standard head position. The resultant, purely articulatory, movements of the measured points were then extracted as plots of time-varying displacements from the neutral facial position in each of the three dimensions. **[Finn88]** produced an automatic optical speech recogniser based on similar principles to those of Brooke. Data was collected through video recordings of the face with highly reflective spots positioned at key articulatory points. From each frame, the co-ordinates of each dot were identified manually and then digitised and stored. The co-ordinate information was converted into a set of distance measurements and each utterance was represented by the overall pattern of these distances across time. Recognition was achieved by comparing the distance patterns from the input images to a set of pre-recorded training tokens.

Brooke's technique has a number of advantages over other existing image recognition systems. It reduces the complex facial system to a limited set of clearly defined articulatory points. This in turn reduces the complexity of image processing required to extract the control signals. The continuously changing data from the recognition of these points is directly related to the visible elements of speech and avoids the segmentation of the speech signal. It remains limited in a number of areas. Firstly, a significant amount of processing is still required to extract the articulatory gestures from both the global head movements and also from each image frame. Secondly, the inter-relationships between the key points, which may be of importance in animation, must be compensated for in the animation model to avoid possible incorrect lip shapes.

2.5 Summary

This chapter has described the research interests in animatronics and, specifically, in the performance control techniques for facial actions. The movements of the lips during speech can provide vital clues about what is being said. The animation of these movements, when synchronised to a given soundtrack, represents an important part in a performance. They must produce a plausible representation of the real articulations that would be required in speaking the words. Of importance to the successful animation is the intensity, timing, rate and rhythm of the actions, as well as the resultant blending effects produced in continuous speech or by other expressive messages. The control of this type of performance in animatronics and computer generated animation has been attempted in numerous ways.

In animatronics, the current method of control is through a manual joystick input system coupled with a purpose built processor that allows each input to be mapped to any number of the output drives. This enables the performer to produce sequences of complex facial expressions through the combined manipulations of the analogue controls. The relationship between input hand movement and output facial expression is not an inherent one and special skills have to be developed by the performer.

A similar approach in computer animation is that of program defined control. This method generates lip animation through the typed descriptions of the programmer. This has a number of limitations since it is based on trial and error techniques, time consuming and inflexible to changes in the overall task.

Both of these techniques rely purely on the pragmatic judgements of the performer or programmer. These decisions are based primarily on their experience of performance and of the control system rather than from a formal rules' system. An alternative approach that can automatically manipulate the natural signals created by actual speech production in order to derive the control input, has the potential to improve the current animation of lip synchronisation.

The most obvious approach is to extract the control signals from the analysis of the actual acoustic soundtrack. The majority of these analysis techniques are based on the segmentation of the continuous speech signal into phonemes or isolated words which result in the loss of vital information from phonetic co-articulation. To overcome

such problems requires a large amount of processing, and other important expressive signals are excluded by these techniques. This results in inferior performances lacking in the variety of action associated with natural speech.

Since the overall objective is the production of visual speech animation, an alternative approach is proposed to extract the desired control signals from visible facial movements. **[Williams90]** and **[Robertson88]** have both highlighted the possibilities of using human facial performance as a control tool for automatic lip synchronisation and expressive performances. **[Brooke83]** has successfully extracted information from the physical changes of the face, through the recognition of a set of key articulatory points. This technique allows the capture of expressive and co-articulatory information from a limited set of points rather than the overall image and avoids the errors resulting from the segmentation of the speech signal.

In summary, it is proposed that the development of a real time technique to accurately sense the motion of the performer's face would enhance the control of an animatronic character's performance. The proposed solution is based on the principle that the complex facial system can be reduced to an optimum set of measurable points at key positions on the facial surface. The sensing of these points can produce sufficient control data to describe the overall visual expression since the movement on the face is a direct result of actual speech production and facial expression. This solution has the advantage of allowing the performer a more innate form of control and in principle should reduce the need for highly specialised skills. The principles of this new technique and the proposed method of solution are discussed in Chapter 4.

[Thomas81a] states that successful animation is specifically attributable to the long hours spent studying living animals, and is directly proportional to how accurately the animator has understood the kinematics, timing and structure of the subject but that it is not the mimicry of live action film. Therefore before developing the above performance control system, the following examination of the human face and it's role in communication is essential:

1. to understand the physiology of the face in terms of how visible actions are produced;

2. to consider the physiology of speech production and the visible facial actions produced as a result;

3. to examine all the factors which can affect these visible elements specifically in continuous speech;

4. to derive the set of primary visual actions associated with, firstly, visible speech and, secondly, with non-verbal communication. This set of actions represents the system criteria for both sensing and animation systems; and

5. to define an optimum set of key points on the face that can provide sufficient information on the defined primary actions to allow the automation of lip synchronisation.

Chapter 3

Facial Communication Systems

Chapter 3

Facial Communication Systems

3.1 Introduction

The previous two chapters have highlighted the research interests in the areas of facial expression and lip synchronisation and emphasised the potential use of visible facial actions as a source of control for animatronic characters. **[Thomas81a]** concluded that "the perception of communication signals from animated characters is based on the viewer's experience of understanding signals from the human form". Therefore, to achieve any form of motion realism, be it animation or simulation, a clear understanding of the underlying anatomy is required. A similar understanding of the method of production of the visible actions related to speech and expression is also necessary. Further to these studies, it is necessary to establish a suitable form of notation to describe the facial actions associated with both speech and expression.

Communication between humans is perhaps the most critical facet of human behaviour and indeed of our existence. Figure 3.1 shows a block diagram to illustrate the concepts of communication that occur through all channels to varying degrees. "The encoded message sent by the speaker is decoded by the listener to enable understanding between the two" **[Massaro87]**. In the ideal case, a normal hearing person in noise-free conditions, even with a clear, full face-to-face view of the speaker, will perceive the majority of speech acoustically through the ear in the form of a learnt language **[Montgomery87]**.

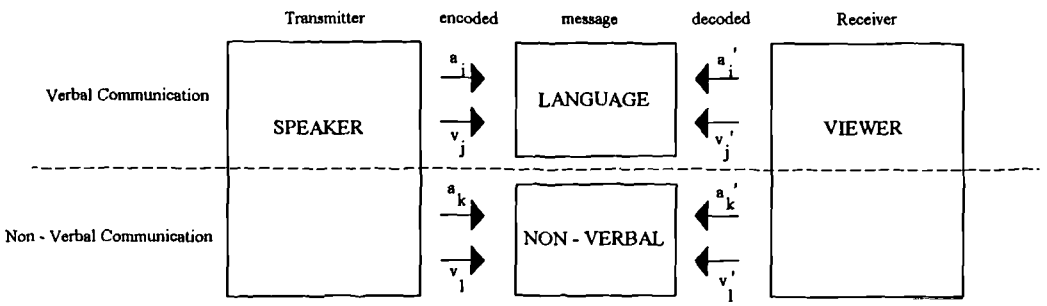


Figure 3.1 Simplified Block Diagram Of Human Communication System

In Figure 3.1 a_i represents the acoustic speech signal, v_j represents the visual speech signal, a_k represents the non- verbal vocalisations and v_l represents the non-verbal visual signals (facial expressions).

Non-verbal communication is defined by psychologists to include a number of distinct methods by which people can convey and decipher messages and meanings without the use of speech and language. These signals can be emotional or conversational and all play a major role in our behaviour **[Argyle88]**. **[Argyle88]** defined the following areas as the main channels for non-verbal communication;

facial expression,

the actions of eyes in direction of gaze or pupil dilation,

gestures,

body actions,

posture and

vocalisations.

Other possible signals are proximity, body contact and appearance.

The animation and recognition of all non-verbal signals are essential in creating the 'illusion of life' but this research has concentrated solely on the area of facial expression. Though the eyes play an important role in facial expression, they are omitted from this research because of the problems of recognising and modelling their

actions and also the lack of work available in defining their actions. The tongue is ignored for the similar reasons though it is acknowledged that it plays a significant role in speech production **[McGrath84]**. Other visible effects on the face not considered include eyelid closure due to the orbicularis oculi and skin colouring due to blood circulation. Each of these areas warrants a full study of its own.

Section 3.2 will describe the physical attributes of the human face. Section 3.3 will discuss how speech is produced, what visible information is produced and how this is perceived by visual means in the form of lip or speech reading. This review will consider both the discrete and continuous properties of speech and the problems that occur as a result. Section 3.4 examines the role of non-verbal facial expressions as communication signals, both as emotive and conversational signals, and presents a possible notation for describing facial actions.

3.2 Physiology Of The Human Face

The human face is defined here as the frontal half of the head, from ears to nose tip, covering the area from the base of the chin to the hairline. The face is a highly complex physical system consisting of skull, muscle, skin, and the body organs; eyes, ears and tongue, **[Warwick74]**, and though the main interest is in the final visible surface displacements there is a need to understand the anatomy of these fundamental elements.

The human skull, as seen in Figure 3.2, can be considered as two major bones; the *cranium* (fixed upper skull) and the *mandible* (jaw) which is freely jointed. The *cranium* is comprised of a number of discrete regions which form a rigid structure over which the skin may slide. The following four regions have relevance to the research:

1. the *frontal* bone in the forehead which forms the eyebrow ridge;
2. the *nasal* bones which form one of the main recognisable features;
3. the *maxilla* which forms the roof of the mouth and the location of the upper teeth; and

4. the *zygomatic* bone which is the prominent cheek bone **[Warwick74]**.

The *mandible* rotates vertically about a horizontal axis near the ear, the *ipsilateral condylion*, which is palpable when rotation occurs. The resulting action of elevation and depression of the mandible play a major role in speech production. The mandible can also protrude, retract and perform lateral, side-to-side, movements which produce idiosyncratic expressions **[Hardcastle78]**. The lower teeth are embedded in the *mandible* and when viewed from the front they, along with the upper teeth, are the only visible elements of the skull. This plays a significant role in speech visibility **[McGrath84]**.

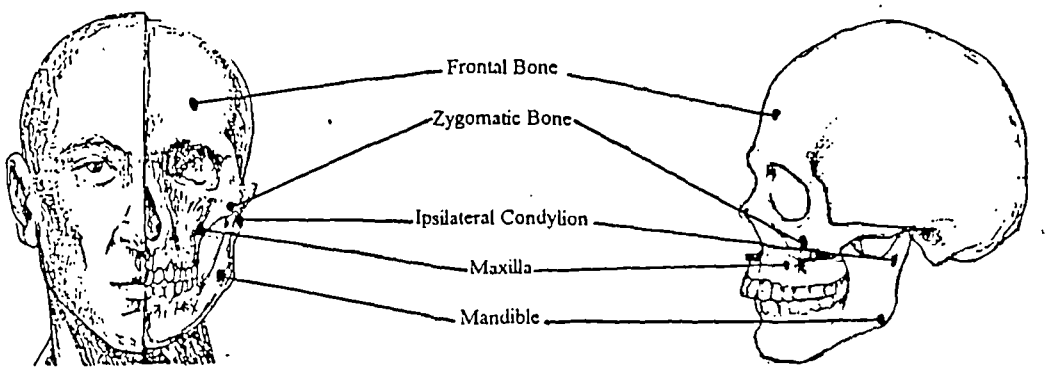


Figure 3.2 The Human Skull

In basic terms, the muscles of the face form the 'driving' elements which force the skin and mandible to move in certain directions and, when the muscle actions are combined, in different ways which result in the visible surface motions. The face contains the most complex system of muscles within the body with 40 individual elements **[Warwick74]**. Their inter-relationships can be seen in Figure 3.3. Each muscle is comprised of a large number of individual fibres which intertwine at certain points to define the muscle's orientation and action. The basic types of muscles in the face are orbital, sheet, perpendicular, horizontal and oblique.

points to define the muscle's orientation and action. The basic types of muscles in the face are orbital, sheet, perpendicular, horizontal and oblique.

In general, the muscles originate in the bones of the skull, and their fibres combine with the internal fatty tissue bonding it to the skin. Only the *orbicularis oris* has no bony attachment. The skin (*epidermis*) is subjected to considerable 'mechanical' stress by these internal and other external forces. It can deform over and around the underlying bone structure causing the visible changes such as facial creases and wrinkles as well as the displacements resulting from the muscle actions **[Warwick74]**.

The two main regions of interest are defined as the upper face, the brow region, and the lower face, the area surrounding the oral cavity. The actions of the brows are primarily the result of medial and lateral parts of the *frontalis* muscles acting perpendicularly. The physical characteristics of the lower face are complicated due to the large number of muscles intertwined with the main orbital muscle, the *orbicularis oris*. The other primary muscles include the *buccinator* and *risorius* horizontal sheet muscles in the cheek region, the *zygomaticus minor* and *major*, the *levator labii superiorus*, the *levator* and *depressor angulii oris* and the *mentalis* oblique muscles, which all contract at some angle to the *oris* (mouth) region **[Warwick74]**. Figure 3.4 shows these primary muscle actions around the *orbicularis oris* with the arrows indicating the direction of their individual actions **[Hardcastle78]**. Appendix A describes in detail the major actions of the lips and their underlying muscle actions.

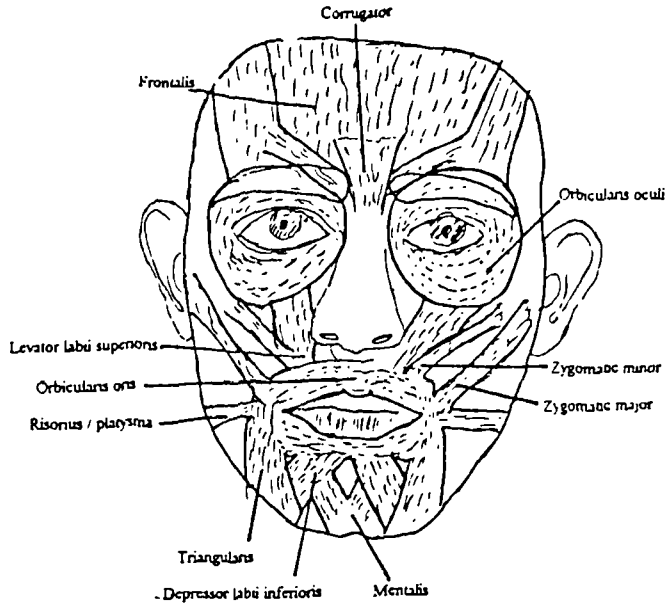


Figure 3.3 The Muscles Of The Face

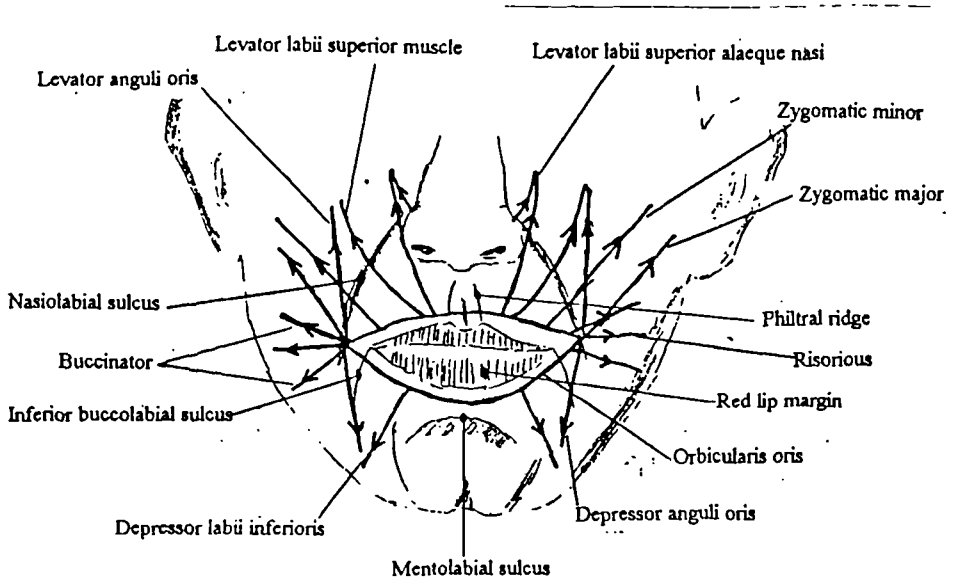


Figure 3.4 Diagram Of The Mouth Region Indicating Primary Muscle Actions.

3.3 Speech Production, Visibility And Lipreading

3.3.1 The Physiology of Speech Production

The basis for the production of all sounds from the human vocal apparatus is the effect of air flowing through the vocal tract, mostly during the exhalation of the lungs. Figure 3.5 shows the vocal tract and the relative positions of the various articulatory organs. The type of sound produced depends on whether the air passes freely through the lower part of the vocal tract (an *unvoiced* sound) or whether it is impeded (a *voiced* sound). The articulators alter the perceived pitch and frequency spectrum of each sound by either, modifying the path of the air stream or by actually interrupting its flow. The principle articulators are the pharynx, the soft palette, the tongue, the lips and the mandible **[Scherer82]**.

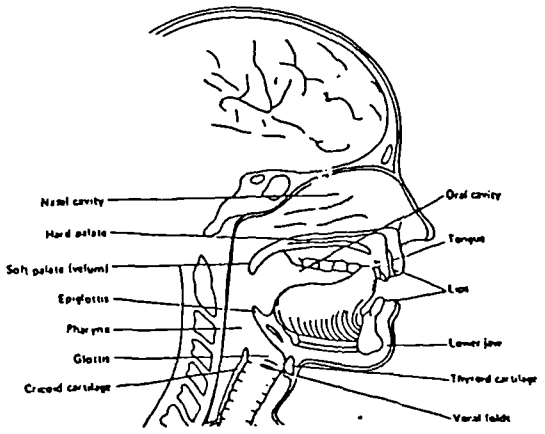


Figure 3.5 The Vocal Tract.

© **[Scherer82]**

Within any language there is a need to provide a description for the meaning of each sound segment. The standard method is to define each discrete unit as a phoneme and then represent the language spoken in terms of streams of phonemes. It should be noted that a phoneme is not a unique acoustic event but a descriptor of a family of sounds **[Jackson88]**. This classification, along with the differences between voiced and unvoiced, produces the two distinct groupings of vowels and consonants. It is important to be aware that these groupings do not necessarily have the same meaning

as the written alphabetic groups. All vowels are *voiced* sounds which are produced by the pressure used in releasing the vocal folds (*glottis*) which sets the air in vibration and is then further altered by the articulators. Consonants are produced when the free air stream is first set in vibration by one of the articulatory organs and then obstructed by another of the articulators. For this reason consonants, many of which are *unvoiced* sounds, are defined by both manner and place of articulation **[Scherer82]**.

The standard notation scheme for describing the majority of known languages is that of the International Phonetic Association (I.P.A.). For the purposes of this research only British English phonemes have been considered, as shown in Figure 3.6 and Table 3.1. See Appendix B for full I.P.A. listing and examples **[Ladefoged78]**.

From Figure 3.5 it is obvious that the majority of articulators are hidden from view. The only visible places of articulation are described on the Labial (lips), Labio-dental (lower lip touching the upper teeth), Linguo-dental (tongue touching the upper teeth) and the Mandible **[Laver80]**. By defining the vocal tract and specifically the oral cavity in terms of latitudinal and longitudinal axis settings, one can consider the way in which different sounds are produced by these articulators. Table 3.2 outlines the main visible articulatory settings.

The neutral setting is defined as upper and lower lips lightly touching with no protrusion and the jaw not unduly closed or open. For the latitudinal settings, **[Laver80]** defined nine simplified lip shapes, including a neutral shape, based on purely horizontal and vertical muscular constrictions and expansions, as shown in Appendix A. In the actual physical state, both axis settings are tied to each other and hence all possible lateral settings could occur with protrusion giving rise to eighteen possible settings. However, within the context of speech articulation the majority of these settings rarely occur and one can reduce the primary settings to rounding (produced by horizontal and vertical constriction), spreading (produced by horizontal expansion) and opening (produced by the *mandible*). In the longitudinal axis, labial protrusion is primarily caused by the *orbicularis oris* and the *mentalis* which results in the lengthening of the vocal tract. Note that the majority of protrusion results in some form of latitudinal change. Labiodental action is defined as the retraction and raising of the lower lip against the base of the upper teeth to shorten the vocal tract whilst the upper teeth retain their setting or are slightly raised. This is usually followed by some form of protrusion **[Hardcastle78]**.

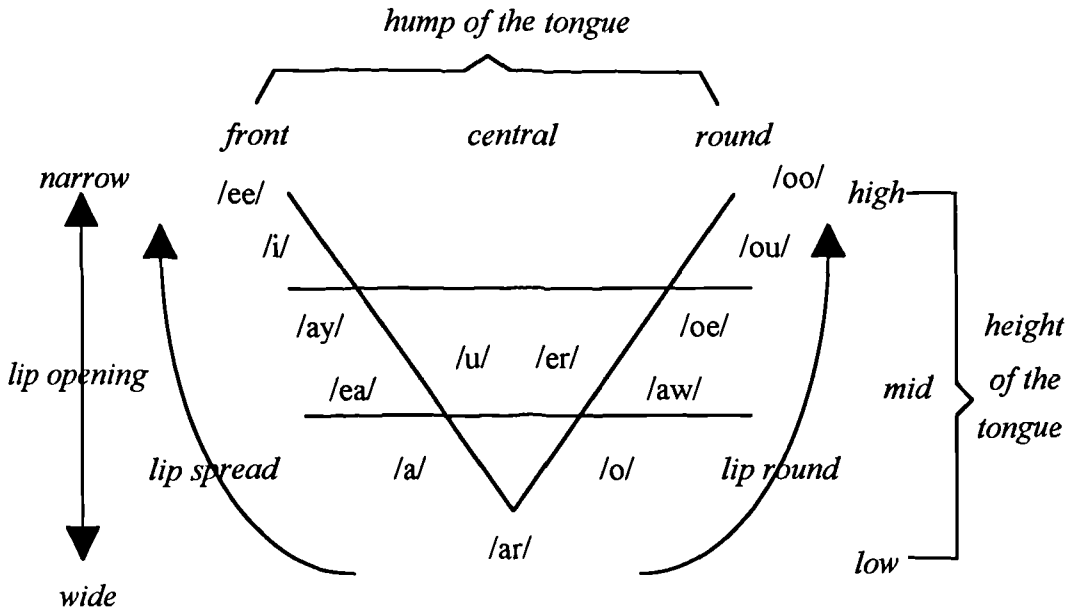


Figure 3.6 I.P.A. Phonetic Vowel Triangle

Manner Of Articulation \ Point Of Articulation	bilabial	labio-dental	linguo-dental	alveolar	palato-alveolar	palatal	velar
Plosive Stop	/p/, /b/			/t/, /d/			/k/, /g/
Nasal Stop	/m/			/n/			/ra/
Frictive		/f/, /v/	/th/	/l/, /s/, /z/	/sh/		
Affrictive					/ch/		
Approximant	/w/			/r/		/j/	

Table 3.1 I.P.A. Phonetic Consonant Chart

axis settings	place of articulation	manner of articulation
Longitudinal	labial	labial protrusion
		labiodentalisation
Latitudinal	labial	rounding (open)
		rounding (close)
		spreading
	mandible	open
		close

Table 3.2 Visible Articulatory Settings

3.3.2 The Visual Elements Of Speech

The concept of a purely auditory process for speech communication is debatable given the contrary anecdotal evidence eg. of people's dislike of phone calls (where there is no visual signal), or poorly dubbed movies. These examples indicate the significance of the visual signal in the perception process. The concept of bi-modal (audio + visual) speech perception was confirmed by the 'McGurk Effect' **[McGurk76]**. McGurk found that when a different auditory signal was presented in synchronisation with a visual signal, subjects perceived a third 'fused' signal. For example an acoustic "ba" synchronised with visible "ga" resulted in fused "da" for 98% of responses. Similarly **[Summerfield82]** drew the conclusion that "visual information on the place of articulation can reduce acoustic confusions even in ideal conditions".

The complimentary nature of the visual element to speech perception becomes more apparent if the signal is degraded by noise. The prime example of this is the hard of

hearing for whom the visual signal becomes the prominent 'message' carrier. **[Berger72]** established that the relative dependency on visual perception of speech is directly related to the viewer's hearing level. Lipreading or Speechreading are the two terms for this form of perception in which the visible actions of the lips help to convey the message. Lipreading was defined by E. Nitchie in 1912 as "the art of understanding a speaker's thoughts by the movements of their lips" **[Nitchie12]**. However research into auditory rehabilitation has lead to the term speechreading which acknowledges and manipulates other visual and linguistic clues, such as contextual or situational information, expressions and gestures, to supplement the observed lips **[Summerfield82]**. As this research is interested in the observed lip actions, with no emphasis on the context of the utterances, the term lipreading is preferred.

The same concept of phonetic grouping is adapted when considering visual speech, due to the different degrees of visibility. Many of the phonemes have similar visual characteristics, due to a hidden place or manner of articulation, and can hence be considered members of the same class. Thus, different phonemes within each unit share the same visible features. **[Nitchie12]** defined the term "homophene" to refer to visually similar phonemes that could not be distinguished by visual cues alone. The term is analogous to homophone, acoustically similar sounds, but the groupings are different. An example of homophenous words using consonantal stops follows **[Berger72]**;

man mat mad

pan pat pad

ban bat bad

[Fisher68] conceived the term visual phoneme or "viseme" to describe this discrete visual perceptual unit. The visibility of the place of articulation is limited to the lips, teeth and tongue tip. For this reason, a significant number of phonemes can not be differentiated visually. This leads to a number-to-one correspondence between phonemes and visemes rather than a one-to-one relationship. This discrepancy in

correspondence leads to confusion in the visual perception of speech and is a major stumbling block in lipreading.

A great deal of research has attempted to define the amount of speech that is visible, with ranges estimated from as low as 11% up to 57% **[Berger72]**. These figures provide an indication of the linguistic redundancy that exists. Acknowledging the existence of the phonetic division between vowels and consonants, one can consider similar distinctions within visemes. These distinctions have been derived experimentally and many factors can affect the classifications, including the research criteria **[Summerfield82]**. A definitive classification is still being sought.

3.3.2.1 Viseme Vowels

Theoretically each vowel is visually distinct since its production is achieved by the unique variation of the inter labial space in conjunction with the tongue's position, see Figure 3.6. Within actual speech, however, viseme classifications can be applied to vowels as "they form visually contrastive groups that are recognised as having distinct movement patterns" **[Jackson88]**. **[Berger72]** states that "the gradation of difficulty in visual identification based along the sides of the 'vowel triangle" with the most visually distinct at the corners, i.e. /i/ : "ee", /a:/ : "ar" and /u:/ : "oo". Table 3.3 shows the vowel groupings based on ideal (slow to normal rate and rhythm), and usual (average to rapid) articulation conditions **[Jeffers71]**. Figure 3.7 shows photographs of the primary static viseme vowels and from these, the possible visual confusions are evident.

3.3.2.2 Viseme Consonants

Given that the place of articulation for a significant number of consonants is hidden from view, confusions exist in the classification of viseme groupings. **[Nitchie50]** defined twelve visemes based on the places of articulation. In experimental conditions, **[Binnie74]** reduced this to a set of only five visemes whereas **[Lesner87]** defined seven using the same stimuli as Binnie. These different groupings can be seen in Table 3.4 listed in descending order of visibility.

These groupings, which were also confirmed by **[Jackson88]**, indicate that bilabial closures are the most visually distinct followed by labiodental and linguodental. These three visemes represent the universally recognisable movements. A possible reason why /w/ fails to be distinguished reliably is its confusion with vowel visemes, due to similar lip rounding, as well as the other approximant /r/. A specific drawback is the problem of the viseme group /t, d, n, s, l/. These are the five most common consonants used in the English language yet they are consistently misinterpreted. In fluent speech, their co-articulation effects will further increase lipreading difficulty **[Pisoni84]**. Figure 3.8 shows photographs of static viseme consonants and again the visual confusion's are obvious.

The discrepancies in the viseme groupings reflect the limitations of purely visual lipreading. Lip readers must make use of other clues available; facial expressions, gestures, the context or the situation of the conversation; if they are to understand the whole message **[Berger72]**.

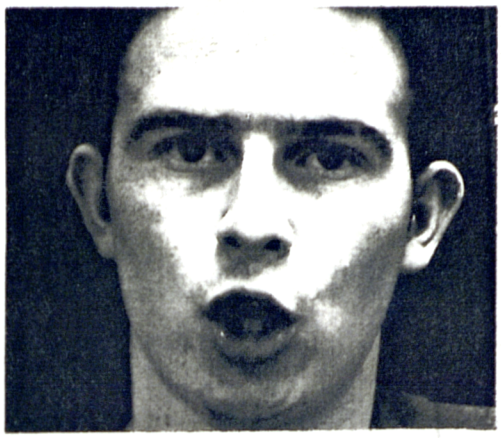
Ideal Viewing Conditions	Usual Viewing Conditions
/oo/, /ou/, /oe/	/oo/, /ou/, /oe/
/ee/, /i/, /ay/	/ar/
/aw/	/aw/
/ar/	/ee/, /i/, /ay/, /ea/, /a/, /o/, /u/
/ea/, /a/, /o/	
/er/	
/u/	

Table 3.3 Viseme Vowel Classifications

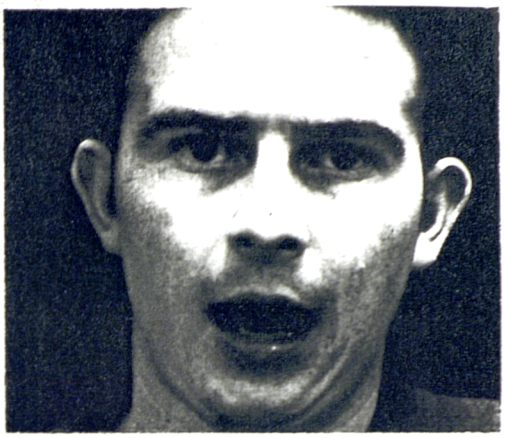
Nitchie, E 1950	Binnie, C. 1974	Lesner, S. 1987
/p, b, m/	/p, b, m/	/p, b, m/
/f, v/	/f, v/	/f, v/
/w/	/th/	/th/
/r/	/sh, ch/	/sh, ch, j/
/th/	/t, d, n, s, z, k, g/	/w, r/
/t, d, n/		/l/
/l/		/t, d, n, s, z, k, g, j/
/s, z/		
/sh, ch, j/		
/y/		
/k, g, ng/		
/h/		

Table 3.4 Viseme Consonant Classifications

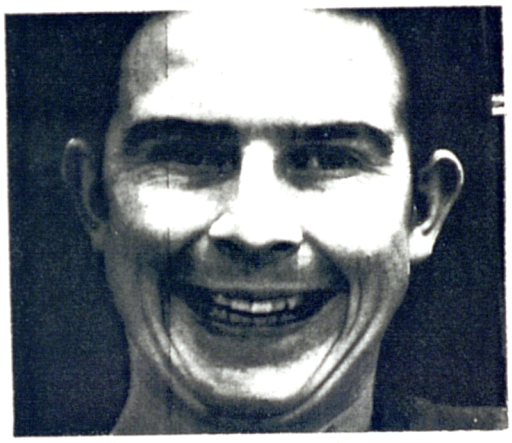
a) /oo/



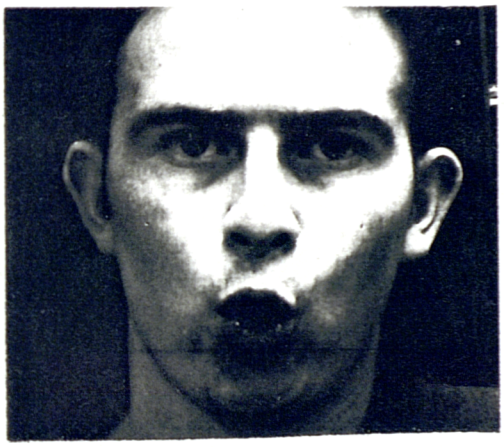
b) /a/



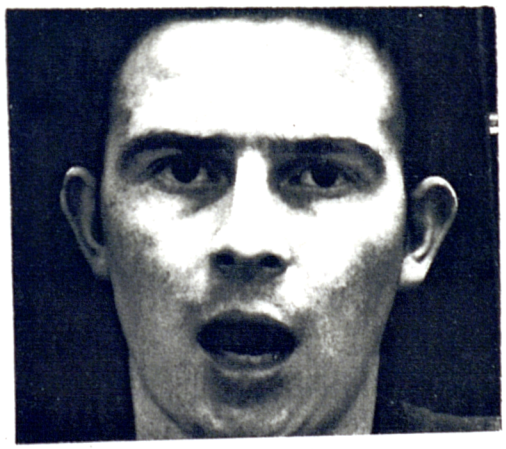
c) /ee/



d) /oe/



e) /ay/



f) /ar/

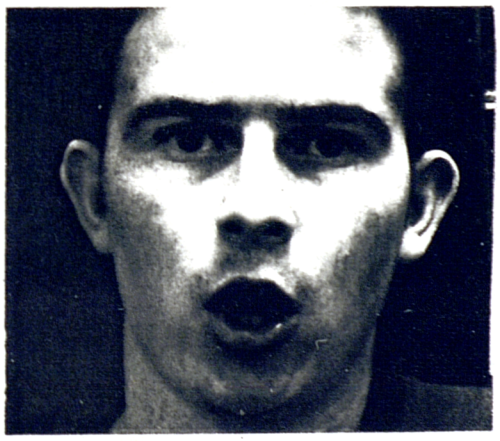
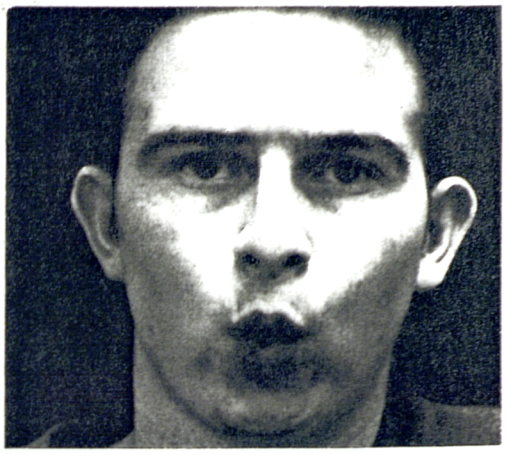


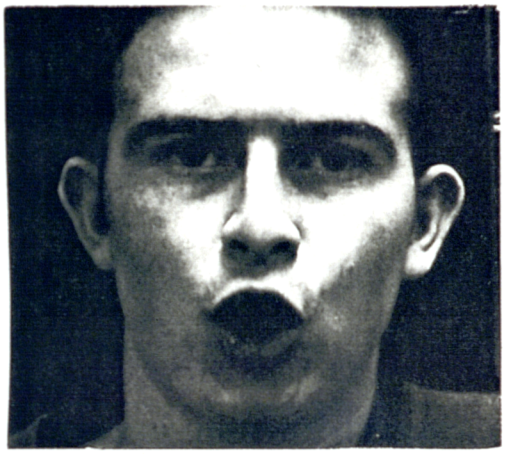
Figure 3.7 Photographs Of Viseme Vowels

a) /t/

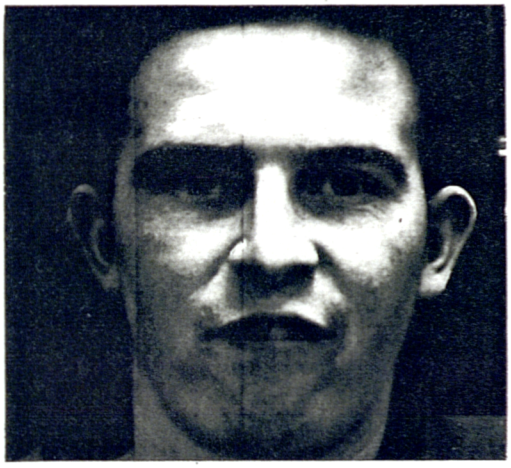
b) /w/



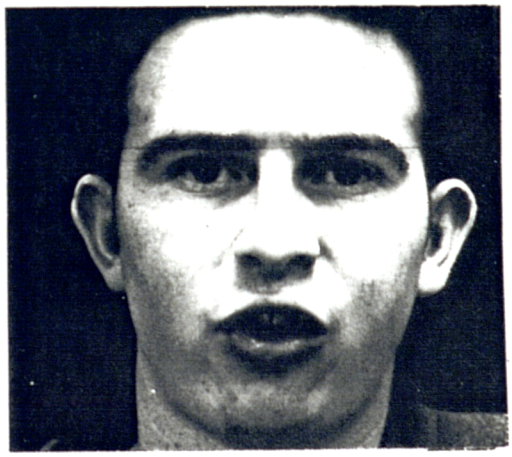
c) /sh/



d) /f/



e) /th/



f) /r/

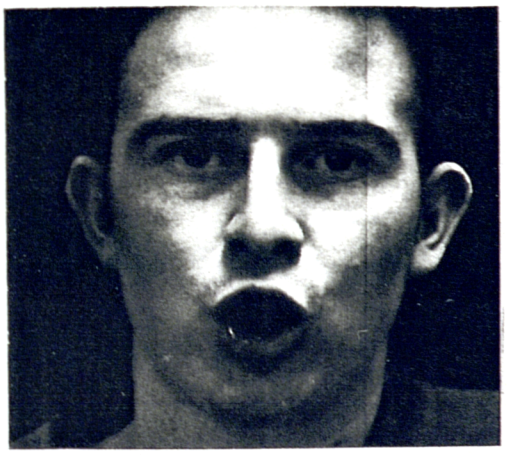


Figure 3.8 Photographs Of Viseme Consonants

3.3.3 Continuous Speech And It's Effects

Whilst phonetic descriptions prove useful in identifying and grouping discrete units of speech, actual continuous speech cannot be simply created by the 'beading' together of phonemes to form words and sentences. In simplistic terms, one cannot insert or delete segments without affecting the surrounding phonemes. In a continuous signal, interactions between articulatory gestures take place resulting in an overlapping of phonemes. The different effects that occur are either co-articulation, accommodation or assimilation.

Co-articulation has the effect of changing the place of articulation of the phoneme, and hence its visual appearance, without actually changing its phonemic qualities. **[Bengerual82]** defined Co-articulation as "the altering of the set of articulatory movements made in the production of one phoneme by those made in the production of an adjacent or nearby phoneme". For example, there is no acoustic change in the consonant /b/ whether it precedes rounded (/ar/), spread (/ee/) or protruded (/oo/) vowels yet the photographs in Figure 3.9 show the resultant change in its visual appearance.

Accommodation is the principle of minimum effort when more than one phoneme is simultaneously produced in a given time without adjustment in articulatory gestures. This can manifest itself in a partial form where the phoneme retains some of its original character or as total accommodation where the phoneme fully resembles its adjacent unit **[MacKay87]**.

Assimilation is a change in the place, manner or voicing of a phoneme such that it appears as another phoneme. An example of this is the word "dogs" where the sound of /s/ changes to a /z/ as a result of the preceding phoneme.

All of these effects can be 'progressive', left to right, 'regressive', right to left, or 'double', and both sides exert similar influence **[MacKay87]** and **[Pelachaud92]**. A different effect of continuous speech is the creation of new vowels known as dipthongs. These are vowels that start as one vowel and finish as another, for example /aU/ in "loud" or /ai/ in "buy". These effects cause further serious problems in visual speech perception. **[Jackson88]** concluded that "the majority of visemes will vary significantly with both vowel and consonant contexts".

Other factors resulting from continuous speech that have an effect on its visual speech perception are the rate, rhythm and size of the signal. **[Lesner88]** and **[Berger72]** both consider the optimum rate for visual reception to be slower than 'normal' rate of articulation to allow the viewer more time to decode the message. Similarly, altering the rhythm of a sentence by the inclusion of pauses will allow the viewer to process previous information as well as providing significant visual punctuation marks **[Lesner88]**. **[Berger72]** also concluded that visual perception is inversely proportional to the size of the signal. This is due to fact that greater processing is required for a sentence than for an individual viseme.

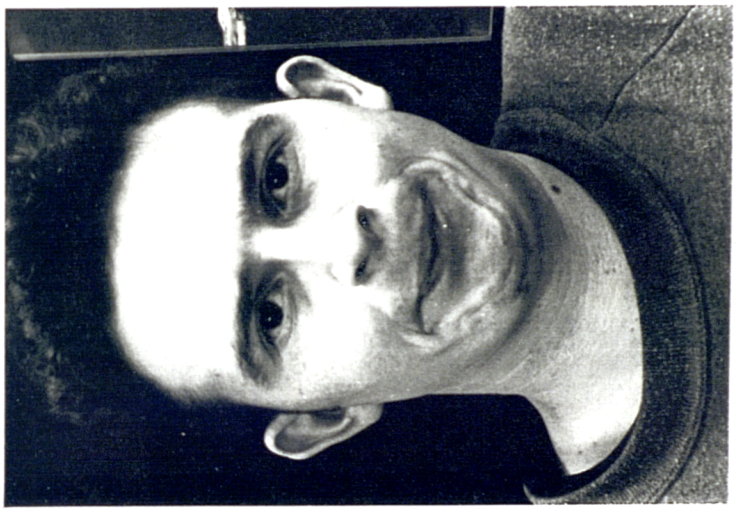
3.3.4 Other Factors Affecting Visual Speech Perception

As well as the effects of continuous speech, a number of the factors can improve or degrade the viewer's visual perception and ultimately corrupt the message. These factors can be defined as speaker and environmental variations **[O'Neill81]**.

Research has shown that different talkers present different degrees of visible articulation **[Kricos82]**. **[Lesner88]** deduced that the number and nature of the viseme categories vary between talkers and **[Montgomery83]** indicated significant differences in lip opening. These speaker variations occur due to differences in the amount of lip movement, in exaggerated speech and in obvious physical limitations such as beards or moustaches. **[Berger72]** suggests that "distinct exaggerated speech compared with an expressionless face is likely to increase perception but careless actions are likely to be a hindrance".

Environmental factors that can have an adverse effect are the distance and angle of the head from the viewer and also the lighting. Another important factor is the amount of the speaker that is visible.

Rounding Effect : "boo"



Spread Effect : "bee"

Protrusion Effect : "bar"

Figure 3.9 Photographic Examples Of Co-Articulation Effects On Bilabial Visemes

3.4 Facial Expression, Actions And Emotions

"The face is rich in communicative potential. It is the primary site for communicating emotional states. It reflects interpersonal attitudes, it provides non-verbal feedback on the comments of others, and next to speech it is the primary source of giving information. For these reasons and because of its visibility, we pay a great deal of attention to what we see in the faces of others" **[Knapp72]**.

Facial expressions are defined as the momentary movements which provide information about the emotional state of the subject. These expressive signals are dependent upon the context; and may vary in intensity, duration, and whether the person is listening, talking, or viewing. **[Ekman73]** confirmed Darwin's theory that the primary set of expressions are universal, not only within racial groups but from culture to culture. This has direct relevance to cinema where there is a desire for viewers of different nationalities to decode the same expressive meaning from the facial images.

Due to their universal nature, there is a requirement for a standard form of notation to describe and measure facial expressions. The identification and classification of the various dimensions of expression is designed to be free of bias or inference about any possible emotional meaning. For example, the description "smile" infers happiness whereas the expression might only be the lips stretching with no emotional message **[Wiggers82]**.

A review of psychological coding classifications (**[Harper78]**, **[Argyle88]** and **[Ekman82]**) indicates that only a few facial action notations have been developed due to problems in defining what to measure on the face. **[Blurton Jones71]** described facial actions, using facial surface landmarks, in three ways. The description uses the locations of shadows and lines, the muscles responsible and the main positions of landmarks, such as mouth corners or brow location. **[Birdwhistell70]** encoded all aspects of body movement to create a large 'vocabulary' of pictorial symbols by using the principles of labanotation. Labanotation is primarily used in dance choreography as a way of recording the motions of individual body parts against time giving a description of gestures in terms of symbols **[Herbison-Evans84]**. This is the same process as the exposure sheet used in traditional animation **[Thomas81a]**. Birdwhistell derived 32 'kinemes', or basic elements of

expression and their symbolic representations are shown in Figure 3.10.This scheme proves useful in supplying visual cues for expressions but it fails to allow for the description of action combinations and avoids the actual 'biomechanics' of the actions **[Waters89]**. Both schemes are limited in their manner of describing facial actions and their relative intensities and in outlining the details of the face.

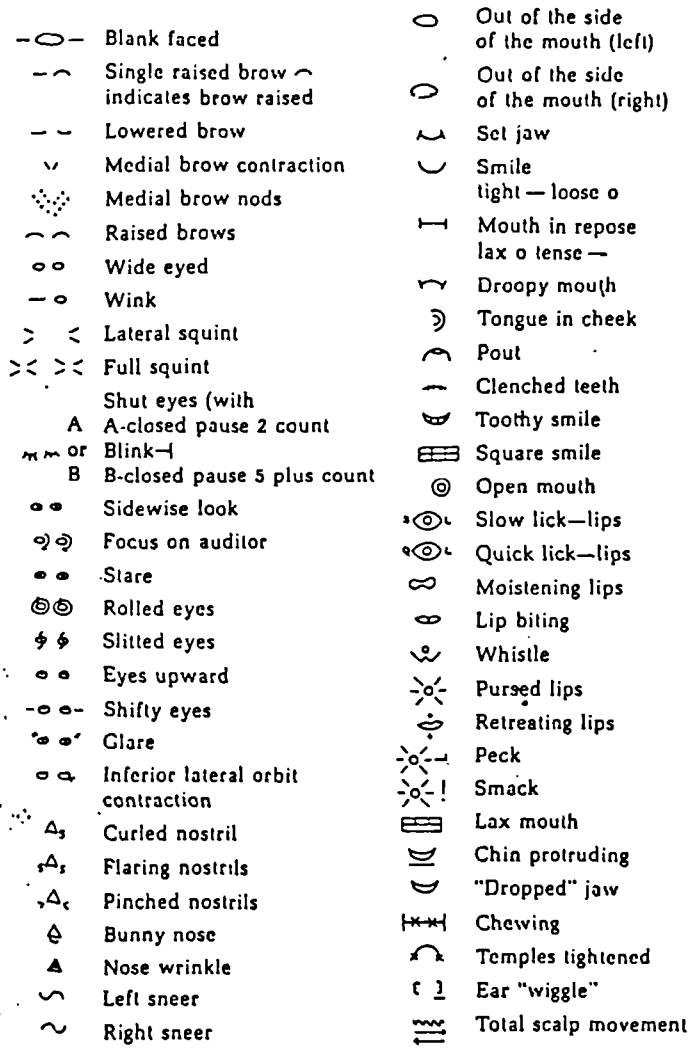


Figure 3.10 Birdwhistell's Facial Kinemes Of Expression

© [Birdwhistell 70].

A superior approach is the work of Ekman and Friesen and their Facial Action Coding System (FACS) **[Ekman78]**. This is a comprehensive system that categorises all possible visually distinguishable facial movements into minimal units, known as Action Units (AU's), which can in combination account for any expression or emotion. The action unit describes either a unique single muscle action or a unique action that is produced only as a result of two or three different muscles acting. It also describes different unique actions caused by the same muscle.

FACS opts for the muscle based 'bio-mechanical' approach. It is the result of muscles causing the temporal changes in the shape of the skin that create visible movements and changes in the location of the landmark features. This approach also overcomes the problems of physiognomic differences where individuals differ in the size, shape and location of their facial features. For example, the shapes or wrinkle patterns produced when the lip corner is pulled upwards are not the same for everybody. Knowledge of the muscular basis of the action and recognition of the movement itself will reduce the possibility of resultant movement being scored incorrectly. It is also free of inference about any possible emotional meaning of the visible actions.

3.4.1 Facial Action Units

There are, in total, 66 independent actions which can be reliably distinguished visually by trained observers **[Ekman82]**. The action units are not concerned with any motion characteristics but with the notation and grading of individual actions from static poses. Table 3.5 lists the facial action units directly related to the present research. In the FACS manual, Ekman and Friesen describe the muscular basis for each AU. Detailed descriptions are given on the observation and grading of the appearance changes and on the production of the A.U. on one own's face.

These independent action units were determined by undertaking a comprehensive study of anatomical texts, judgements of posed images, pulpating their own muscles and comparing in a mirror, observing the changes in the facial surface when a needle was placed in the muscles and then delivered electrical current and monitoring electrical activity via a needle when the muscles were moved.

AU	FACS name	Muscular basis
1	inner brow raiser	medial portion of Frontalis
2	outer brow raiser	lateral portion of Frontalis
4	brow lowerer	Corrugator, depressor glabella and/or depressor supercilli
8	lips towards each other	Orbicularis oris
10	upper lip raiser	Levator labii superioris, caput infraorbitalis
11	nasolabial furrow deepener	Zygomatic minor
12	lip corner puller	Zygomatic major
15	lip corner depressor	Triangularis
16	lower lip depressor	Depressor labii inferioris
17	chin raiser	Mentalis
18	lip puckerer	Incisivii labii superioris and/or inferiori
20	lip stretcher	Risorious
22	lip funneler	Orbicularis oris
23	lip tightener	Orbicularis oris
24	lip pressor	Orbicularis oris
25	lips part	Depressor labii or relaxation of either Mentalis or Orbicularis Oris
26	jaw drop	Masseter with temporal and internal Pterygoid relaxed
27	mouth stretches	Pterygoids; digastric
28	lips suck	Orbicularis oris

Table 3.5 Table Of Facial Action Units

3.4.1.1 Eyebrow Action Units

[Ekman79] concentrates specifically on the eyebrow region and it's limited set of actions. This is due to the straightforward nature of the musculature and the ability to investigate them in isolation from the other areas of the face. This is advantageous

when considering expressions that occur at the same time as speech articulations. Figure 3.11 shows the brow action units and the possible combinations.

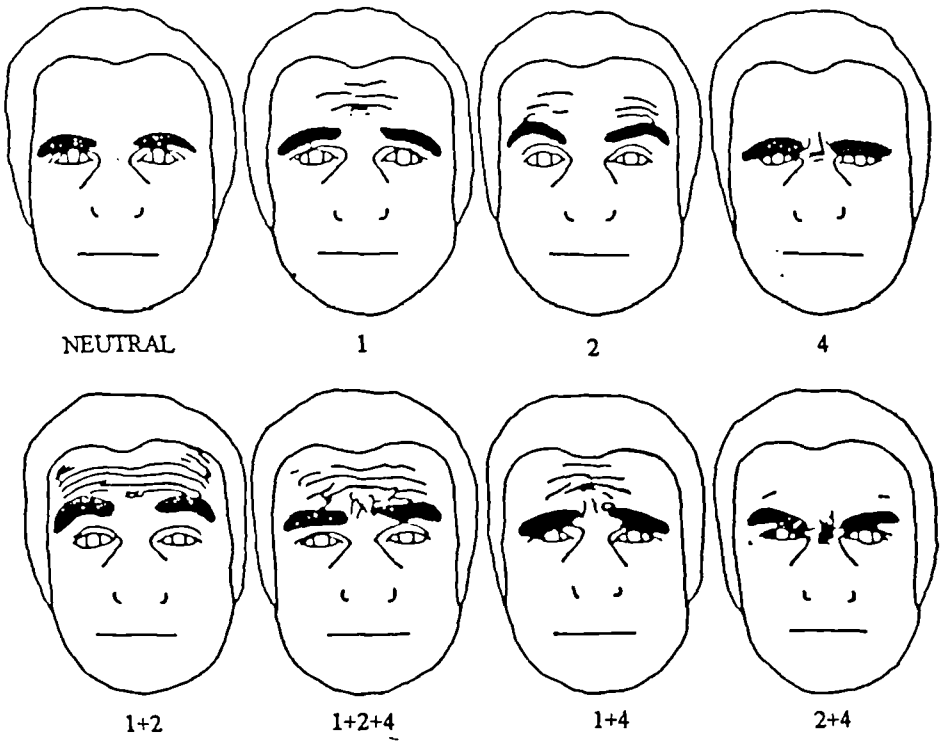


Figure 3.11 Action Units Of The Brow Region

© **[Ekman79].**

3.4.1.2 Timing Of Action Units and Expressions

Action units are limited by the fact that their definitions are based on static poses or images which are harder to identify when viewed in actual continuous communication in a manner similar to the phonetic problems in speech already discussed. **[Ekman82]** deduced that each action has start and stop points and a plateau where there is no further increase observed in the muscle action. From the start until the muscle movement reaches the plateau is the onset time. Apex time is the duration of the plateau and offset time is from the end of the apex to the point where the muscle is no longer acting. Each of these time periods can vary in duration and smoothness and at present research has failed to deduce any common functions between expressions.

3.4.2 Facial Expressions to Convey Emotional Signals

From these single action units, a large number of combinations of facial expression can occur. **[Goleman81],** in reviewing FACS, claims that some seven thousand are possible. From these permutations, there is a desire to define the combinations of actions that form recognisable emotional signals. "Darwin started studies of facial expression in the 1870's and defined that different expressions were used for different emotions" **[Argyle88]**. More recent research has found that observers can only discriminate between a small number of broad groups classed as "universal expressions of emotion" **[Ekman73]**.

From his work, **[Ekman82]** defined the following six distinct emotional groupings, ordered in terms of greater percentage correct recognition;

1. happiness.
2. surprise.
3. fear.
4. sadness.
5. anger.
6. disgust / contempt.

[Harper78] also reported recognition of interest, shame, pain, startle, puzzlement, amusement, boredom and impatience. These groupings are of limited use since research has failed to present clear definitions on how best to reproduce them or how to distinguish them reliably from the main six.

[Ekman82] and **[Wiggers82]** have both attempted to 'construct' the six primary emotions in terms of the specific action units required. Table 3.6 presents data compiled from their results. The 'emotional predictions' are based on the judged observations of static facial poses, which are independent of any temporal changes.

The photographs in Figure 3.12 show the attempts to realise the universal emotions based on Ekman's action unit definitions. Photograph b) is acknowledged as being a 'poor' representation of the emotion "Sadness". This highlights a limitation in the researcher's production of certain universal emotions.

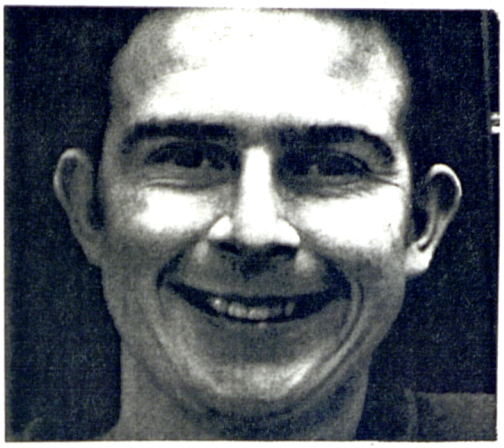
The probable reason for the confusion's between emotions is the fact that they share the same or similar action units. For example, surprise and fear are confused due to similar brow raises. In an attempt to overcome these confusion's, **[Boucher75]** suggested that certain parts of the face are dominant for particular emotions, even though all three areas of the face are involved in each emotion. He defined the following; happiness can be derived from lower face or lower face with eyes, surprise from brows or brows with eyes or lower face, fear from brows and eyes and sadness from eyes and lower face. Anger could not be decoded from any one single area.

The definitions in Table 3.6 are for idealised poses of purely one emotion. **[Ekman92]** correctly argues that "in real life people are more likely to express blends or combinations of two or more emotions at any one time". When this occurs it is likely that different emotions are expressed in different parts of the face **[Argyle88]**. Examples of these combinations are a 'pleasant surprise', a blend of happiness ('open smile') and surprise ('raised brows'), or the 'receipt of bad or frightening news', a blend of sadness and fear **[Ekman73]**. Blends can also occur when an attempt is made to conceal the 'true' emotion. For example, a 'false smile' when concealing one's anger.

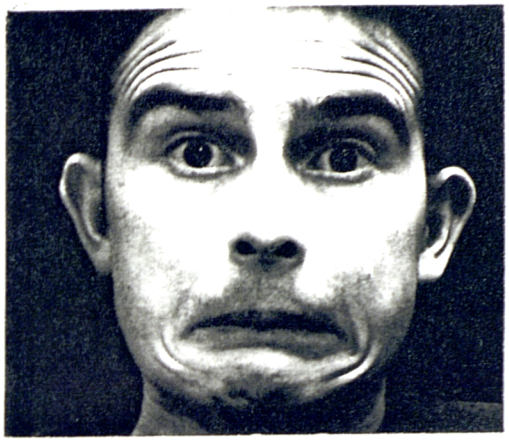
Emotion	Prediction in A.U.'s	Description	Percentage Correct Recognition	Confusions
HAPPINESS	12+6 or 12+6+25/26	Lip corner puller with cheek raiser or with lips parting or jaw dropping	64-100	NONE
SURPRISE	(1+2)+26	combination of both brows, inner and outer, raising with jaw dropping	77-100	FEAR
FEAR	(1+2)+5+20 or (1+2)+5+20+25/	combination of both brows, inner and outer, raising, upper eyelids and lips streching or with lips parting or jaw dropping	64-95	SURPRISE, ANGER, CONTEMPT
SADNESS	(1+4)+15 or (1+4)+15+6+2	combination of inner brow raise and drawing together, lip corn depressor with or without lower eyelid raising causing slight chee and with lips parting	67-95	FEAR
ANGER	{4 or 4+5} + {10 or 1 or 25}	brow lowering and drawing together with or without upper lid rai lower face, lip tightening or upper lip raise or lower lip depress wi parting	69-100	FEAR (major), SURPRISE and DISGUST (minor).
DISGUST	10+9+25/26 or 10+9+17	combination of upper lip raiser, nose wrinkler and lips parting or ja or chin raiser. Can also include tongue thrust (AU19)	64-100	ANGER, CONTEMPT
CONTEMPT	{(1+2) or (1+2)L or 2 {(10L+25) or 14L or	brow, inner and outer, raising either bilateral (L+R) or unilateral (L unilateral outer brow raise only. In lower face, unilateral upper lip with lips parting or unilateral inward corner puller or chin rais	72-100	SURPRISE, DISGUST

Table 3.6 Construction Of Primary Emotions From Facial Action Units

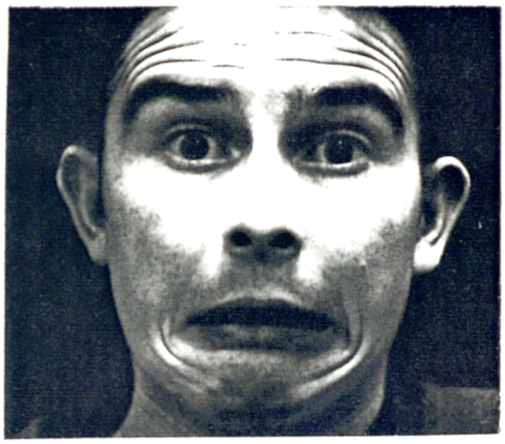
a) Happiness



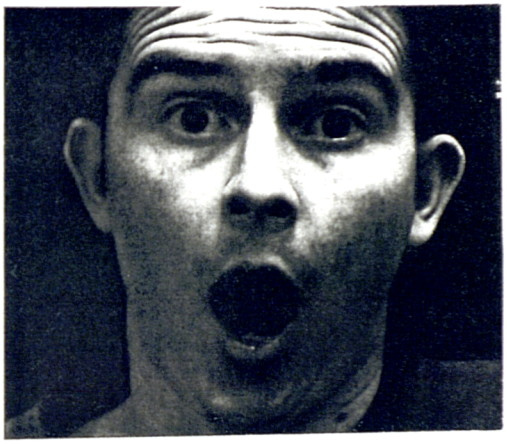
b) Sadness



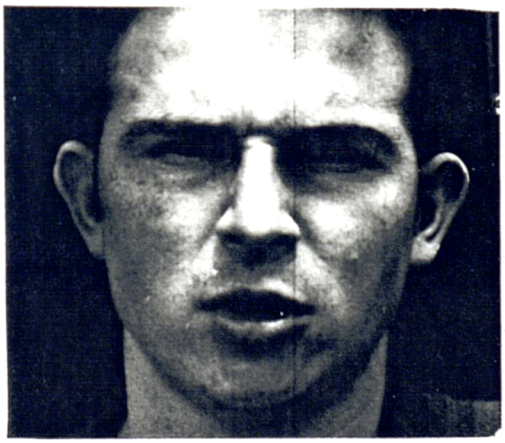
c) Fear



d) Surprise



e) Anger



f) Disgust

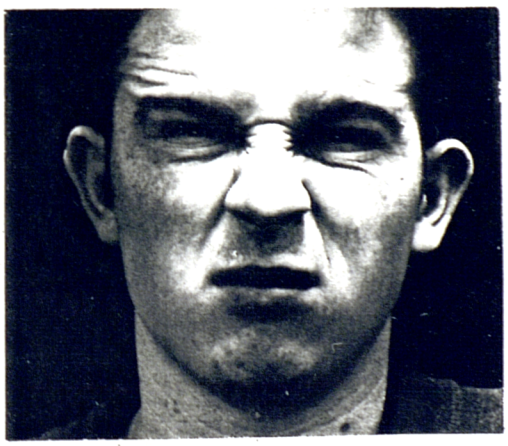


Figure 3.12 Photographs Of Primary Emotions Constructed From Facial Action Units

3.4.3 Facial Expressions in Conversation

A whole area of research exists on the non-verbal signals from the paralanguage and vocalisations that can occur in conversation. These pauses, stresses, changes in pitch and intonation, all affect the meaning of the message. However less consideration has been given to the type of facial expressions that accompany speech and the visual messages they convey. **[Ekman82]** has investigated these actions with particular reference to the brow region, reasoning that their actions though not operating in isolation are independent of the speech articulatory actions of the lower face. For example the raising of eyebrows frequently accompanies the stressing of key words in spoken sentences. They can also act as indicators of punctuation, question marks and pauses in speech. Tables 3.6 and 3.7 outline the descriptors and action units associated with speaker conversational signals and conversational signals without words **[Ekman82]**. It should be noted that these actions are ambiguous, and liable to mis-interpretation, unless one knows the context in which they occur.

For conversational signals without words, **[Ekman82]** states that individual eyebrow actions cannot solely convey a message. However one might like to consider a quote by the actor Roger Moore on his acting technique "left brow raise, right brow raise".

Name Of Descriptor	Definition Of Action	Prediction in A.U.'s	Type Of Brow Actions	Other Body Actions
BATON	emphasis on a single word	(1+2) or 4 or (1+4)	most common is both brows raised. next common is inner brows together then inner brows raised and together.	commonly seen as a hand action
UNDERLINER	emphasis over more than one word	(1+2) or 4	most common is both brows raised or inner brows together .	
PUNCTUATION	emphasis at end of an utterance as a pause(similar to an exclamation mark).	(1+2) or 4	brows raised in context of something 'amazing'. brows together at a juncture implying seriousness.	
QUESTION MARK	indication of a question being asked	(1+2)	brows raised.	
WORD SEARCH	indication that utterance is not complete.	(1+2)	brows raised.	similar to "um" or "ah" in vocalisation.

Table 3.7 Speaker Conversational Signals From Brows

Name Of Descriptor	Prediction in A.U.'s	Type Of Brow Actions	Other Body Actions
FLASH	(1+2) + 5	both brows raised.	upward tilt of head and upper lid raise
DISBELIEF	(1+2) +15+17+10	both brows raised.	lip corner depressor, lower lip raise, upper lip raise and head motion side to side.
MOCK ASTONISHMENT	(1+2) + 5 + 26	brows raised.	raised upper eyelid and jaw drop. Also head tilted to one side.
AFFIRMATION	(1+2)	brows raised.	head thrown back.
NEGATION	4	brows lowered and drawn together	head movements side to side.

Table 3.8 Conversational Signals Without Words From Brows

3.5 Summary

The human face, with its complex relationships between skin, muscle and skull, has the ability to produce thousands of facial expressions through the combination of multiple muscle actions. In order to recognise or animate some of these actions, there is the need for a notation to describe them in terms of their visibility. Ekman and Friesen's Facial Action Coding System, with its description of expressions in terms of their component Action Units, is recognised as the most comprehensive and adaptable devised to date. It is free of any emotional bias, it originates from an anatomical basis with every unit based purely on visible actions and it allows the definition of complex combinations such as the universal emotions in terms of individual units. FACS has been successfully adapted in a number of projects in computer animation **[Waters91]**, **[Patel91]**, **[Guenter89]** and **[Choi90]**.

From the examination of speech production, research has shown that certain elements of speech can be perceived visually from the articulatory actions of the lips and jaw. The viseme groupings represent a suitable notation to describe these discrete elements as it accounts for the redundancies that exist due to hidden places of articulation. The perception of visual speech is, however, further degraded by the effects of a number of other factors, primarily the changes resulting from continuous speech. **[Summerfield82]** wryly pointed out that

"Hollywood's conception of the secret agent who perfectly divines her opponents plan by observing a muttered conversation across a dimly lit cafe is pure fantasy. Research suggests that the agent may fare little better if she could persuade her adversary to face her at a distance of 1.5m in bright illumination so that his arms and torso were visible in addition to his face, to remove any disguises such as a beard or a moustache that might obscure his mouth, to speak naturally but slowly, and possibly to wear lipstick to emphasise the patterns of his lip actions. Even then the agent would probably find that the twenty four acoustically distinguishable English consonants fell into at most twelve, and maybe as few as four, visually distinguishable groups".

The effects of coarticulation and other message factors, such as rate and rhythm, produce important visual cues that are vital in the animation of synchronised lip movements. Techniques that attempt to segment the soundtrack into visemes in order to produce this animation, often fail to account for the loss of visual information regarding the changes on and between visemes. **[Massaro87]** indicated that the viewer's, like the listener's, perceptual processes do not attempt to 'dissect' each word into its component visemes and then reconstruct it within the mind. For good lip synchronisation, the objective is not the actual recognition of individual speech elements but the production of the changing lip movements in time with the acoustic signal.

In conclusion, the objectives for a performance control system are as follows:

1. to sense, consistently, the primary elements of visible speech in isolated conditions;

2. to sense, in isolation, the expressive facial action units and the universal emotions, formed by their combination; and

3. to automatically produce control signals from the visual motion of the lips and face in continuous speech, avoiding the segmentation of the input signal and the resultant loss of coarticulatory and expressive information.

Similarly, the objectives for an animatronic model are as follows:

4. to animate the primary sets of visual articulatory and expressive actions in isolated conditions;

5. to animate the continuous motion of the lips in synchronisation with a given soundtrack to produce perceptually correct visual speech; and

6. to animate the primary emotions with differing intensities in isolation and blended with speech signals in overall performance.

Chapter 4

Hypothesis Of Facial Control: System Design And Method Of Solution

Chapter 4

Hypothesis Of Facial Control: System Design And Method Of Solution

4.1 Introduction

As described in previous chapters, this research has focused on the animation of facial expressions and the synchronised lip movements of animatronic characters. Specifically, research has concentrated on the techniques used to control such performances. Research in Chapter 2 described the limitations of the present techniques used in animatronics and computer animation. It concluded that a more natural and thus superior form of control was possible if the correct input signals could be extracted automatically from the visible facial movements of the performer. Chapter 3 investigated the human facial communication system. It deduced how the visible actions are produced and defined what information is perceived from these actions. It indicated that only a limited set of distinct actions actually conveys relevant information about the communicated message, even though the face is capable of a large number of actions. The research also indicated that the complex facial system can be reduced to an optimum set of measurable points at key positions on its surface.

The following chapter presents the hypothesis for performance control through the automatic, real time sensing of facial actions. Section 4.2 explains the principles of the technique to achieve this form of control that could overcome the problems inherent in other existing techniques. A method of solution is proposed to evaluate both individual elements and the overall system is described. Section 4.3 describes the functional theory for the mapping within the system and the derivation of control and drive parameters. Section 4.4 describes the research conclusions made from the two previous Sections. The primary set of the facial actions necessary to produce successful facial performance is derived along with the optimum set of facial points necessary to fully describe these actions. A photogrammetric investigation of the key points on the human face is presented to provide a detailed description of the action displacements. Section 4.5 considers the results from this investigation and produces the final design requirements for the sensing and drive sub-systems.

4.2 Research Hypothesis of Facial Action Sensing System

4.2.1 The Human Face As A Source Of Performance Control

The review of present performance control techniques, in Chapter 2, highlighted the problems that exist in the creation of accurate lip synchronisation. Each of these methods are limited in certain ways that affect or complicate the final performance.

The manual controls used in animatronics are restricted by the complex and unnatural processes required to create the time-varying lip movements. Successful performance is possible only by highly skilled performers. The techniques for producing performances by user-defined programs are constrained by a number of factors. Firstly, simultaneous control input and performance output are not possible preventing any real time application. Secondly this method, and that of manual control, is reliant upon the pragmatic judgements of the performer or programmer, based primarily on their experiences of performance.

Alternative approaches exist that extract the control signals from the analysis of the actual acoustic soundtrack. These are limited by the fact that the analysis is based on the segmentation of the speech signal into phonemes or isolated words. This segmentation results in the loss of vital information on phonetic co-articulation that affects the visual appearance of the lips. Other important expressive signals are excluded by these techniques which result in inferior animation that lacks the variety of action associated with natural speech.

From the review in Chapter 2, it was concluded that a more innate method of control based on the extraction of signals from the visible speech actions, had the potential to improve the present animation of lip synchronisation.

Chapter 3 examined the physiology of the human face in communication and concluded that it is a highly complex system capable of thousands of different visible actions. With any complex system, there is a desire to describe it with the minimal amount of information possible and this is true of the face **[Hardcastle78]**. **[Ekman78]** defined all the possible permutations of the facial changes, in terms of a reduced set of primary actions. The problem that exists within visual sensing, is to recognise these action units, automatically, in terms of a limited set of data variations.

In image processing, several techniques have extracted information on the visible actions from reduced descriptions of the face, **[Brooke83]**, **[Terzopoulos90]** and **[Finn88]**. **[Brooke83]** successfully extracted facial parameter changes automatically during articulation, in terms of key point movements. This principle of recognition allows the capture of expressive and co-articulatory information from a reduced set of points rather than the overall image. It avoids the errors resulting from the segmentation of the speech signal whilst providing correct information on the natural rate, rhythm and intensity (magnitude) of the actions associated with speech.

The present image techniques have a number of drawbacks in their possible use as performance control tools in animatronics. A significant amount of processing is necessary to track and extract the correct information from each frame and consequently limits its application in real time control. The methods to acquire the data from the performer require either restrictive head positioning or the application of complex algorithms to normalise each frame.

4.2.2 Proposed Method Of Control Using Optical Sensors

The alternatives that exist in visual sensing techniques were considered. The specifications for an alternative were based on the need to have sufficient sensitivity to recognise physical changes of less than 25 mm **[Brooke83]**, **[Fromkin64]**. It was concluded that a novel optical technique using proximity sensors satisfied the design criteria.

The proximity sensor is composed of an infra-red emitting diode and a spectrally matched photo transistor housed in the same package **[RSData83]**. Proximity sensors are typically used as optical counters or switches **[Ohba92]**. They are also used in the detection of an objects' presence, within a defined field of view **[Todd86]**.

The photo transistor produces a photocurrent, I_C, which is proportional to the magnitude of the emitted radiation from the emitter, only when a reflective surface is present within the field of view. The amount of radiant power from a reflector is dependant upon a number of different variables; the size or area of the reflector, the type of reflective surface, the orientation between the sensor and reflector and the distance between the sensor and the reflector. The final property of the reflected power can be adapted to produce a continuous output signal, that is directly proportional to the varying distance between the sensor and the reflector (c.f. Section 5.2 for practical system). This is based on fact that the reflectors motion lies within the physical field of view and principally along the focal z axis. Figure 4.1 illustrates the proximity sensor operation.

Reflector motion across the field of view, would also result in a change in the output signal resulting in an incorrect control signal. This type of motion, along with the effects of changes in orientation, should be constrained where possible.

There are a number of different types of reflection. Specular reflection describes a surface where the angle of reflection of a light ray is equal to the angle of incidence. This is not of use as it is dependant upon the exact positioning of the emitter, receiver and reflector. Retro-reflective surfaces reflect light rays back along the same path as the incident ray and with the same amount of radiant power. Diffuse reflection describes surfaces that reflect the incident rays at all angles to the surface. The amount of reflective power in the reflective rays is dependant on the surface. A white

matte surface is defined as having a greater amount of diffusion than a black shiny surface **[Marston88]**. In conclusion, diffuse materials offer the most suitable form of reflection for the proposed sensing system

If the reflector was to move too close to the sensor, as shown by point X in Figure 4.1, the signal falls to zero. This results in the sensor being able to generate the same I_c for two different positions. The region of interest should therefore be restricted to the area indicated in Figure 4.1, where the signal characteristics were considered sufficiently constant. This allowed I_c to be defined as approximately linear. The final output signal, from the necessary conversion circuitry, represents a zero order measurement of the changes in the position of the reflective surface, relative to the sensor. A zero order system is defined as a measurement instrument that has no dynamic qualities **[Pallás-Areny91]**.

The advantages of using this type of optical sensors were as follows:

1. the production of a continuous, analogue signal directly proportional to displacement changes of a surface point;
2. the sensor was sensitive over the range of displacements required;
3. the reduction in the amount of signal processing required thereby allowing real time control;
4. the generated signals would allow straightforward interface with the existing Henson System; and
5. they were low cost and lightweight.

The overall advantages of the proposed system were defined as the following: the generation of automatic control signals based on the naturally produced actions of the human performer; the continuous sensing of the important visual changes that occur as a result of co-articulation and expressive blending as well as the changes in rate and rhythm; and the production of signals based directly on variations in the intensity of the actions, in terms of the varying magnitude of the point displacements relative to some neutral position.

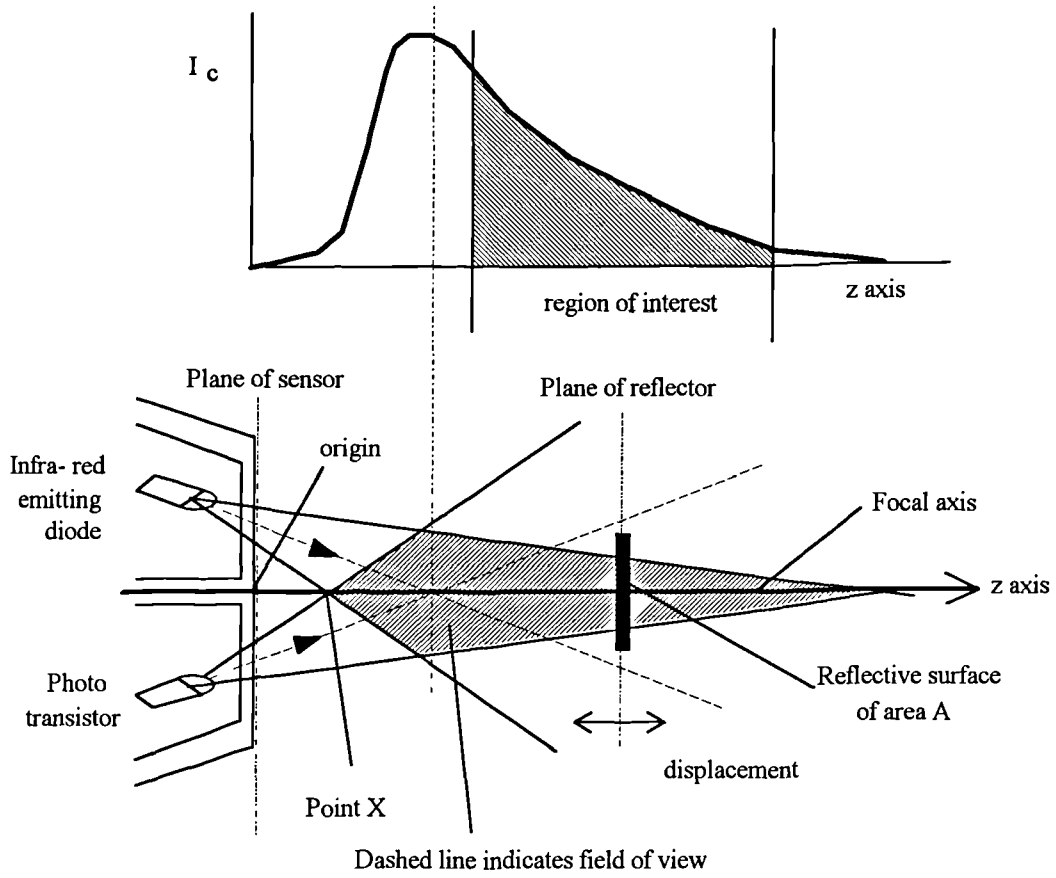


Figure 4.1 Principle Of An Optical Proximity Sensor For Range Measurement

To achieve this method of facial sensing, a number of theories were derived based on the research of Chapter 3.

1. A primary set of visibly distinct actions exists that, if correctly animated in synchronisation with an audio soundtrack, can convey the same desired message to the viewer.

2. An optimum set of visible key points exist on the face which provide sufficient information to accurately describe the primary actions in terms of displacement changes.

3. The displacement of each key point can be reduced to specific and constant paths of motion. In other words, each point can be defined as having specific degrees of freedom. Ideally, all of the displacements of a point can be sensed along a single trajectory or through the combined sensing of distinct paths.

The assumption that the motion of each point is purely along specific axes was based on initial subjective conclusions from the speaker's reproduction of the primary actions. The photogrammetric investigation into the displacements of the researcher's face (c.f. Section 4.4), produced results that confirmed this theory. It is acknowledged that the human face is capable of producing a number of idiosyncratic actions which do not fall into the above categorisations and would cause the system to fail. It was concluded that these actions do not occur simultaneously with articulatory gestures in normal speech.

4.2.3 Possible Methods Of Analysis Of Proposed Control Technique

It was important to derive a suitable method to fully analyse the proposed system's capabilities. Both physical and perceptual techniques were considered. In terms of engineering, physical methods were considered first due to the belief that subjective methods were open to misinterpretation. The overall assessment of the system could be considered as the evaluation of the following theories:

1. the optical technique could produce accurate and consistent measurement, with sufficient sensitivity, of the key point displacements in space;

2. the facial action sensing system could produce information not only on static expressions but on the temporal changes associated with continuous speech;

3. the defined set of point displacements could provide sufficient data on the changes in facial shape, in terms of magnitude and rate, to produce improved control for the final animation. The resultant performance should be of the similar actions with identical timing and of comparable magnitude;

4. the primary set of actions are sufficient to produce a successful and life-like performance.

Each of these theories was, in some way, related to the others which caused a number of problems in the evaluation of the system. **[Chatfield83]** stated "In the analysis of a signal, its variation must be interpreted against some known quantity. To achieve this, steps must be taken to isolate the effects of interest from all other possible variables. This will allow an evaluation of the signal's precision and accuracy". This type of analysis also allows the identification of the nature of the errors that exist and draws conclusions on their source **[Barry78]**. Therefore analytical methods had to be developed to separate each of the categories from the effects of the others to allow correct evaluations.

The analysis of the sensing instrumentation could be undertaken in controlled conditions in order to evaluate the optical sensing characteristics and possible errors in the final measured signal (c.f. Section 5.3). The produced results would have little relevance to the assessment of the system's ability to actually sense facial actions.

The need to analyse the overall system posed the question of what other 'known quantities' exist ? The only two possibilities were the acoustic signal from the soundtrack or the repeated recordings of defined tasks.

Comparison between the individual sensor variations and the recorded speech signal would only generate conclusions on the timing of the displacements based on the subjective evaluation of graphical variations. It would not produce conclusions on the intensity or accuracy of the sensing system nor would it assess the overall effect of the information derived from all the points.

Comparison between repeated recordings of the same task, by a performer, poses a number of drawbacks in analysis. These exist due to the problems of physically controlling the face to generate identical sequences of test data. The resultant variations could easily be misinterpreted as errors leading to the wrong conclusions being drawn. Only through practice, and with careful monitoring, is the performer likely to produce actions of the same rate, rhythm and intensity. The reproduction of identical facial actions is not considered a criterion for the majority of control systems as it is acknowledged that no two performances are identical in every way. The

results of such comparisons, if identical inputs could be ensured, would provide an indication of the sensing system's ability to recognise facial changes but it could not produce an assessment of the overall objective as a source of control.

Both of these methods fail to provide a full evaluation of the overall system objectives of enhanced performance control.

The overall objective of any animatronic system is the successful performance of a pre-defined task. In this case the task is the production of synchronised lip movements and primary expressive actions necessary to convince the viewer that the actions are life-like. The success of the system is therefore based on the subjective opinions of the viewer. These decisions are based on their cognitive experiences of facial actions coupled with their knowledge of the message to be reproduced or transmitted **[Massaro87]**. The use of subjective analysis is open to misinterpretation and experiments have to be carefully controlled to isolate the main message or variable from other factors. Problems are likely to occur if an animatronic fantasy character is used. The conformation and appearance of the facial model may influence the perception of the viewer. The model may also be unable to produce the desired output actions due to its design or it may require different mapping approaches thereby reducing the assessment of the key point theory. As a consequence, subjective analysis, alone, will fail to provide a thorough assessment of the system and, in particular, the accuracy and success of the sensing system. It was concluded that the generation of both subjective and objective data was necessary if any comprehensive conclusions were to be determined.

4.2.4 The Proposed Method Of Solution

A methodology of analysis was developed with the capacity to produce both perceptual and physical data. This method was based on similar principles to those applied in robotic tele-operation systems **[Todd86]**, **[Thring83]** and **[Coiffet86]**. Within tele-operation systems, the output arm is designed to produce identical movements through the same degrees of freedom as that of the input control systems. This allows the operator to generate identical or mirrored actions, at some remote location, through the production of natural arm actions. These are commonly known as master: slave systems. The proposed solution was to design and construct an

animatronic face of the exact shape and dimension as that of the researcher's in the 'expressionless' (neutral) state.

For the clarity of the text, the animatronic model will be described as the replica face and the researcher as the live face. The replica was defined as a 'mirror' representation of the live face based on an adaptation of the principle of key points. The drive system in the replica was designed to produce the same displacements, in magnitude and trajectory, at an identical set of key points. The assumption was made that the drive actions at each point were distinct from those at the other points.

With the motion of each point being independently defined there exists the likelihood of the production of unnatural actions. The key points on the replica do not operate under the same kinematic rules that exist in the human face. Consequently it would be possible for the corners and centres of the lips to act in opposite directions, such as, 'corner stretches' with 'centre protrude'. The following argument allows this possible error to be eliminated: the control of the replica is derived from an actual human face which obeys the kinematic rules defined by the physiology of the facial muscles. If the live face cannot achieve these unnatural actions, and if the control system correctly senses actual facial changes, then the replica cannot be driven to produce them. Also, within the design of the replica, certain mechanical compensations were constructed to improve the overall facial changes at the areas other than the key points.

The consequence of the above argument was to reduce the complex input action to a series of individual control signals that, when correctly mapped to the drive system within the replica, will produce an identical set of individual displacements. Consequently, the resultant changes in the replica should produce a perceptually similar overall action. To achieve this, the sensing system must have the capacity to generate an equivalent set of input signals to equal the number of degrees of freedom (DOF) of motion possible at each key point. Similarly the drive system must be capable of animating the same number of DOF.

The development of this theoretical design should generate data suitable for both objective and perceptual analysis. The overall system diagram in Figure 4.2 shows the proposed method for analysis. A technique for data acquisition was to be developed to allow the recording and storage of the measured data from the sensing systems for subsequent analysis (c.f. Section 5.2 for the description of design). The following

sub-sections discuss the proposed methodology for the assessment of the hypotheses presented within this section.

4.2.4.1 Analysis of Principle of Key Point Facial Sensing

The development of a data playback system (c.f. 5.2) would provide direct software control over the individual drive actions and ensure that the replica could reproduce identical actions. The performer would have the capacity to control specific output actions in a more precise way. Analysis would be possible through the recording of the repeated displacements at each point by positioning the sensing system on the replica. The position of the sensors and key point reflectors could be altered and the generated signals compared since the input motion would be identical in rate and magnitude. This type of investigation would allow the assessment of the system as a facial action sensing system.

This method was based on the assumption that the replica displacements were comparable with those of the live face. Subjective evaluation of each point moving through all possible positions would enable conclusions to be drawn on this assumption.

4.2.4.2 Objective Analysis Of System As A Source Of Control And Key Point Theory

The comparison of objective data would be possible through the postulation that the key point displacements were identical on both the live and replica. Signal analysis of each individual channel would be achieved by the extraction of data from an identical sensing system on the replica. By recording signals from the replica displacements at the same time as the input displacements, statistical evaluations would be possible on the differences in rate and intensity for each key point. This would allow conclusions to be drawn on the consistency of individual channels and the overall theory of "mirror" correspondence.

4.2.4.3 Subjective Analysis Of System As A Source Of Control And Key Point Theory

The visual analysis of the actions generated by the replica, with respect to the known actions of the live face, would allow assessment of the individual key points and of the overall animation. Subjective assessments of the visible displacements at each key point would enable further conclusions to be drawn with respect to the performance of each individual channel. Viewing the overall action would produce conclusions on the ability of the overall system, and specifically the principle of key points, to successfully animate visible speech signals and primary emotions. Evaluations could be drawn on the ability of the primary set of actions to convey the correct messages through subjective testing of independent viewers.

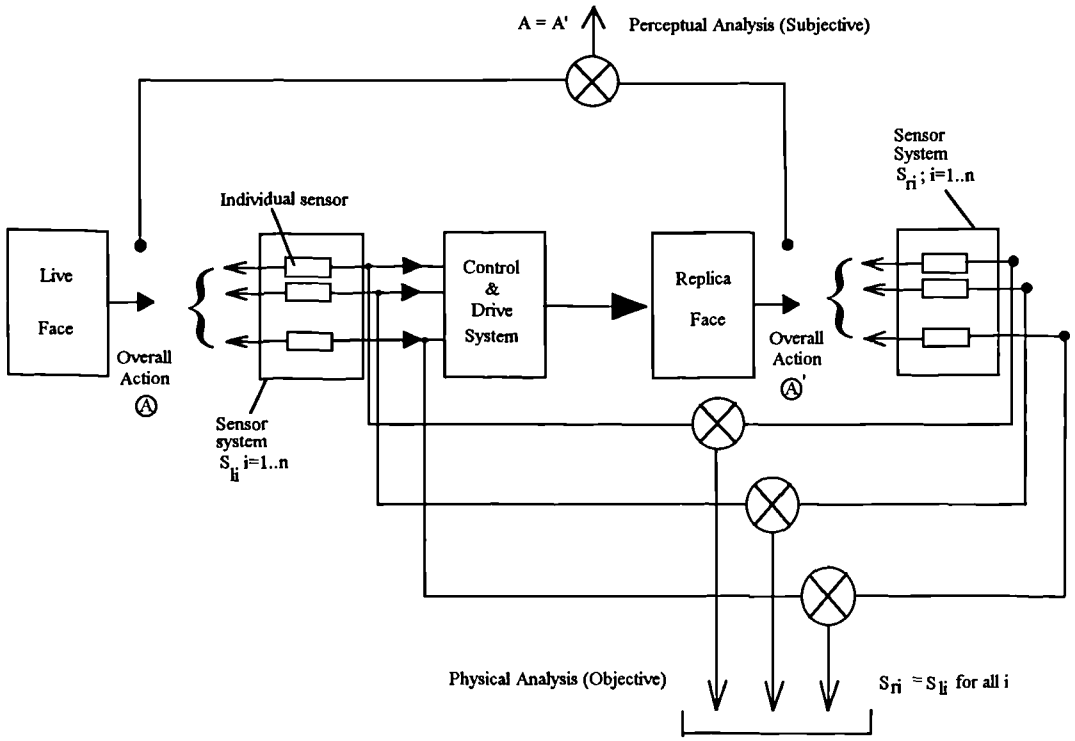


Figure 4.2 Block Diagram Of Overall System Analysis

4.3 Functional Theory Of Proposed System

4.3.1 Linear Theory Of System Design

This section concentrates on the development of a theoretical model for the design of the system, shown in Figure 4.3, in terms of three distinct sub-systems; sensing, control (or mapping) and drive.

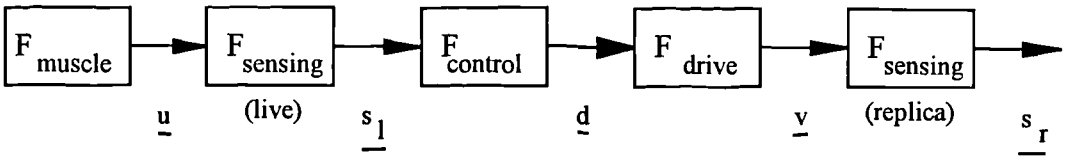


Figure 4.3 Block Diagram Of Overall System

The proposed system was based on the principle that the overall facial actions produced by the live head could be considered as the individual actions of a distinct set of key points $\underline{u}$, where $\underline{u}$ represents a matrix of n vector displacements. These displacements were sensed by the sensing system which produced the input signals represented by the matrix $\underline{s_l}$. The output drive signals $\underline{d}$, where $\underline{d}$ represents a matrix of m drives, were generated through the defined relationships between $\underline{s_l}$ and $\underline{d}$ in the control system. The subsequent positional changes of the drives, when mechanically linked to the skin of the replica face, should displace an identical set of key points, at the same rate and with the same magnitude. The final animation was then the result of the recombination of these individual displacements $\underline{v}$, where $\underline{v}$ represents a matrix of n displacements. This was defined as a mirror correspondence where all relationships are considered as steady state to remove the complexities inherent in the design of dynamic systems.

The overall system could then be defined by the following equation

$$\underline{v} = \underline{F_{total}}(\underline{u}). \qquad \text{Equation [4.1].}$$

where $\underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_n \end{bmatrix}$, $\underline{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_n \end{bmatrix}$ and the overall function of the system is defined as

$\underline{F_{total}} = \begin{bmatrix} F_{11} & F_{12} & F_{1n} \\ F_{21} & F_{22} & F_{2n} \\ F_{n1} & F_{n2} & F_{nn} \end{bmatrix}$. In it's full representation, the system is defined as

$$\begin{bmatrix} v_1 \\ v_2 \\ v_n \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} & F_{1n} \\ F_{21} & F_{22} & F_{2n} \\ F_{n1} & F_{n2} & F_{nn} \end{bmatrix} \left(\begin{bmatrix} u_1 \\ u_2 \\ u_n \end{bmatrix} \right). \qquad \text{Equation [4.2].}$$

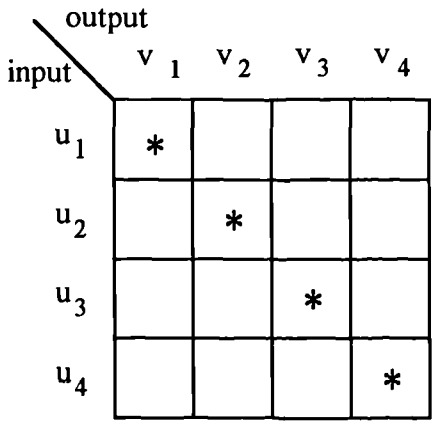
Given the proposed hypothesis that the overall facial action is the result of the summation of all the individual displacements, the following relationship could be defined between the input and output. This is valid for all outputs.

$$v_n = F_{n1}(u_1) + F_{n2}(u_2) + \ldots\ldots + F_{nn}(u_n). \qquad \text{Equation [4.3].}$$

In Section 4.2, the principle of key point correspondence was proposed. Each input displacement was said to have distinct control over its respective output displacement. This type of relationship reduces the overall function matrix $\underline{F_{total}}$ to a diagonal representation as shown in Figure 4.4 where each relationship can be described as mutually exclusive. This allowed Equation [4.3] to be reduced to

$$v_n = F_{nn}(u_n) \text{ for all key points } n. \qquad \text{Equation [4.4].}$$

From this overall representation of the system, the specific elements were developed to produce the linear characteristics. The final system is shown in the function diagram of Figure 4.5. The individual characteristics are examined on the following sub-section.



For $n = 4$.

Figure 4.4 Diagonal Representation Of The Key Point Theory

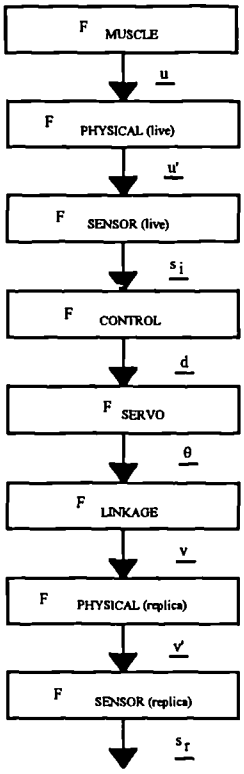


Figure 4.5 Functional Representation Of Overall System

4.3.1.1 Description Of The Individual System Functions

The function F_{muscle} represents the physiological functions in the human face which produce muscular changes and their characteristics are difficult to define.

The function $F_{physical(live)}$ represents the difference between actual vector displacements of the live face ($\underline{u}$) and the measured vector displacements by the sensing system ($\underline{u}'$). It was assumed to be linear and equal. The assumption was made that the displacement of each key point was purely along the focal axis of its respective sensor. The same argument was expanded to the function $F_{physical(replica)}$ for the key point displacements on the replica.

The function $F_{sensor(live)}$ was defined as the transfer function of the sensing system on the live face. In terms of it's design, this was defined as a linear relationship between the displacement input ($\underline{u}$) and the produced output signal ($\underline{S_l}$). The proposed type of sensor should be considered as a zero order measurement system, where its output was related to its input by means of an equation of the type $y(t) = k \cdot x(t)$ **[Pallás-Areny91]**. Its behaviour was characterised by its static sensitivity k and remained constant regardless of the input frequency. Consequently, its dynamic error and delay were both definable as zero. In order to have this type of relationship, it was important that the system did not include any energy storing elements. The measurement could then be considered as instantaneous. The same assumptions are made for $F_{sensor(replica)}$.

The function $F_{control}$ represents the mapping function of the system between the control input ($\underline{S_l}$) and the positional offset positions for the drive output ($\underline{d}$). $F_{control}$ was constructed within the HPC System as a series of linear curves between definable control and drive parameters. For full details of its operation refer to Section 4.3.3.

The function F_{servo} represents the electro-mechanical transfer function of the digital control servo motor. At any instant in time, the output motor shaft position ($\underline{\theta}$) was assumed to be linearly proportional to the input motor offset signal ($\underline{d}$). This signal was in the form of the variable width of a pulse. Within the servo drive, circuitry is incorporated to feed back information on the output position. This ensures that sufficient damping is applied to prevent any non-linearity in the output due to

overshoot. In design terms, the servo was considered as static and linear whilst it was acknowledged that the electro-mechanical conversion may produce time delays which affect the motor response.

The function $F_{linkage}$ represents the mechanical linkage between servo output shaft, ($\underline{\theta}$) and the final facial displacement on the replica, $\underline{v}$ and was defined as linear in this theoretical model. The actual characteristics were created through the physical design and construction of the animatronic model. The mechanical linkage was designed to produce the connection between drive and face that overcomes the physical effects of friction and the latex skin. These actual effects were difficult to model until the final head is constructed.

Using the same type of representation as before, each of these functions can be stated in the following way;

$$\underline{u'} = \underline{F_{physical(live)}}(\underline{u}). \qquad \text{Equation [4.5].}$$

$$\underline{S_l} = \underline{F_{sensor(live)}}(\underline{u'}). \qquad \text{Equation [4.6].}$$

$$\underline{d} = \underline{F_{control}}(\underline{S_l}) \qquad \text{Equation [4.7].}$$

$$\underline{\theta} = \underline{F_{servo}}(\underline{d}) \qquad \text{Equation [4.8].}$$

$$\underline{v} = \underline{F_{linkage}}(\underline{\theta}) \qquad \text{Equation [4.9].}$$

$$\underline{v'} = \underline{F_{physical(replica)}}(\underline{v}) \qquad \text{Equation [4.10].}$$

$$\underline{S_r} = \underline{F_{sensor(replica)}}(\underline{v'}) \qquad \text{Equation [4.11].}$$

In summary, the parameters of the functions F_{sensor} and F_{servo} were defined by their physical characteristics, $F_{linkage}$ were defined by the animatronic designer (c.f. Section 5.4) and $F_{physical}$ were defined by the physical relationship between facial and sensed motion. The parameters of the function $F_{control}$ were definable by the performer and are considered in Section 4.3.3.

4.3.2 Theory Of Objective Analysis

From the equations above in Section 4.3.1.1, the overall system equation [4.1], can now be defined as

$$\underline{v} = \underline{F}_{linkage}\left(\underline{F}_{servo}\left(\underline{F}_{control}\left(\underline{F}_{sensor(live)}\left(\underline{F}_{physical(live)}(\underline{u})\right)\right)\right)\right) \qquad \text{Equation [4.12].}$$

For the proposed objective analysis of the system, as discussed in Section 4.2, the following equation describes the relationship between the measured signals of live and replica faces.

$$\underline{S_r} = \underline{F}_{sensor(replica)}\left(\underline{F}_{physical(replica)}\left(\underline{F}_{linkage}\left(\underline{F}_{servo}\left(\underline{F}_{control}(\underline{S_l})\right)\right)\right)\right) \qquad \text{Equation [4.13].}$$

The assumptions were made that the sensor systems were identical in construction and that the physical displacements at input and output were along the same trajectories, i.e. $\underline{F}_{sensor(replica)} = \underline{F}_{sensor(live)}$ and $\underline{F}_{physical(replica)} = \underline{F}_{physical(live)}$. Therefore

$$\underline{S_r} = \underline{F}_{total}(\underline{S_l}). \qquad \text{Equation [4.14].}$$

From Section 4.3.1, for the function $\underline{F}_{total}$ to be the same in Equations [4.1] and [4.14], the assumption was made that a linear relationship existed between the live and replica faces. Ideally, $\underline{F}_{total}$ =1 and therefore $\underline{v} = \underline{u}$ and $\underline{S_l} = \underline{S_r}$.

$$\underline{F}_{total} = \underline{F}_{linkage} \cdot \underline{F}_{servo} \cdot \underline{F}_{control} \cdot \underline{F}_{sensor} \cdot \underline{F}_{physical} = 1. \qquad \text{Equation [4.15].}$$

Analysis between the measured signals for each key point should, therefore, allow assessment of the overall system's ability to sense, map and reproduce the key point displacements.

4.3.3 The Design Parameters For The Control System

Having established the theoretical model to represent the overall system, it was necessary to define the technique with which to automatically map the control input to the drive output. The Control system consisted of three distinct functions; the mapping function (F_{map}), the summation function (F_{sum}) and the motor conditioning function ($F_{motorcondition}$). All of these were produced within the Henson Performance Control (HPC) system and the function diagram of its operation is shown in Figure 4.6.

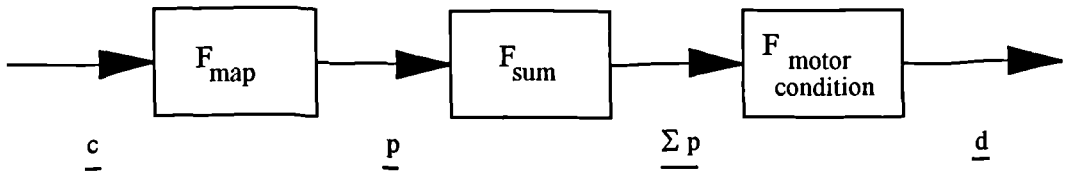


Figure 4.6 Function Diagram Of The Henson Control System

Using the same representation as the previous section, the operation of the control system could be defined by the following equations;

$$\underline{p} = \underline{F_{map}}(\underline{c}) \qquad \text{Equation [4.16].}$$

$$\underline{\Sigma p} = \underline{F_{summation}}(\underline{p}) \qquad \text{Equation [4.17].}$$

$$\underline{d} = \underline{F_{motorcondition}}(\underline{\Sigma p}) \qquad \text{Equation [4.18].}$$

The function $F_{motorcondition}$ enabled the mechanical designer or performer to set positional limits for the motion of each drive. These limits were important as they prevent physical damage from occurring to the skin, linkage or motor. They were defined by the visual assessment of the full scale displacement possible for each drive, and its related linkage, and the effect on the skin. These limits were defined as D_U

(upper limit) and D_L (lower limit). A third reference parameter, D_N, was also defined which represents the neutral or datum position for each drive and it's value is always set to zero. Consequently, any change in the drive signal would result in the production of an output action relative to this datum. This is comparable to the expressions of the face which are distinguished relative to the neutral or expressionless face (c.f. Section 3.4).

The output signal of $F_{motorcondition}$ is $\underline{d}$, the desired motor offset position, in the form of a pulse width modulation signal, relative to D_N within the ranges $D_L \leq \underline{d} \leq D_U$. The function of F_{sum} is the production of $\Sigma \underline{p}$ which is defined as the total change in the individual motor offsets $\underline{p}$ from F_{map} resulting from changes in all the input signals. This can be explained by the following example of a 4 X 3 matrix equation where M_{mn} represents the individual mapping functions;

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = \begin{bmatrix} M_{11} M_{12} M_{13} \\ M_{21} M_{22} M_{23} \\ M_{31} M_{32} M_{33} \\ M_{41} M_{42} M_{43} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}. \qquad \text{Equation [4.19].}$$

where $p_m = \sum_{mn} M_{mn} C_n$ Equation [4.20].

or in expanded form, $p_1 = M_{11}C_1 + M_{12}C_2 + M_{13}C_3$,

$$p_2 = M_{21}C_1 + M_{22}C_2 + M_{23}C_3,$$

$$p_3 = M_{31}C_1 + M_{32}C_2 + M_{33}C_3 \text{ and}$$

$$p_4 = M_{41}C_1 + M_{42}C_2 + M_{43}C_3.$$

F_{sum} produces a linear sum of the individual maps, $\Sigma \underline{p}$, and allows the possibility to link all the control inputs, $\underline{c}$, to all the drive outputs, $\underline{p}$. In the present HPC system, there are 24 control inputs available and upto 32 drives can be connected, i.e. $F_{control}$

represents a 24 X 32 matrix of M_{mn}. The majority of control developed in this research was greatly reduced from this, typically as (1 X 1) or (1 X 3) or (2 X 2) relationships. Where the control has no effect on a drive, the drive value is held constant at D_N, i.e. zero. The neutral position always ensures that the drive takes a defined position rather than some arbitrary value.

The individual mapping functions (M_{mn}) define the relationship between each individual control and drive. The typical map produced by Hensons is shown in Figure 4.7. The drive reference points are defined by the performer through the subjective analysis of the final output displacements, $\underline{v}$. They represent the desired positions for the final displacement and are, typically, the full scale deflection.

The control parameters; C_U, C_N and C_L, are defined by known reference positions in the range of the control. In the proposed system, they could be defined by the full scale displacement of each key point on the live face, $\underline{u}$, but, as stated before in Section 4.2, the exact and consistent production of specific displacements on the live face is difficult. The solution to this was to derive the reference limits from the replica as the proposed system would enable exact and repeatable control over the key point displacements. This was based on the theory that $\underline{v} = \underline{u}$ and $\underline{S_r} = \underline{S_l}$. Consequently, the control reference values for F_{map} were defined by $C_U = S_{rU}$, $C_L = S_{rL}$ and $C_N = S_{rN}$ where S_r represents the measured signal at known output positions on the replica.

Between these known reference limits, a mathematical model is produced to define the overall relationship. In this particular system, the map defines the function characteristic as two distinct linear sections bounded by the stored control and drive limits. All other values are determined by a straight line interpolation. This has the advantage of requiring the storage of only a minimal number of reference parameters from which the overall characteristic could be defined.

The mapped link between the neutral reference values, C_N and D_N, is commonly used in Henson control systems and was initially adopted within this research. It is used to ensure that each control, of any type, has an easily definable position which will always result in the drive moving to its neutral state.

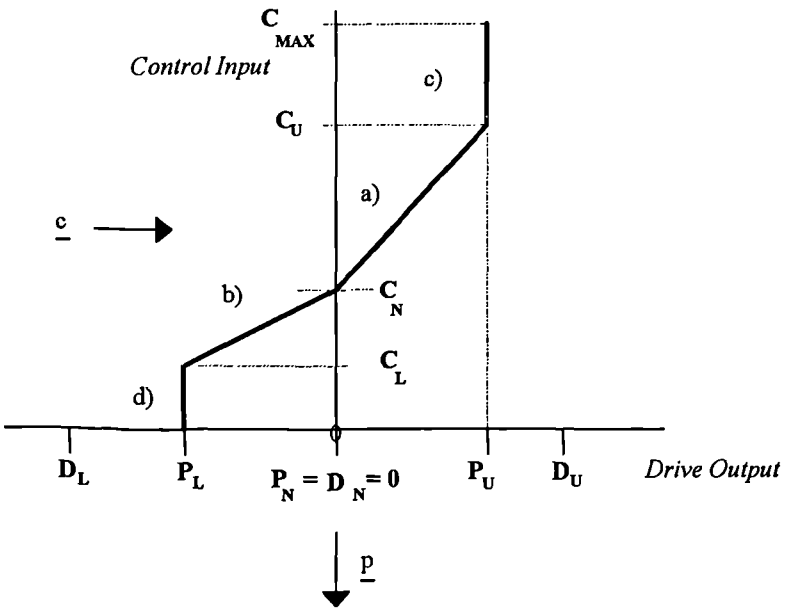


Figure 4.7 Diagram of Two Part Linear Mapping Function

F_{map} contains a number of distinct regions. Consider the one to one correspondence between drive p_1 and control c_1. The linear relationship could be defined by the equation

$$p_1 = F_{11}(c_1) = g \cdot c_1 + k. \qquad \text{Equation [4.21].}$$

where g is the slope of line and k is the intersection on y axis. For this type of map, k represents the value when p_1 was at neutral, $p_n = 0$, which has been defined as the neutral control position or datum C_N. The following equations defined each region.

Region a). where $C_N \le c_1 \le C_U$, $p_1 = g_a \cdot c_1 + k_a$,

$$g_a = \frac{P_U - P_N}{C_U - C_N} \text{ and } k_a = C_N.$$

Region b). where $C_L \le c_1 \le C_N$, $p_1 = g_b \cdot c_1 + k_b$, $g_b = \frac{P_N - P_L}{C_N - C_L}$ and $k_b = C_N$.

Region c). where $C_U \leq c_1 \leq C_{MAX}$, $p_1 = P_U$.

Region d). where $0 \leq c_1 \leq C_L$, $p_1 = P_L$.

Regions c) and d) define when the control input exceeds its reference parameters and result in fact that change in c_1 would produce no change in resultant drive position, p_1. There also exists the possibility that the drive parameters, P_U and P_L, may be set at values greater than the defined physical limits, D_U and D_L. This would produce similar regions where the control had no effect on the drive.

4.3.4 Conditioning Of Input Control Signals

4.3.4.1 Two Part Linear Conditioning

Within the overall system there was a desire to condition, or amplify, the input signals firstly, to increase the sensitivity of the mapping function to allow a more precise definition of the output position and secondly to use the HPC system as a "stand alone" tool once the initial set of reference values have been defined. This would allow for a straightforward inter-face between different control types by maintaining F_{map} constant and applying relevant conditioning.

For the previous F_{map}, this resulted in the control parameters being set to full scale values of $C_U = 255$, $C_L = 0$ and $C_N = 128$ with all other values determined by the sectional interpolation.

The conditioning function ($F_{controlcondition}$) in Figure 4.8 was defined by the following equations for the two part linear F_{map}.

Region a). where $S_{rN} \leq s_1 \leq S_{rU}$, $c_1 = g_c \cdot s_{11} + S_{rN}$,

$$g_c = \frac{C_U - C_N}{S_{rU} - S_{rN}} \text{ and } k_c = S_{rN}.$$

Region b). where $S_{rL} \leq s_{I1} \leq S_{rN}$, $c_1 = g_d \cdot s_{I1} + S_{rL}$,

$$g_d = \frac{C_N - C_L}{S_{rN} - S_{rL}} \text{ and } k_d = S_{rL}.$$

Region c). where $S_{rU} \leq s_{I1} \leq S_{rMAX}$, $c_1 = C_U = 255$.

Region d). where $0 \leq s_{I1} \leq S_{rL}$, $c_1 = C_L = 0$.

4.3.4.2 One Part Linear Conditioning

As an alternative to the two part F_{map}, a single line function bounded only by the upper and lower limits could be defined. The neutral link would be removed from both F_{map} and $F_{controlcondition}$. A typical representation of $F_{controlcondition}$ for this type of mapping is shown in Figure 4.8. It's overall performance would be dependent on the linearity of the system. The main region, $S_{rL} \leq s_I \leq S_{rU}$, would be defined by

$$c_1 = g_e \cdot s_{I1} + S_{rL}, \text{ where } g_e = \frac{C_U - C_L}{S_{rL} - S_{rL}} \text{ and } k_e = S_{rL}.$$

4.3.5 Practical Compensation Technique For Non-Linearity In Overall System

Having developed the ideal theoretical model, it was important to consider the practical system. This led to the following compensation technique which was designed to reduce the non-linear effects present in the final system.

From the Equation [4.13],

$\underline{s_r} = \underline{F}_{sensor(replica)}\left(\underline{F}_{physical(replica)}\left(\underline{F}_{linkage}\left(\underline{F}_{servo}\left(\underline{F}_{control}\left(\underline{s_I}\right)\right)\right)\right)\right)$, by removing the input control, $\underline{s_I}$, and replacing it with a different control signal, $\underline{c}$, which could

produce every drive output position, the measured signals, $\underline{S_r}$, would then describe the overall characteristic of the system.

$$\underline{S_r} = \underline{F_{sensor}}\left(\underline{F_{physical}}\left(\underline{F_{linkage}}\left(\underline{F_{servo}}\left(\underline{F_{control}}(\underline{c})\right)\right)\right)\right), \quad \text{Equation [4.22].}$$

$$\text{i.e. } \underline{S_r} = \underline{F_{total}}(\underline{c}). \quad \text{Equation [4.23].}$$

As principle has defined that $\underline{S_r} = \underline{S_I}$ so

$$\underline{c} = \underline{F_{total}}^{-1}(\underline{S_r}) = \underline{F_{total}}^{-1}(\underline{S_I}) \quad \text{Equation [4.24].}$$

Therefore, this inverse compensation function $\underline{F_{total}^{-1}}$ could be created as a "look up" table from the measured characteristic, $\underline{S_r}$, where each value describes the corrected control input required to produce the desired output **[Trankler89]**. An example is shown in Figure 4.9. A disadvantage of this method is the amount of storage required to keep the large number of reference points necessary to obtain the sufficient accuracy in the signal conditioning.

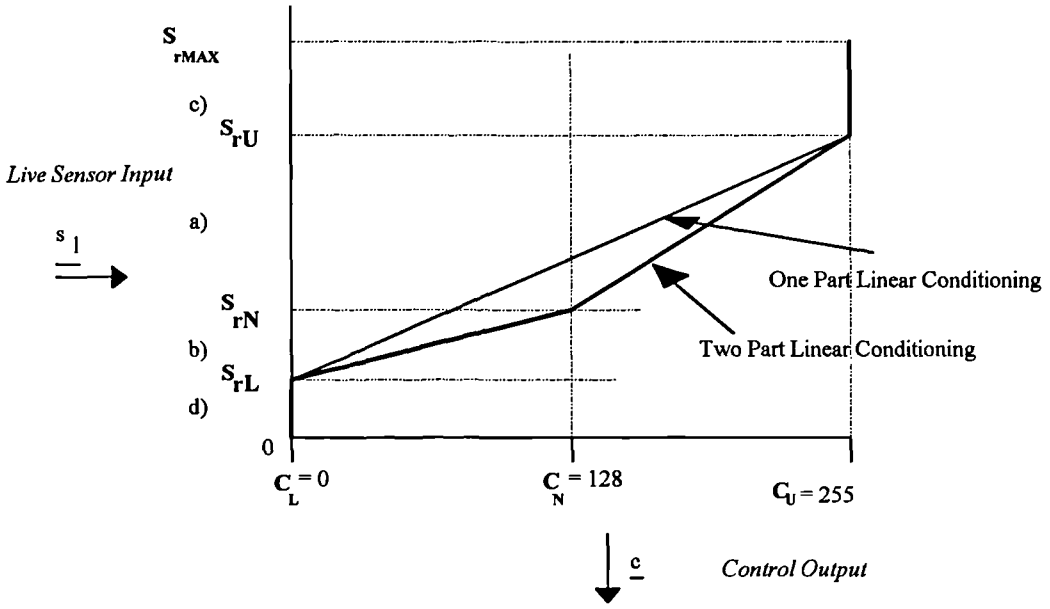


Figure 4.8 Graphical Representation Of Conditioning Of Control Inputs

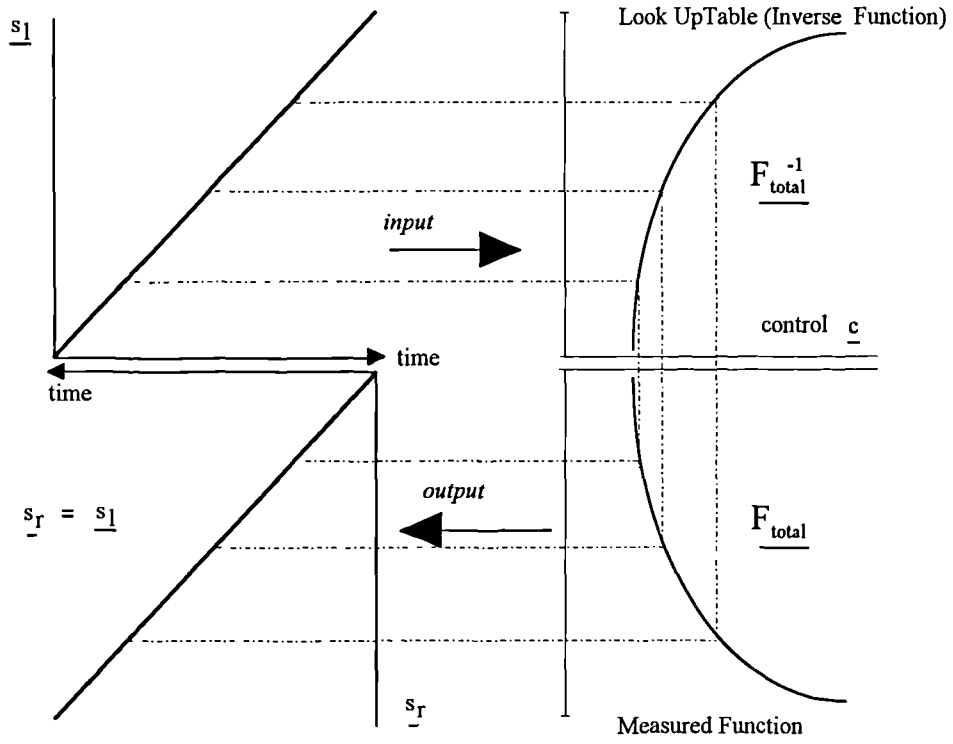


Figure 4.9 Example Of Non-Linear Compensation Technique

4.3.6 Summary Of Function Objectives

To achieve an operational system based on the theoretical model, the following criteria had to be realised.

1. The creation of a sensing system of the zero order type to ensure linearity of measurement.

2. The development of a technique to hold sensors in fixed axes to allow measurement of displacement vectors along linear paths.

3. The design and construction of the replica which was capable of the production of linear displacements at specific key points.

4. The subjective setting and storage of both sets of drive parameters, those for the physical limits and those for the desired range of displacements.

5. The correct measurement of control parameters from the replica head for input signal conditioning.

6. The development of software to produce the different conditioning and compensation functions.

7. The development of a data acquisition system for the recording and playback of control signals to allow both repeated playing of sensor control and also separate software control.

4.4 Definition Of Primary Visible Actions And Key Point Derivation

Having established the hypothesis and method of solution in the Section 4.2, it was necessary to draw a number of conclusions from the review of facial communication in Chapter 3. Firstly, it was proposed that there exists a fundamental set of visible

facial actions that satisfy the criteria for the production of realistic lip synchronisation and primary expressions for animation. These defined actions represent the test data for the sensing system and the design of the animatronic head. Secondly, it was necessary to define the optimum set of key points on the face to provide sufficient information on the primary actions which could be both sensed and animated. For each defined key point it was important to deduce its properties from the live face. The measured information on the individual displacements and trajectories during the primary actions represented the final design criteria for the sensing system (the input) and the animatronic model (the output).

4.4.1 Definition Of Fundamental Set Of Visible Facial Actions

Within Chapter 3, the thesis established the different groupings and notations used to describe visually distinct facial actions. The main groupings are visemes, facial action units and visually articulatory actions. It was necessary to derive an optimum set of actions from these groupings, using a standard notation, which represent the basis for the production of realistic life-like animation. In Appendix A, **[Laver80]** simplifies the description of the visible articulatory gestures to purely horizontal and vertical muscular constrictions and expansions with or without protrusion. These descriptions are too vague to be of practical use but they highlight the desirability of a reduced set of descriptors. Experiments in visual speech perception have established that viewers can only distinguish a limited set of distinct actions **[Jackson88]**. Though Ekman has stated that the face is capable of over 7000 unique visible actions **[Goleman81]**, their production is based on a significantly smaller set of distinct actions. This set can be reduced further in facial performance given that many are too subtle for recognition by untrained viewers.

The definition of the final set of actions is shown in Table 4.1. A large number of combinations, notably all the primary visemes, are produced from this set of individual actions. Not all of these actions can be produced in combination due to the physiology of the face, refer back to Section 3.2.

The research into viseme vowels, in Section 3.3, suggested that the "vowel triangle", **[Berger72]**, could be reduced to a fundamental set based on the vowels 'positioned' at the corners. These were defined as /oo/, ("boot"), /ee/, ("bee") and /ar/ ("bar").

The primary viseme consonants were defined as merely /b, p, m/ and /f, v/. This overall set represents the minimum information required for the system to achieve successful animation. The assumption was made that the remaining visemes, vowels, and consonants could be considered as either visually similar to this set but of a proportionally smaller degree of articulation (magnitude) or they are produced through some other combination of primary action units.

As stated in Section 3.1, this research has not considered the actions of the eyes and eyelids or the actions of the tongue. The omission of the tongue results in the system's inability to reproduce linguo-dental consonants, such as /th/, where the tongue is clearly visible. Research suggests that, when produced, they are likely to be confused with the labio-dental consonants, such as /f/, since they share similar articulatory gestures **[Jackson88]**.

Table 4.2 defines the fundamental set of visemes in terms of the combination of actions required for their production. Similarly in Section 3.4, Table 3.6 describes the construction of the universal emotions from the facial action units.

4.4.2 Derivation Of Optimum Set Of Key Points For System

Given the proposed method of facial control, it was necessary to derive the optimum set of points required to provide sufficient information on the facial changes that result from the production of the primary actions defined above.

All points are described at the neutral position of the face, defined as the state when all the muscles are in a relaxed state which can be described as "expression-less". **[Laver80]** defined the neutral setting of the mouth as "the lips lightly touching each other with no stretch or protrusion where the jaw is neither closed or unduly open". All displacements of the face are then described relative to this neutral setting. Though the upper and lower lips are physically connected through the same muscle, the *orbicularis oris*, it was necessary to consider them as distinct since they are manipulated largely by separate muscles that allow independent motion **[Hardcastle78]**.

Along with the research into the human facial system in Chapter 3, the points defined in the work of **[Brooke83]**, **[Choi90]**, **[Finn88]**, **[Morishima91a]** and **[Pearce86]** were considered in the derivation of the optimum set of key points. Brooke distinguished inner and outer margins for the lips at the centres and corners of the mouth. In terms of sensing, the exact definition of such distinct points is limited by possible confusions with the teeth and tongue and in the difficulty of deciding where the inner margin begins. Therefore the points around the lips are situated at the outer margins defined by the vermilion, or red lip, margin. The derived set are listed in Table 4.3.

	Description	Action Unit Number or Equivalent
Lower Face	Jaw Open / Close	AU26 Jaw Drop
	Lips Stretch	AU20
	Lip Corner Puller	AU12
	Lip Corner Depressor	AU15
	Lips Protrude	AU18 Lips Puckered
	Lips Part	AU26
	Lips Pressed	AU24
	Lower Lip Tuck	no AU definition; possible combination of AU17 Chin Raiser With Lower Lip Restricted By Upper Teeth
	Upper Lip Raiser	AU10
Upper Face	Brows Inner Raiser	AU1
	Brows Outer Raiser	AU2
	Brows Together and Lower	AU4

Table 4.1 Table Of Fundamental Actions Necessary For Facial Performance

Description	Action Unit Equivalent	Viseme Equivalents: Vowels	Viseme Equivalents: Consonants
Open / Close	AU26 Jaw Drop	/ar/	
Spread	AU20 Lips Stretch + AU12 Lips Part	/ee/	/s/, /z/
Protrude 1: Close Rounding	AU18 Lips Puckerer	/oo/	/sh/, /w/
Protrude 2: Open Rounding	AU22 Lip Funneler or AU18 Lip Puckerer + AU25 Jaw Drop	/o/	/r/
Labio-Dental	Combination of AU10 Upper Lip Raiser + AU17 Chin Raiser With Lower Lip Restricted By Upper Teeth		/f/, /v/
Lips Relaxed with Opening	AU26 Lips Part		/t/, /d/, /n/, /k/, /g/
Lips Pressed	AU24		/b/, /p/, /m/
Lip Corner Depressor	AU15		
Brows Inner Raiser	AU1		
Brows Outer Raiser	AU2		
Brows Together and Lower	AU4		

Table 4.2 Construction Of Visemes From Primary Actions

Point Number	Defined Key Points
1 & 4	Outer Brows, Left and Right
2 & 3	Inner Brows, Left and Right
5 & 8	Lip Corners, Outer Margin, Left and Right
6	Upper Lip Centre, Outer Margin
7	Lower Lip Centre, Outer Margin
9	Mid Point Between Corner And Centre Of Upper Lip
10	Mid Point Between Corner And Centre Of Lower Lip
11	Tip Of Jaw / Chin

Table 4.3 Defined Set Of Key Facial Points

4.4.3 Photogrammetric Analysis Of Facial Key Point Actions

Whilst the work in the computer generated animation has defined different sets of points, few have measured the actual physical changes in the points. The majority rely on the decisions of the programmer to define the maximum and minimum limits of displacement.

Along with the definition of this optimum set of points, measurements of the displacements and trajectories of each key point was essential to define the requirements of the sensor system and the design of the animatronic model.

A number of research projects have investigated the displacement changes of the face, **[Abry86]** and **[Fromkin64]**. Their measurements, along with the results from Brooke's system **[Brooke83]**, provided useful indicators of the degree of motion of the facial parameters. As stated in Section 3.3.4, there are a number of factors which produce variations in visible speech by different speakers. Thus there was a necessity to measure the researcher's head in articulation to establish the exact changes that occur.

The aims of the investigation were to measure the displacements of the previously defined set of key points in three dimensions. Photogrammetric measurements in three dimensions were derived from static facial images for the different primary actions at maximum intensity and also for the natural intensity articulations of the primary visemes. By adapting the techniques of **[Fromkin64]** and **[Brooke83]**, standardised simultaneous frontal and lateral view photographs were obtained in sufficiently natural speaking conditions.

[Fromkin64] described the necessity for apparatus to remove any slight head movements which could lead to measurement errors. The apparatus was designed to cause minimum discomfort to the subject. It placed the subject's head (using a head rest) in the same fixed position for every photograph, with regard to the tilt of the head and the distance from the camera and lighting. The apparatus is illustrated in Figure 4.10. A plane mirror was placed adjacent to the subject's head with its horizontal axis at 45 degrees to the principal axis. Each key point was indicated by a small reflective disc of 5 mm diameter glued to the skin. The camera was fixed at a distance of 1.5 m away and was focused at the subject's nose.

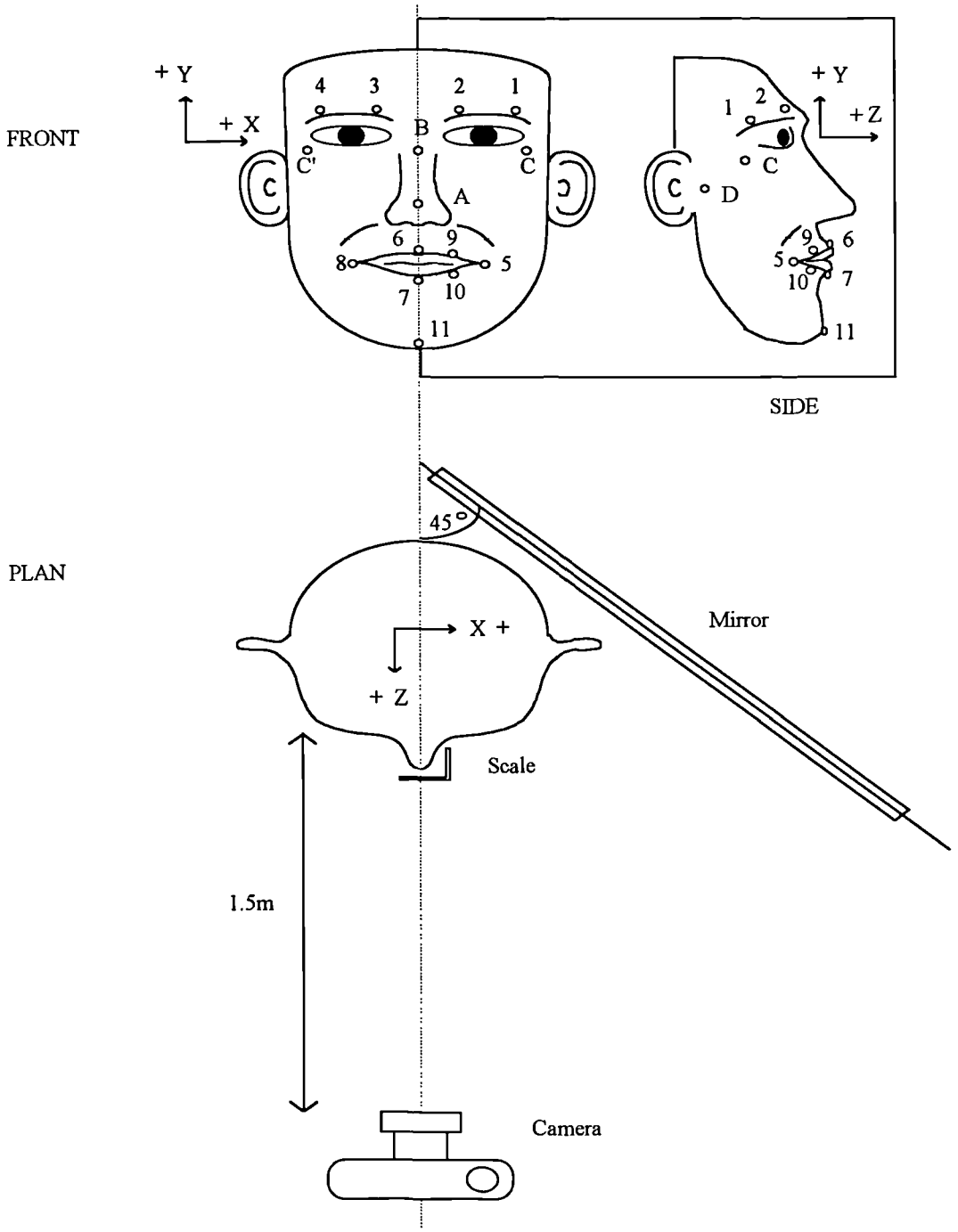


Figure 4.10 Photogrammetric Experimental Arrangement

To calibrate the photographs, horizontal and vertical scales were placed in the same position and distance from the camera as the lips at the neutral expression. The longitudinal scale was placed in same vertical axis as centre of the lips, i.e. at the same distance away from the camera as the nose tip. The lateral scale was placed in same horizontal axis as the lips in neutral position. This can be seen in the photographs of Figure 4.11.

The displacements of each key point were derived relative to this initial datum. A set of points were defined at the rigid areas of the face and they were used to correlate the displacements of the articulatory points. The rigid points were defined at the tip of the nose (A), the bridge of the nose (B), the corners of the cheek bones (C) and at the pivot point of the jaw (D). The use of these points compensated for movement of the head position in each photograph.

The key expressive points on the face are listed in Table 4.3 and the input actions are listed in Table 4.4. Photograph numbers 1 to 8 represent the primary facial actions at maximum intensity. These levels of expression are unlikely to occur in the majority of speech production so a further set of photographs, numbers 9 to 13, were taken of the primary visemes with a moderate degree of action intensity. Each viseme photograph was taken at the mid-point of articulation which represented the most visually perceptible expression.

Tracings were made of each key point displacement from the series of photographs relative to the neutral datum. By tracing each point in frontal and side views, measurements were possible in each of the Cartesian co-ordinates. The side (longitudinal) axis is represented by z co-ordinates and the frontal (lateral) by x and y co-ordinates. The positive direction of motion is defined in Figure 4.10. In the side view, the measurements of lip protrusion and jaw rotation were made using the contour features of the lips and chin.

The actual physical changes in each point's position were determined using the calculated scaling factor from the neutral photographs (these measurements are listed in Appendix C). Figures 4.12 and 4.13 show the resultant key point displacements for the primary actions and the visemes respectively.

The measured dimensions for each of the displacements could only be considered as estimates rather than exact measurements for a number of reasons. Firstly, the physical production of static expressions was not a natural process. Despite the efforts to maintain the head in the same position and orientation, slight variations occurred which were likely to affect the readings although they were compensated for by use of the rigid points. The set of fixed points were not completely rigid and could be affected by the actions. The indicators of each point were 5 mm in diameter and even though considerable effort was taken to ensure that the centre was traced, variations were likely to occur.

Photograph Number	Action Description
1	neutral closed
2	inner and outer brow raise
3	brows together and lower
4	lips puckered, jaw closed
5	lips puckered, jaw open
6	lips stretched, jaw closed
7	lips stretched, jaw open
8	jaw wide open
9	/f/
10	/b/
11	/oo/
12	/ee/
13	/ar/

Table 4.4 Table Of Actions Used in Photogrammetric Experiment

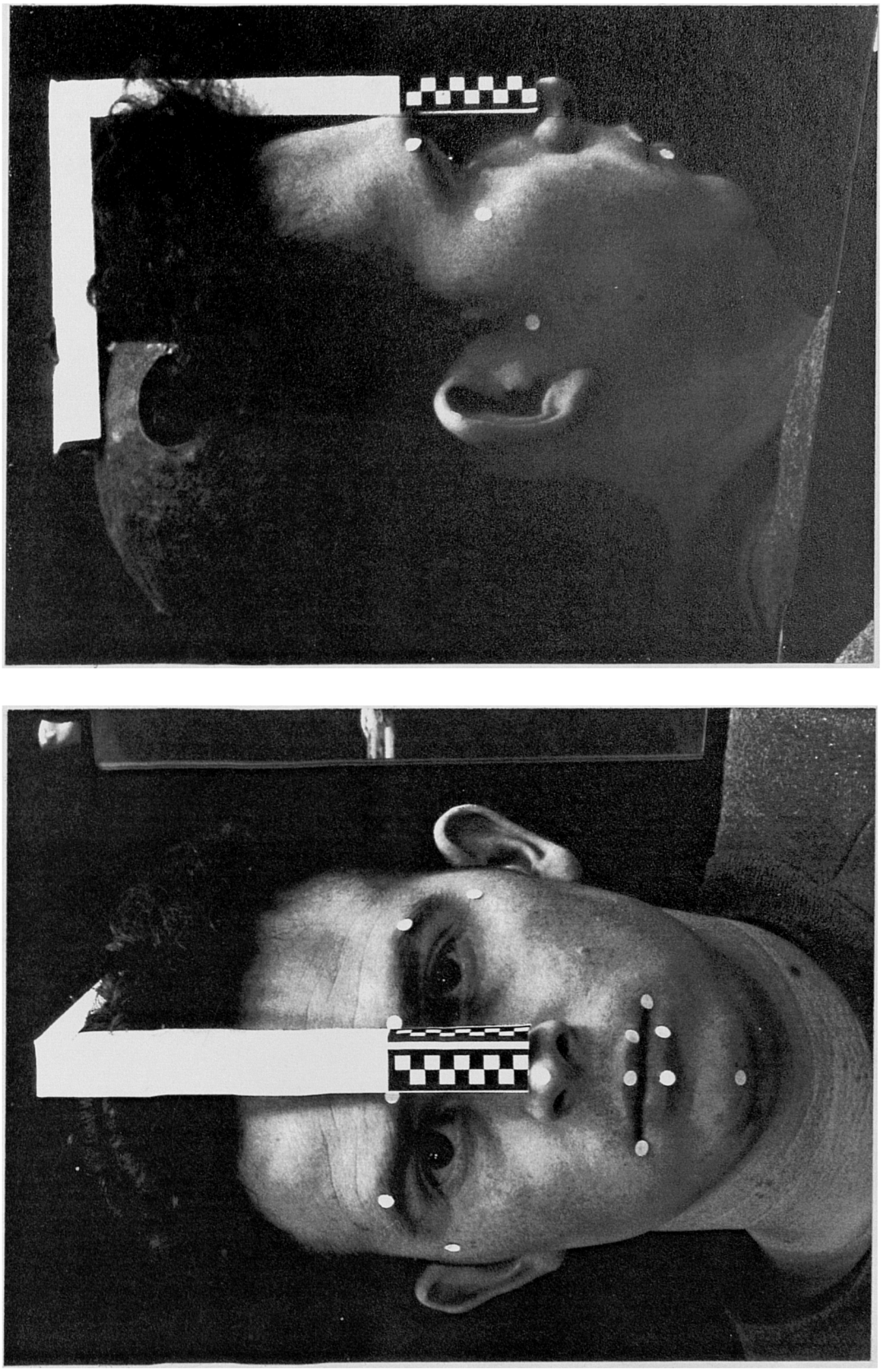


Figure 4.11 Photographs Of Neutral Face In Photogrammetric Measurement

Figure 4.12 Resultant Displacements Of Key Points For Actions Of Maximum Intensity

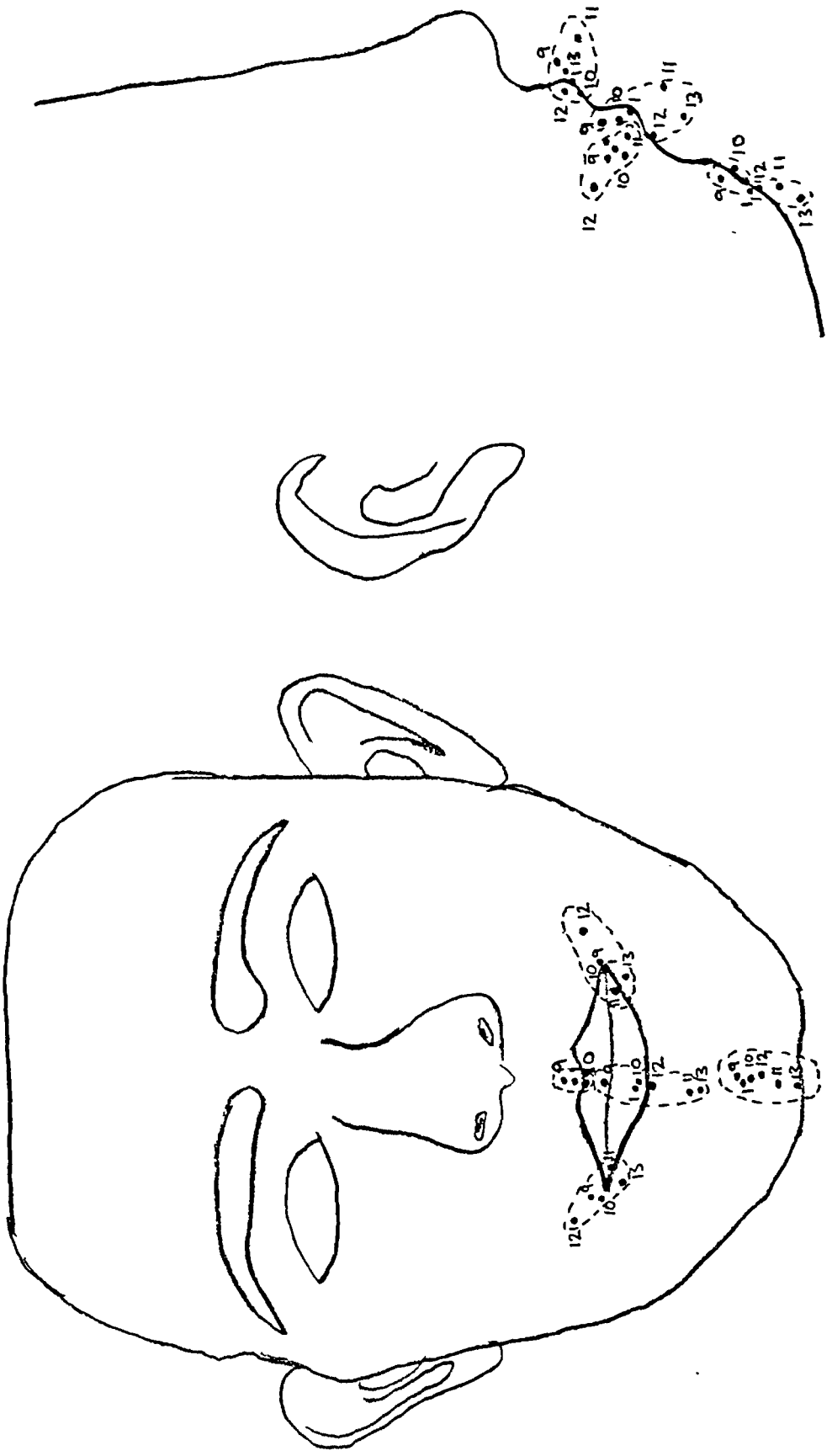


Figure 4.13 Resultant Displacements Of Key Points For Static Visemes

4.5 Principles Of Actual Design For Overall System

In Section 4.2, research proposed that the primary actions of the face could be reduced to a set of key point displacements. From the photogrammetric assessment of these key points, it was concluded that the majority of the displacements occur along trajectories on or around the face. For the correct measurement of the defined points, the sensors must be held in the same axes.

The objective of this section is to describe, firstly, the proposed technique to position sensors in their respective trajectories, secondly, to describe the specifications for the sensing of each individual displacement, thirdly, to describe the proposed design for the drive system and finally to discuss the design of F_{map} for each point.

4.5.1 The Design Of The Sensor Support Mask

A major objective in the realisation of the final system was the development of a system to hold sensors in their pre-defined axes to allow the sensing of the primary facial actions.

The displacement ranges of the facial actions are small relative to the size of the head. Therefore, the measurement of these actions must be separated from the effects of global movements of the talker **[Brooke83]**. **[Brooke83]** defined two methods of experimental analysis to eliminate global body movements; firstly, by restraining the head to a fixed position, which is likely to cause physical discomfort, and, secondly, by allowing the head and body freedom to move and employing mathematical correction techniques to extract required data. Both of these approaches are applicable only to image analysis techniques where the camera is remote from the face. The proposed method of action sensing in this research is dependant on measuring movements along paths that can occur in the plane of the face and not along Cartesian axes.

[Petajan88a] developed an alternative approach, for the measurement of the lips, by use of a head mounted camera. This allows the speaker freedom to move their head whilst retaining the camera in the same position relative to the mouth's position. It was concluded that an adaptation of this technique to mount the sensors about the

head, relative to the face, would allow the successful sensing of the defined key points.

The following requirements were important in the production of a suitable mounting system:

1. to hold and maintain the sensors in the desired position with respect to the face and irrespective of the global head movements;

2. to allow unrestricted facial movement with minimal discomfort which should reduce the possibility of the production of unnatural facial actions; and

3. for repeated performances or tests, the system should consistently place the sensors and reflectors in the identical positions to retain the same transfer function between action and control signal, $F_{physical}$.

The other requirement for a successful sensing system was the development of a method to position reflectors at the defined key points. The shape and orientation of these reflectors should allow the measurement of the relative displacements of the points.

The system developed to realise the above requirements was a facial mask, made in lightweight fibreglass, of the exact dimensions of the researcher's face and head. Figure 4.14 shows photographs of the final sensor mounting system. Using the cast of the researcher's head, the mask was constructed to have the distinct, neutral, facial features imprinted onto its' inner surface. When the mask was worn by the speaker, it fitted exactly on the face and then only in one unique orientation. This solved the problem of placing sensors in same position for repeated tests. The mask was held in position by clamping the front piece to the rear portion to utilise the rigidity of the rear skull. This coupled the mask to the global actions of the head and as a result maintained the specific relationships, $F_{physical}$, between sensor and reflector.

The rigid facial features in the mask; the nose, cheek bones and chin, were retained but the areas affected by facial expression were removed. This allowed the sensors to be mounted in their defined orientations whilst not restricting the speaker's facial movements. The sensors were mounted on the mask using a framework of brass rods

as shown in Figure 4.14. These allowed the position and orientation of the sensor to be altered where necessary. A sensor (S_{nose}) was mounted on the mask and directed, via a hole in the mask, at the tip of the nose. This should provide an indication of any change in the mask's position relative to the performer's face.

4.5.2 The Jaw

Results from analysis in Section 4.4.3, indicated that the range of jaw displacement is 20 to 30 mm. This displacement was likely to exceed the range of the proposed sensor, hence an alternative measurement technique was derived.

During speech production, the jaw action could be considered as a purely rotational action about an axis located at the *ipsilateral condylion* close to the ear, as shown in Figure 3.2. As a result, the drive system was designed as a single drive and linkage to rotate the replica's jaw about this defined axis. The method to measure this rotation used a linear potentiometer positioned on the same axis with its sliding arm affixed to the actual jaw. The body of the potentiometer was held rigid to the upper part of face. Diagram a) of Figure 4.15 shows the principle of design for one to one mapping with only two parameters; D_N with jaw closed and D_U with jaw at maximum open.

Using calliper measurements from the researcher's face, both points of rotation could be determined on the inside of the support mask. Great care was taken to ensure that these were the correct points as an incorrect definition of the axis would produce incorrect measurements as well as physical discomfort to the wearer. Diagram b) of Figure 4.15 shows the final measurement system using a linear potentiometer to measure the jaw rotation about the defined axis. The control signal produced can be applied directly to the jaw drive as the drive system is designed on same principle of rotation.

The proposed design for $F_{control(jaw)}$ was defined by following equation.

$$d_{jaw} = F_{control(jaw)}\left(S_{1(jaw)}\right). \qquad \text{Equation [4.25].}$$

This could be reduced to a linear interpolation function bounded by the upper limits (P_U and S_{rU}) and the lower limits (P_N and S_{rN}) as shown in diagrams' c) and d) of Figure 4.15.

4.5.3 The Brows

Both inner and outer brows move primarily in a single path along the same plane as the forehead. The brows of the live face do not have the same flexibility as those defined by **[Ekman79],** and are limited to purely raise and lower.

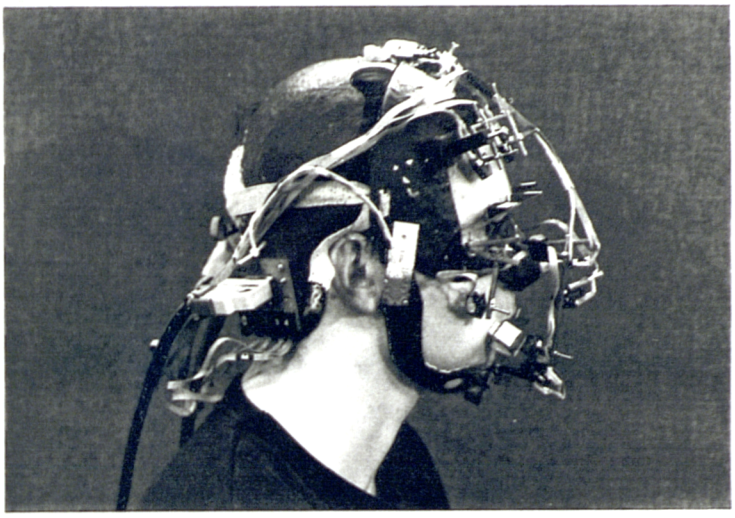
The range of motion is approximately ± 8 mm in raise and lower about the defined datum. The trajectory of these displacements poses problems in tracking with the optical sensing system. The following argument provided a viable solution to this problem. The frontal bone in the human forehead was defined as a flat surface over which the skin moved freely. The motion of the brows could then be sensed in the proposed apparatus shown in Figure 4.16. A reflector, positioned at the key point with its reflective area perpendicular to the facial surface, would transpose the motion of the surface point to a path parallel to the forehead. A sensor held in the same trajectory would measure the changes in the reflector position that correspond to the surface point position.

Both the inner and outer points in the brows were driven by individual drives. The linkages were designed to move the point on the surface through the same trajectory and range as that of the live point. This reduced $F_{control}$ to a one to one correspondence between the control and drives for each point as shown in Figure 4.17 and defined by the following equation.

$$d_{brow} = F_{control(brow)}\left(S_{l(brow)}\right). \qquad \text{Equation [4.26].}$$

For the two part linear maps, the drive parameters were defined at maximum upper (P_U) and lower displacements (P_L) and at the neutral position (P_N). The control parameters were measured from the replica at these full range displacements; S_{rU}, S_{rL} and S_{rN} respectively. Between these limits, all other values were derived through linear interpolation.

Photograph a)



Photograph b)

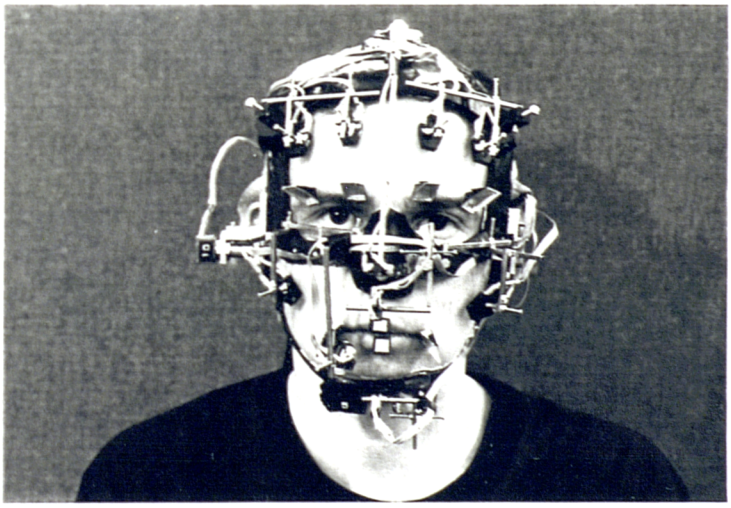
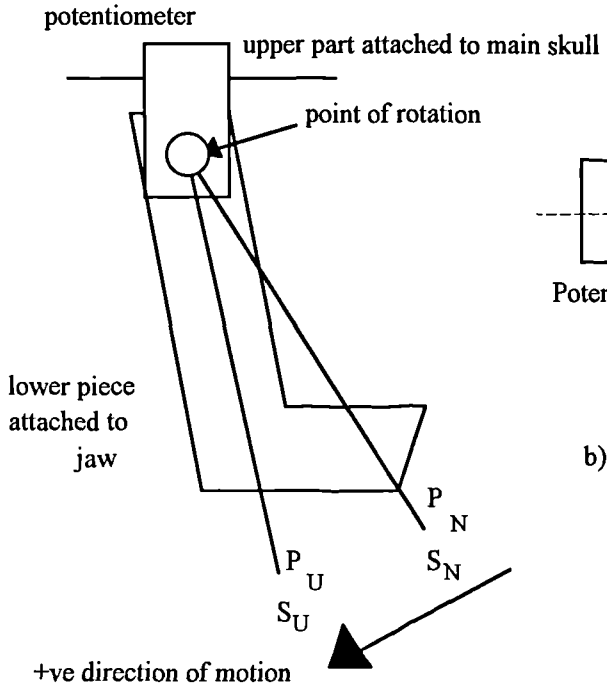
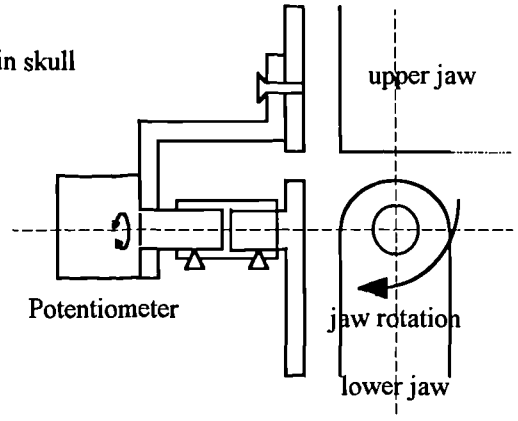


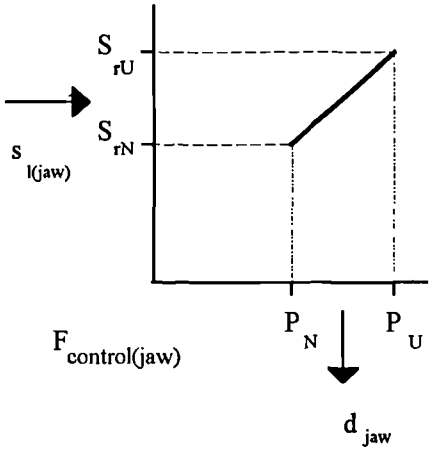
Figure 4.14Photographs Of Sensor Support Mask



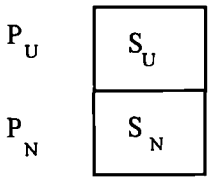
a) Proposed Design: Sensing And Drive



b) Proposed Design: Sensing And Drive



c) Mapping Function



d) Permutation Representation

Figure 4.15 Design For Jaw Sensing, Mapping And Drive Systems

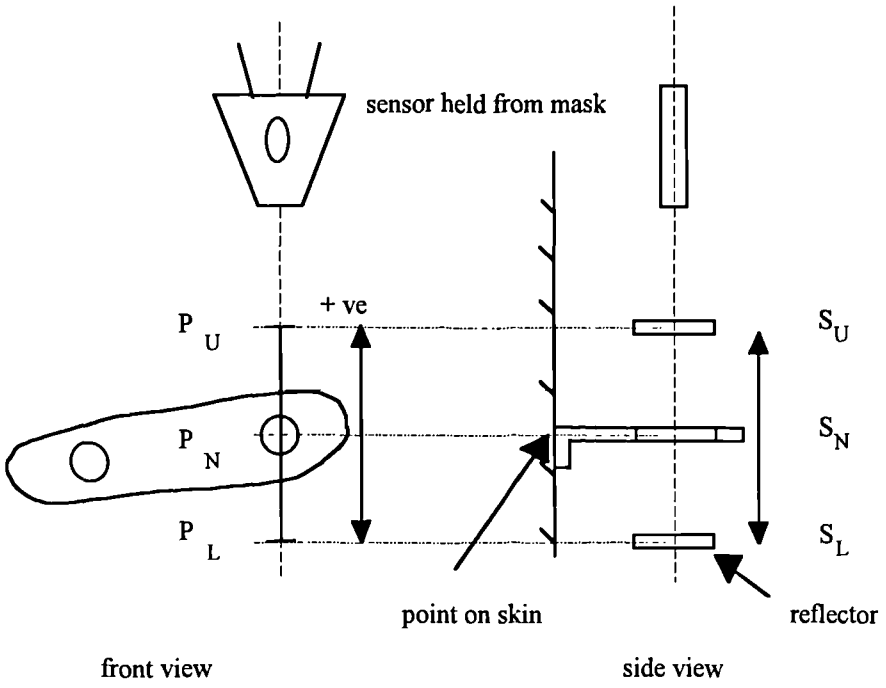
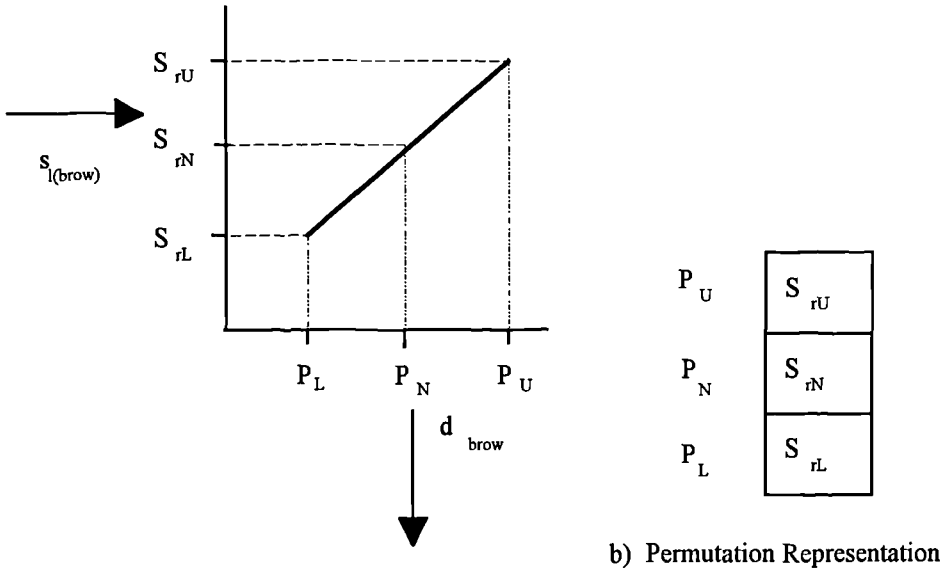


Figure 4.16 Design For Drive And Sensing Systems in All Brows



a) Mapping Function

b) Permutation Representation

Figure 4.17 Mapping Theory For Each Of The Brows

4.5.4 The Upper Lip Centre

The photogrammetric measurements suggest that the upper lip centre exhibits changes in position along a single path primarily in the z axis. The limits of displacement are defined by protrusion and stretch against the teeth. The upper lip-centre position is unaffected by the action of the jaw. Similarly, the centre displays no displacement in the x axis. The approximate range of displacement is - 5 mm to + 10 mm in z axis and - 5 to +2 mm in y axis. It was proposed that a sensor, held along the defined axis, would accurately measure these displacements, if a reflector was positioned at the defined point.

Since the range of positions covers more than one axis of motion, the proposed driving system consisted of two drives; drive (D_8) to produce z axis variations and drive (D_9) to produce y axis displacements. Given the proposed linearity of design, the displacement of each drive (D_8 and D_9) can be reduced to three reference parameters: P_U; P_N; and P_L. The actual point of interest could therefore be defined at nine different reference points. The overall argument was based on the assumption that linear interpolation between these maximum parameters would define the correct trajectory of motion. The results from Section 4.4.3 suggested that only a single trajectory, defined by a combination of only three of these permutations, was likely to produce the correct perceptual changes.

As shown in Figure 4.18, the permutations could only be predicted before actual construction. These final positions must be defined by subjective analysis.

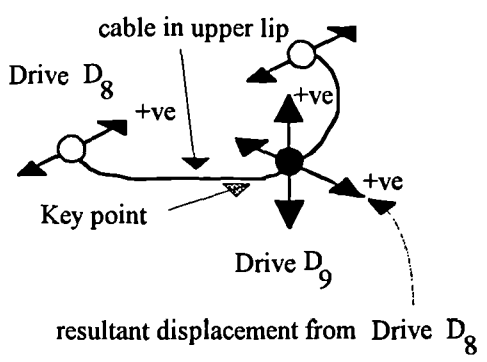
The following equations define the action of the upper lip.

$$s_{l(upperlip)} = F_{sensor(upper)}\left(u_{upper}\right). \qquad \text{Equation [4.27].}$$

$$d_8 = F_{control(8)}\left(s_{l(upperlip)}\right). \qquad \text{Equation [4.28].}$$

$$d_9 = F_{control(9)}\left(s_{l(upperlip)}\right). \qquad \text{Equation [4.29].}$$

$$v_{upper} = F_{drive(8)}(d_8) + F_{drive(9)}(d_9). \qquad \text{Equation [4.30].}$$



a) Different Drive Positions

Drive D_9 \ Drive D_8	P_L	P_N	P_U
P_U	X	X	S_{rU}
P_N	X	S_{rN}	X
P_L	S_{rL}	X	X

b) Predicted Permutation Map

Figure 4.18 Design For Upper Lip Centre Key Point

4.5.5 The Mid Points Of The Lips

The displacements of the mid-points of both lips were small, relative to those at the centres and corners. Their dimensional changes were a direct result of the actions at the centre and corner points rather than through any independent action at each point. It was concluded that these points were of relevance in the conformation of the lips but that their motion was directly linked to the motion of other points and hence their animation could be produced from these points.

4.5.6 The Lower Lip Centre

The analysis of the lower lip displacements indicated that its vertical displacements are directly related to the action of the jaw. For every photograph, a line was drawn between jaw pivot point to measured chin position. The lower lip position was then measured relative to this axis. From this plot, the displacements move, principally, along the same trajectory. The range of displacements was - 5 mm to + 10 mm in z axis and -2 mm to + 5 mm in y axis.

The proposed sensor system was based on the principle that a sensor held in a position unaffected by the jaw rotation, i.e. affixed in some way to the lower jaw, and in the defined trajectory would enable measurements of the relative lip displacements. Figure 4.19 illustrates the technique proposed.

The proposed drive system was designed in an identical way as the upper lip centre with two drives to achieve the combined degree of freedom. To produce lip movements unaffected by the action of the jaw, the drives were held rigid on the lower jaw piece of the replica. Equations [4.27] to [4.30] are identical for the lower lip.

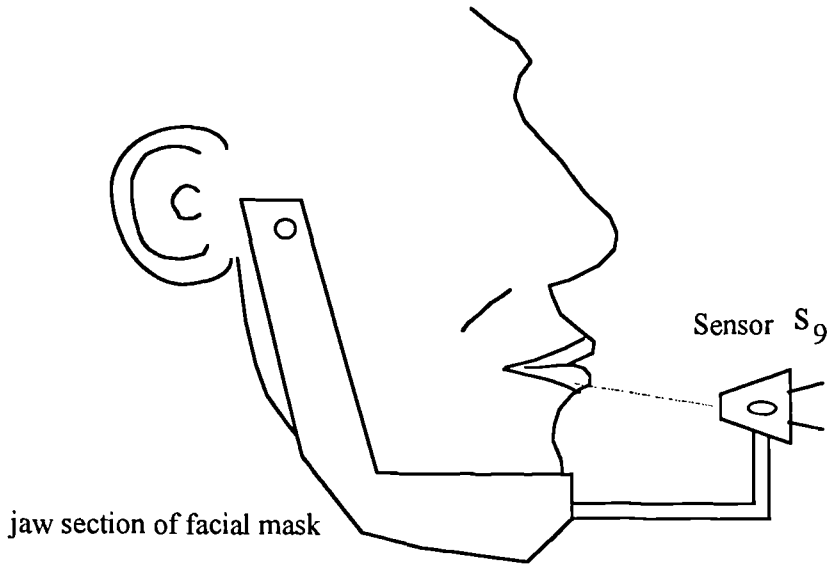


Figure 4.19 Design For Lower Lip Sensing System

4.5.7 The Corners

From the physiology of the human face, (c.f. Section 3.2), it was shown that the corner is the most complex region of the mouth due to the various inter-connections between muscles. The photogrammetric analysis confirmed that the corner is capable of taking numerous distinct positions during expression and articulation. From these measurements it was concluded that the overall motion of the corner could be

reduced to three distinct actions; corner stretch, corners protrude (or together) and corner drop, all relative to the neutral position. It was reasoned that all possible positions of the corner could, therefore, be defined by some combination of these three actions. The assumption was made that each of the maximum displacements was a result of a distinct combination of these actions and that all positions between displacements were definable by linear interpolations between these limits. This reduction is shown in diagram a) of Figure 4.20. For the purposes of this project, the action of "corner depressor" was ignored as it plays no major role in the production of visible speech. It was acknowledged that its action is vital in facial expressions such as "sadness" (c.f. Section 3.4) and this omission represents a possible limitation in the final performance.

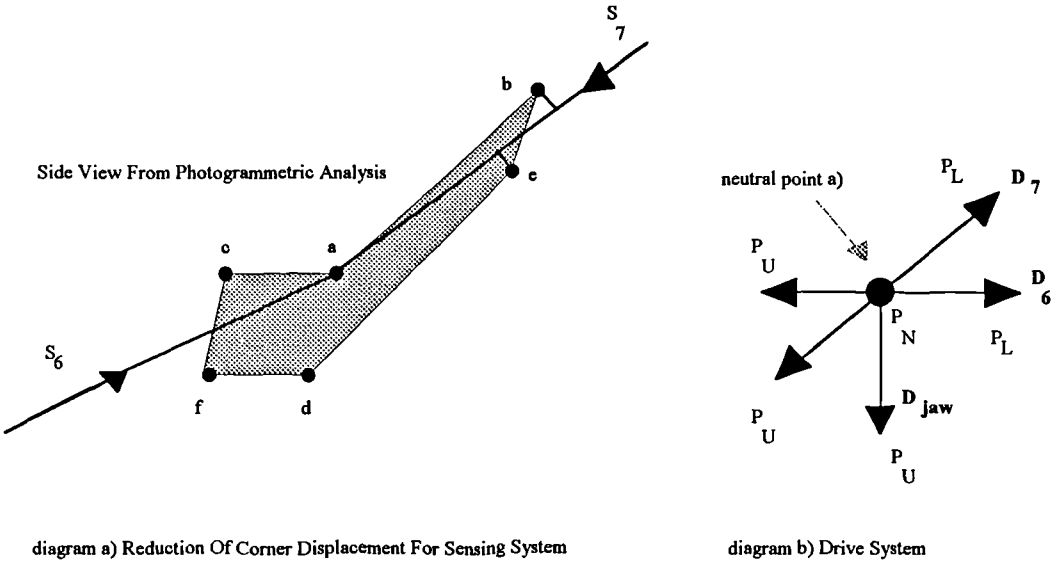


diagram a) Reduction Of Corner Displacement For Sensing System

diagram b) Drive System

Figure 4.20 Diagrams Of Corner Drive And Sensing Systems

From the analysis of the facial actions in Section 3.4, the following conclusions were drawn on the actions of the corner. Firstly, the actions of protrusion and stretch are mutually exclusive, i.e. cannot occur simultaneously. Secondly, the protrusion action is symmetrical about the centre of the face and, thirdly, the action of the stretch action is asymmetrical allowing idiosyncratic actions such as the emotion "contempt" to be produced. Finally, the action of the corner drop is direct result of the action of the jaw acting through the skin during jaw rotation. Similar reductions have been

used in computer animation **[Magnenat-Thalmann89a]**. As a result of these conclusions, the proposed design was identical for both corners with the action of corners protrude only sensed at one corner but mapped to both.

Table 4.5 indicates the permutation of drive and control signals required for the production of the six distinct positions, shown in diagram a) of Figure 4.20.

		Corners Protrude	Corner Stretch	Jaw	Horizontal	Vertical Raise	Jaw
Displacement Point	Action Description	Sensor S6	Sensor S7	Sensor Sjaw	Drive D6	Drive D7	Drive Djaw
a	neutral	N	N	N	N	N	N
b	lips stretch, jaw closed	L	L	N	L	U	N
c	lips pucker, jaw closed	U	U	N	U	L	N
d	corner drop	N	N	U	N	N	U
e	lips stretch, jaw open	L	L	U	L	U	U
f	lips pucker, jaw open	U	U	U	U	L	U

Table 4.5 Table Of Reduced Corner Actions For Final Design

From this table, the drive system at the corner was designed to produce only the two distinct actions of stretch and protrude as it was concluded that the action of corner drop results purely from the effect of the jaw rotation through the skin. This is shown in diagram b) of Figure 4.20.

The specific corner actions were achieved by the combined actions of two drives; horizontal (D_6) and vertical stretch (D_7). The mechanical linkages at the corner were designed to be capable of motion in the vertical axis due to the influence of the jaw action, i.e. the connection at the skin should not restrict the motion of the corner drop action. The linkage was also designed to have the ability to simultaneously

produce stretch or protrusion actions with or with out the corner drop. This was achieved by hinging, or pivoting, the drive linkage of D_6 to be passive to vertical displacement whilst still capable of production of horizontal actions at any vertical axis (refer ahead to Section 5.4 for practical design).

The design can be defined by $\underline{v} = \underline{F_{drive}}(\underline{d})$ which, at the corner is stated as,

$$v_{corner} = F_{corner6}(d_6) + F_{corner7}(d_7) + F_{jaw}(d_{jaw}). \quad \text{Equation [4.31].}$$

where $F_{corner6}$ represents the function of drive D_6 at the corner and v_{corner} is the final displacement in space of the corner point.

As stated earlier in Section 4.2, every degree of freedom at any key point, requires measurement and hence control of its individual displacement. Three signals are, therefore, required at the corner, for the measurement of displacements along the three degrees of freedom and each sensor should only have effect in its respective region and prevent any drive system from trying to act against each other. Using similar design for reflectors as those in the brow region, the motion of the point was transposed to a parallel axis away from the surface of the face. The area of these reflectors should be sufficiently large to ensure that a measurement was taken for all positions along the defined axis, irrespective of the action of the jaw.

The mapping relationships at the corner are defined from the equation $\underline{d} = \underline{F_{control}}(\underline{s_1})$. By use of $\underline{F_{sum}}$ (c.f. Section 4.3), the following equations ensure that the actions of corner stretch and corner protrude are mutually exclusive.

$$d_6 = F_{66}(s_{16}) + F_{67}(s_{17}). \quad \text{Equation [4.32].}$$

$$d_7 = F_{76}(s_{16}) + F_{77}(s_{17}). \quad \text{Equation [4.33].}$$

where F_{66} represents the mapping function between drive D_6 and sensor S_6.

This is shown in Figure 4.21 which displays the mapping relationships. From these functions, it is clear that provided S_6 and S_7 are not in their upper ranges at the same instant then the overall actions produced will be one or the other.

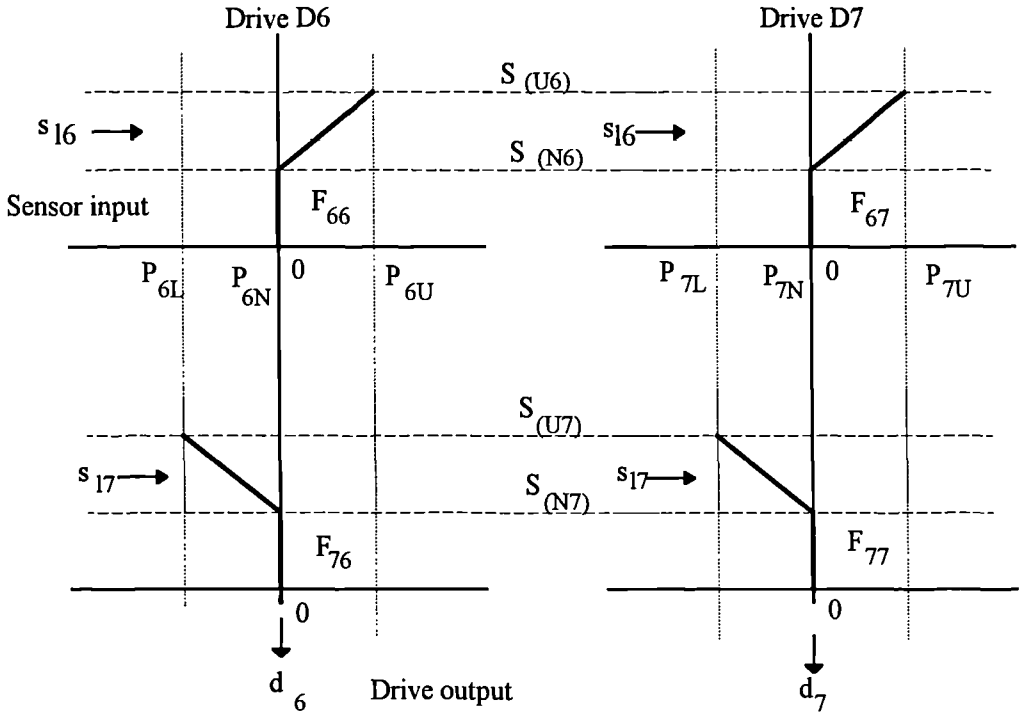


Figure 4.21 Diagram Of The Mapping Functions For Control And Drive At The Corner

4.6 Summary

This chapter has detailed the theoretical principles behind the design of a novel optical technique to automatically sense facial actions, for use as the source of control for animatronic performances. The design was based on the reduction of overall facial changes to an optimum set of displacement measurements derived from a group of key points.

A method of solution was developed to generate data suitable for both physical and perceptual analysis. This was to be achieved by the construction of an animatronic face, identical in size and shape to the researcher, based on the principle of key point displacement. It was designed to produce the same individual displacements in magnitude and trajectory, to an identical set on the live face. These points are shown in diagram a) of Figure 4.22. This should result in a simplification of the mapping correspondence between sensed input and driven output signals. It was proposed that the recombination of the individual displacement at the replica would generate perceptually similar actions to those produced by the live face at the input.

This would allow objective data to be derived from measurements using the sensing system of both input and output displacements. Perceptual data would be generated through the visual evaluation of the output actions. This type of evaluation would result from, firstly, the viewer's experience of visual speech recognition and, secondly, through the visual comparison between the live and replica facial actions.

The final design of the system to realise the animation and sensing of the desired facial actions has been described based on the previously defined key point principle. The design of the replica drive system is shown in diagram c) of Figure 4.22 and the design of the proposed sensor system is shown in diagram b) of Figure 4.22. The overall function of the system, $\underline{F_{total}}$, is shown in the diagnonal representation of Figure 4.23. The system is defined as a 4X4 matrix in the upper face and a 6X9 matrix for the lower face.

Diagram a) Defined Key Points

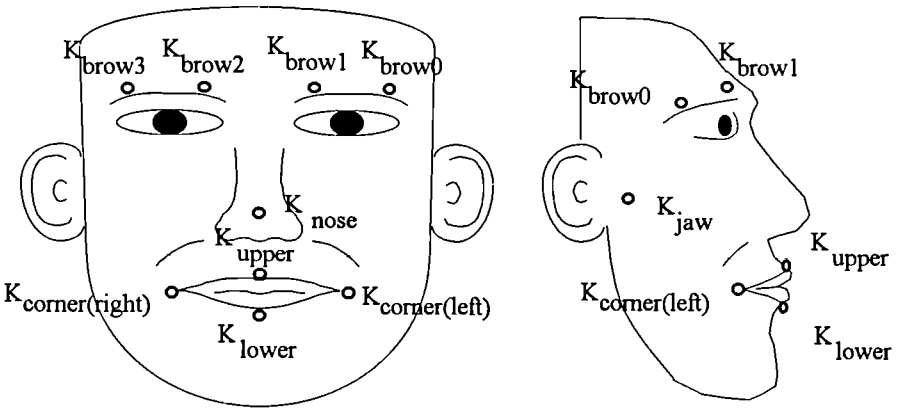


Diagram b) Defined Sensor Positions And Trajectories

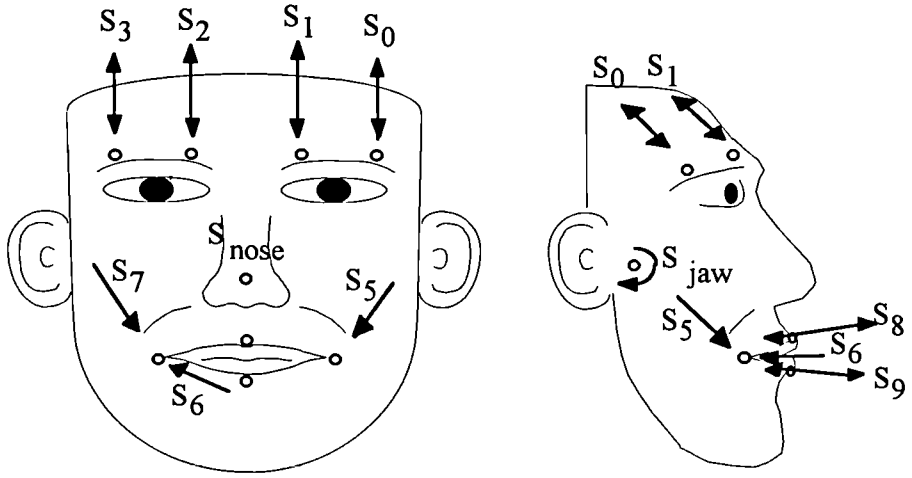


Diagram c) Defined Drive Actions

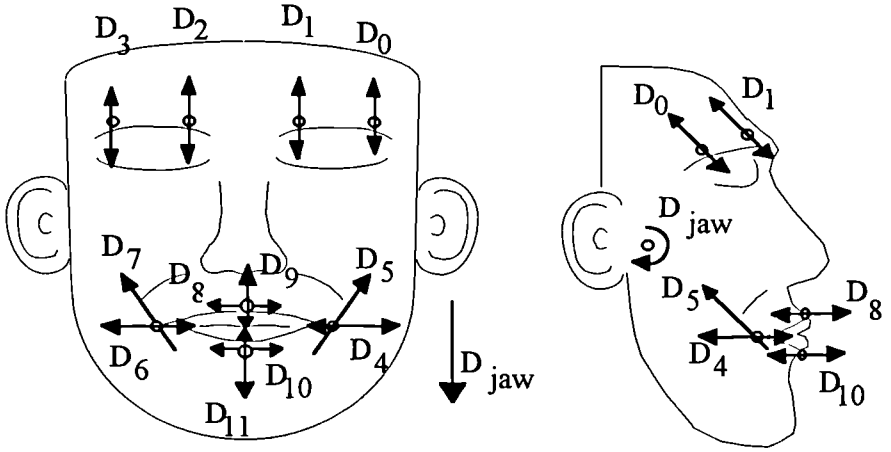


Figure 4.22 Diagrams of Final Design Of Facial Action Sensing And Animation Systems Based On The Key Point Principle

Diagram a) Upper Face

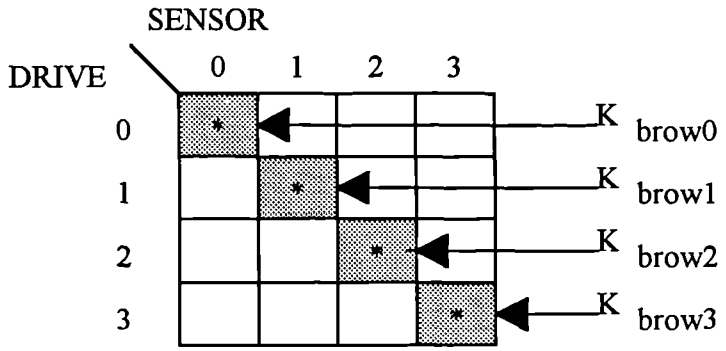


Diagram b) Lower Face

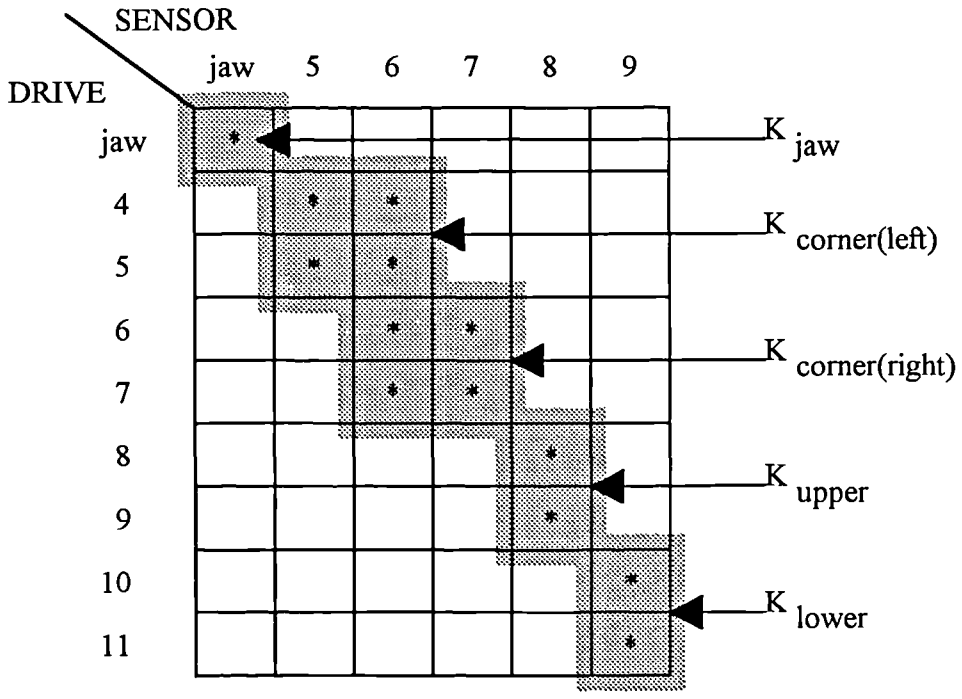


Figure 4.23 Matrix Representation Of The Final System

To realise the overall system, the following engineering tasks have been undertaken:

1) the development, by electronic means, of converting the optical measurements into continuous analogue voltages directly proportional to the physical displacements. These signals should be of the form that can be input to either the HPC System directly, or to the computer for data conditioning;

2) the development of a data acquisition system capable of data recording, handling, and playback at sufficient rates to allow real time control;

3) the development of software procedures to: firstly, control the data acquisition system for record and play; secondly, to control file handling; and thirdly, to apply conditioning techniques to input data;

4) the development of software to allow separate control of the replica's drive system for certain analytical experiments;

5) the physical design and construction of mechanical linkages to a 3-D animatronic replica capable of identical facial changes to those of the live face. This includes the design of mechanical linkages to achieve the final deformations in replica skin.

6) the construction of the system to support sensors on and about the face. This method must: firstly, maintain sensors in desired positions (and thereby retain the correct reference points); and secondly, separate the motion of the key points from the effects of the global head movements.

These represented a significant set of engineering tasks that are discussed in detail in the following chapter.

Chapter 5

Practical Aspects Of Final System

Chapter 5

Practical Aspects Of Final System

5.1 Introduction

The purpose of this chapter is to describe the various practical elements developed to realise the objectives defined in Chapter 4. Section 5.2 describes the overall design of the final system for control data transfer, storage and analysis. The data acquisition system to allow multiple input/output and the software created to control its operation is described. Section 5.3 describes the design of the infra-red sensing system to produce continuous voltage signal based on the relative displacement of a point in space. A comprehensive investigation into the system characteristics is presented. Section 5.4 describes the design principles and physical construction of drive and linkage system to create the animatronic replica face.

5.2 Overall Design for Data Control System

The final design is shown in Figure 5.1. It was constructed to handle data from either sensor, hand, or computer control in the production of the final output animation.

As stated in Section 4.3, the HPC System was used as a "stand-alone" processor for the following tasks:

1) the production and storage of F_{map} for each input signal by user defined control and drive parameters (reference limits);

2) the production of F_{sum} for the individual drive outputs;

3) the definition of overall drive parameters to prevent physical damage ($F_{motorcondition}$); and

4) the production of output pulse width modulation signals for up to 32 individual servo drives. These signals are output as a sequence of pulses that are then demultilexed to provide positional control for each drive.

The HPC system was considered satisfactory for its proposed use within the final system, and therefore no research was undertaken into the evaluation of its performance. The HPC system has 24 channels for analogue input which are converted at a rate of 100 Hz with 8 bit resolution.

The input signal requirements to the HPC System were restricted to analogue voltages in the range of 0-5 volts (dc). The relative sampling frequency for data manipulation was defined by the "Nyquist Frequency" **[Sheridan92].** This frequency is defined as "the sampling of a continuously changing signal at a rate to allow reproduction of the highest frequency component present, the bandwidth, thereby ensuring that no information is missed." Where the bandwidth of the input signal is ω then the required frequency of sampling must be at least 2ω. This sampling frequency must be sufficiently large to allow the stored digital representations to be defined as continuous signals. Within this research, the bandwidth of the input signal was defined by consideration of its use within film and television. Television restricts the number of visual frames to 50 per second (UK), and in film this is reduced to 24 frames. Consequently, the sampling rate was required to be at least 50Hz, preferably 100Hz or greater.

The other requirements for the signals produced within the system are the reduction of the effects of noise, the consistency and accuracy of signals based on the type of input, and a suitable interface with the existing HPC system.

5.2.1 Development Of A Data Acquisition System

From the proposed methods of solution and analysis, it was important to develop a method of handling the control data produced within the system. The requirements for the system were as follows;

1. the acquisition of input control signals (d.c. analogue voltages) at the defined sampling rates;
2. the storage of the recorded signals for later analysis;
3. the conditioning of the input signals by any one of the proposed methods (c.f. Section 4.3); and
4. the output of stored files containing control signals at same sampling rate;

The overall system should also have ability to produce all of the above in real time, and for multiple input and output channels.

The final system is shown in Figure 5.1. It allowed control of the HPC system in a number of ways; through direct hand control, through direct sensing control and also via the computer. The system allowed data to be recorded whilst direct control was applied to the HPC system or it recorded the control signals for off-line conditioning via the processor. It permitted recorded files to be played back at a later stage for repeated testing.

In order to have easy interface with the HPC System, the data acquisition system was designed to provide 16 channels of analogue input and 16 channels of analogue output. This dual function board was designed and constructed by a member of staff, Dr. Booth, to be compatible with IBM computers.

The acquisition was achieved through two 8 channel analogue to digital converters that produced parallel inputs to the processor of 8 bit resolution with a conversion time of 2.5 μseconds. The input range was 0 to 5 volts d.c, which converted to 0-255. The analogue output was achieved through four quad digital to analogue converters producing 0 to 5 volt analogue outputs from 0 to 255 digital values of 8 bit resolution. The analogue output level operated in a write mode and responded

purely to the activity of the digital inputs, remaining at a constant level until the digital value was altered.

Within the processor, the input and output data values were mapped directly to defined addresses in the memory I/O block. The rate of read and write of these memory addresses, and hence the sampling rate for both conversions was produced purely by software control.

To allow for the data acquisition from different sources, an interface board was developed to remove the need to constantly change the actual connectors. The design of this board is described in Appendix E. Practical problems occurred on the analogue output signals as a result of noise generated by the power supply in the computer. These were reduced, through the use of low pass filters on the individual channels, to a level where the output signal took the desired value ±1. This was deduced by feeding the output back via the A/D converter to the computer, and assessing the difference between input and output. The results are not presented but are available on request.

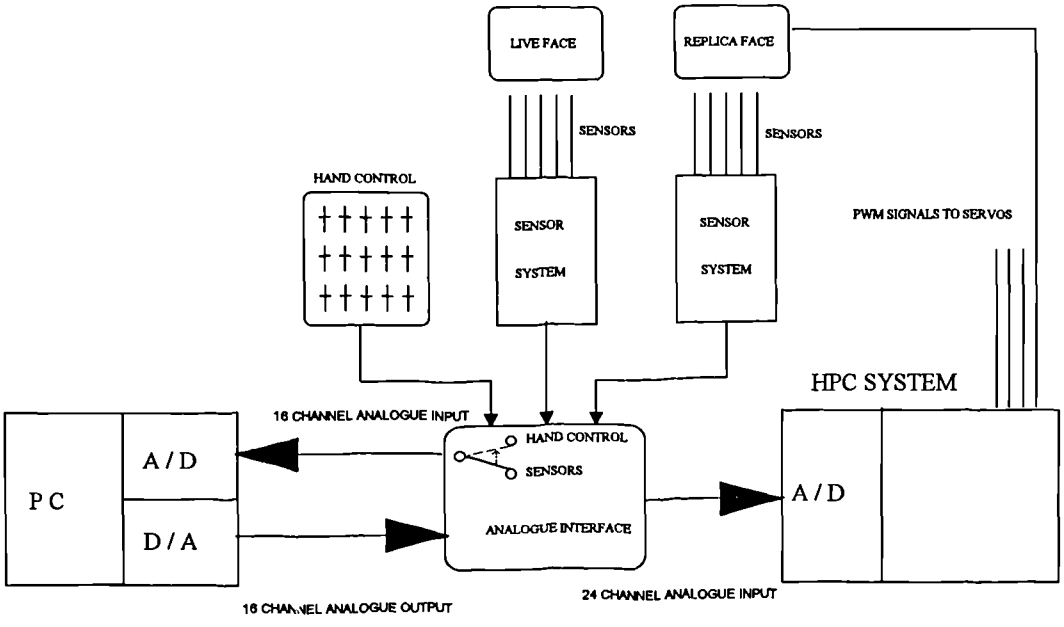


Figure 5.1 Diagram Of The Overall Practical System

5.2.2 Software To Control Data Acquisition System

To achieve these different tasks, procedures were developed to produce the necessary recording, sampling, file handling, conditioning, and playback. These procedures were written in "Borland C". The flow diagram of the program PLAYRECORD is shown in Figure 5.2. This represents an example of the typical procedures undertaken. Key functions of the program are listed in Appendix D.

Data sampling was achieved through the application of *dos* clock present in the processor. The principle is shown in the Flow Diagram of Figure 5.2 and in the listing of Appendix D.5. The period of record or play procedures were sufficiently fast to be considered incidental in overall periods. By defining the frequency of 100Hz, the delay was therefore set at 1/100 = 10msec.

The technique of read then wait produced sampling at sufficient rates. Using a stopwatch function, available in Borland C, with an accuracy of 1μsec, the period of record was given for each operation. From the returned time it was evaluated that the use of the present procedure was accurate over repeated tests to within ±0.05sec. It was concluded that this timing was acceptable for the present system, but that the possible errors resulting from the time variations should be taken into account in subsequent analysis.

Given that proposed techniques for the analysis of data were likely to be actual product packages, either DaDisp, Excel, or Minitab, the most straight forward format for data handling was to use ASCII text files. The limitation of the format was expensive memory requirements, but, given that the average length of recording was 5 seconds at 100Hz, the data storage requirement was not excessive. The advantage of this format was its portability between packages and within C programming. Each input and output file was stored as 16 columns. Each column represented an individual control signal, and each row represented the set of samples taken at any one instant.

Listings of the conditioning functions are shown in Appendices D.2, D.3, and D.4. Linear 2 part conditioning applied the equations derived in Section 4.3.3. The input parameters were stored as a single column in a text file. The output parameters were

defined as 0,128, and 255 values for the respective lower, neutral and upper limits. Linear 1 part conditioning applied the equations derived in Section 4.3.3.

To achieve non linear conditioning, a text file was produced to store the individual variations, recorded when each key point was driven through every position (c.f. Sections 4.3.3.2 and 6.2). These variations represent the overall function of each point. The look up table was produced as an array of 16 columns by 255 values by the inversion of these individual functions. For final conditioning, the input value in the play file points to a specific value in table and that value represents the required output control value. Appendix D.4 shows the actual code used to produce the look up table.

5.2.3 Other Software Produced For Research

The following programs were developed for various applications in the final system:

1. PLAY — plays out stored text files with user defined conditioning at 10,50,100, and 200Hz.;

2. RECORD — records control data at the same frequencies, and stores data as a text file;

3. PLAYRECORD — simultaneous play back of control data, with or without conditioning, to replica and record measured changes by the sensor system;

4. LIVE — records control signals, applies desired conditioning and outputs signals to replica in real time at 50Hz;

5. KEYREC — records values at A/D input when keyboard pressed.

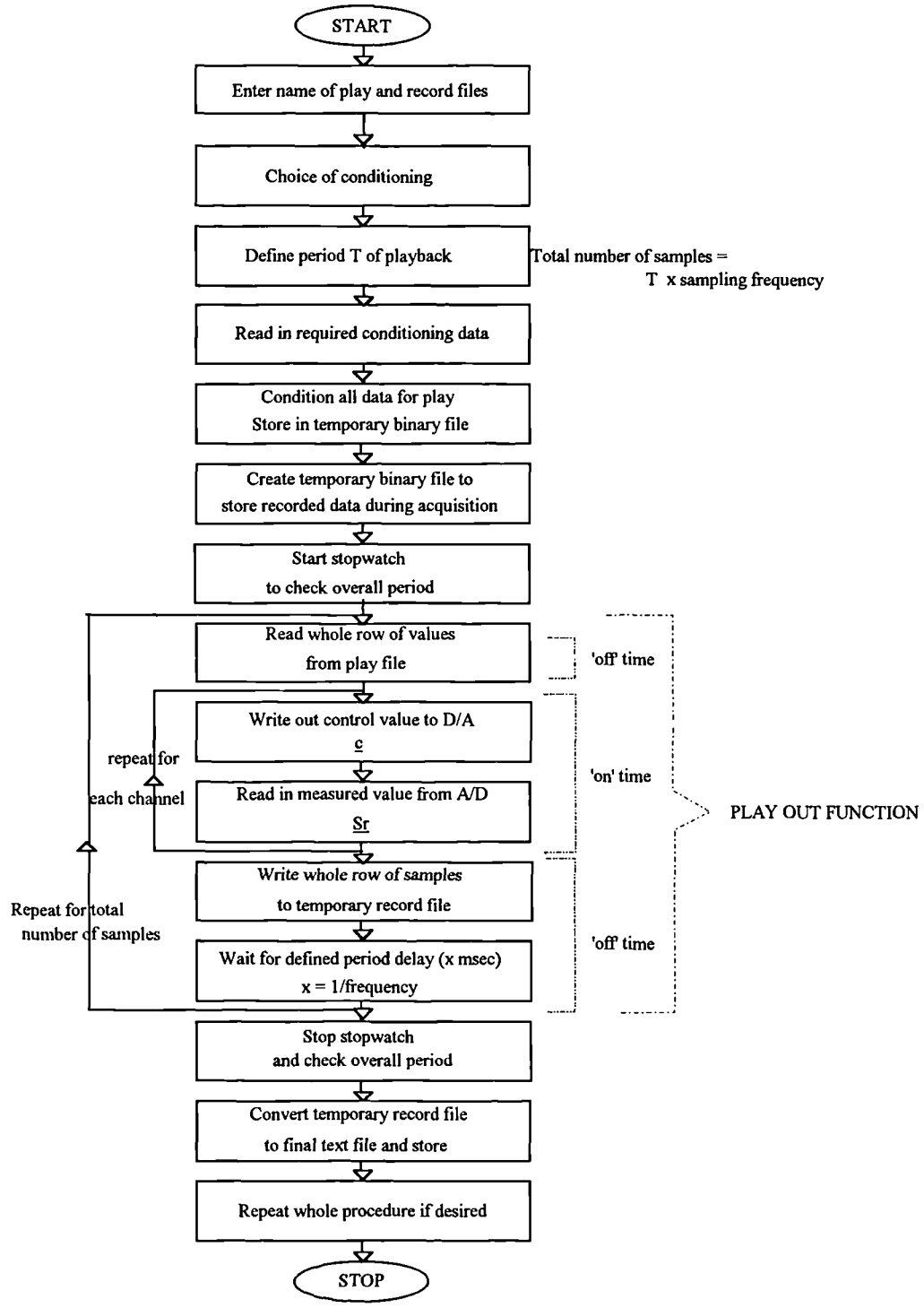


Figure 5.2 Flow Diagram Of Software Procedures To Play And Record Control Data

5.3 Design And Analysis Of Infra-Red Sensing System

5.3.1 Construction Of The Infra-Red Sensing System

From the theory of Section 4.2. It was determined that infra-red proximity sensors, **[RSData83]**, were capable of producing a photocurrent I_c proportional to the changing distance between the sensor and reflector where the sensitivity of the final output signal was defined by the optical characteristics of the detector. The sensor was composed of an infra-red emitter housed in the same package as a spectrally matched phototransisitor.

The following circuitry was designed by Hensons to produce an analogue dc voltage output signal from the detected photocurrent. The principle of the design is shown in the block diagram of Figure 5.3. Each board was constructed for five sensors.

The main problem that exists with optical sensors is interference from background lighting and from stray rays of other sensors **[Todd85]**. An infra-red filter in the sensor package reduces the effects of ambient illumination. An improved method of overcoming this variation was the sequential pulsing of the emitter/detector pairs. The pulsing of the emitter allowed a much higher peak power than if powered continuously. The pulsing of the detector enabled detection of only the reflected signals produced by its own emitter during an "on" time of the pulse. The fluctuations caused by other illuminations were eliminated during an "off" time by a.c. coupling of this signal. The overall period was 1msec with individual "on" times of 0.2msec. Consequently the sensor produced a pulsed signal (V_{in}) of varying depth, directly proportional to I_c and, hence, the amount of reflected power.

The requirement for the sensing system was the measurement of the displacement of the key point from a relative neutral position. It was neither desirable nor practical to have a measurement of the actual distance between sensor and reflector as this would require identical positioning for every recording.

The practical technique to achieve this was the application of an automatic gain control (AGC) circuit with a "sample and hold" feedback control. The voltage V_{reset} was reset to half the output level whenever the switch was manually operated. This

resulted in the output signal at the reset being reduced to a level proportional to the input signal and V_{reset}.

When the reset switch was released, the level at V_{reset} retained half the output level. All subsequent changes in the input signal produced output levels relative to this voltage and, hence, to the output at the datum.

If the system was reset at a distance closer to the sensor, the reset level of the output was increased due to the higher component of V_{in} in the final signal. Similarly, when reset at a distance away from the sensor, the output level was dependent primarily on the reference level, as the input signal was significantly reduced due to low reflected power.

The rectifier block was used to restore the full d.c. level to the output signal during the "off" state of the signal, i.e. it produced a constant level signal proportional to the amplitude of the pulse output of the AGC. The buffer amplifier converted this signal into an amplified d.c. level and the signal limiter, in form of a zener diode restricted the variation of the final output signal to a range of zero to V_z (zener voltage).

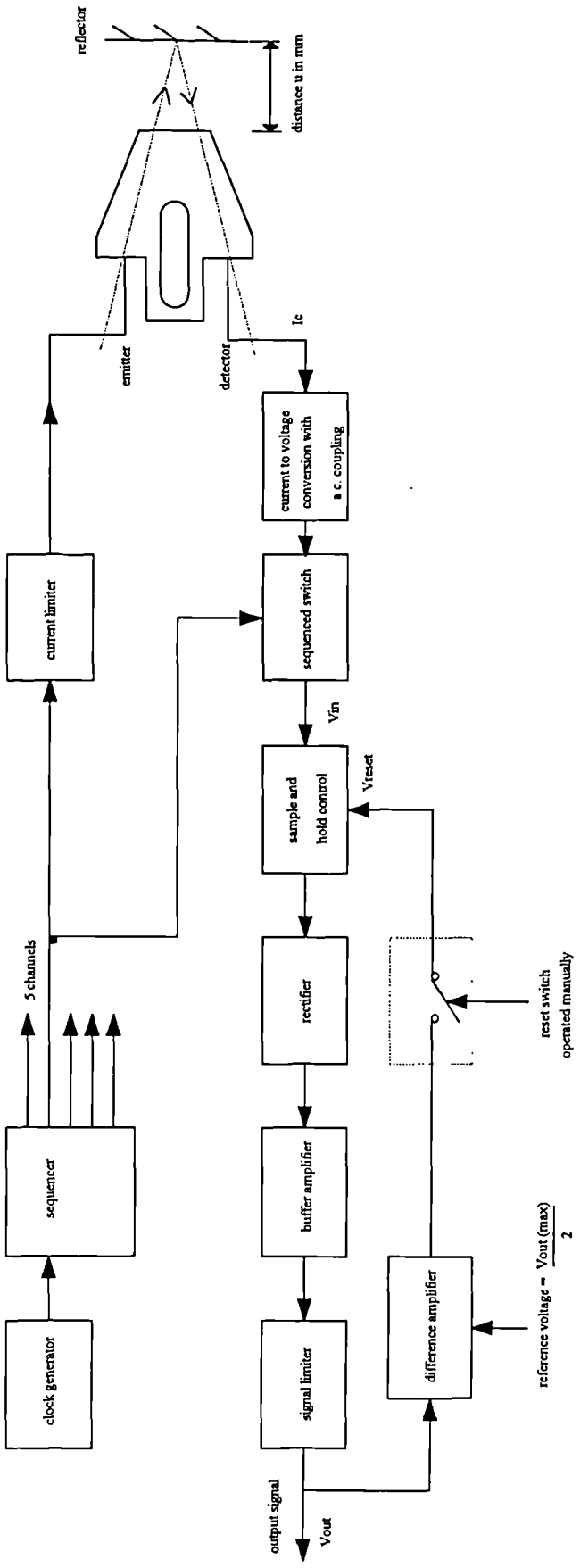


Figure 5.3 Block Diagram Of The Infra-Red Sensing System

5.3.2 Analysis Of Sensing System Characteristics

Before the final sensor system could be applied in facial action sensing, it was important to examine its characteristics in ideal conditions. The optical theory discussed in Section 4.2 stated that the photocurrent, I_c, at the receiver is dependent upon the two main factors; the area of the reflector, A, present in the sensor field of view and the linear displacement, u_z, of a reflective point in space along a single focal axis relative to the sensor.

Ideally, the final sensed signal, S, should be proportional purely to the displacement, u_z. In reality, S is proportional to a number of variables, as shown in Figure 5.4. These variables were defined as the following;

1. u_x and u_y are the displacements of the reflector along their respective axes perpendicular to the focal axis;
2. r is defined as the distance between sensor and reflector at which the system is reset. It represents the datum point for subsequent measurement;
3. A defines the area of the reflector;
4. δ is defined as the angular displacement between the planes of the sensor and reflector. Variation in δ will affect the relative area of the reflector in the sensing range;
5. α is defined as the angular displacement at which the system is reset; and
6. M represents the different types of reflective material.

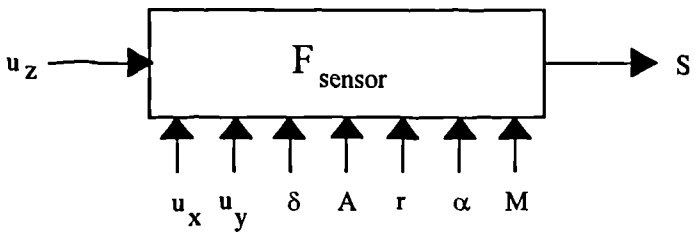


Figure 5.4 Function Diagram Of Sensing System

As stated in Section 4.2, the sensor system can be defined by the following equation

$S = F_{sensor}(u)$. In ideal conditions, it can be reduced to $S = k \cdot u + S_n$ where k is the linear sensitivity of the system and S_n is the sensed value at the reset. In the practical system it was concluded that $S = F_{sensor}(u_z, u_x, u_y, A, \delta, r, \alpha, M)$.

Therefore it was important to investigate the effects of each variable in isolation to establish the overall characteristic of the sensing system with reference to its use in sensing facial point changes.

5.3.3 Experimental Procedure And Apparatus

The following apparatus were constructed to analyse each of the variables in ideal conditions. The apparatus shown in Figure 5.5 was used for experiments investigating linear changes in displacement; u_z, u_x and u_y. It was constructed from a microscope stand with mm scales along each of three Cartesian axes. Manual increments were then made in the reflector's position relative to the fixed sensor. The design held the plane of the reflector perpendicular to the plane of the sensor, ensuring that δ and α had no effect.

The experimental apparatus of Figure 5.6 was used with the scales marked out in degrees for investigations into possible angular variations; δ and α. The design held the reflector at a constant distance from the sensor along the focal axis, ensuring that r, u_z, u_x and u_y remained constant.

For both set-ups, the sensed value, S, was recorded after each increment using the program KEYREC (where researcher pushed key return to record value).

In the following text, reflector positions within the linear apparatus were defined as co-ordinate positions (x,y,z). The origin represented the point directly in front of the sensor. The focal axis was defined by both x and y =0. In the angular apparatus, the angles δ and α represent the difference between the orientation of sensor and reflector with the ideal position being defined as 90°.

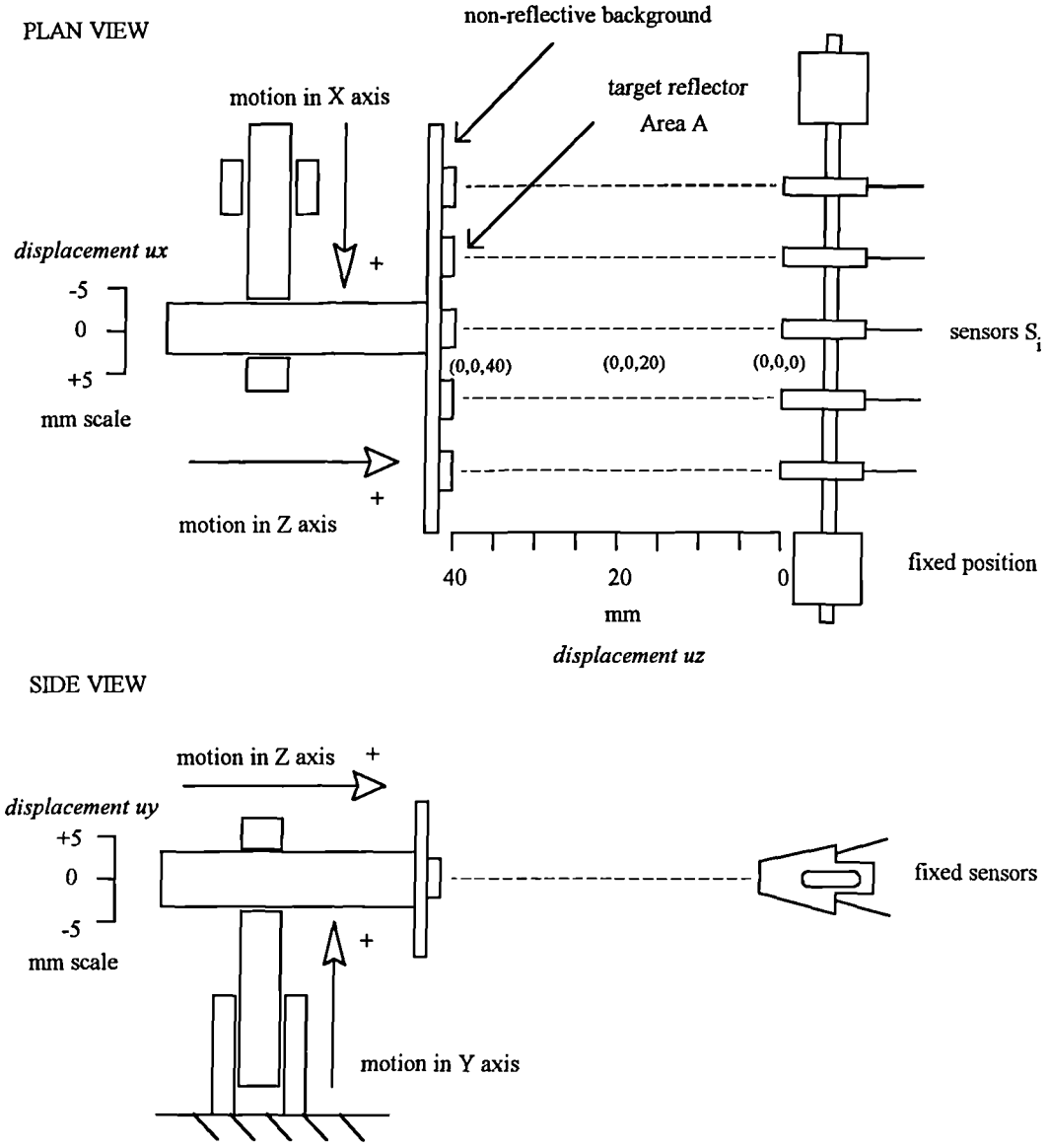


Figure 5.5 Apparatus To Examine Effects Of Distance On Sensed Signal

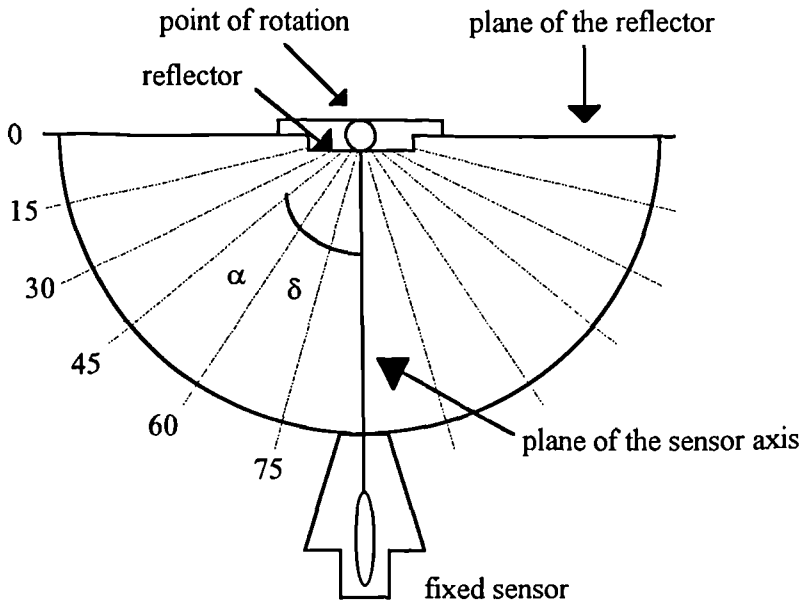


Figure 5.6 Apparatus To Examine Effects Of Angular Displacement On The Sensed Signal

5.3.3.1 Principles Of Error Analysis For Objective Examinations

With reference to the principles of error analysis discussed in Appendix D, the error theory was adapted to produce the following analysis measurements. Tests were repeated i times and for each test, n readings or samples were taken.

For each set of readings, the mean was determined where

$$x_{mean}(n) = \frac{\sum_i x_i(n)}{n} \qquad \text{Equation [5.1].}$$

The standard error, $\sigma_m(i)$ for each test was determined by

$$\sigma_m(i) = \frac{1}{n}\Sigma_n\left(x_i(n) - x_{mean}(n)\right)^2. \qquad \text{Equation [5.2].}$$

This value was an indication of the precision of the test *i* in comparison with the other tests. Therefore the overall tolerance measurement for repeated tests was defined by

$$\pm\sigma_{mT} = \sqrt{\frac{1}{i}\sum_i \left(\sigma_m(i)\right)^2}. \qquad \text{Equation [5.3].}$$

5.3.3.2 Investigation Of System Consistency For Constant Position

Prior to the specific examinations of the individual variables, it was important to investigate the accuracy of the system in the production of consistent signals proportional to fixed displacements. The reflector was positioned at set distances from the sensor, r, the system was reset and multiple readings were taken at 50 Hz over a period of 10 seconds. The results are shown in Table 5.1.

The tests were repeated n times, and the mean value and standard error for each test were generated. The mean of the mean values and the respective standard error were determined. The derived value for $\pm 2\sigma_{mT}$ was defined as the maximum permissible error, 95.5 % certainty of the true value lying within that range of values. These results indicate the ability of the system to maintain a consistent level for the output signal within the defined limits and represent a measure of the accuracy of the system. From these results, it was defined that the true value for all subsequent recordings, at all distances, was $S_{true} = S_{mean} \pm 3$ *values* or $\pm 1.18\%$. This represented the accuracy of the sensor system.

datum at r mm	mean value	$\pm\sigma_m$	$\pm 2\sigma_m$	$\pm 3.29\sigma_m$	range of permissible values (rounded) (95.5%)
5	132	0.24	0.49	0.80	1
10	127	0.17	0.35	0.57	1
15	122	0.55	1.11	1.82	2
20	115	0.85	1.70	2.80	2
25	112	0.72	1.45	2.38	2
30	110	0.75	1.50	2.47	2
35	106	0.81	1.62	2.67	2
40	102	1.13	2.25	3.71	3

Table 5.1 Table of Error Analysis For Recorded Measurements

5.3.3.3 Investigation Of The System And Experimental Consistency For Changing Input

Having established that the system could produce consistent readings at set distances, it was necessary to determine whether system could produce accurate measurements for changing variables.

Using the linear apparatus of Figure 5.5, a white diffuse reflector of area 10 mm square was positioned at a datum distance, on the focal axis, of r = 25 mm and the system was reset. The reflector was then moved along focal axis from co-ordinates (0,0,40) through to (0,0,0) with a reading taken at every mm change and then repeated *n* times. Figure 5.7 shows a graphical example of three tests.

Using the analysis techniques discussed in Appendix D, σ_m was determined for each individual test, and σ_{mT} was then determined for the whole series of readings. This

was repeated for all of the sensors in the system. It was concluded that, although specific values were different, these overall findings were consistent throughout.

From this analysis, the value of σ_{mT} equalled ± 0.77 (rounded to 1), the value of $2\sigma_{mT}$ equalled ± 1.54 (rounded to 2) and the value of $3.29\,\sigma_{mT}$ equalled ± 2.53 (rounded to 3). Given the desire for accurate measurements, the tolerance limits for each test were fixed at $\sigma_m = \pm 2$ (95.5%). This was maintained through all of the following experiments to ensure precise measurements and to help identify errors should they occur. This value represents the accepted tolerance for the experimental procedure used. If the value of σ_m, for any test, exceeded the tolerance then its results were ignored. The main source of errors resulted from human error in the increment procedure.

5.3.4 Examination Of Linear Displacement On The Sensed Signal

Having established the criterion for accurate and precise measurements, examination was undertaken to define the function of the sensor system in terms of the ideal displacement along the focal axis (c.f. Section 5.3.1 for practical description of reset theory).

The reflector; white, diffuse 10 mm square; was placed at a series of distances, *r*, from the sensor along the focal axis. The system was reset at each *r* thereby defining it as the datum for following measurement. The reflector was moved through (0,0,40) to (0,0,0) with sensed value, *s*, recorded at every increment. The test was repeated a number of times to derive mean variation and ensure signal consistency. The mean variations for the different datum settings are shown in Figure 5.8. A number of points can be made from these plots.

1. The system has the ability to produce output signals proportional to the changes in position of a reflector, relative to the sensor, within a significant field of view, approximately 40 mm.
2. When the datum is reset at a distance less than 15 mm, the range of measurement is reduced.

4. When datum is reset in excess of 15 mm, the signals produced show clearly defined curves based on reflector displacement.
5. The upper cut-off limit is the result of the clamping zener diode in the conversion circuitry, refer to Section 5.2.1.
6. The distinct lower limit is a result of the reflector moving away from the sensor, relative to it's datum, causing a reduction in the photocurrent to a level where the radiant power is equal to the level produced by ambient illumination.
7. The plots indicate that the sensitivity of the curves, $\partial s/\partial u_z$, and hence the characteristics of the overall function varies with the reset distance.

Figure 5.9 shows the normalised plots for each curve relative to its datum. The differences are a result of the reduction in the radiant power, and hence I_C, as the distance between reflector and sensor increases, as shown in Figure 4.1. This reduced output signal results in changes to the measurement value at the datum and subsequently alters all measurements relative to it. The sensitivity varies as it is proportional to this reduced reflected power.

Each of the resultant curves suggest that they have linear characteristics. Using the least squares method, straight line approximations were developed for each curve, as shown in Figure 5.10. Table 5.2 displays the results from these approximations. The sensitivity of each curve is simplified to the gradient of the straight line approximation. In the equation $y = k.x + b$, k is defined as the sensitivity and b is the output value at the datum.

The least squares method is based on the production of a line approximation where the error measurement, given in Table 5.2, is defined as maximum residual error difference, positive and negative, between straight line and actual curve.

In Table 5.2, the measured range is approximated from Figure 5.10 as the region between the upper and lower limits.

A number of conclusions were drawn from this analysis. At reset distances closer to the sensor, the measurements have a higher sensitivity but more significant non-

linearities at the 'tails' of the curve which reduce the overall range of linear measurement. At reset distances further from the sensor, the sensitivity of the system decreases but the function tends towards a straight line approximation resulting in changes to the characteristic of $\underline{F_{sensor}}$. The sensor produces minimal measurement of displacement away from the sensor but increases range towards the sensor.

The consequences of this are firstly, the sensor must be held in approximately the same position relative to the face and must be positioned in the same place every time for the characteristic of $\underline{F_{sensor}}$ to remain constant.

Secondly, the system has the ability to act as two distinct types of sensor:

a. high sensitivity sensor, capable of equal measurements of reduced range either side of the datum position; and

b. low sensitivity sensor, capable of increased range of measurement towards the sensor from the datum position, of higher linearity.

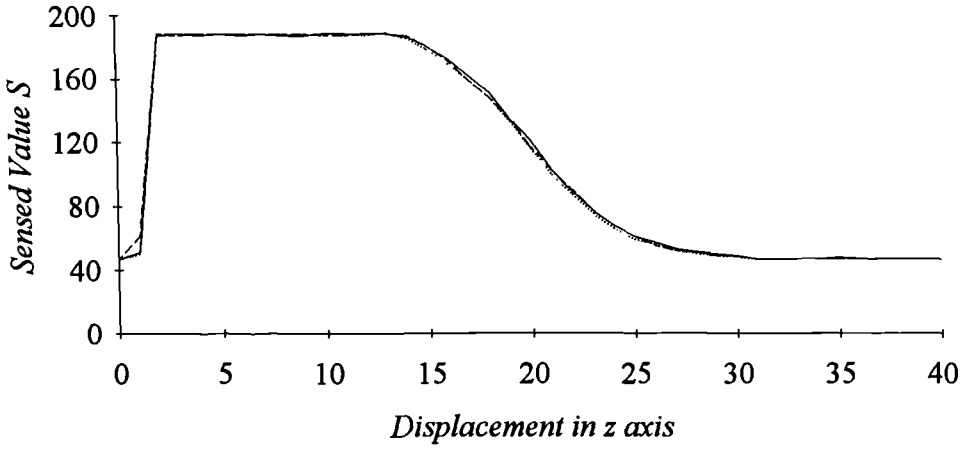


Figure 5.7 Graphical Example of Consistency of System Measurements

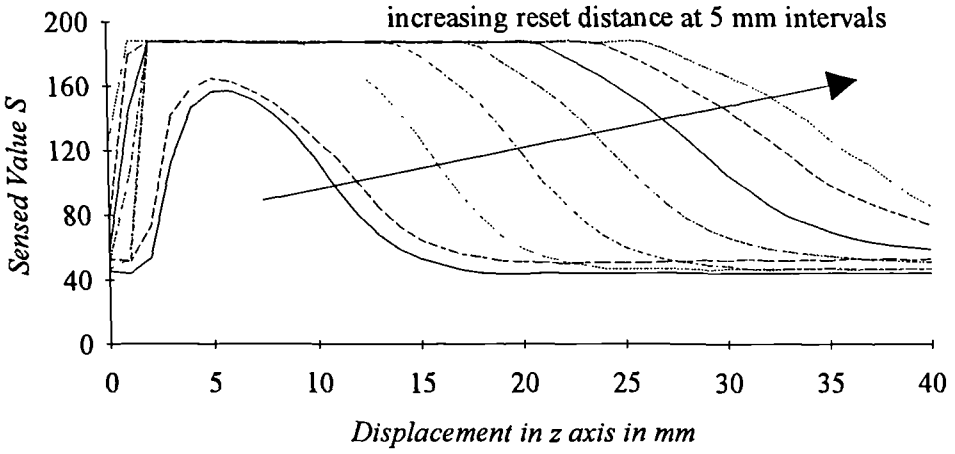


Figure 5.8 Plots Of Signal Variations For Linear Displacements At Different Datum Distances

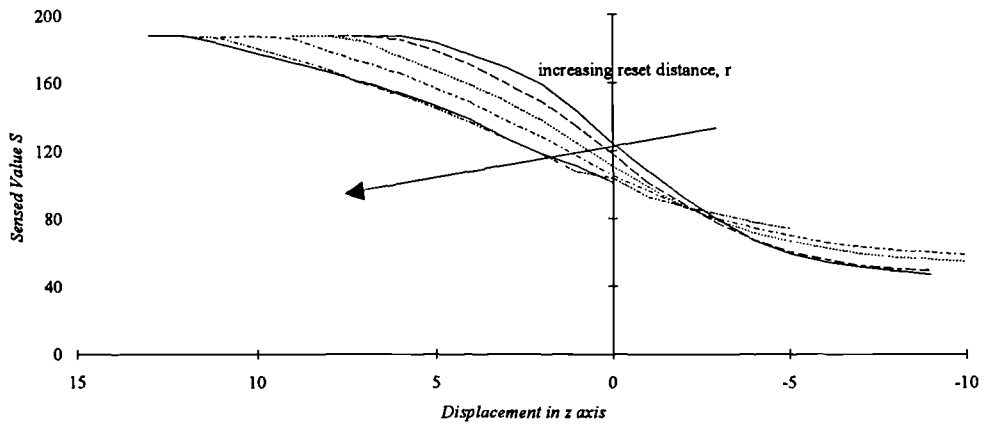
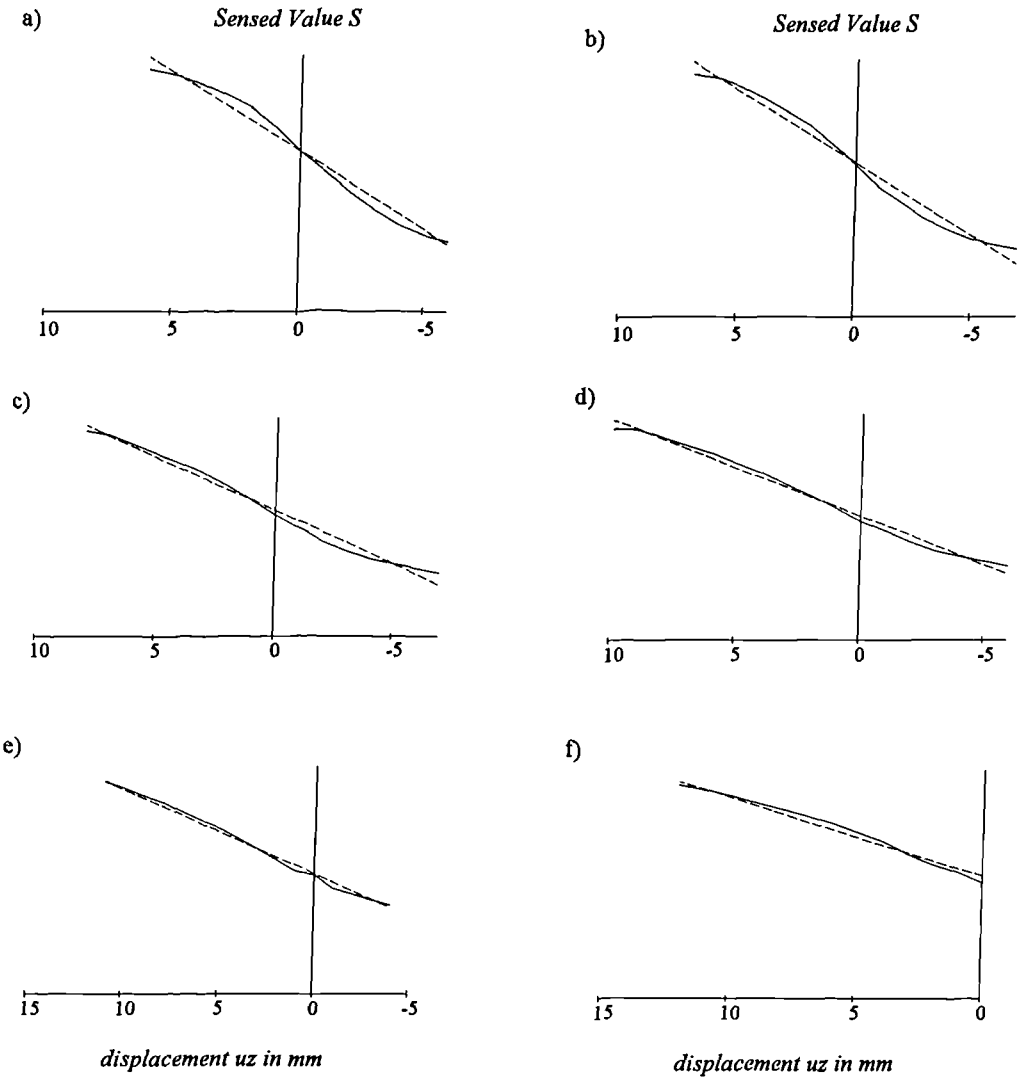


Figure 5.9 Normalised Plot of Signal Variations For Different Datum Positions

				measured range	
Reset distance, r, in mm	Sensed Value, s, At Datum	Sensitivity (values/mm)	maximum residual error (values)	positive	negative
15	124	12.19	7.81	6	6
20	119	11.23	7.21	7	7
25	115	9.67	6.02	8	7
30	110	8.54	4.53	10	6
35	105	7.41	3.57	11	4
40	102	6.91	4.68	12	0

Table 5.2 Table Of Analysis Of Straight Line Approximations



a)reset at 15 mm b) reset at 20 mm

c) reset at 25 mm d) reset at 30 mm

e) reset at 35 mm f) reset at 40 mm

zero on plots indicates datum position.

Figure 5.10 Plots Of Actual Variations Against Least Squares Approximation For Different Datum Distances

5.3.5 Examination Of The Effects Of Different Reflective Areas On The Sensed Signal

These experiments were designed to evaluate the following effects of different reflective areas: Test 1 examined the sensed signal for variations along the focal (z) axis. Figure 5.11 displays the resultant signals for white diffuse reflectors with system reset at r = (0,0,25). Test 2 examined the resultant signal for displacements perpendicular to focal axis in both x and y directions and the resultant plots are shown in Figure 5.12 for white diffuse reflectors. The system was reset at r = (0,0,20) and held constant in z axis.

Test 3 considered the effects of varying the angle between the sensor and the reflector and test 4 examined the resultant signal for different angles of reset. Figure 5.13 displays the resultant plots for both of these variations in angular displacement for a white diffuse reflector of area 15mm square.

The plots in Figures 5.11, 5.12 and 5.13 confirmed a number of points on the characteristics of F_{sensor}:

1. the area of the reflector must exceed 10mm square to ensure measurement over a maximum range;

2. motion across the focal axis of the sensor will produce variations in the output signal and as a consequence result in possible errors of the sensing system;

3 a region exists within the field of view where the signal remains constant, which is directly proportional to the area of the reflector;

4 the x and y fields of view and hence orientation of the sensor are similar (this is shown by the fact that plots Ux and Uy are similar);

5 from results not shown, it was confirmed that similar constant regions occurred at r = 15, 20 and 30mm.

Consequently, the actual measurement of a reflector towards the sensor, but displaced from the focal axis will remain identical, thereby allowing flexibility in

positioning. This is dependent upon the perpendicular displacement being less than half the area of the reflector.

The plots a) and b) of Figure 5.13 confirmed that F_{sensor} is affected by significant variations in the angle between sensor and reflector. This results from change in the relative area of the reflector present within the field of view of the sensor.

The results suggest that a region exists (±15°) about the focal axis where the radiant power, and hence, output signal is constant. The actual size of the constant region is defined by the physical angle of the emitter and detector in the sensor relative to the focal axis.

This was confirmed by Test 4 where the system was reset at a different angle and the reflector was rotated about the axis. When reset occurs at $\alpha<\pm15°$, the overall measurement was considered identical. This is shown in plot b) of Figure 5.13.

This consequently allows a certain flexibility in the angular relationship between sensor and reflector for positioning in the final system, i.e. provided the rotational change does not exceed ±15°.

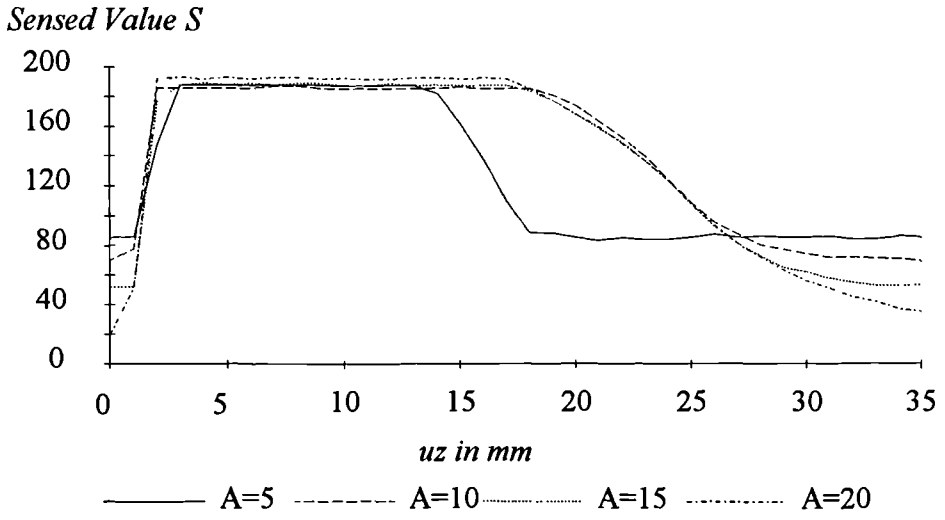
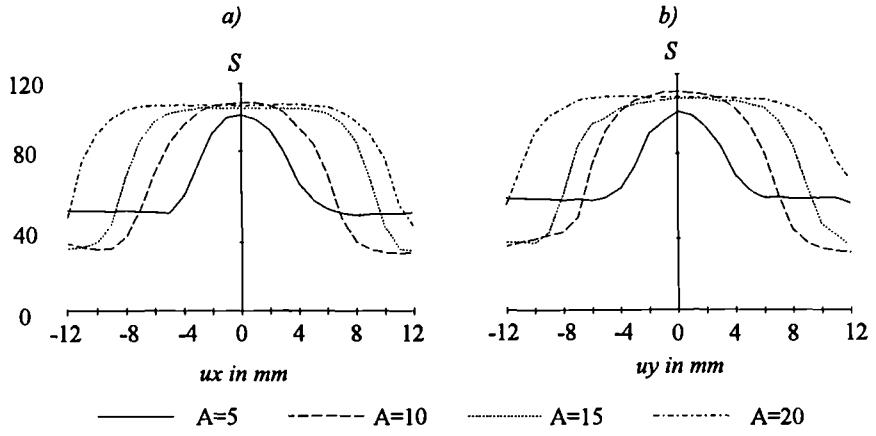


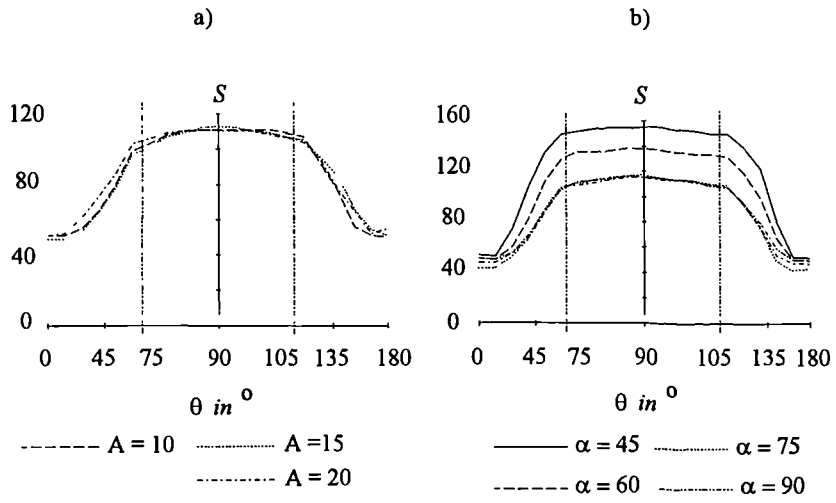
Figure 5.11 Plot Of Linear Displacement In Focal Axis For Different Reflective Areas



Plot a) Variations in u_x for different areas A

Plot b) Variations in u_y for different areas A

Figure 5.12 Plots Of Linear Displacements Perpendicular To Focal Axis For Different Reflective Areas



Plot a) Variation in θ for different Areas A (α =90°)

Plot b) Variation in θ for different reset angles α (A=15 mm)

Figure 5.13 Plots Of Angular Displacements About The Focal Axis For Different Areas And Different Reset Angles

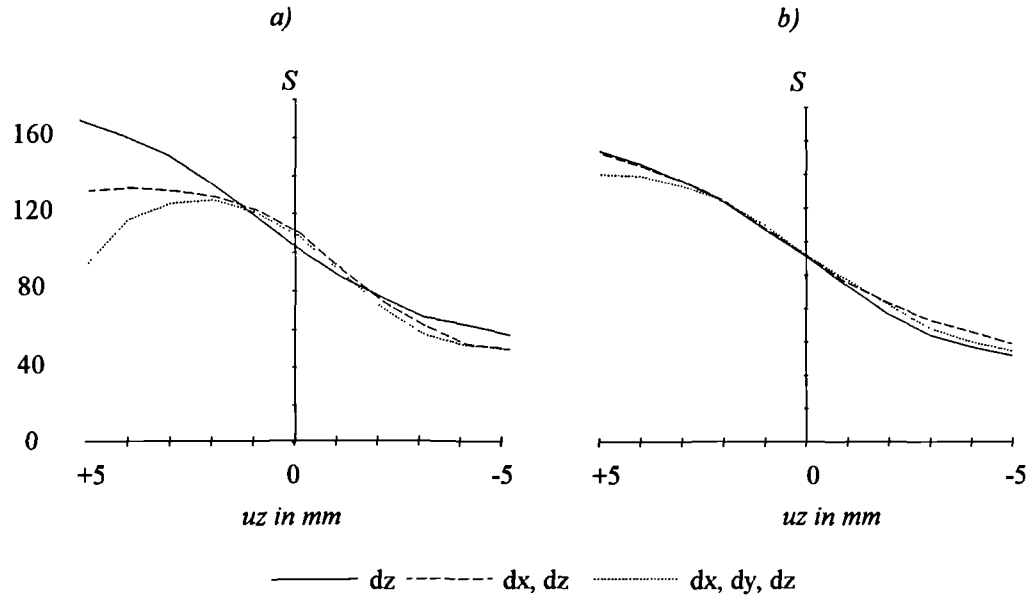
5.3.6 Examination Of Reflector Motion Across The Sensor Axis

This experiment was designed to evaluate the possible effects of reflector motion in directions other than purely the focal axis. Two white reflectors, of two different areas, were displaced along paths that crossed the z axis. The system was always reset with the reflector at (0,0,20). Plots shown in Figure 5.14 display resultant signals for the tests compared with purely z axis displacement.

The resultant plots suggest that the transfer function, F_{sensor}, is invariant to motion across the defined focal axis provided that the perpendicular displacement, in x or y axis, does not exceed the boundary defined by the reflective area. The cross-axis motion of the tests was sufficient to move the reflector area of 10mm diameter, out of sensing range. This resulted in the variations of the measured signal, as shown in Plot a) of Figure 5.14, which would produce significant errors for the control system. To prevent these errors from occurring, the area of the reflector must be of sufficient size to produce the same correct output signal, as shown in Plot b) of Figure 5.14. As a result of this invariance to non-parallel motion, a certain amount of flexibility exists in initial positioning of sensor and reflector for each key point.

5.3.7 Examination Of Different Reflective Materials On The Sensed Signal

In order to establish the effects of different reflective materials on the sensing system characteristics, two sets of experiments were undertaken. The first experiment evaluated linear displacements in purely the z axis for different reflective materials; white paper (highly diffuse), retro-reflective, black photographic paper and a section of latex rubber. The area was set at 10 mm square and the system was reset at 25 mm before moving through (0,0,30) to (0,0,0). Test 2 evaluated the angular displacements for white diffuse and retro reflective areas of 10 mm square with the system reset at $\alpha = 90°$. The plots shown in Figure 5.15 display the resultant sensed signals for both tests. From these results, it was established that the preferred material was white diffuse paper which was capable of reflection over significant range in U_z and invariant to angular displacement over the range of $\pm 15°$.



Plots a) variation in u for Area 10mm

Plots b) variation in u for Area 15mm

Figure 5.14 Plots Of Sensor System Transfer Function Resulting From Non-Parallel Motion

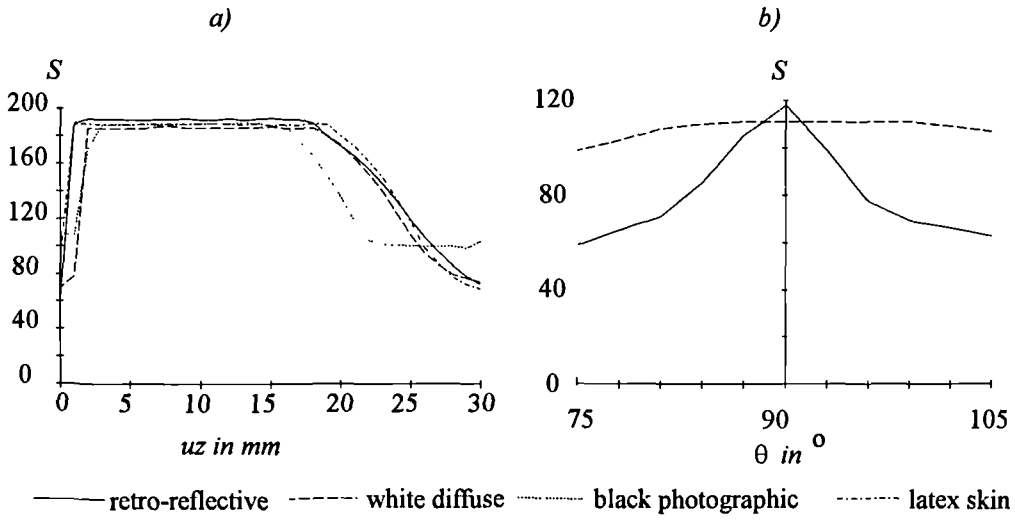


Figure 5.15 Plots Of Signal Variations Resulting From Different Reflective Materials

5.3.8 Summary Of Sensing System Characteristics

In ideal conditions with experimental control over each of the variables, the examinations detailed in the previous sections have drawn a number of conclusions.

The system can consistently recognise reflector displacements, U_z, along its focal axis which can be considered as a linear function.

If the reflector remains at a fixed position up to 40mm from the sensor, the sensor system produces a constant signal, within an accepted tolerance of ±3 values.

The characteristics of the function F_{sensor} is dependent on the actual distance between sensor and reflector when the system is reset. The results of Section 5.3.2 suggest that a range of datum positions exist, 20 mm $\leq r \leq$ 30 mm, where the overall characteristics were sufficiently similar, given the tolerance range of the measurements. Consequently, the same displacement about the different datum would still produce the same output signals. This conclusion allows for the unavoidable variations likely to occur in the repeated positioning of the sensor and reflector on the face. This is shown in diagram 1) of Figure 5.16.

A possible reason for errors in the final measured signal would result from the setting of the datum at an incorrect distance, thereby altering the relative displacement. This is shown in diagram 2) of Figure 5.16. When reset at the desired neutral position a), and then displaced by u_a, the ideal output S_a is produced. If the system was reset with the reflector at position b) the same displacement u_a would produce a different output S_b. Consequently, steps should be taken to ensure that all points are reset at the desired neutral expression on both live and replica faces.

The area of the reflector should be at least 10 mm square to ensure that sufficient reflected power is received at the sensor. The area should also be of sufficient size to minimise any fluctuations resulting from motion across the focal axis.

The system is invariant to angular changes between the plane of the sensor and reflector of less than ± 15° about the focal axis.

The preferred type of reflector is highly diffuse white card or paper.

The system is invariant to motion across the focal axis provided the cross-axis displacement is less than half the reflector's area.

The choice of reflector area, reset distance and orientation must remain constant in the final system.

In conclusion, this investigation proved that the sensor system was capable of the production of output values proportional to the displacement of a reflector about some definable datum position. This was considered as a linear relationship provided that the other factors were restrained.

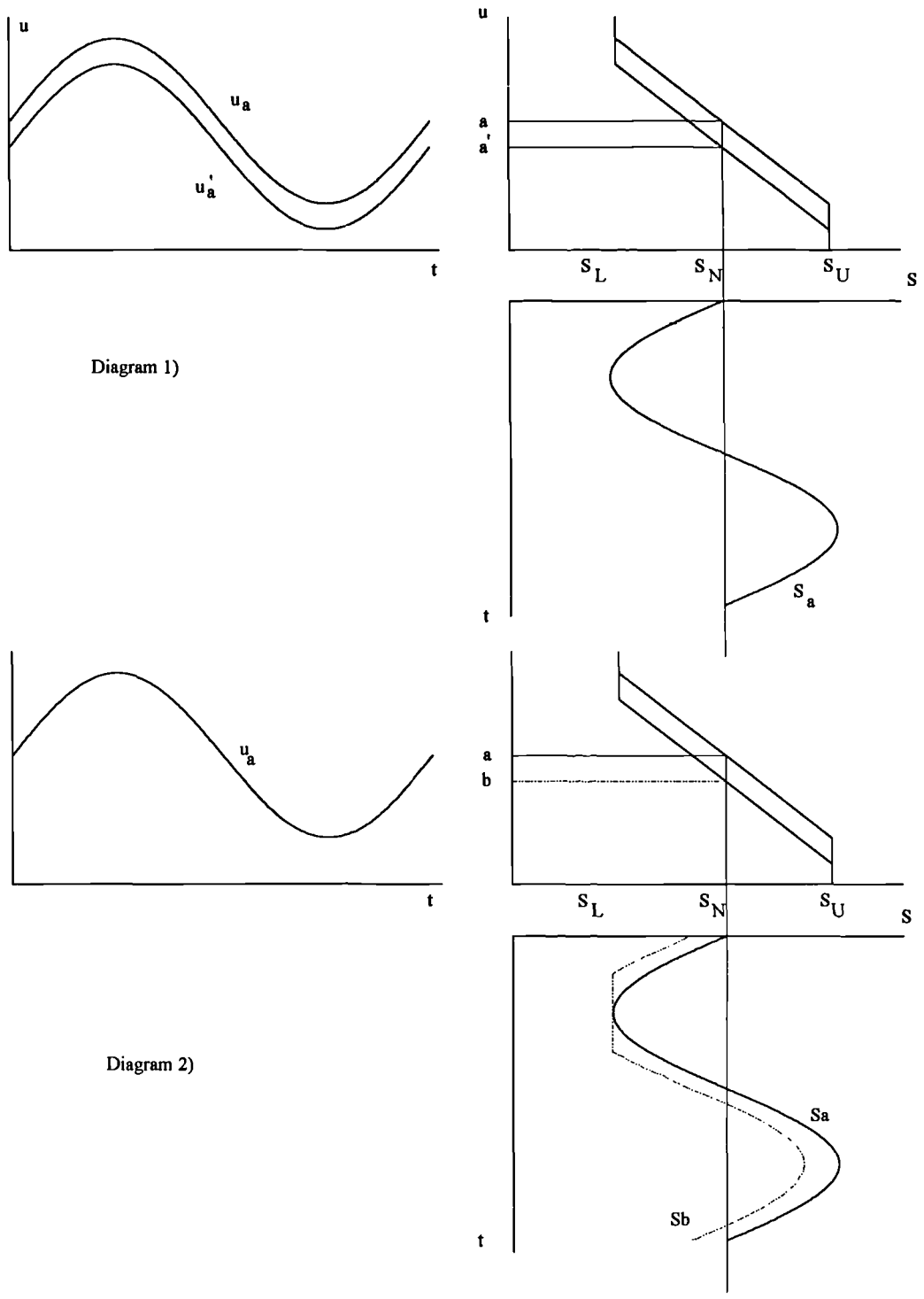


Figure 5.16 Examples Of Possible Variations In Final Signal As A Result of The Reset Function

5.4 Design And Construction Of The Replica Face

5.4.1 Design Principles

This section details the design and construction by the researcher, of the final replica head in 3-D. The following design was developed through a series of prototype versions which were considered as unacceptable for use in analysis due to their inability to produce distinct displacements or realistic motion. As **[Kepler73]** stated

> "there are few concrete rules to guide the kinematic designer; he must rely largely on the his experience with similar mechanisms. This phase of the design requires much ingenuity, inventiveness, imagination and patience".

From Section 4.2, a novel method was developed to analyse the proposed facial action system. This was based on the creation of an animatronic replica face to animate identical actions to those produced by the performer. The practical design and construction of this face was dependent on the following criteria. The physical conformation of the face must be identical in size and shape to the live face at its "expressionless" or neutral face. The skin must be sufficiently flexible to simulate the characteristics of human skin and it also should be of similar appearance and colour to that of the live face. The inner skull must be rigid and retain the skin in position, where necessary, it should also not impede the movement of the skin in expressive areas of the face. The overall drive system must produce the required displacements at a sufficient rate and smoothness of action to produce realistic motion. Finally, the mechanical linkages and drives must be hidden from view.

For the practical design of each mechanical linkage between the servo drive and the skin at the key point, a number of factors had to be taken into consideration. These included the method of physical connection to the skin, the effects resulting from its physical characteristics, and the required interactions, where necessary, between linkages at the key points. Other factors that affected the final design included the positioning of the servo drives and the position of other existing linkages. Further to the proposed design of Section 4.5, certain compensations were also added to the design to animate, where necessary, the interactive effects that occur in areas other than the key points.

From the initial design of Section 4.5, and using the physical measurements of the live face as the overall objectives, the design of each point mechanism was first produced as a kinematic displacement diagram **[Kepler73]**. This type of representation provides a method for the determination of all the possible positions of the various parts in the linkage.

In each of these design diagrams, the position indicated by the solid line represents the position of the mechanism at rest and the light or hatched lines represent the extremes or limits of displacement. This type of design ignored the effects of the forces produced by the elastic effects of the skin. Once the design of each linkage satisfied its displacement requirements then a further set of diagrams were produced to define the actual sizing and construction of each individual element. From these plans the components were constructed in a variety of materials, typically, from brass rods, hinges or sections and Bowden cable. The practical construction diagrams are not presented in the thesis as they repeat the information provided in the idealised line diagrams.

From these designs, the final system was realised. Video sequence V.2 shows the final mechanical construction of the replica and indicates the complexity of the final design.

5.4.2 The Conformation Of The Head

The following sub-section considers the physical elements that were used to produce the final replica model. All of the techniques described were developed by Hensons and were considered satisfactory for the purposes of this research.

5.4.2.1 Casting Of The Live Face

A full head cast was taken of the researcher to ensure that the dimensions of the replica face were identical, using a technique to capture all the details of the specific areas of the face. From this cast, a skin and inner shell were produced. For full discussion of these procedures refer to **[Kehoe85]**. This method of casting produced a number of errors in the final cast as a result of the high degree of discomfort to the researcher.

Despite the best intentions of the researcher to remain expressionless, a certain amount of "fear" was visible on the cast resulting in the lack of definition in the brows, a large number of creases around the eye region and in the mouth where the lips were tightly closed preventing any definition of their shape. These errors in conformation resulted in a number of limitations to the overall appearance of the replica. It was acknowledged that these visible errors could have an adverse effect on the ability to produce perceptually correct lip actions and facial expressions.

5.4.2.2 The Skin

The skin used in this project was made from latex rubber, of average thickness of 10 mm. The latex skin acts like a flexible elastic surface which the drives can manipulate and deform in various ways to produce similar changes to those of the human skin. This type of skin has limited degree of elasticity which if exceeded by the forces acting on it will result in visible tears to its surface. The creation of a realistic simulation of human skin is a major research area that clearly goes beyond the boundaries of this research.

It was concluded, that restrictions exist in the conformation of the skin which could possibly have adverse effects on the visual perception of the actions produced.

5.4.2.3 The Skull And Teeth

The skull was constructed from a fibreglass shell. It's shape and dimensions are similar to the facial cast but reduced by the thickness of the skin, i.e. 10 mm. The skull was configured to resemble the human skull. The skin was attached to the skull in similar positions to the connections in the human face (c.f. Section 3.2) and the remaining areas of the skull were cut away to allow the skin to move unimpeded during the articulatory actions. As described in Section 4.5.2, the jaw of the replica was constructed from this inner shell to be of comparable shape.

The final element of the structure was the production and positioning of a set of teeth. The replica teeth were again cast from the researcher. They were positioned and attached to the skull of the replica, by careful subjective comparisons with the live face, to ensure that they were in the correct central position at the correct height,

at the correct depth relative to the lips and the correct relationship existed between upper and lower set.

5.4.2.4 The Servo Drive Mechanisms

The drives used in this research were digital control d.c. servo motors with a full scale deflection of $\theta = \pm 90°$. The preferred operating range was defined as $\theta = \pm 45°$ to ensure against the drive moving to "dead points" where it has no effect on the output. The definitions of the type of drive for the specific actions required were based on the advice and experience of the engineers at Hensons. Their performance and the rate of the pulse width drive control signals were considered satisfactory for this project. It is acknowledged that the specification of drives is an important area of research but that it was outside the requirements of this research.

5.4.2.5 The Mechanical Support Frame

A metal framework at the rear of the face was constructed to support the face and servo drives. It also provided a frame to support, where necessary, sections of the mechanical linkages. Its rigidity ensured that the possible effects resulting from drives or linkages changing position were minimised. It allowed the majority of the drives to be held at positions remote from the restricted inner area at the rear of the face and had the advantage of allowing easier access to them.

5.4.3 Design of the Jaw

As proposed in Section 4.5.2, the jaw was produced from the fibreglass skull as a separate piece which was free to rotate about the defined axis of rotation. The principle of the design is shown in the free-body diagram in Figure 5.17

The displacement of the driving arm of servo D_{jaw} was designed to produce an angle of rotation identical to the measured angle from photogrammetric analysis (c.f. Section 4.4.3). The drive, itself, was held rigid to the overall frame at the rear of the face with its axis of rotation at the same height and depth as the replica jaw. The effects of the skin were not considered in this design, due to the complexity of modelling their effects which exceed the boundaries of this project. It was

acknowledged that their effects represented possible errors in final animation. The magnitude of torque required for production was chosen on the advice of animatronics engineers at Hensons, and was deemed satisfactory for present requirements. Again, an investigation into this design area represents future work.

5.4.4 The Brows

The principles of design have been described in Section 4.5.3 and the displacement diagram of a single brow is shown in Figure 5.18. The design was identical for inner and outer actions on both brows. The practical system was constructed along the plane of the replica "forehead" with an area of the skull cut away beneath the brows area to allow free motion. A small section of fibreglass was glued to the inner surface of skin to form a rigid attachment between the drive and skin. The linkage between drive and brow was achieved with a length of Bowden (multicore) cable. This type of cable is capable of exerting forces in both directions of motion over short distances (i.e. can act as a "push/pull "rod). This property allows points on the brow to be considered as "sliders" capable of vertical displacement in raise and lower actions along a single axis. The cable was linked to the brow using tubing of a comparable diameter to restrain cable action to the desired vertical displacement. .

5.4.5 The Upper Lip Centre

From design principles set out in Section 4.5.4, it was established that the centre point of the lip required displacement in both horizontal and vertical axes. The freebody design for horizontal displacement is shown in Figure 5.19 . It was based around a similar principle to that of a pair of scissors which results in equal and opposite forces (from the same source) being applied to the length of Bowden cable defined by A and A'. This cable section (A- A') was conformed to resemble an approximate shape of the lip, and glued into the skin in the upper centre lip region. This design had the advantage of incorporating the displacements of the upper mid points.

The force exerted on the cable should, if restrained from other possible paths of displacement, result in it being displaced along the Z axis resulting in lip protrusion. Similarly, when the driving action was reversed, points A - A' would move away from

each other until a point was reached where the cable action was restrained by the teeth. This should produce the desired stretch action associated with the upper lip. This also represented the physical limit of the linkage. If D_8 produced further output displacement, the linkage between A-A' would break as a result of the resistance of the teeth.

This action drive is shown in diagram a) of Figure 5.20. Its design and construction was based on the same technique developed for the brows using the Bowden cable to produce the push/pull (slider) action at the centre of the lip. The flexibility of the cable allows vertical action for different horizontal displacements as shown in diagram b) of Figure 5.20.

From this initial design, visual recognition of the action suggested that the vertical displacement was not purely at a single point but throughout the whole central lip area. Pivots were, therefore, added to the horizontal output arms to allow overall vertical displacement, due to D_9, as well as the primary horizontal protrusion action due to D_8. This is shown in diagram a) of Figure 5.20.

In practical terms this produced the following problem: The final output arms, due to their physical construction, were held at different vertical heights (B - A and B' - A'). When restricted from vertical displacement,by the action of D_9 , the lip cable would act purely in protrusion. With no such restraint, a protrusion action would result in the production of a moment about the centre of the cable, creating a non linear and visually incorrect action in the skin.

To overcome this problem, a mechanical compensation method was developed to prevent vertical displacement of differing magnitude in the output arms. This is shown in Figure 5.21. The principle of its design was that both output arms are connected with restraining arms to the same set of pivots. If arm B-A is raised by ymm, as a result of D_9 action, restraining pivots will move, forcing arm B' -A' to raise by an identical ymm. In summary, the drive combination was designed to produce displacements in both horizontal and vertical axes comparable to those measured from the face.

5.4.6 The Lower Lip Centre

The design of the lower lip centre was identical in principle to that of the upper lip. Limitations in the physical space available resulted in two drives acting together through separate linkages to produce the final protrusion action. To produce lower lip actions of protrusion and stretch, unaffected by the action of the jaw, the drives and linkage were physically held on the replica jaw.

5.4.7 The Corners

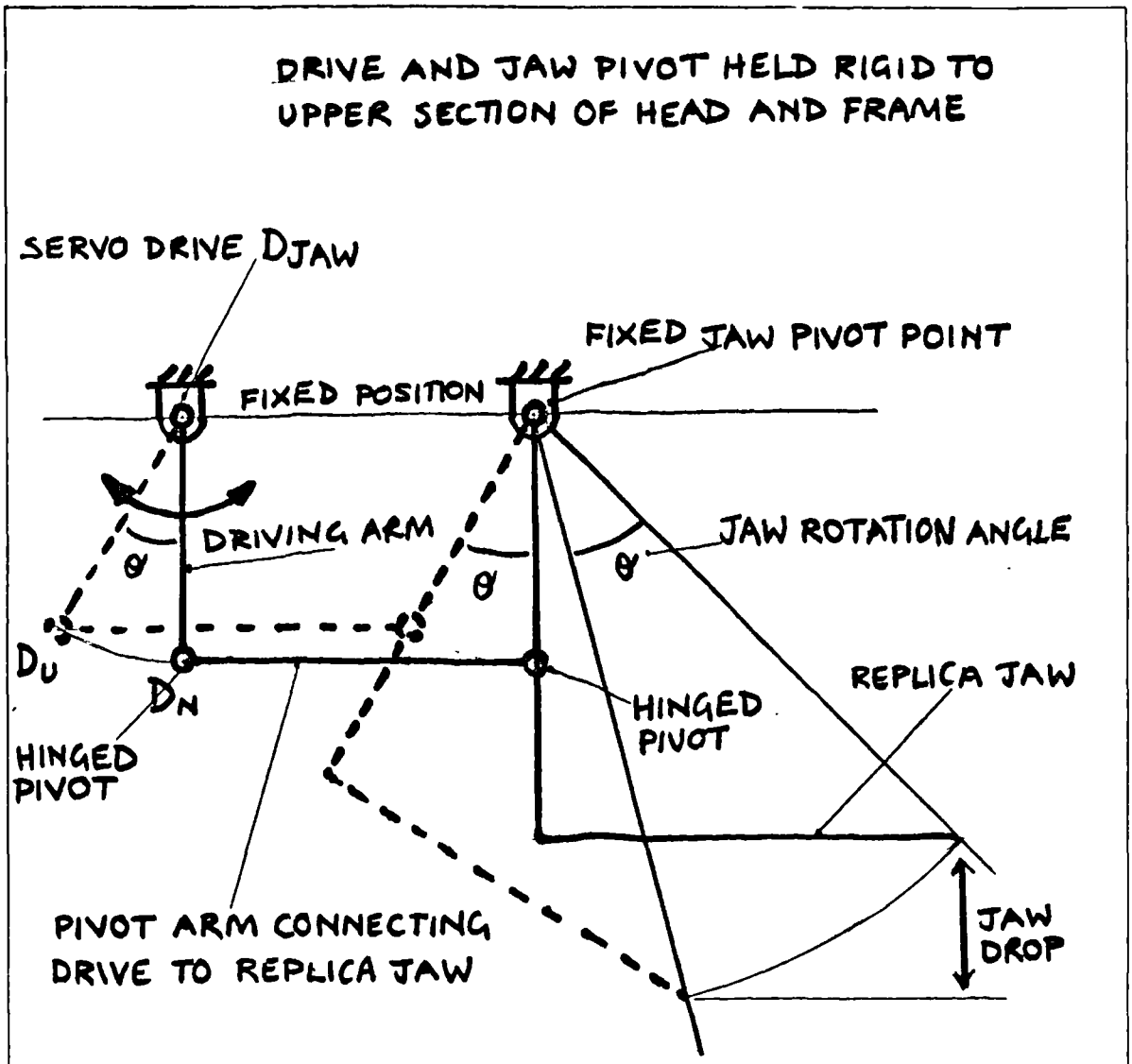
As stated in Section 4.5.7, the design of the corner drive system was a complex problem that should achieve the following criteria:

1. horizontal action in both directions around the teeth;
2. vertical stretch action;
3 vertical action resulting from jaw drop; and
4 a further action of horizontal in/out action resulting from skin displacement against teeth was also desirable.

The horizontal action of both corners is shown in the displacement diagram of Figure 5.22. The final attachment to the skin was over an area rather than at a specific point in order to reduce stress on the skin. To allow vertical displacement at the corner in both stretch and drop, and to allow the skin to fall against the teeth, the output arms OC and OC' were pivoted in two places, as shown in diagram a) of Figure 5.23.

These pivots enabled the corner to move horizontally at any vertical displacement defined by the actions of corner stretch or drop, relative to the neutral position. The vertical stretch action was achieved using a separate drive as shown in diagram b) of Figure 5.23. The linkage to the corner was achieved with a section of cable. Its properties allowed force to be exerted only in the vertical raise action to prevent any possible restriction to jaw drop.

In summary, the mechanical design produced the desired complex combination of stretch, horizontal and drop actions, allowing the displacement of the corner in the region defined by Section 4.5.7.



			ACTUAL			
No. off	Item	Description	Size	B.S. No.	Material	Note
Drn.		JAW ACTION (FREEBODY DIAGRAM) K_{JAW}				Scale: 1:1 Weight.
Date						
App						Drawing No
Date	1994					FIGURE 5.17

Figure 5.17 Free-body Diagram For The Construction Of The Jaw Rotation

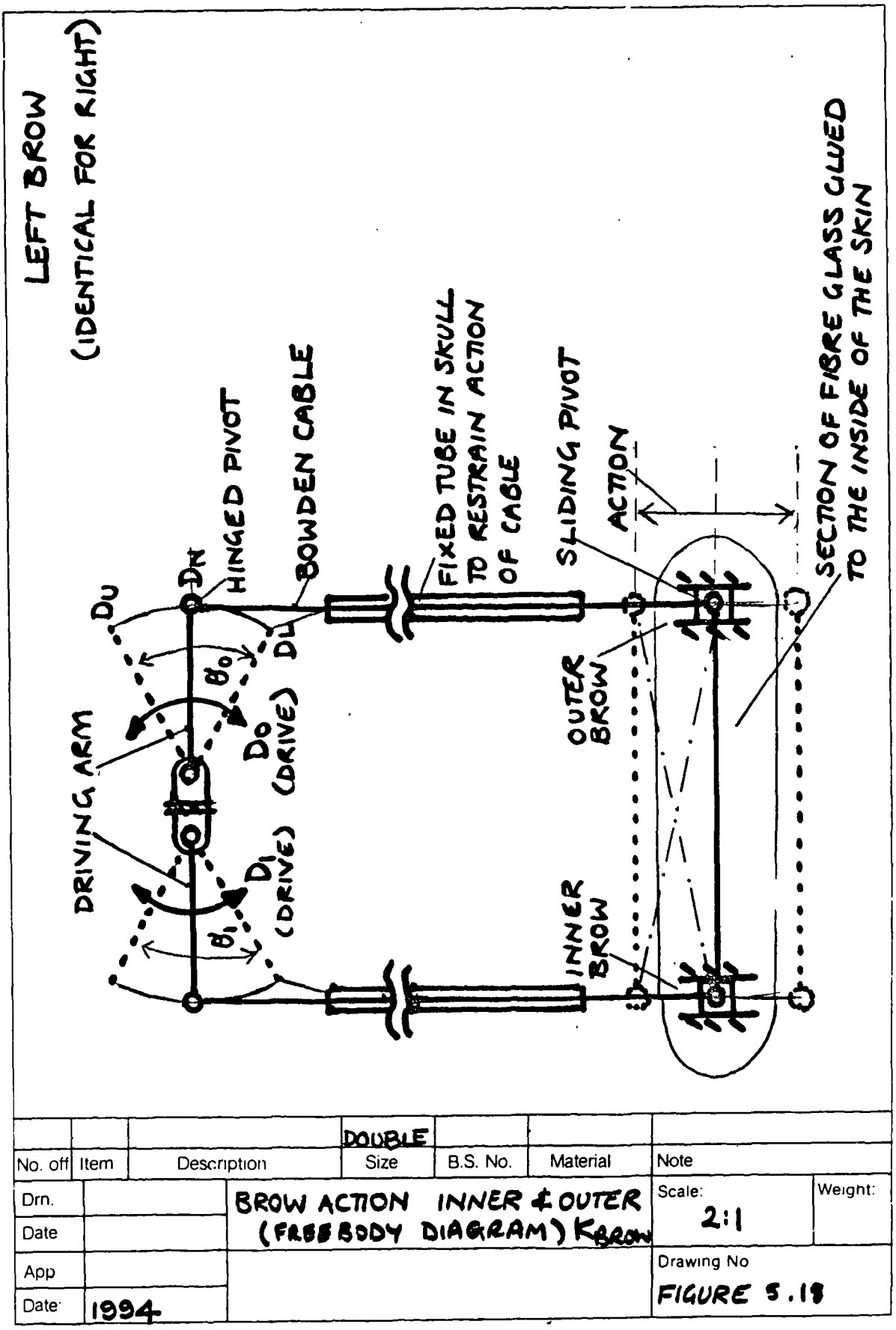


Figure 5.18 Free-body Diagram For The Construction Of The Inner And Outer Brow Actions

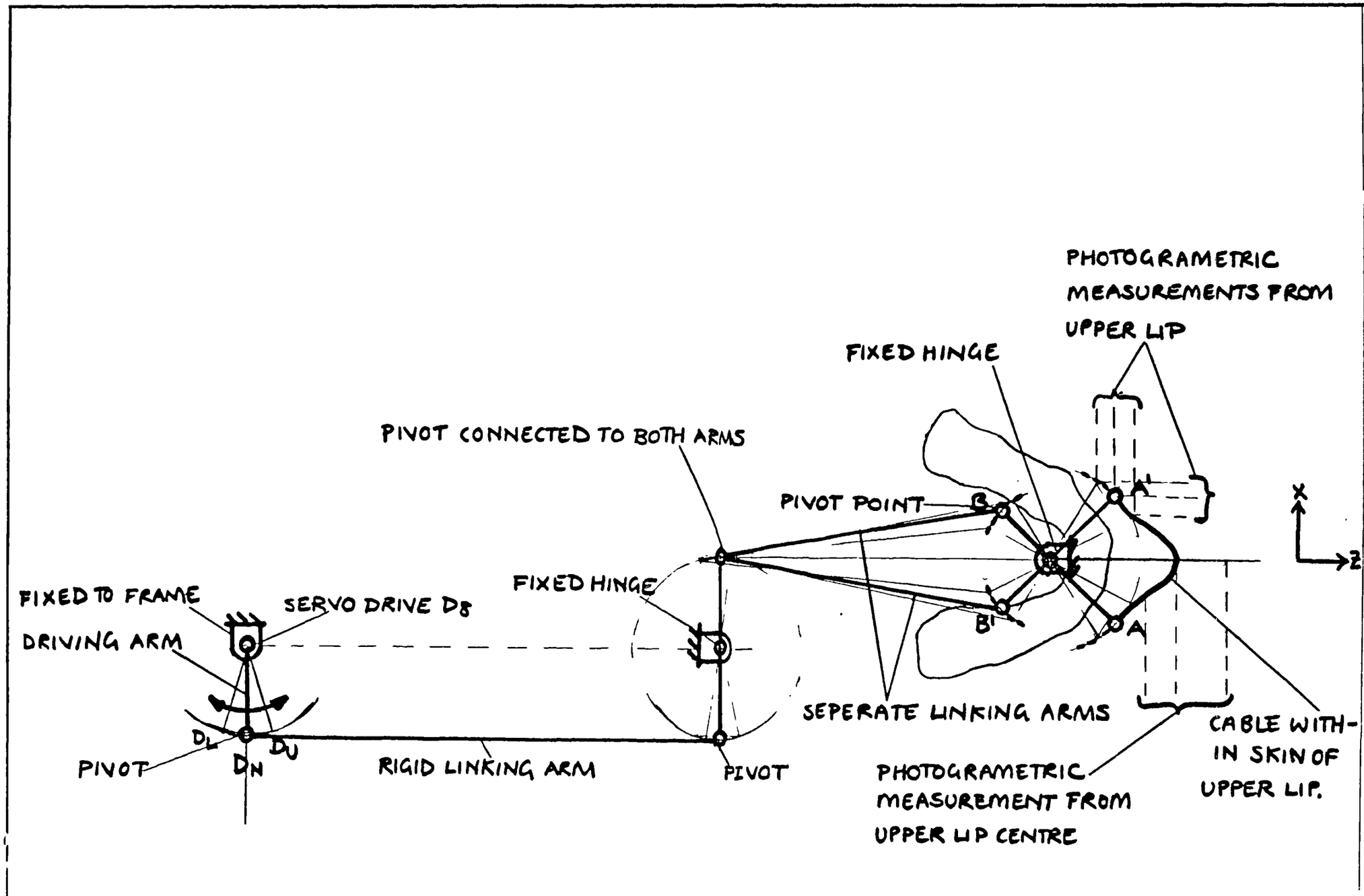


Figure 5.19 Free-body Diagram For The Construction Of The Horizontal Action Of The Upper Lip, Centre

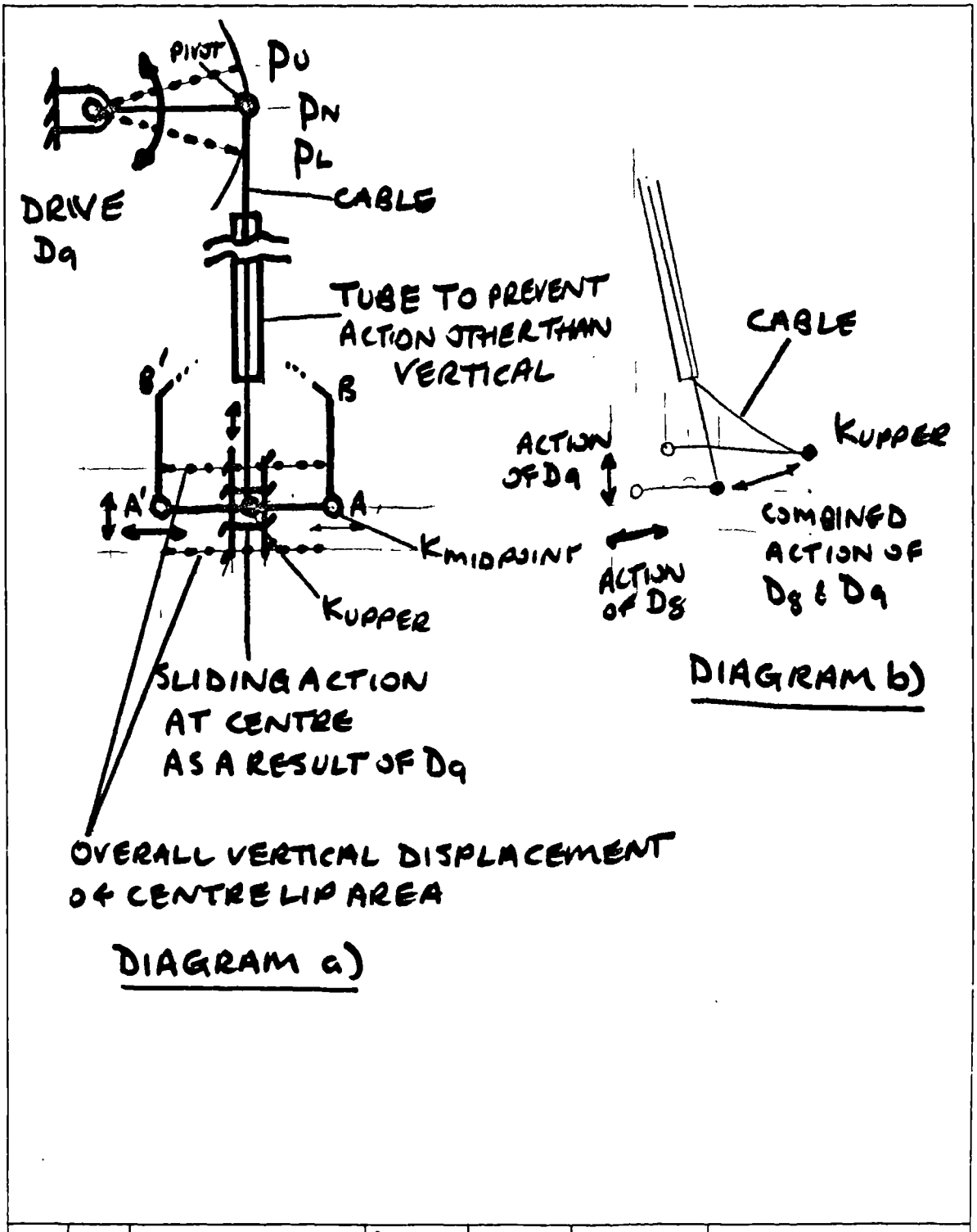


Figure 5.20 Free-body Diagram For The Construction Of The Vertical Action Of The Upper Lip, Centre

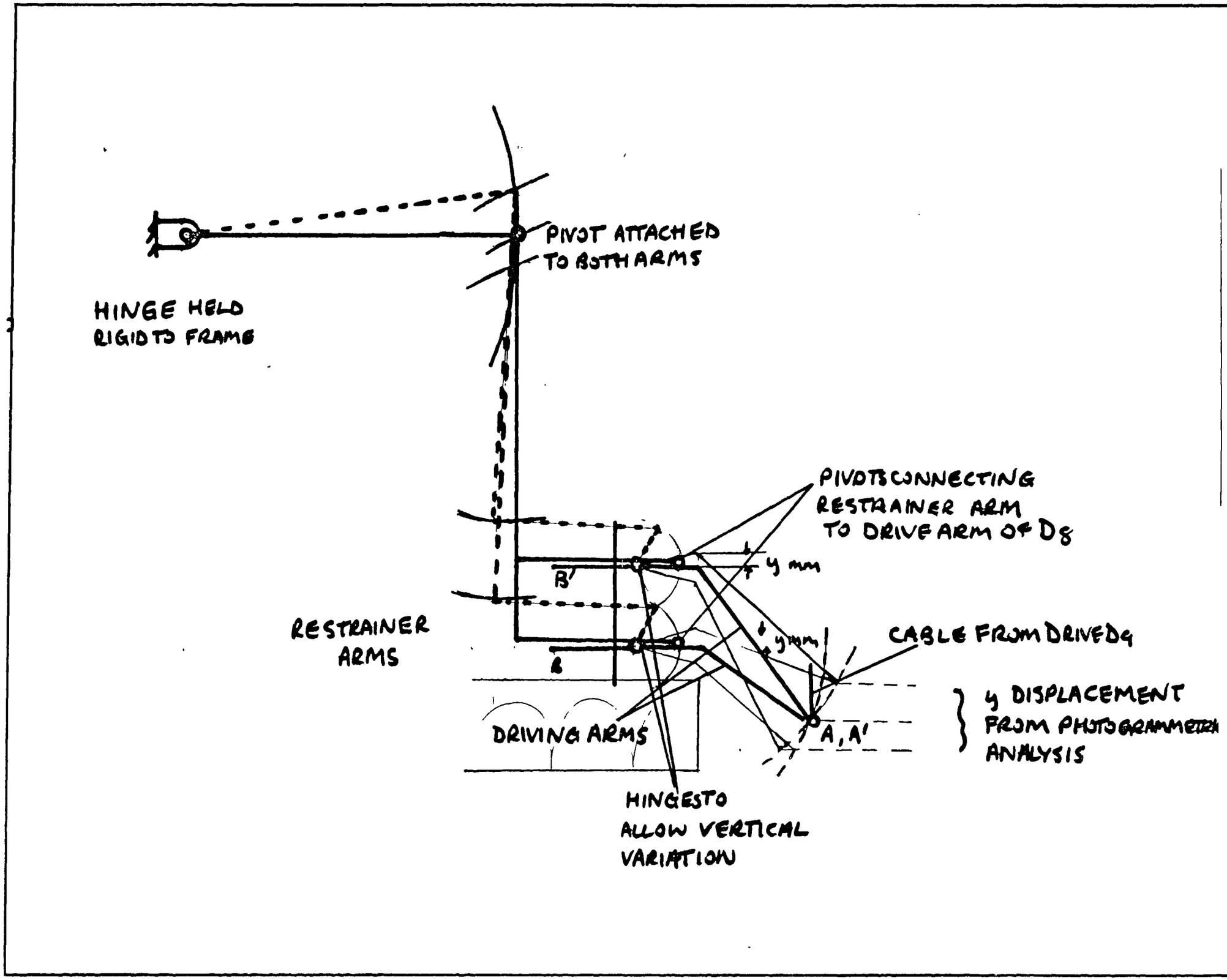


Figure 5.21 Free-body Diagram For The Construction Of A Vertical Action Restrainer At The Upper Lip, Centre

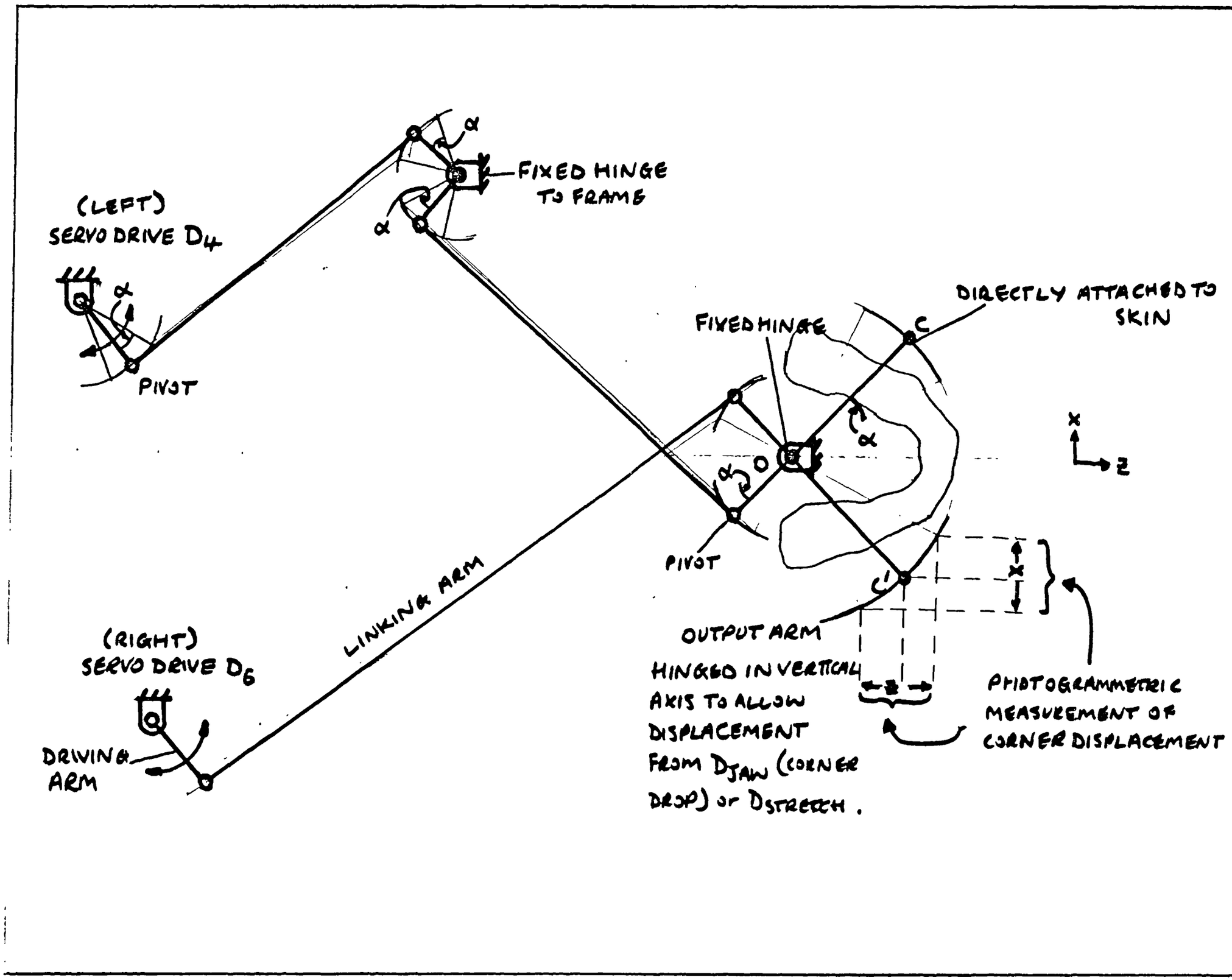


Figure 5.22 Free-body Diagram For The Construction Of The Horizontal Actions Of The Corners

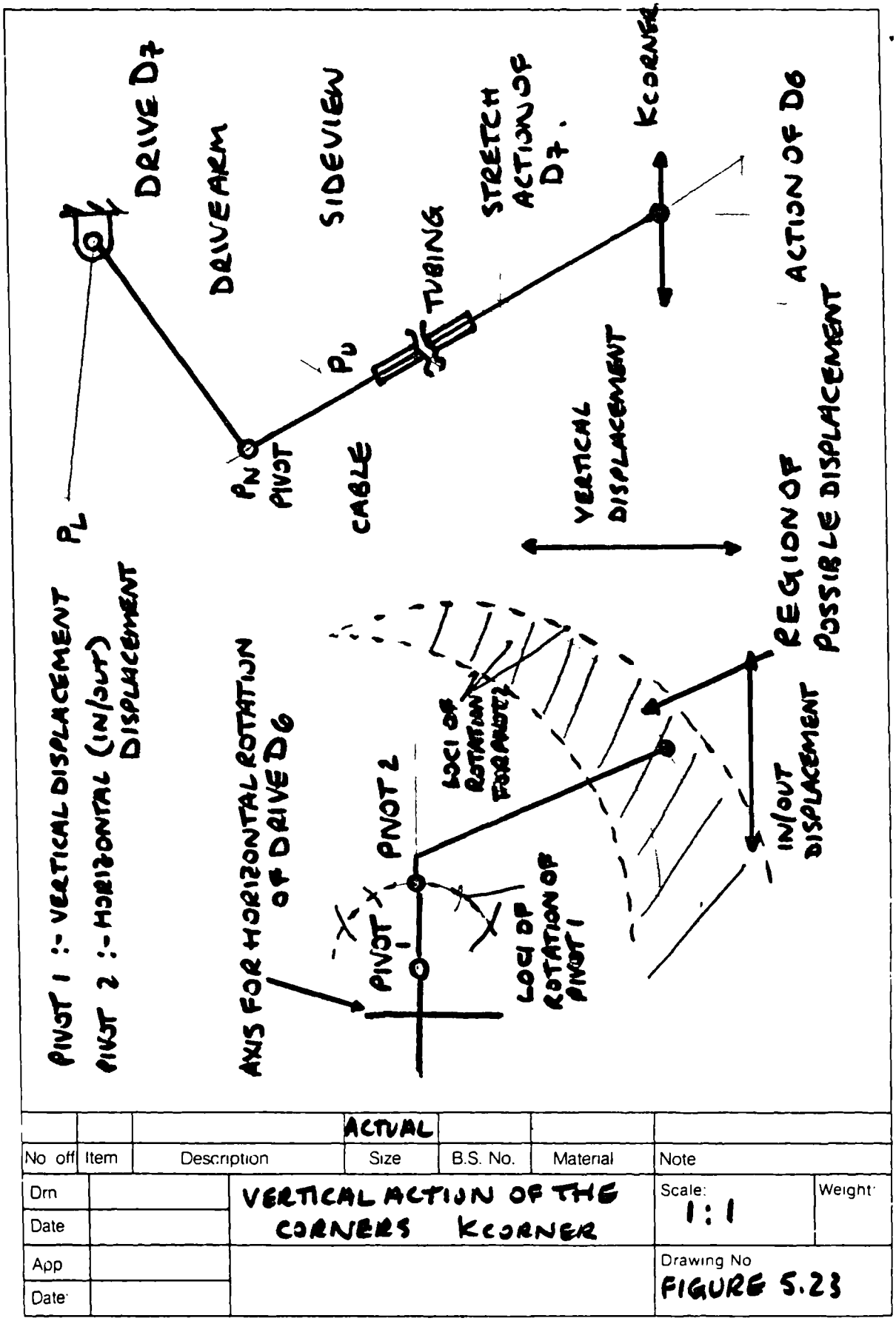


Figure 5.23 Free-body Diagram For The Construction Of The Vertical Actions Of The Corners

5.5 Summary

This chapter has described the practical realisation of the individual elements of the system for their subsequent application in facial action performance control. It was shown that the sensor system could produce analogue signals, directly proportional to the displacement of a reflective area, relative to some definable datum position.

The investigation of the sensing system in Section 5.3, produced a number of conclusions on its overall characteristics. The system fulfils the measurement requirements in terms of range and sensitivity for the majority of possible facial displacements as derived in Section 4.4. It was acknowledged that maximum displacement at certain key points will exceed the range of the sensor, and that this would produce limitations in the overall ability of the system to sense accurately all key point changes. It was concluded that it was important to take the following precautions to ensure against possible errors:

The repeated positioning of reflectors on both faces in similar positions with comparable orientation to maintain the relationship between sensor and reflector, $F_{physical}$, and to prevent subsequent variations resulting from alterations to the function F_{sensor}. It should also ensure that the datum distance was defined at the same neutral position for each recording in F_{sensor} and, hence, in the conditioning parameters (c.f. Section 5.3).

Excessive head actions should be minimised to prevent any changes to the support mask position to the face. An examination of head motion (results available, but not presented here) by the performer suggested that the position of the mask relative to the face was altered by vigorous motion, producing errors in measurements. Consequently the performer endeavoured to avoid such actions.

The Data Acquisition System achieved a number of tasks. These were as follows:

1 The recording of control input data from 16 channels of digital input;

2 the data storage in text form for subsequent manipulation and analysis;

3 The development of algorithms to individually condition the control channels;

4 The playback of control data files via 16 channels of analogue output;

5 The production of definable and consistent sampling rates within an acceptable range;

6 The playback of control data, $\underline{c}$, to the replica with ability to simultaneously record the measured changes from the sensor system $\underline{S_r}$.

7. The real time operation of the above tasks.

This facility also enabled direct and definable software control for exact and repeatable positional changes of specific and definable drives or key points.

The *dos* clock, within the processor, produced the timing for the software procedures. This was a satisfactory method, given the low sampling frequency required. Inconsistency occurred in the record and playback time which were acknowledged as being significant only if playback data was to be produced simultaneously with a recorded acoustic signal.

The final construction of the replica was in the opinion of the researcher satisfactory for application as an output system for visually distinct performances. The construction achieved the design requirements of individual displacements. The replica also animated certain kinematic inter-relationships between key points, specifically in the corner region to allow corner drop. A number of drawbacks existed as a result of the overall appearance which were considered detrimental to the final visual performance. These were the visible skin tears at the corner of the mouth, the lack of expressive creases in the brow region, and the lack of eyes and hair.

In summary, a fully operational prototype system was developed by the researcher to enable the analysis of the research hypothesis proposed in Chapter 4. The following chapter will present the results and conclusions of this analysis.

Chapter 6

Results And Analysis Of Facial Action Control System

Chapter 6

Results And Analysis Of Facial Action Control System

6.1 Introduction

This chapter presents the results and analysis of the final system. Descriptions are given of the test procedures undertaken to assess the capabilities of the system based on the proposed method of analysis from Section 4.2.

The analysis in ideal conditions of the optical sensing technique, in Section 5.3, has established that the proposed method was capable of the measuring the displacement of a reflective area along its focal axis, relative to some definable datum. This was true over a series of ranges and provided that certain criteria were maintained. The system was therefore suitable for application as a facial action sensing system.

In summary, the overall objective of the proposed system was the consistent and accurate production of control information not only for static expressions but for the temporal changes associated with continuous speech and expressive facial actions.

The hypothesis developed to achieve this objective was evaluated by the application of the methodology described in Section 4.2. This chapter describes the test procedures to evaluate the following theories;

1. an optimum set of visible key points exist on the face which can provide sufficient information to accurately describe the primary actions of the face, in terms of displacement changes.

2. the defined set of point displacements can provide sufficient data, in terms of magnitude and rate, to produce accurate control for the final animation.

3. the final animated performance should be of the similar actions with identical timing and of comparable magnitude; and

4. the primary set of visible actions, are correctly animated in synchronisation with an audio soundtrack, to convey the same desired message to the viewer.

This chapter is organised in the following way. Sections 6.2 and 6.3 discuss the results generated through the controlled experiments of the replica actions under software control. Investigations into firstly, the ability of the sensing system to recognise all of the key point displacements on a facial surface and secondly, consideration of the capacity of the animatronic replica to produce visually distinct actions, through subjective analysis, are described.

Section 6.4 examines, both objectively and subjectively, the overall production of facial animation through the direct control from a performer's face. Graphical, photographic and video analysis of specific examples are presented for the production of isolated syllables and for the overall performance of real time, continuous speech animation.

Section 6.5 presents an evaluation of the overall ability of the system to achieve its objectives and considers the validity of design principles. The present limitations are described and, where necessary, proposed solutions to these problems are discussed to improve later performances.

6.2 Practical Analysis Of The Key Point Principle

The following series of procedures were developed to achieve a number of aims.

1. To draw conclusions on the ability of the sensor system to consistently measure facial changes at the key points.
2. The deduction of the optimum sensor and reflector positions at each point.
3. To evaluate the effects of key point interaction in the replica through objective measurements.
4. The subjective analysis of the replica actions at the individual key points to determine the individual drive parameters; P_U, P_L and P_N, for $\underline{F_{map}}$. Where necessary, this included the derivation of the correct vector displacement, $\underline{v}$, when more than one drive acts on the key point.
5. The derivation of control parameters; S_U, S_L and S_N, for $\underline{F_{condition}}$ for each key point.
6. The derivation of $\underline{F_{total}}$, through the measurement of $\underline{S_r}$ of all positions in each key point displacement. This led to the generation of the look-up tables defined by $\underline{F^{-1}_{total}}$ (c.f. Section 4.3.3).

To achieve this analysis, the replica was manipulated directly by software control using the data playback system. Using the program PLAYRECORD, described in Section 5.2, stored files were inputted to provide user defined and exact control signals, $\underline{c}$, directly to the input of $\underline{F_{control}}$. This allowed the animation of individual drives and key point drive clusters to be analysed in isolation. The mask was positioned on the replica face to record the displacements, $\underline{v}$, for each investigation. The jaw section of the mask and the reflectors were attached to the skin, with a make-up adhesive, with care taken to ensure that the positions were similar to those defined in Chapter 4.

Analysis of $\underline{S_r}$ measurements drew conclusions on the ability of the sensing system to consistently sense the actual actions produced in facial displacement. This argument was valid provided the replica actions were subjectively evaluated as being visually similar to those of the live face.

The following sub-sections discuss the results derived from the specific analysis of the key point at the upper lip (K_{upper}). The procedures and analysis methods were undertaken for the other key points and the conclusions from those tests are described in the following section, Section 6.3. These results were not included in this thesis but are available upon request.

6.2.1 The Analysis And Reduction Of Drives Parameters To Single Trajectory Motion.

The need to subjectively reduce the range of drive outputs to distinct trajectories of motion was based on the fact that the majority of actions produced by the drive cluster are not related to the perceptually correct actions of the live face (c.f. Section 4.2). The subsequent reduction of drive outputs led to the definition of the specific drive parameters for $\underline{F_{control}}$ required to produce the desired trajectory of motion. By measuring the resultant displacements of K_{upper} with the sensor S_8, the control parameters for the subsequent control by live face were derived; i.e. S_L, S_N and S_U for $\underline{F_{condition}}$, the conditioning function.

From the proposed design theory in Section 4.5.4, drives D_8 and D_9 at K_{upper}, could act on the skin to produce distinct motion in two degrees of freedom. The upper lip centre, however, moves through, primarily, a single axis of motion. This axis was defined by visual identification of the correct permutation of drive positions. The resultant measurements from sensor S_8 indicated that the output displacement from the specific drive permutations produced sensed values that satisfied the requirements for control signals. This was based on the assumption that the reflector was positioned correctly (c.f. Section 6.2.2).

Diagram a) of Figure 6.1 shows predicted parameter values for the sensor measurements, diagram b) indicates the actual measured values by sensor S_8 which

show that displacements other than the desired trajectory will produce changes in the measured signal. Diagram c) shows a representation of the final drive limits for the upper centre determined through visual analysis and diagram d) shows the final mapping functions ($\underline{F_{map}}$) for the key point displacement of K_{upper}.

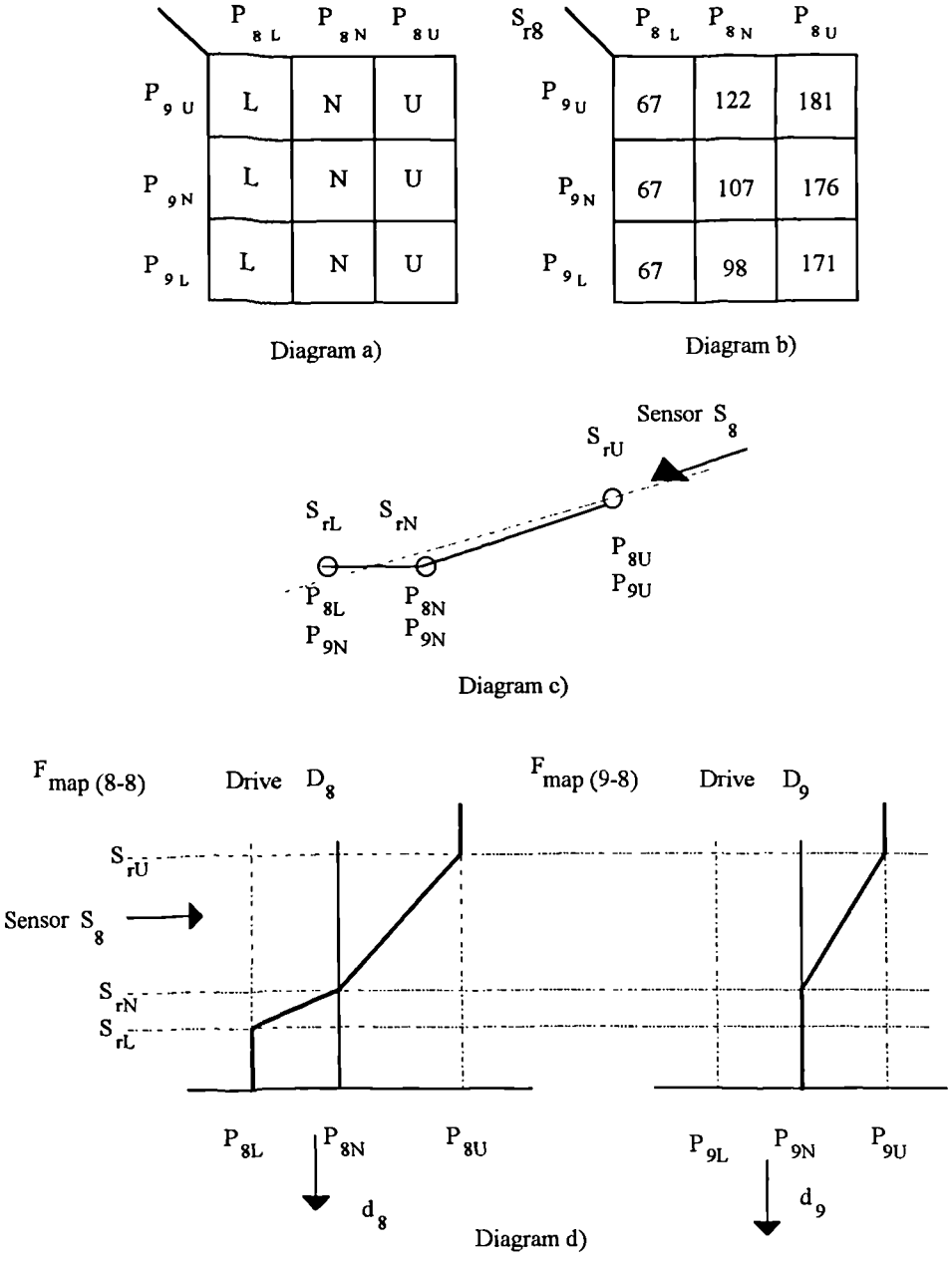


Figure 6.1 Example Of Drive Cluster Reduction For Upper Lip Centre.

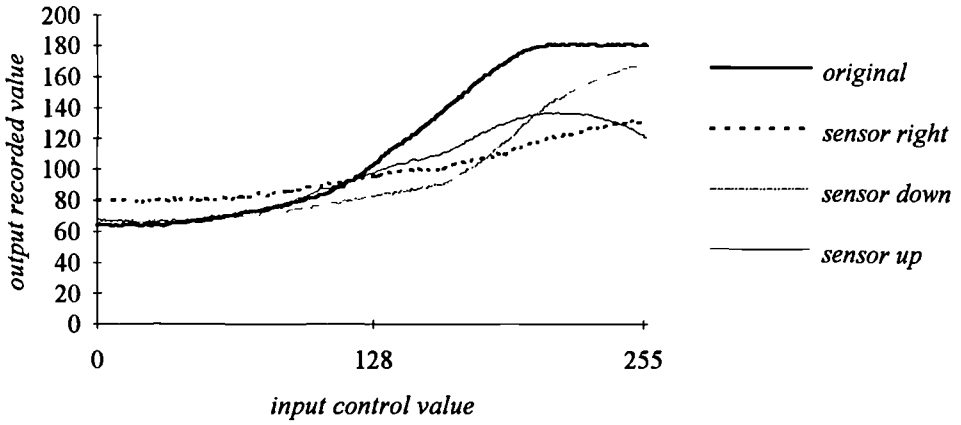
In summary, these investigations resulted in the reduction of each key point to a single trajectory of displacement bounded by a minimal set of reference limits. All other positions were determined by linear interpolation. The control parameters for each $F_{condition}$ were defined by the measured values of S_8 at these defined limits. Their definition was dependent on subsequent analysis of the following section.

6.2.2 Assessment Of Sensor And Reflector Positioning

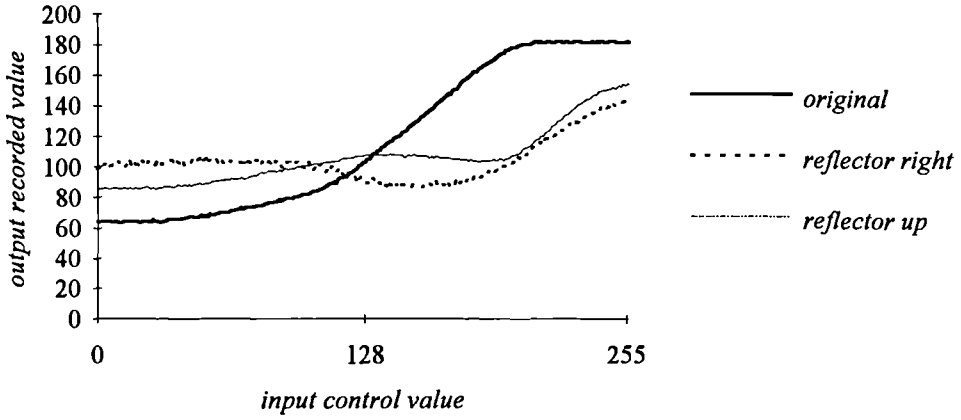
The initial settings for each reflector and respective sensor were defined in the conclusions of Chapter 4. From these settings, the positions were altered, where necessary, to produce variations which the researcher defined as satisfying the signal criteria of sufficient range and sensitivity. This trial and error procedure was a direct consequence of the problems inherent in the use of the infra-red sensor where the emitted rays are invisible to the human eye. The subsequent examination evaluates the possible variations resulting from the incorrect positioning of the sensor and reflector at the upper lip centre, K_{upper}. The different positions were defined at 10 mm offsets in the horizontal and vertical axes. The resultant measurements are shown in Figure 6.2. The same procedure was carried out for all other points.

In Figure 6.2 and all subsequent figures, the control value '0' represents C_L, and hence P_L, the value 128 represents C_N (P_N) and the value 255 represents C_U (P_U) for each drive cluster.

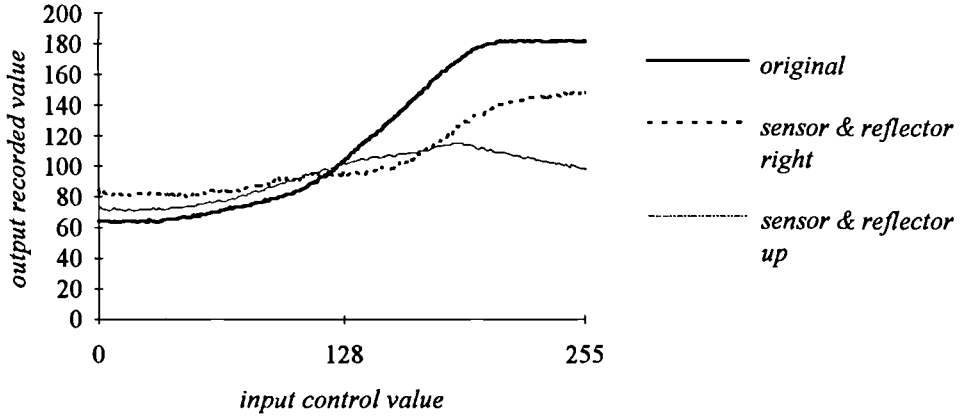
The recorded measurements suggest that changes in the position of the sensor and reflector with respect to the point of interest of 10 mm or greater would result in alterations to the key point function $F_{total(upper)}$. Consequently, differences would occur in the final measurement and lead to errors in the conditioning of the control signals. In conclusion, the research endeavoured to define and retain the same key positions of the reflector and sensor on both the live and replica faces.



Plots a) Variations In The Position Of The Reflector



Plots b) Variations In Sensor Position Only



Plots c) Variations In The Position Of Sensor And Reflector

Figure 6.2 Plot Of Measured Characteristics For Variations In The Position Of The Sensor And/ Or Reflector At The Upper Lip Centre

6.2.3 Analysis Of Key Point Motion Through Its Full Range Of Displacement

Following the definition of each $\underline{F_{map}}$, it was important to examine the variations of the key point over its full range of displacements. This was achieved through the measurement of $\underline{s_r}$ for all drive positions. The reasons for this examination were, firstly, to confirm that the sensor was capable of linear measurement over the full range of point displacement and, secondly, to generate the values for the look-up table for subsequent control compensation (c.f. Section 4.3.5).

This was achieved by using the program PLAYRECORD, to input a ramp signal, (0 to 255), directly as control input, $\underline{c}$, to the specific drives at the point of interest to move it through all its possible positions along the defined trajectory. At every increment, the sensor S_8 measured the output displacements and the recorded signals represented the key point function $\underline{F_{total(upper)}}$. All other points were held at neutral to ensure that only the effect of the desired point was measured.

Following the same procedure as that of Section 5.3, the tests were repeated *i* times and the standard error measurement was calculated for each test. From the previous experiments, the accepted tolerance range for evaluation of test consistency was increased to $\pm 3.29\sigma_m$ (±5 values), based on the fact that the practical technique for production of point displacements was likely to be less consistent than previous operator defined method. The subsequent measurements at all key points were therefore considered consistent for the repeated tests within this tolerance.

Figure 6.3 displays both the signals recorded by S_8 over the full range and the predicted function based on only the measured parameters for K_{upper}.

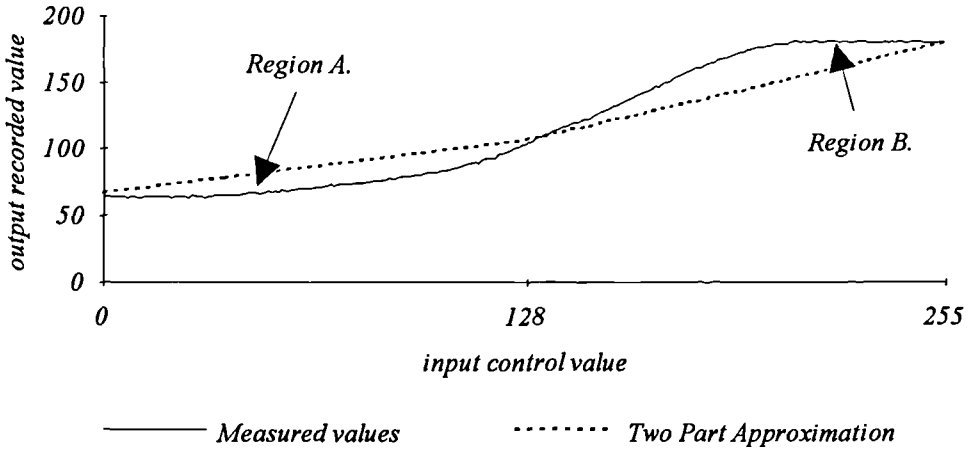


Figure 6.3 Plots Of Actual And Predicted Key Point Function At Upper Lip Centre

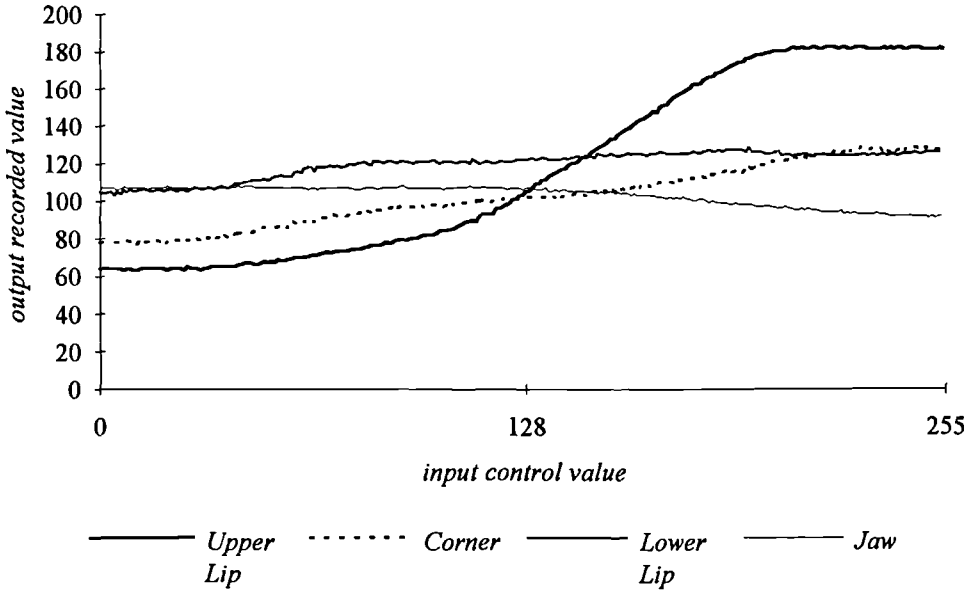


Figure 6.4 Measured Effect of Key Point Interaction At Upper Lip Centre On Replica

The results highlighted certain non-linearities between the desired input and actual output. Regions A and B, in Figure 6.3, existed at the limits of full scale deflection and as a result changes in the control input fails to produce any resultant change in the sensed output. This could result in the production of differences with linear conditioning. There were a number of possible reasons for these non-linear regions;

a) the actual displacement, $\underline{v}$, does not produce any relative displacement, $\underline{v'}$, along the focal axis of the sensor.

b) the incorrect definition of the drive parameters at full scale deflection. The practical technique for defining the drive parameters P_U and P_L was reliant on the visual identification of these limits. If these defined limits exceed the physical limits of the drive; D_U and D_L, non linear regions would then exist. (c.f. Section 4.3.1).

c) the full scale displacement of the point exceeds the measuring range of the sensor. As a result the sensor fails to measure all key point displacements. This represents a current limitation of the sensor system which could affect the subsequent results of the final system.

6.2.4 Evaluation Of The Effects Produced By Other Key Points

Having established that the sensing system was capable of consistent measurement of key point changes in isolation, it was important to assess the effects on $\underline{F_{total(upper)}}$ due to the interactions with other key points. This allowed evaluation of the diagonalisation theory proposed in Section 4.3. For the upper lip centre, the effects of the corner, lower lip and jaw displacements were examined. This was achieved by measuring the resultant displacement at K_{upper} with the drives D_8 and D_9 held in the neutral position, and the other points moved through their full range of displacements. Figure 6.4 plots the results which highlighted a number of points.

Variations in the recorded signal from corner displacement result from the mechanical interactions that occur through the skin and linkages in the lips of the replica. The effects of this interaction should be avoided in the final system due to the following argument. In the sensing of the live face, this type of interaction would not occur due

to the physiology of the face. This prevents the production of, say, corners together without some form of centre protrusion taking place and consequently, the system would not produce control signals necessary to cause this type of interaction.

The resultant effect of the lower lip motion is due to the physical contact produced between the lips which could not be avoided.

The action of the jaw rotation has no physical effect on the position of the upper centre. The resultant change was, therefore, the product of some other effect to alter the amount of reflected power. The conclusion was drawn that due to the proximity of the lower lip, stray reflections from its surface were sensed by S_8. When the lower lip moved away from its datum position, S_8 sensed a change in the intensity of the reflected power which changed the output signal accordingly.

This could be solved by adjustment of the sensor orientation though that could reduce the sensitivity of measurement of the key point itself. As the maximum variation in the sensed signal was less than 20, which equates to approximately 2 mm from sensitivity measurements of Section 5.3, the present research concluded that the effect was minimal and made no attempt to rectify the problem. It did, however, represent a possible drawback in the final system and its possible effects were taken into account in final analysis.

6.2.5 Summary

In conclusion, the animation of the upper lip actions were considered successful in the production of visually distinct positions. One particular criticism of the mechanical production was the lack of visibility of the teeth when the upper centre was stretched against the teeth. The results shown in this section proved that the technique to sense the key point actions was capable of the production of an output signal proportional to the magnitude of displacement. The data recorded indicated that the sensor system was capable of producing consistent output signals proportional to the changes in facial displacements. The results for all points suggested that signals of sufficient range and sensitivity were possible from the defined key point displacements on the face.

6.3 Analysis Of Individual Key Point Actions

Using the same test procedures as those in the previous section, the following evaluations were drawn on the sensing and animation of actions at the other individual key points. Examples of the final animation of individual points, under software control, can be seen in video sequences V.3 and V.4.

6.3.1 The Brows

The following conclusions were drawn on the actions of the brows. The visual analysis and measured displacements confirmed that the proposed design for drive and control could successfully realise the animation and sensing of brow actions through the defined set of independent points. The video sequence V.3, shows that the brows could produce a wide variety of distinct shapes. It was concluded that, in fact, the replica was capable of greater variety of expression than the actual live face. These animated actions in the replica brow were not ideal in terms of their trajectory as they were inclined to move the reflector across the defined sensing axis rather than along it. The use of reflectors with sufficiently large areas (A=15 mm square) nullified any possible effects on the final sensed signal. The technique for the attachment of reflectors to transpose the key point motion to an axis displaced from the surface of the skin proved successful, and allowed consistent measurements of the point displacements.

The plots in Figure 6.5 display the measured characteristics $\underline{S_r}$ for the individual points. These were produced by applying a ramp input to both points on each brow at the same instant. Each characteristic displays sufficient range and sensitivity to be considered effective in the measurement of brow displacements.

6.3.2 The Jaw

The plot shown in Figure 6.6 displays the mean characteristic for jaw displacement compared with the predicted linear characteristic. It was concluded that the proposed technique to produce measurements of the jaw was of sufficient sensitivity and linearity to produce accurate measurements over the full scale.

A possible problem that existed in the accurate measurement of the jaw was the attachment of the lower piece of the support mask to the skin of the replica. The skin at the chin is not a rigid object and is free to move, to a certain degree, over the surface of the inner skull which could result in variations in the jaw measurement. To minimise this possible effect, the lower piece was conformed to the shape of the chin and jaw bone, and care was taken in maintaining the same position of attachment.

Visual analysis of the replica jaw action concluded that the final animation was visually distinct and of comparable range to the human action. Problems arose on the replica due to the excessive strain on the skin, tearing it at the corners. It was concluded that the non-linearities in the measured characteristic result from errors in the mechanical linkage at small openings and from the resistive effect of the skin as the jaw rotation increases.

6.3.3 The Lower Lip Centre

The animation of the lower lip was a result of a combination of three drives and the visual assessment of their resultant animation proved satisfactory for the production of the distinct actions of protrusion and of stretch against the teeth.

The proposed technique to measure the displacements of the lower lip, through a sensor attached to the mask jaw section, was shown to be successful provided a number of factors could be restrained. The plots in Figure 6.7 indicate that variations which occured in the measured signal as a result of the displacements at the key points in the mouth region. The effects of the corner and upper lip on the lower lip were due to the same reasons explained for the upper lip in Section 6.2.3.

The differences in the measured signal, due to the action of the jaw, were the result of the varying degrees of change in the relative positions of the reflector and the sensor. This was due to differences in the vertical displacement between the skin and the mask. This represented a limitation in the accurate sensing of the lower lip position. From the repeated measurements, the maximum range of values resulting from this jaw displacement was recorded at +25 values which approximates to a change of 3 mm in the lip protrusion, again based on the sensitivity measurements in Section

5.3.4. Care was taken in the positioning of the mask, and of the sensor, and the possible effect of this factor was taken into account in final analysis.

6.3.4 The Corners

The animation and sensing of the corners represented the most complex problem of the production of the final system. The video sequence V.3 shows that the final drive system was capable of the production of a variety of actions in the corner region as a result of the combined actions of drives D_6 and D_7. The final actions of corner stretch and corner protrude (together), with and without corner drop, were recognised as being visually distinct, refer to video sequence V.4.

Through the investigation of individual drive displacement with step responses from the neutral position, the values recorded by Sensors S_6 and S_7 showed inconsistencies in the return to the correct reset value. The mean range of measured values was 30 which equates to variations of 3 mm in position. The reason for this fluctuation was the fact that the mechanical linkage of Drive D_6 was pivoted to allow for vertical actions of stretch or drop. This resulted in the corner being capable of taking any number of positions around the neutral position due to mechanical inconsistency. Examination of the corner stretch action measurements indicated that the skin produced damping effects on the action returning to the neutral position. These limitations could affect the final performance of the system and were taken into account during the final analysis of Section 6.4.

From the proposed mapping functions of Section 4.5.7, each sensor has control only over its own distinct region. The plots in Figure 6.8 show that, when positioned on the replica with the jaw closed, the individual sensors were capable of measuring their respective displacements. The range of C_L to C_N represents the action of corner stretch and the region C_N to C_U represents the action of corner protrude.

The results showed that regions exist at the limits of deflection where there was no discernible change in the measured signal. It was concluded that the physical displacement of the corner exceeds the range of the sensor, resulting in these cut-off regions. This was classed as a limitation of the present system and it was acknowledged that it would have an effect on the final system.

An investigation into the effects caused by different jaw positions produced the resultant plots shown in Figures 6.9 and 6.10. In conclusion, the resultant signals measured by corner sensors S_6 and S_7 confirm the design principle of Section 4.5.7 that the sensors would produce reduced measurements for actions affected by the action of the corner drop.

This, however, indicates a limitation in the control of the final animation. The reduced level of measurement from a live face will result in the application of a reduced control signal consequently producing a smaller displacement on the replica. This will prevent the production of a fully realised animation. Through time restrictions, a solution to this problem was not implemented but its theory is described in point 7) of Section 6.5.

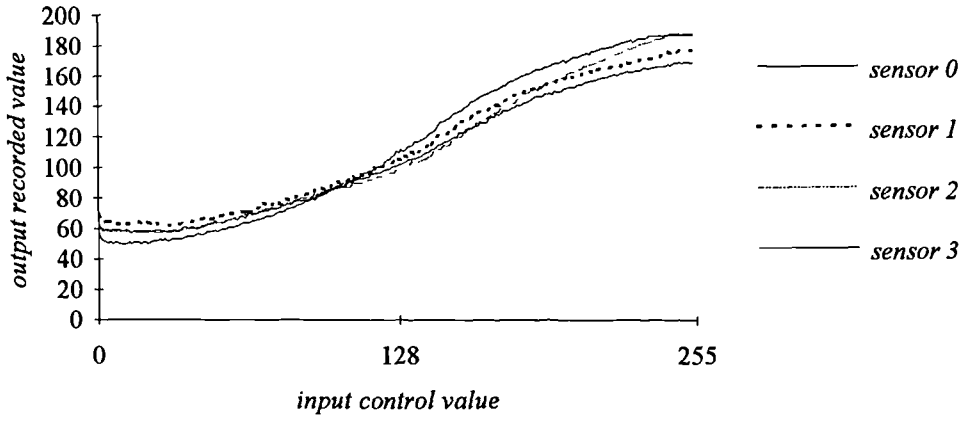


Figure 6.5 Plot Of Measured Characteristics For The Brows

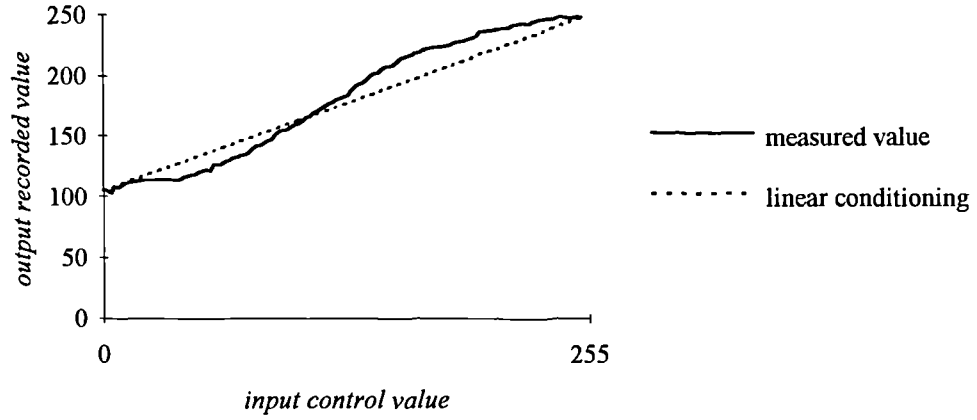


Figure 6.6 Plot Of Measured Characteristic Of The Jaw

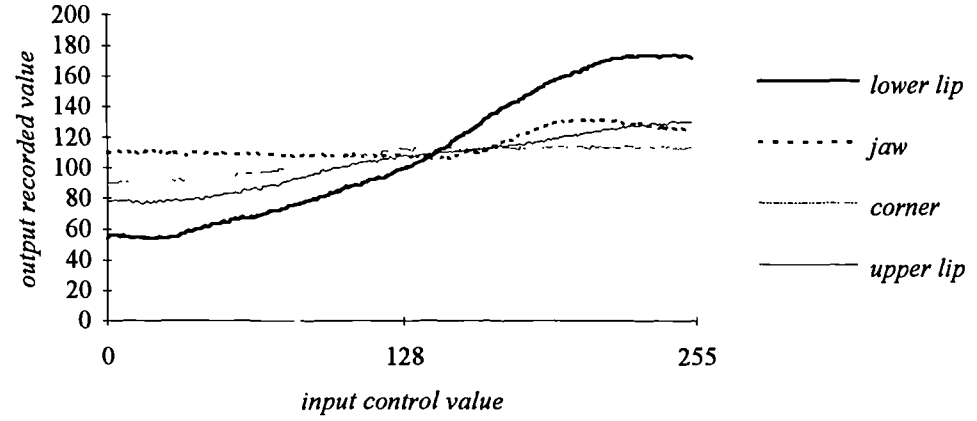


Figure 6.7 Plot Of Measured Characteristics Of The Lower Lip Centre

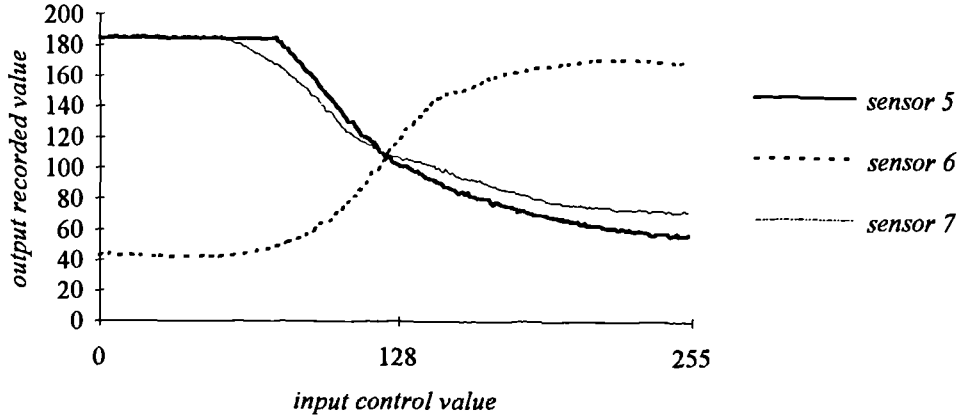


Figure 6.8 Plot Of Measured Characteristics Of The Corner

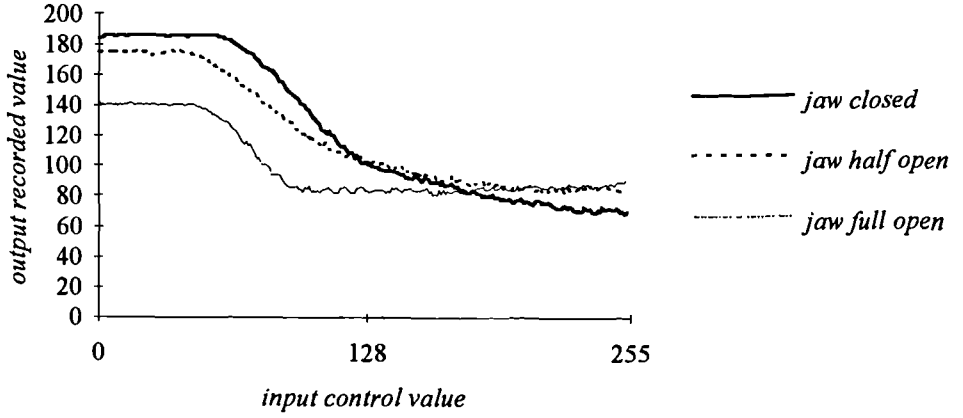


Figure 6.9 Resultant Differences In Corner Stretch For Variations In The Jaw Position

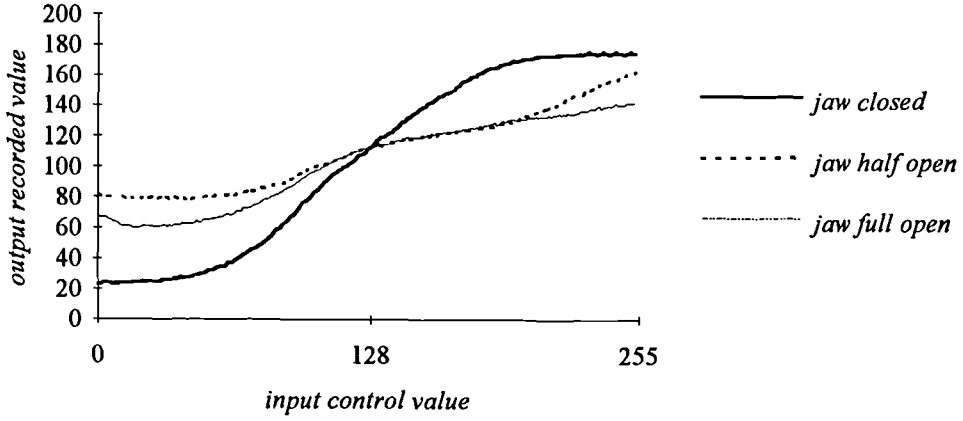


Figure 6.10 Resultant Differences In Corner Protrude For Variations In The Jaw Position

6.4 Experimental Analysis Of Facial Action Control System

6.4.1 Experimental Procedure

The results from Sections 6.2 and 6.3 indicated that the proposed sensing system had the capacity to consistently produce distinct signals from the defined key point displacements on the replica face, under controlled conditions. The following section presents the final analysis of the overall system for the production of facial animation through the automatic sensing of the performer's facial actions.

As stated in Section 4.2, the construction of the replica face enabled objective and subjective data to be generated for subsequent analysis. Objective data was achieved through the comparison of the individual key point displacements from both the live and replica faces. The visual analysis of the actions generated by the replica, with respect to the known actions of the live face, produced assessments on the ability of the system to produce visible point motion and of the overall animation.

A list of the stimulus data used in the analysis of the overall system is shown in Table 6.1. This set of tests represented a comprehensive investigation into the specific action units, emotions and visemes defined in Section 4.4.

The choice of test input was based on a number of factors. The investigation of the basic action units, as defined by Ekman (c.f. Section 3.4), was undertaken as they represented the fundamental elements necessary for all animated performances. The failure of the system to sense or animate these actions, in isolation, would reduce the capacity of their production in the final continuous performance. A similar argument was adopted in the use of viseme nonsense syllables. Each viseme grouping, Vowel-Consonant-Vowel (VCV) and Consonant-Vowel-Consonant (CVC), represents the minimal units required for visual perception. This technique is used in research to assess lip-reading performance **[Jackson88]**, **[Lesner88]**, **[Montgomery87]**. For the system to achieve continuous speech animation, it must be able to achieve these primary elements. The actual production of these syllables is the result of the complex combinations of the primary action units and they, therefore, represent suitable inputs to assess the system. The investigation of static emotional expressions was designed to evaluate the system's ability to produce different combinations of facial expressions which are visually perceptible as having some distinct meaning.

For objective analysis, the following procedure was undertaken. From the theory of Section 4.3, assumptions were made that the sensing system (F_{sensor}) and the physical relationships between sensor and reflector ($F_{physical}$), on both replica and live faces, were identical. To maintain this assumption and minimise any practical errors resulting from the differences in the construction of individual masks, a single sensor system was used. Provided the reflectors were attached to both faces at similar positions then the use of the same mask retained the sensor positions, defined from the procedures described in Section 6.2.

For the initial recording of the input actions, the system was placed on the live face and the series of tests listed in Table 6.1 were performed and recorded. Subjective opinions were taken at the recording to determine whether the input actions produced by the live face were a suitable representation of the desired facial action. This was an attempt to ensure that any variations in the final performance were the result of the overall system rather than an incorrect input. The input signals, $\underline{S_l}$, were recorded at 50 Hz using the software program RECORD. Input recordings were repeated a number of times on different occasions to ensure consistency in the technique.

For playback, the mask was then placed on the replica and using the program PLAYRECORD, the conditioned control signals, $\underline{c}$, were output to the replica. In parallel, the signals, $\underline{S_r}$, were recorded from the key point displacements of the replica. Using the same procedure as defined in Section 5.3, each test playback and record was repeated a sufficient number of times to generate a consistency of measurement.

The mean recorded signals, $\underline{S_r}$, were then compared objectively with the input signals $\underline{S_l}$ to generate statistical and graphical results. Subjective analysis was made on the replica playback with regard to the individual point changes and the overall perceptual change.

The objective analysis was developed to draw conclusions on the ability of the individual key points to produce displacements of comparable intensity at the correct rate and rhythm as the input actions from the live face. Intensity comparisons were derived from the graphical analysis of the magnitude differences, at any instant,

between the individual signal pairs. Graphical analysis also produced indicators for evaluation of the errors and inconsistencies of the individual channels. This analysis is described in Section 6.4.3. The application of standard time series analysis drew conclusions on the similarities of the individual pairs in terms of rate and rhythm. This is described in Section 6.4.2. Subjective analysis of the overall facial performance is discussed in Section 6.4.4.

6.4.2 Time Series Analysis Of Recorded Data

From initial graphical comparisons of the recorded data, it was clear that a time lag existed between the output and the input. The plots shown in Figure 6.11 illustrate this time difference. The data was recorded for the viseme test "/a/ /f/ /a/" at the upper and lower lip centres. The statistical function of cross-correlation was applied to evaluate the time lag and to also derive a measure of the similarity that exists in time and shape between the individual sets of input and output signals. The correlation function used in the analysis produces a Pearson Moment coefficient which is defined as a dimension-less index value that ranges from -1.0 to +1.0 inclusive and reflects the linear relationship between two data sets. It is independent of the differences that exist in the actual magnitude of the two signals **[Chatfield83]**.

Table 6.2 presents examples of the results derived from this investigation for each key point in each test. The first row shows the value of correlation measured at zero lag. Zero lag is defined as the true temporal relationship between input and output. The lag factor was derived by measurement of the cross correlation between the input and the output, shifted in time with respect to each other and is defined as the time shift required to produce maximum correlation between the two. Figure 6.12 shows an example of the cross-correlation between the input and output for the upper lip centre for test "/a/ /f/ /a/". This analysis was undertaken for all signals in all tests to assess the consistency of the lag factor firstly, between tests and secondly, between sensors. The following conclusions were drawn from this analysis.

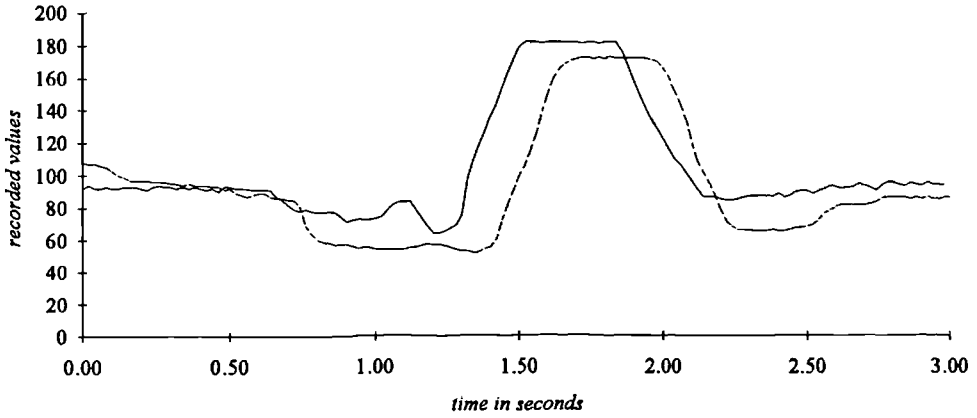
1. The measurement of cross-correlation at zero lag provided little indication of the similarities that existed.

2. The results displayed in Table 6.2 suggest that the measured lag was consistent throughout the experiments for the individual sensors, and for the overall system, in the range of 4 to 7 readings or 80 to 140 msec.

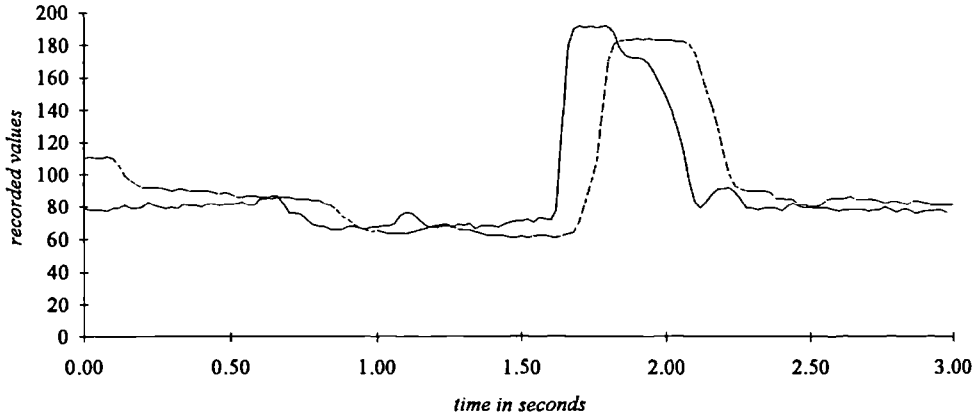
It was concluded that the primary reason for the lag was the dynamic delay present in the electro-mechanical conversion between the drive signal, $\underline{d}$ and final surface displacement on the replica's skin, $\underline{v}$. This was the combined result of the resistive effects of the skin and of the insufficient sampling frequency for the input signals.

Test Number	Test Stimulus
1	brow raise
2	brow lower
3	brows random
4	lips stretch, jaw closed
5	lips stretch, jaw open
6	lips protrude, jaw closed
7	lips protrude, jaw open
8	random jaw open
9	Primary Emotion : Happiness
10	Primary Emotion : Surprise
11	Primary Emotion : Anger
12	Primary Emotion : Sadness
13	Primary Emotion : Fear
14	VCV syllable : /oo/ /p/ /oo/
15	VCV syllable : /ar/ /p/ /ar/
16	VCV syllable : /ee/ /p/ /ee/
17	VCV syllable : /ar/ /f/ /ar/
18	VCV syllable : /ar/ /th/ /ar/
19	CVC syllable : /p/ /ee/ /p/
20	CVC syllable : /p/ /ar/ /p/
21	CVC syllable : /p/ /oo/ /p/
22	Combination Of brow raise with CVC syllable : /p/ /oo/ /p/
23	Combination Of brow lower with CVC syllable : /p/ /oo/ /p/
24	Combination Of brow lower with CVC syllable : /p/ /oo/ /p/

Table 6.1 Table of Input Actions Used In The Final Analysis Of The Overall System



Plot a) Upper Lip Centre



Plot b) Lower Lip Centre

Figure 6.11 Examples Of The Lag Present In Final Results

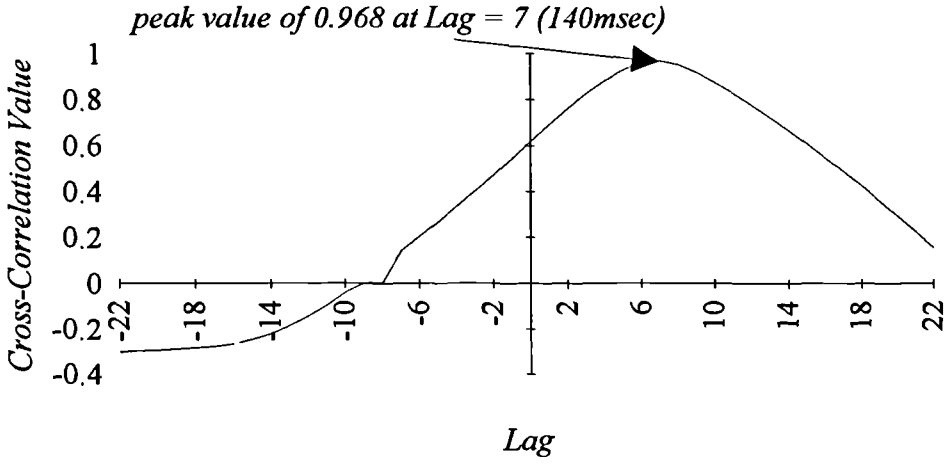
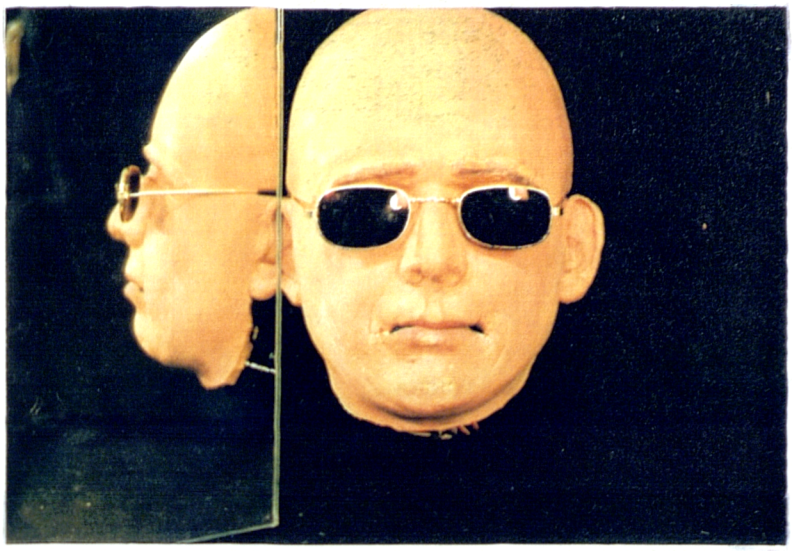


Figure 6.12 Example Of Derivation Of Time Lag For Upper Lip Centre During "/a//f//a/" Test

		s5	*s6*	*s7*	*s8*	*s9*	*s10*	*mean*
afa	*cross-correlation at zero lag*	0.623	0.696	0.617	0.769	0.618	0.841	0.694
	measured lag	6	6	6	7	7	5	6
	cross-correlation at measured lag	0.851	0.793	0.875	0.968	0.969	0.922	0.896
opo	*cross-correlation at zero lag*	0.209	0.254	0.574	0.676	0.700	0.697	0.518
	measured lag	7	7	7	5	5	5	6
	cross-correlation at measured lag	0.797	0.759	0.879	0.887	0.858	0.922	0.850
epe	*cross-correlation at zero lag*	0.795	0.790	0.691	0.805	0.272	0.711	0.677
	measured lag	5	5	6	5	4	5	5
	cross-correlation at measured lag	0.854	0.831	0.915	0.917	0.340	0.917	0.796
lippra	*cross-correlation at zero lag*	0.690	0.878	0.822	0.868	0.897	0.546	0.784
	measured lag	5	4	5	5	5	5	5
	cross-correlation at measured lag	0.751	0.927	0.874	0.913	0.946	0.750	0.860
mean	*cross-correlation at zero lag*	0.579	0.655	0.676	0.780	0.622	0.699	0.668
	measured lag	6	6	6	6	5	5	5
	cross-correlation at measured lag	0.813	0.828	0.886	0.921	0.778	0.878	0.851

Table 6.2 Table Of Lag Factor Results

Photograph a)



Photograph b)

Photograph c)

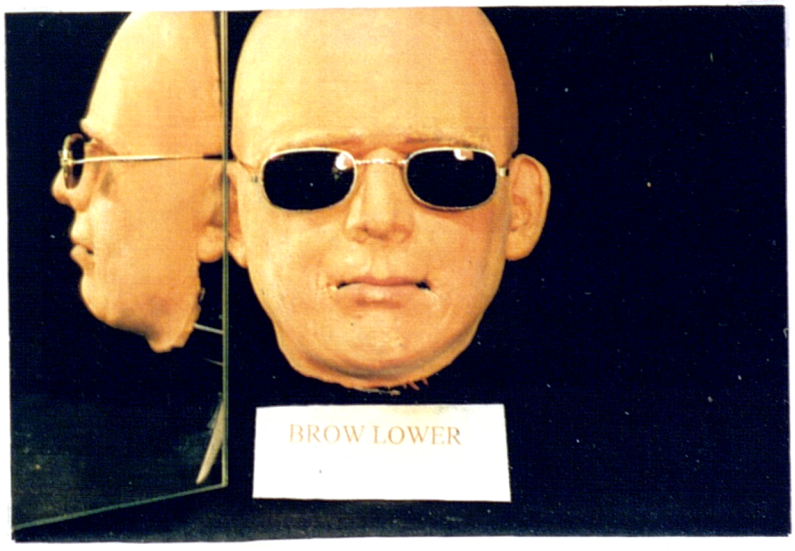


Figure 6.13 Photographs of Action Units On The Replica

Photograph a)

Photograph b)

Photograph c)

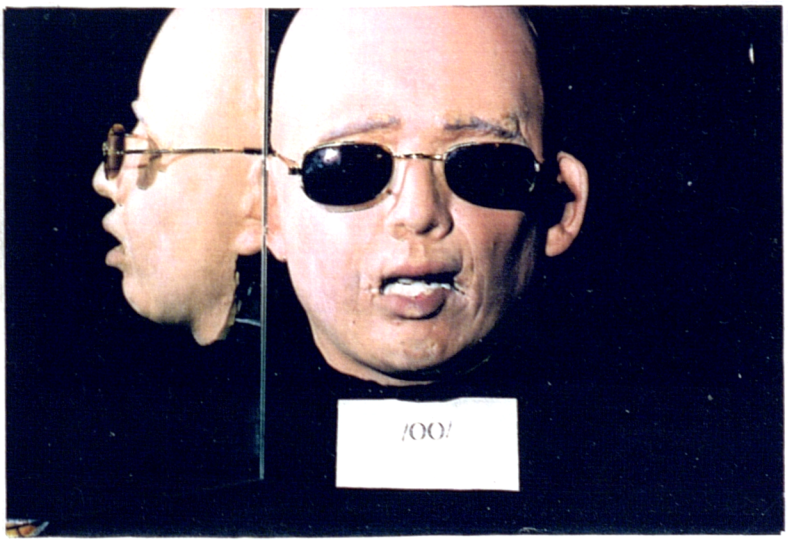


Figure 6.14Photographs Of Viseme Vowels On The Replica

Photograph a)

Photograph b)

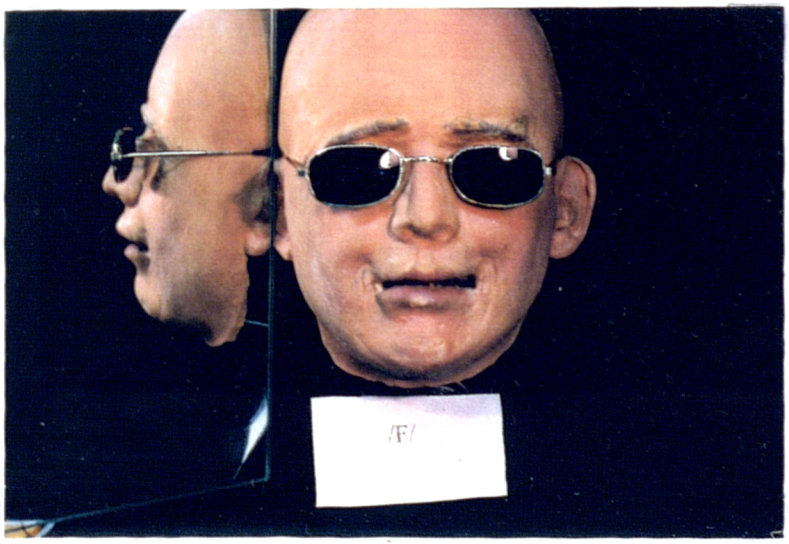
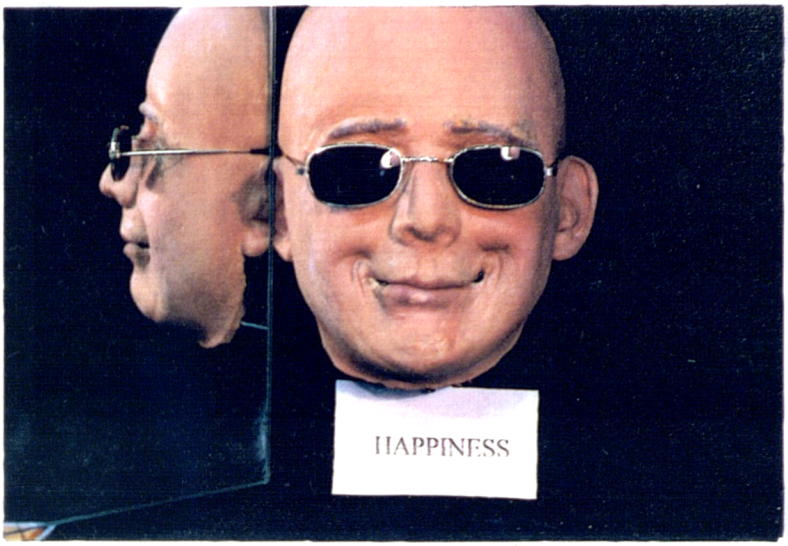


Figure 6.15Photographs Of Visemes Consonants On The Replica

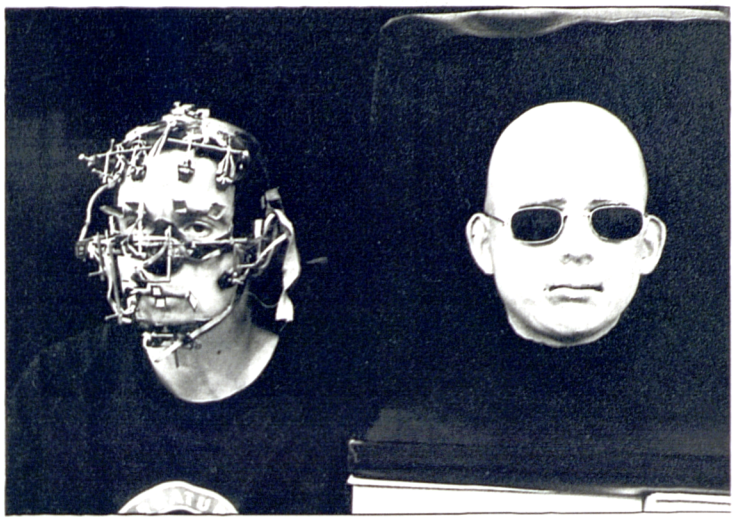
Photograph a)



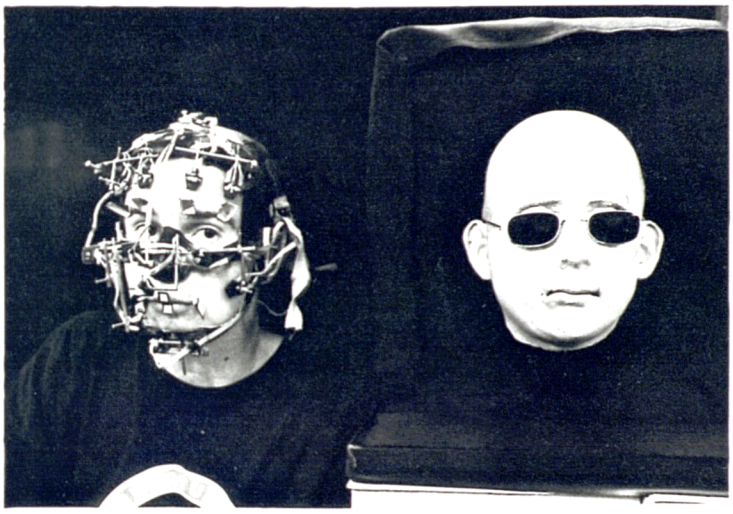
Photograph b)

Figure 6.16Photographs of Primary Emotions On The Replica

Photograph a)



Photograph b)



Photograph c)

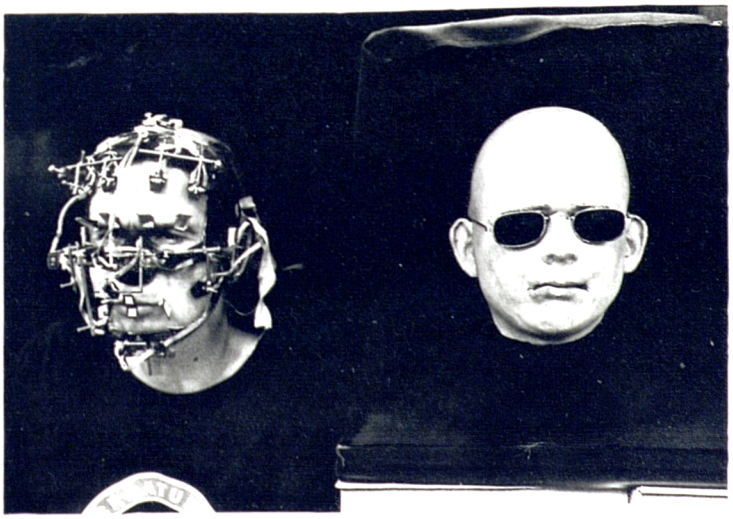
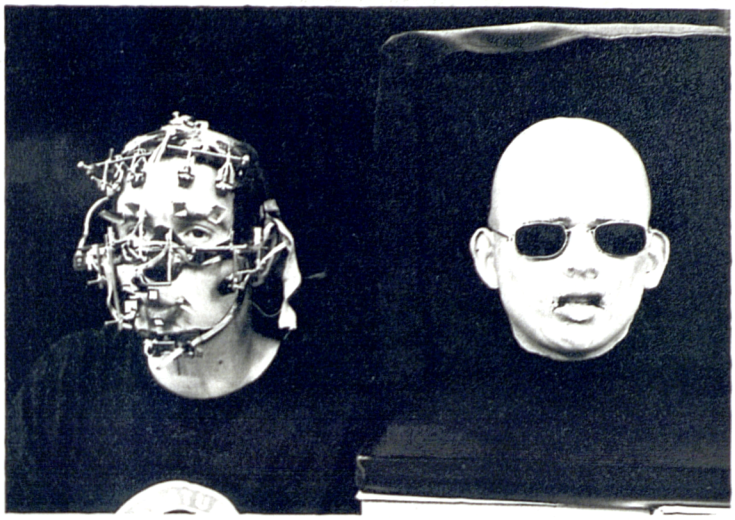
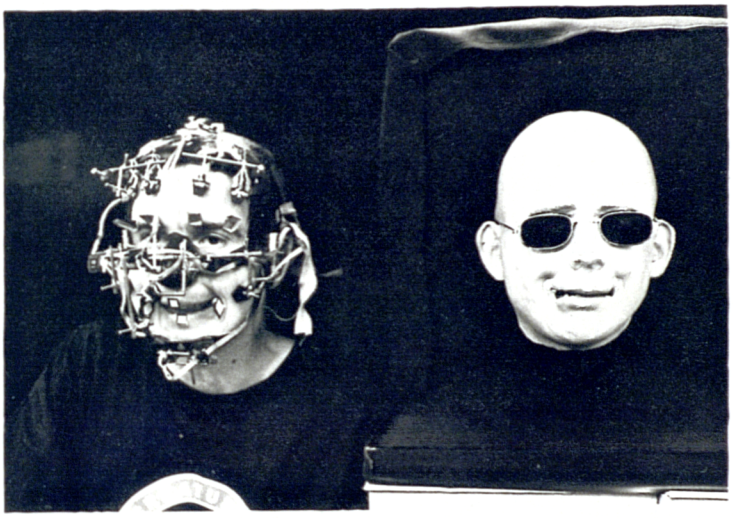


Figure 6.17Photographs of Live Control Of Action Units On The Replica

Photograph a)



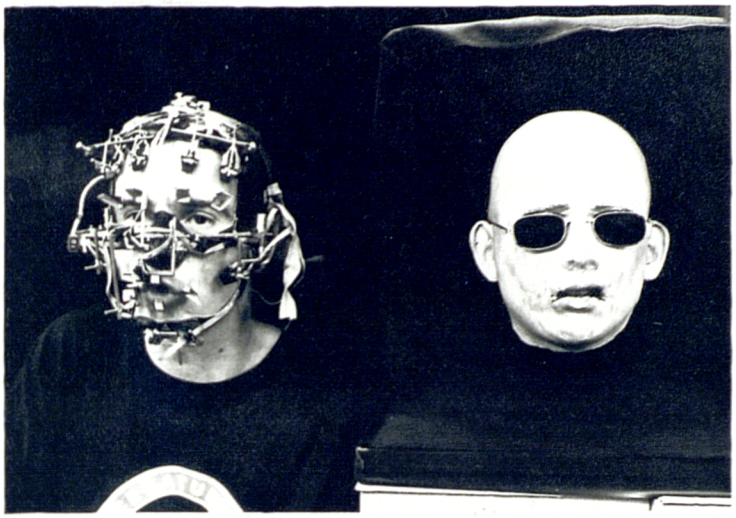
Photograph b)



Photograph c)

Figure 6.18Photographs Of Live Control Of Viseme Vowels On The Replica

Photograph a)



Photograph b)

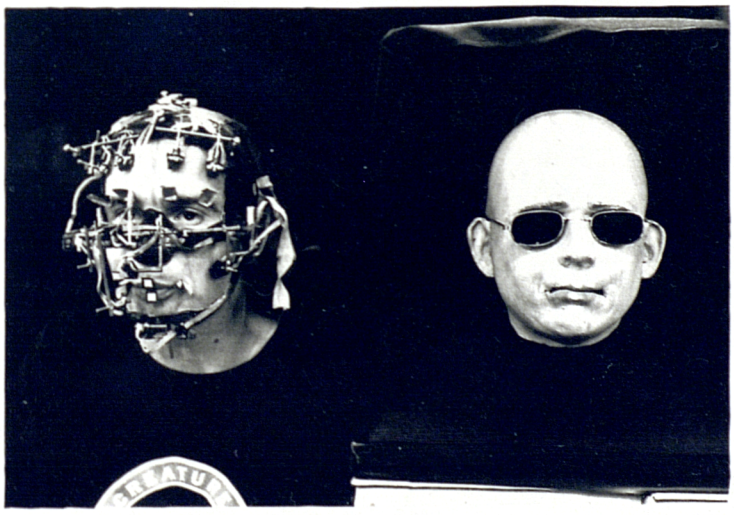
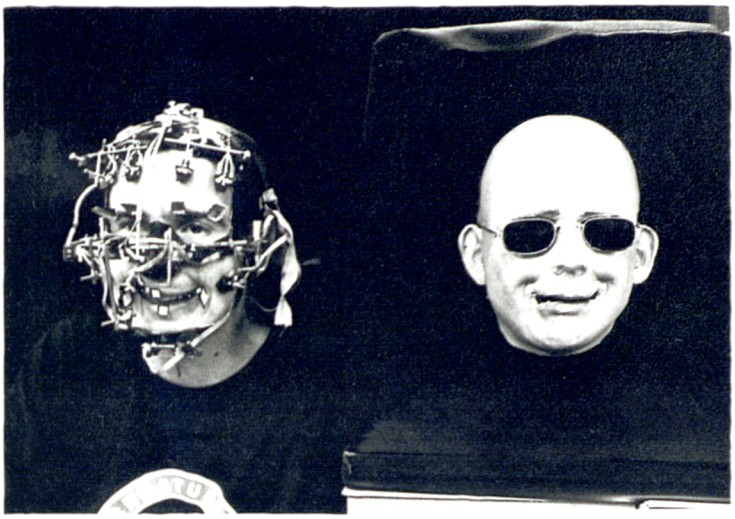


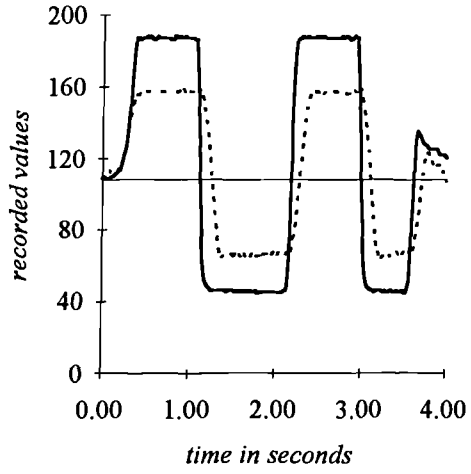
Figure 6.19Photographs Of Live Control Of Viseme Consonants On The Replica

Photograph a)

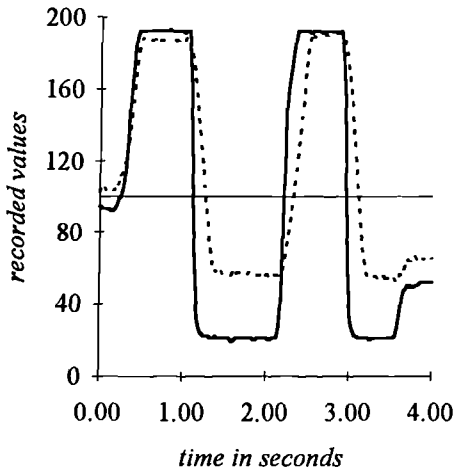


Photograph b)

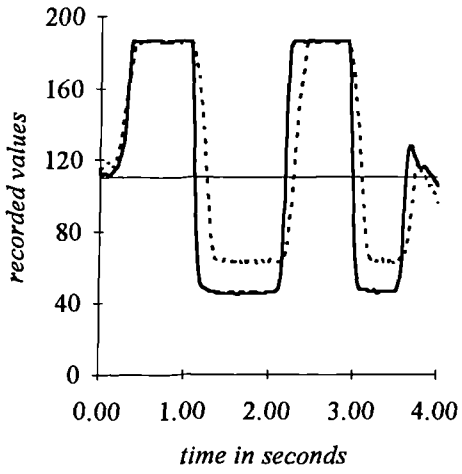
Figure 6.20 Photographs Of Live Control Of Emotions On The Replica



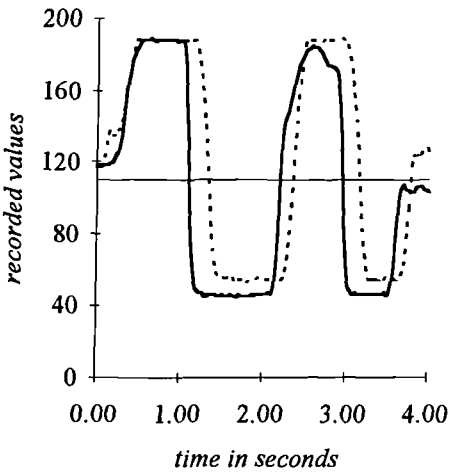
a) left outer brow



c) right outer brow



b) left inner brow



d) right inner brow

input output neutral

Figure 6.21 Plots Of "Brow Random" Test Results With Lag Compensation

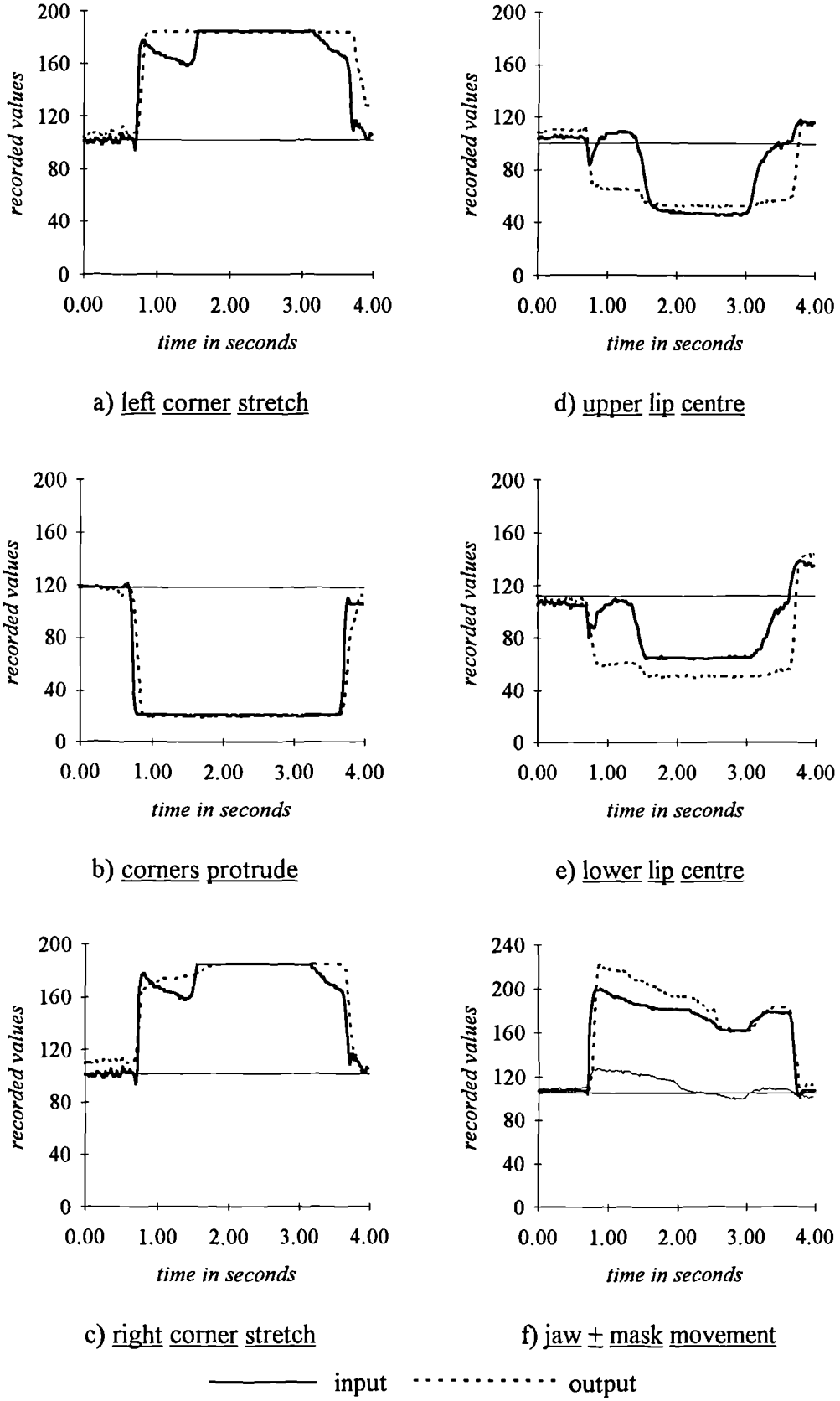


Figure 6.22 Plots Of Lips Stretch, Jaw Open Test Results With Lag Compensation

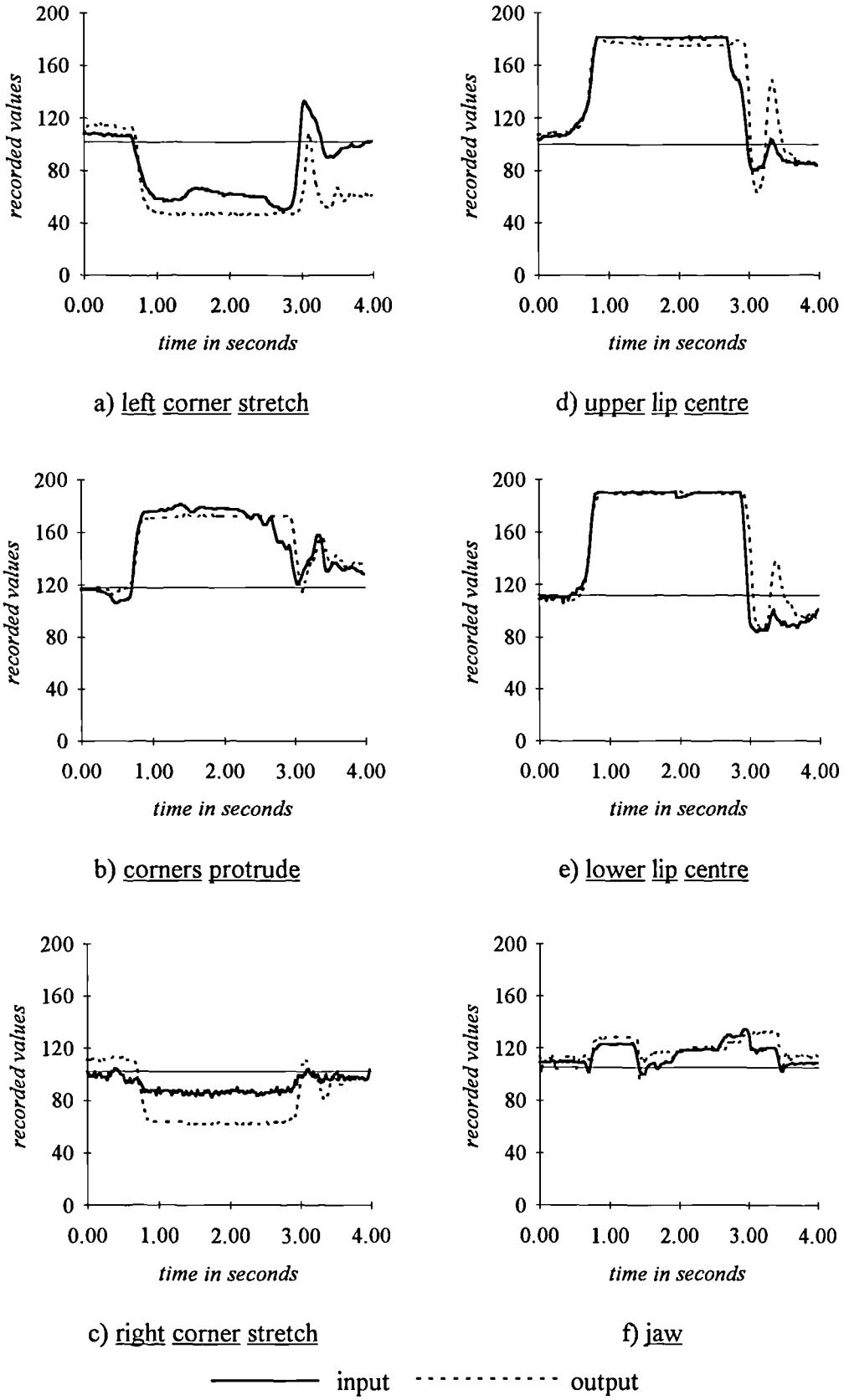


Figure 6.23 Plots Of Lips Protrude, Jaw Closed Test Results With Lag Compensation

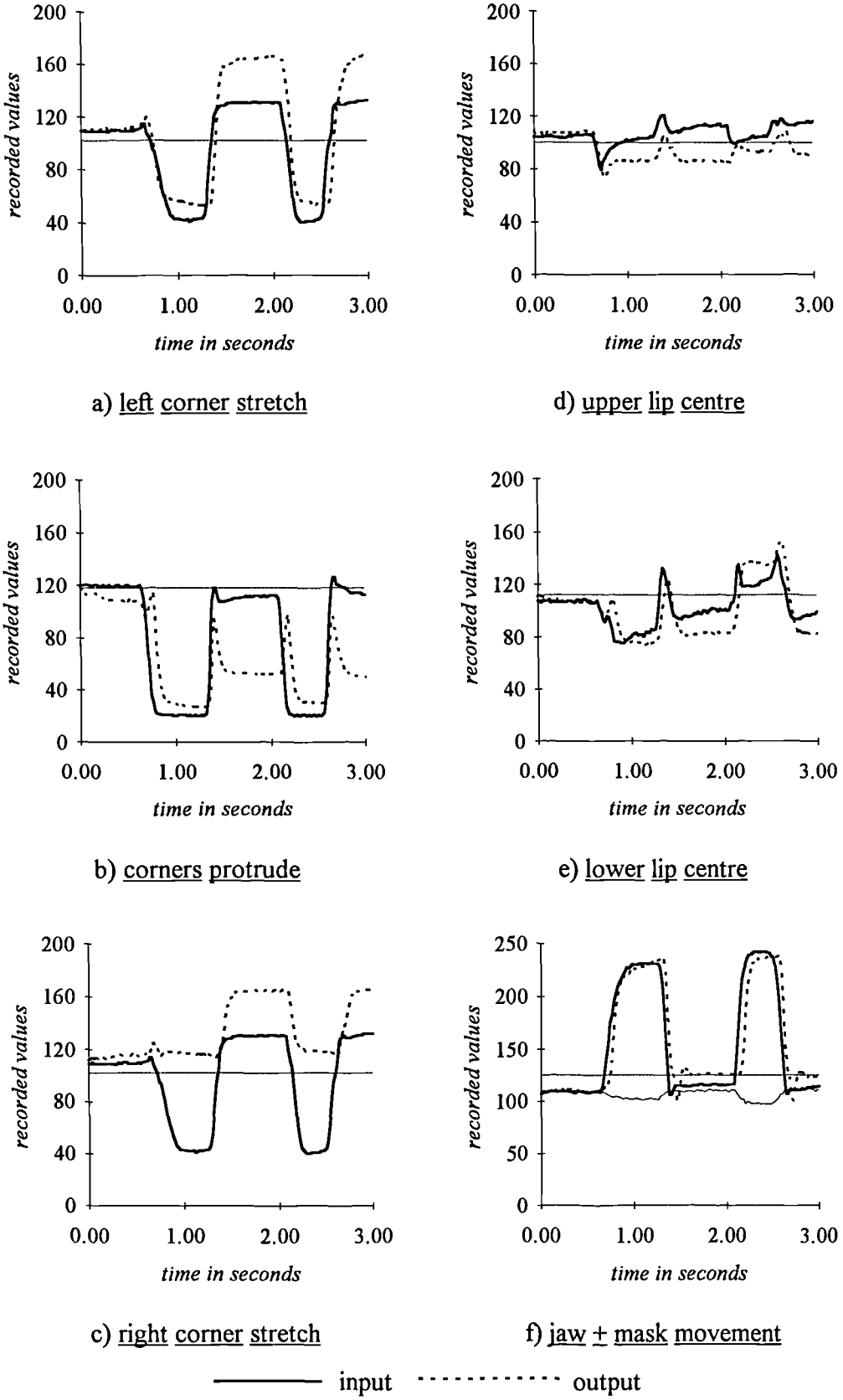
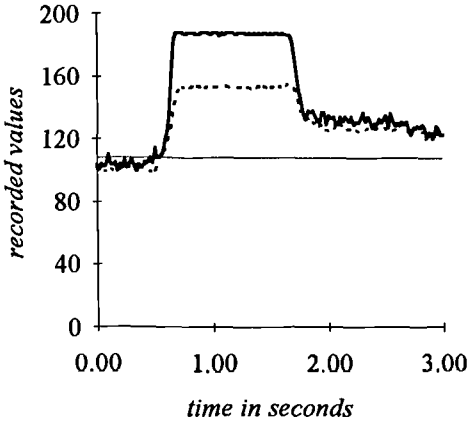
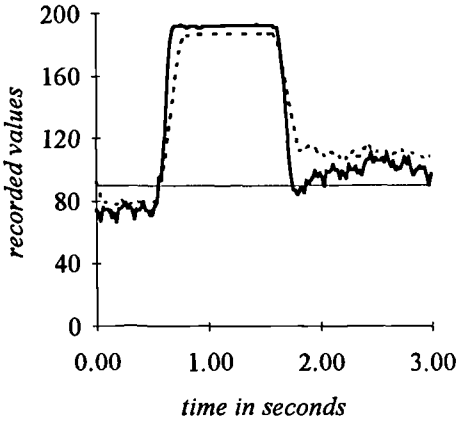


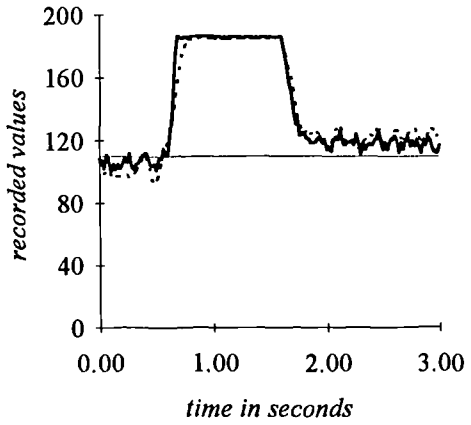
Figure 6.24 Plots Of "Random Jaw Open" Test Results With Lag Compensation



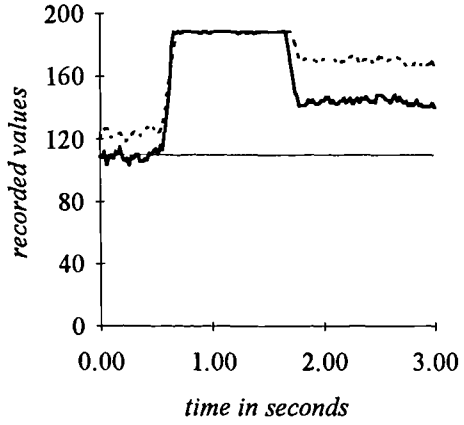
a) left outer brow



c) right outer brow



b) left inner brow



d) right inner brow

——— input ····· output

Figure 6.25 Plots Of Brow Actions In "Surprise" Test Results With Lag Compensation

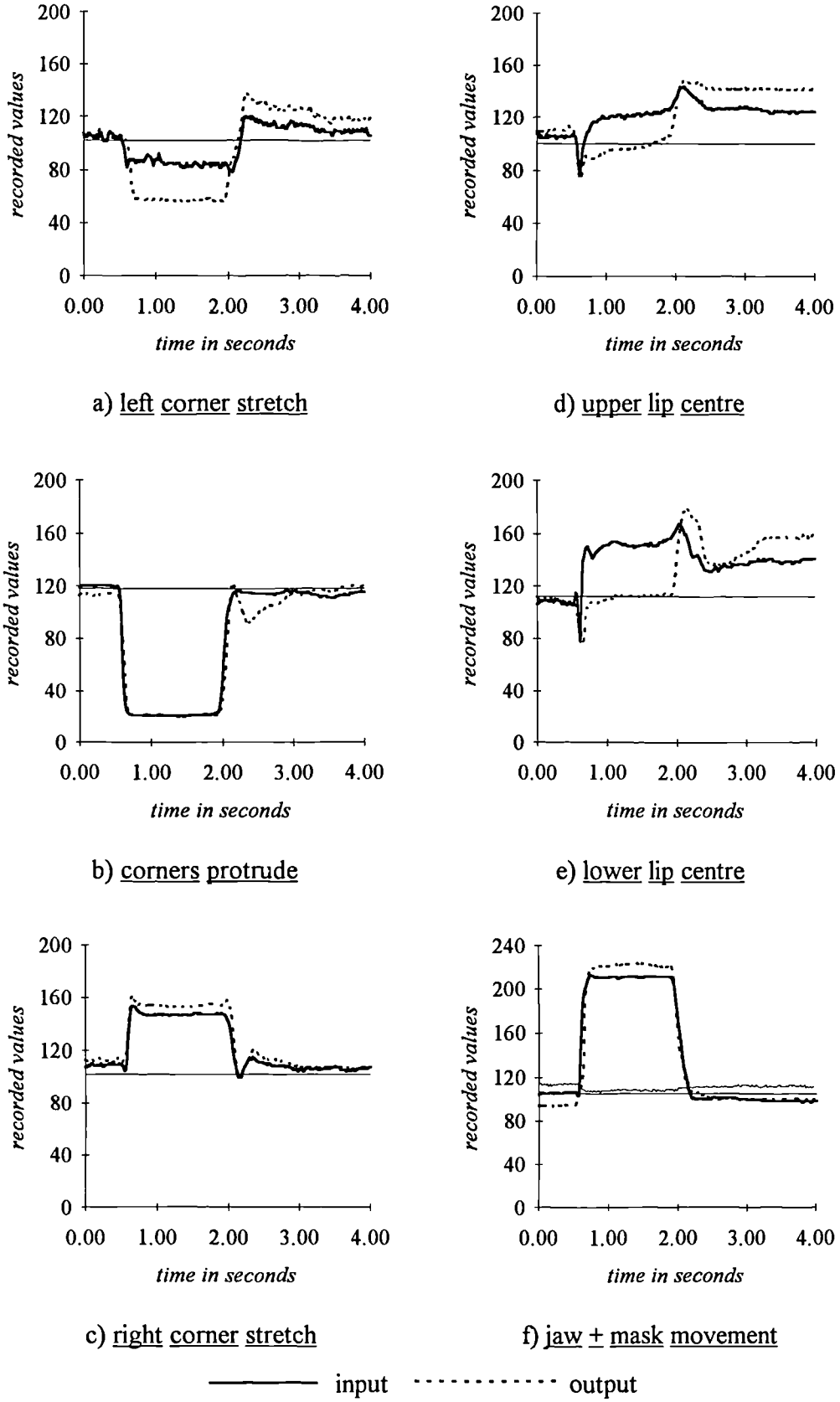


Figure 6.26 Plots Of Lip Actions in "Surprise" Test Results With Lag Compensation

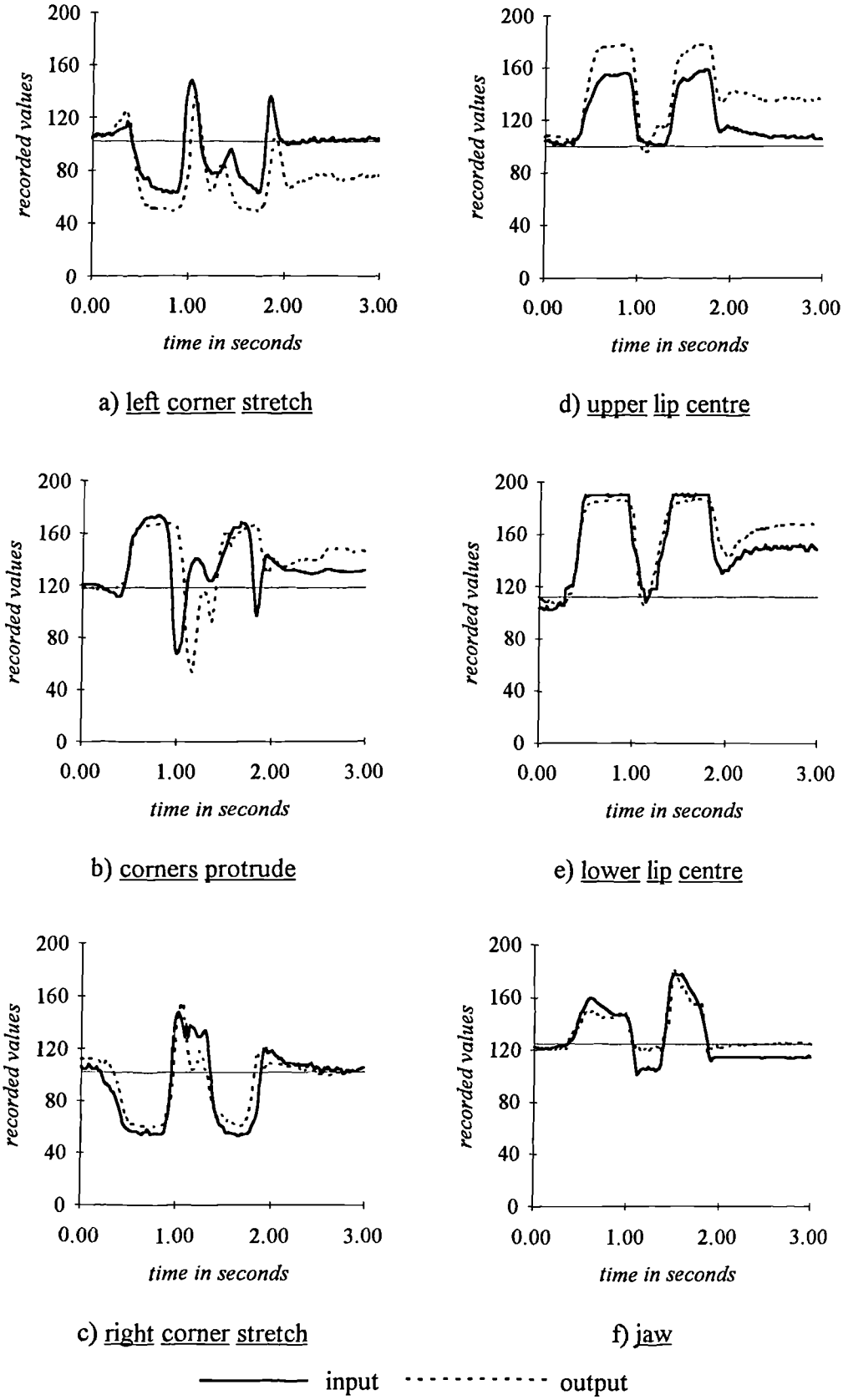


Figure 6.27 Plots Of "/OO/ /P/ /OO/" Test Results With Lag Compensation

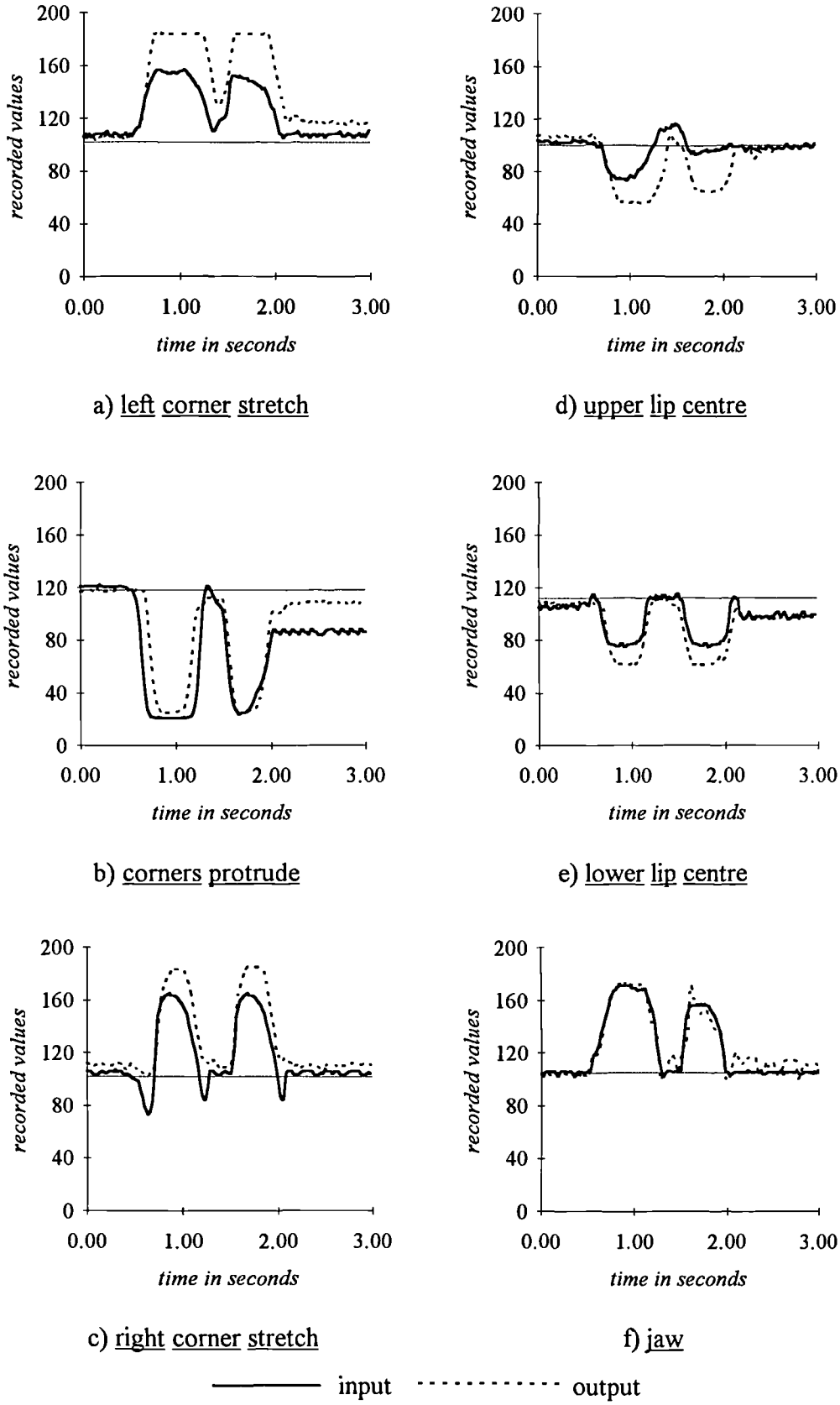


Figure 6.28 Plots Of "/A/ /P/ /A/" Test Results With Lag Compensation

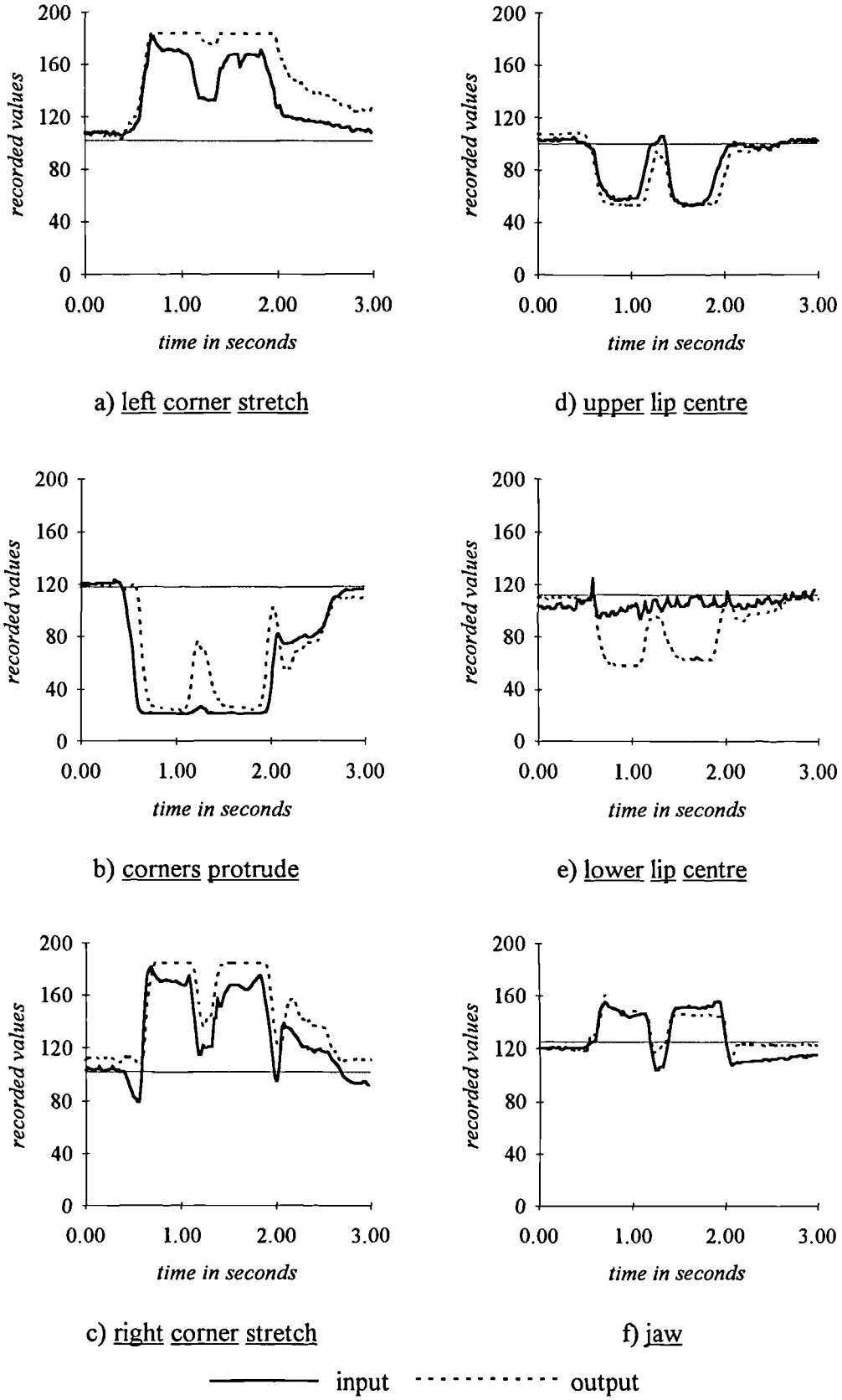
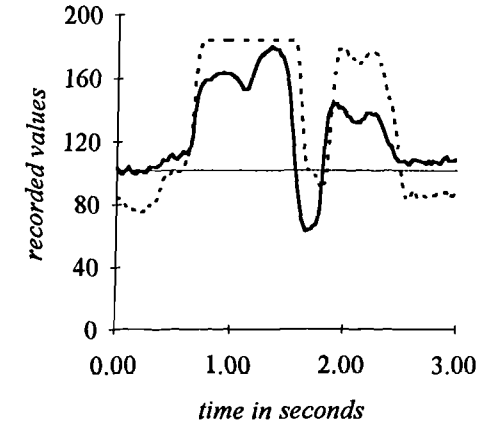
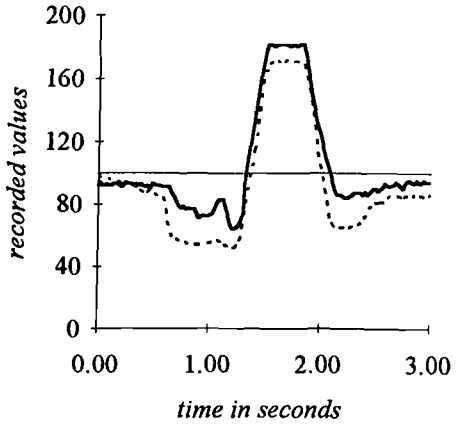


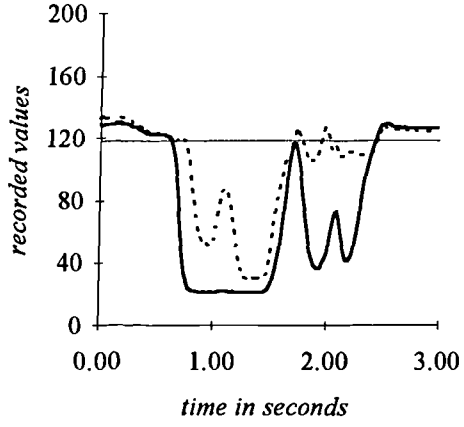
Figure 6.29 Plots Of "/EE/ /P/ /EE/" Test Results With Lag Compensation



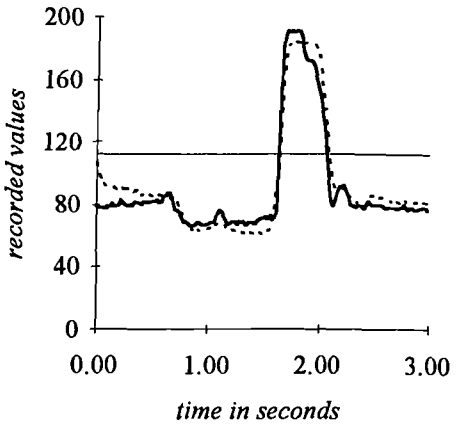
a) left corner stretch



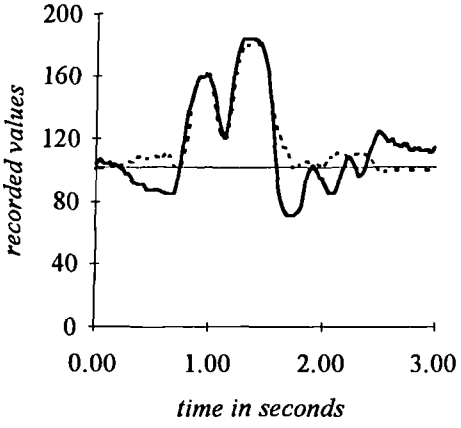
d) upper lip centre



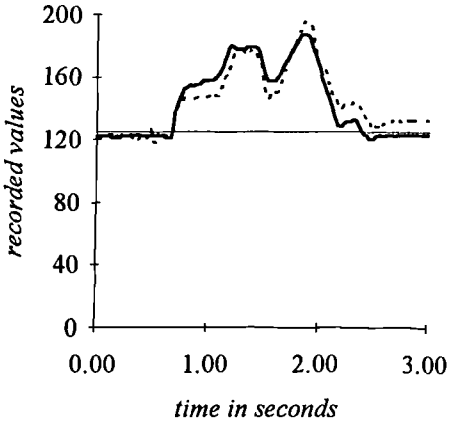
b) corners protrude



e) lower lip centre



c) right corner stretch



f) jaw

——— input ····· output

Figure 6.30 Plots Of "/A/ /F/ /A/" Test Results With Lag Compensation

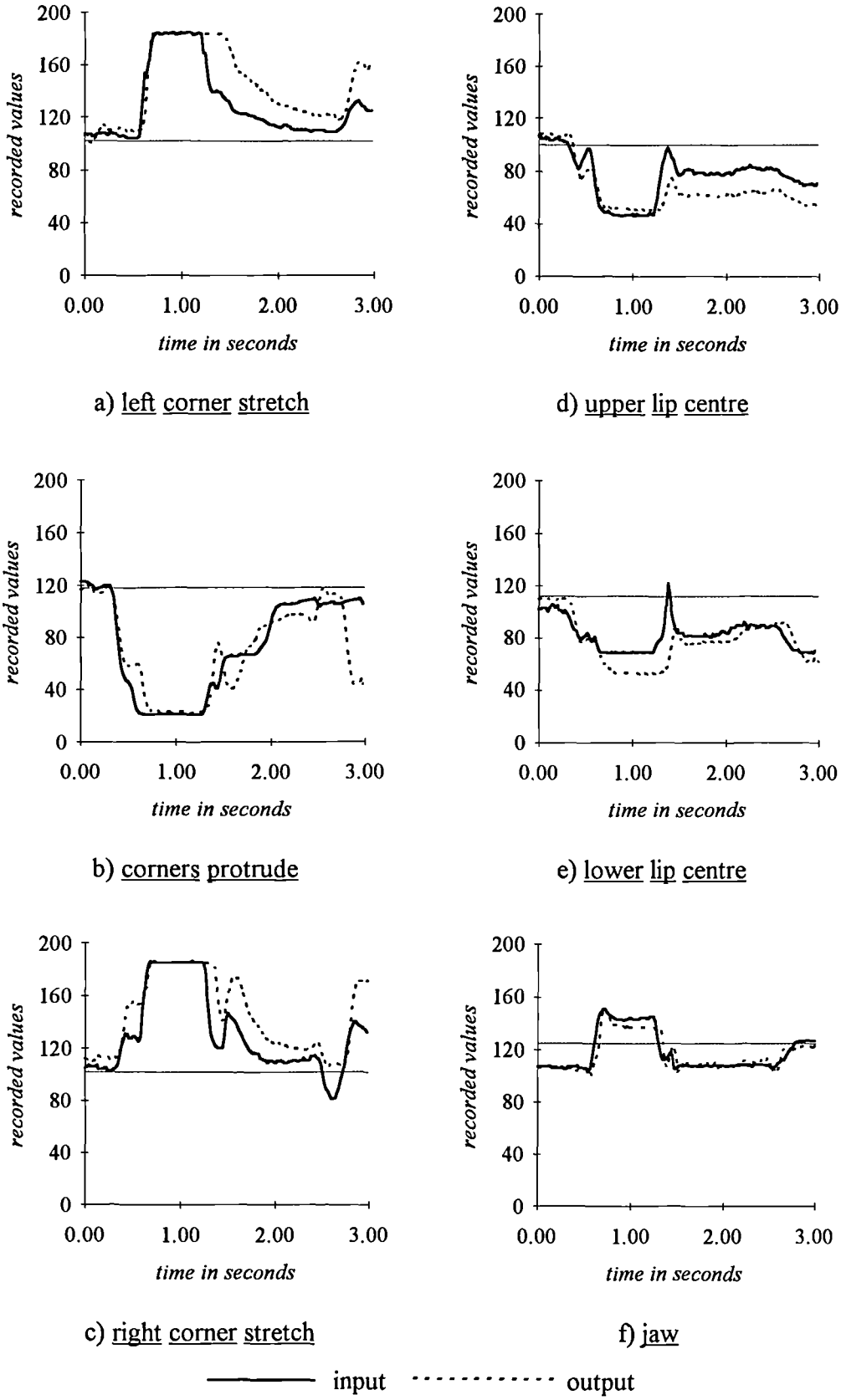


Figure 6.31 Plots Of "/P/ /EE/ /P/" Test Results With Lag Compensation

Test Number	Test Stimulus	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	mean
5	lips stretch, jaw open	0.850	0.922	0.898	0.645	0.707	0.907	0.822
6	lips protrude, jaw closed	0.751	0.927	0.874	0.913	0.946	0.750	0.860
8	random jaw open	0.906	0.624	0.654	0.148	0.797	0.927	0.676
10	Primary Emotion : Surprise	0.946	0.985	0.994	0.630	0.253	0.986	0.880
14	VCV syllable : /oo/ /p/ /oo/	0.797	0.759	0.879	0.887	0.858	0.922	0.850
15	VCV syllable : /ar/ /p/ /ar/	0.940	0.892	0.947	0.750	0.935	0.971	0.906
16	VCV syllable : /ee/ /p/ /ee/	0.854	0.831	0.915	0.917	0.340	0.917	0.796
17	VCV syllable : /ar/ /f/ /ar/	0.851	0.793	0.875	0.968	0.969	0.922	0.896
19	CVC syllable : /p/ /ee/ /p/	0.862	0.875	0.902	0.806	0.810	0.940	0.866

		Sensor 0	Sensor 1	Sensor 2	Sensor 3	mean
3	brows random	0.823	0.851	0.876	0.881	0.858
10	Primary Emotion : Surprise	0.966	0.980	0.924	0.972	0.961

Table 6.3 Table Of Cross-Correlation For Example Results At Corrected Lag

6.4.3 Graphical Analysis Of Experimental Data

This section presents examples of the graphical data recorded from the experimental analysis of the isolated action units, emotions and visemes listed in Table 6.1. Photographic examples of certain tests are shown in Figures 6.13 to 6.16. These were taken at the mid-point of articulation to present the most visually perceptible expression. The photographs in Figures 6.17 to 6.20 display the live face in real time control of the replica. The plots shown in the Figures 6.21 to 6.31 are all lag corrected by the factor measured in the previous section (c.f. Section 6.4.2). This was to improve the assessment of the individual point relationships. The output signals represent the mean variations produced from repeated playback tests using the bi-linear conditioning favoured by Hensons and described in Section 4.3. The measured neutral control parameter, S_{rN}, is depicted on the individual plots by the straight line. The data from these plots, along with the cross-correlation measurements between input and output at the corrected lag as shown in Table 6.3, enabled a wide range of evaluations to be drawn on the key point relationships and the performance of the individual elements of the system.

The overall conclusion, drawn from the graphical results presented at the start of this section, suggested that the key point sensor system was capable of the continuous extraction of control information from the actions of the performer's face at the same rhythm and of similar intensity to the perceived idea of the expected variations resulting from visual speech production. These predictions, along with ideas on what happens at each key point during the articulations, were also incorporated to the visual assessment at the time of recording the input actions and the subsequent subjective analysis. For example, the variations at K_{upper} and K_{lower} in the production of "/oo/ /p/ /oo/" (plots d] and e] in Figure 6.27) have measured peaks which correspond to the predicted variation resulting from the protrusion of the lips that accompanies the production of "oo". Refer to Section 4.4.1 for description of the other relationships between visemes and facial actions.

From the comparison between the input and output signals it was clear that the final animated displacements were produced at similar rates, rhythm and intensity to the input signals. Specific examples of the similarity in magnitude and rhythm include the following;

at $K_{corner(right)protrude}$ for test "lips protrude, jaw closed" (plot b] of Figure 6.23);

at $K_{corner(right)stretch}$ for test "/ar/ /p/ /ar/" (plot c] of Figure 6.28);

at K_{upper} for test "/ee/ /p/ /ee/" (plot d] of Figure 6.29);

at K_{lower}, for test "/ar/ /f/ /ar/" (plot e] of Figure 6.30); and

at K_{jaw} for test "/ar/ /p/ /ar/" (plot f] of Figure 6.28).

These conclusions of high similarity between input and output were confirmed by the cross-correlation measurements of Table 6.3. The measured correlations are, on the whole, of sufficiently high value to suggest significant similarities exist between the live and replica signals.

These results suggested that firstly, the sensor system was capable of sensing the majority of the primary actions, secondly, the overall control and drive system was capable of the production of comparable displacements at equivalent rates and thirdly, it was deduced that synchronisation would be possible with the acoustic signal.

Analysis of the plots indicated that although these overall conclusions hold true for most cases, a number of differences were visible, firstly, between the input signal and the perceived idea of variation and secondly, between the input and output signals. These inconsistencies prevent the present system from fully realising identical key point animation.

It was concluded that these inconsistencies were a combination of different variants in the sensing, drive and control systems. The difficulty of analysis lay in the determination of the exact source of the error. This difficulty was highlighted by the inconsistency of the errors firstly, at each point for individual tests and secondly, between the different sensors in a specific test. For example, at K_{upper}, the data recorded for the test "/ee/ /p/ /ee/" (plot d] in Figure 6.29) was highly correlated between input and output and the predicted variation yet the data recorded for test "/ar/ /p/ /ar/" (plot d] in Figure 6.28) or for test "/oo/ /p/ /oo/" (plot d] in Figure 6.27) show significant differences. Similarly, in the same test, certain signals show high correlation whilst the others may show significant fluctuations, for example the data in test "/ar/ /p/ /ar/" (in Figure 6.28).

From the results in Sections 5.3, 6.2 and 6.3 and the assessments drawn on the practical nature of the final system, the following problems were defined as the primary causes of the discrepancies in the final results.

6.4.3.1 Inconsistencies Due To The Sensor System

Within the sensing system, variations resulted from any of the following;

1. the magnitude of facial displacement exceeded the measurement range of the sensor;

2. changes occured in the relative position or orientation between the sensor and reflector; and

3. the system was reset at the incorrect datum.

The effect of the limited range of the sensor was clearly visible in the actions of the brows as shown in Figure 6.21. The action of the live brows clearly exceeded the range of the sensors and resulted in the "step-like" response. The consequence of this was a reduction in the ability of the sensors to produce graded measurements over the whole range. This also had the effect of producing the non-linear areas of the key point functions, previously discussed in Section 6.3.

The positional changes occurred as a result of either a change in the position of the sensor as a result of the support mask or as a result of a change in the orientation of reflectors attached to the facial surface.

The design requirements for the support mask were defined in Section 4.5. From the personal experience of the researcher and as shown in the photographs of Figure 4.14, the prototype mask was not ideal in its physical design. It was claustrophobic to wear, restrictive to certain facial actions, specifically for the actions of corner stretch and jaw open. The movement of the mask also produced differences in the measured signals of the sensors. The restriction of the jaw was due to the fact that the action of opening of the mouth is not purely one of rotation. At large openings this resulted in a vertical rather than rotational force being applied to the mask. This action was sufficient to displace the mask from its original position causing changes in the position of the sensors relative to the face resulting in incorrect changes in the control

signals. An example of this is shown in the variations at K_{upper} and K_{lower} for test "random jaw open" (plots d] and e] in Figure 6.26). The signal from S_{nose} was used to highlight any change in the position of the mask relative to the face by measuring any variation at the tip of the nose and is shown in plot f] for tests where its value was monitored. Another example of the resultant change in sensor position was seen in the resultant actions of the brows, which is visible from the video evidence V.8 and V.9, where the resultant motion of the sensors toward the reflector produces animation of brow raise.

Changes in the orientation between sensor and reflector in excess of ±15° about the defined focal axis would result in the variations to the measured signal (c.f. Section 5.3). It was concluded that the cause of the fluctuations in the measurement of the corner displacements were produced by the changes in the reflector's orientation relative to the sensor. This was due to the physical attachment to the skin . The skin, both latex and flesh, does not move along a plane axis but through a curved axis. The action of corner stretch results in the build up of fatty tissue at the large displacements causing the cheeks to bulge around the nasio-labial ridge on the face. Consequently, this area acts against the body of the reflector preventing its linear motion. An example of this effect is visible for the input fluctuations of test "/ee//p//ee/" (plots a], b] and c] of Figure 6.29).

The final source of variation was the effect of resetting the system at an incorrect datum, i.e. when the facial points are not at their neutral positions. This effect has been described in Section 5.3. An example of the resultant effect on the measured signal is visible at K_{lower} in the test "/ar/ /f/ /ar/" (plot e] of Figure 6.30) where the neutral value was significantly less that its expected value.

6.4.3.2 Inconsistencies Due To The Drive System

It was concluded that despite the best endeavours in the construction of the replica, physical differences existed between the two faces such that

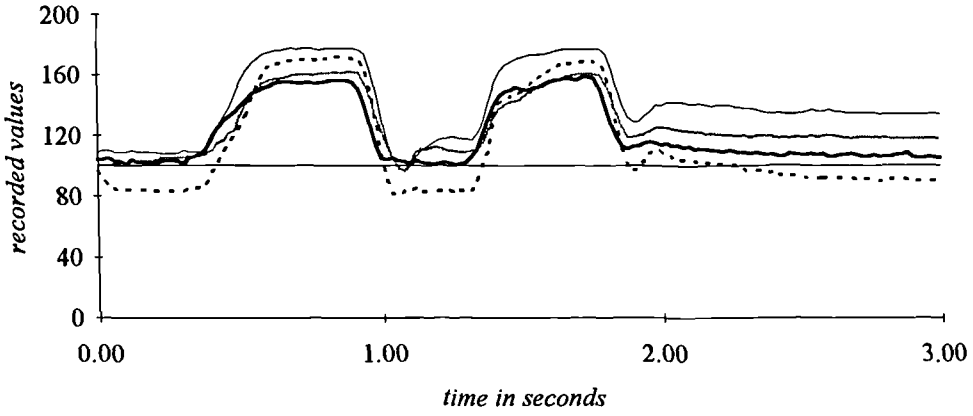
$\underline{F_{physical(live)}} \neq \underline{F_{physical(replica)}}$ thereby reducing the accuracy of the conditioning system. The effect of such differences was difficult to establish purely from the graphical data.

The data produced in Sections 6.2 and 6.3 showed the possible interactions between key points due to mechanical coupling via the skin or physical interconnection. This coupling was unavoidable due to the combined tasks required in deforming the lips. The principle argument to prevent these possible effects was dependent on the sensing system being capable of exact measurement of all live facial displacement and on the control system providing the correct conditioning. In the practical system this argument was no longer valid.

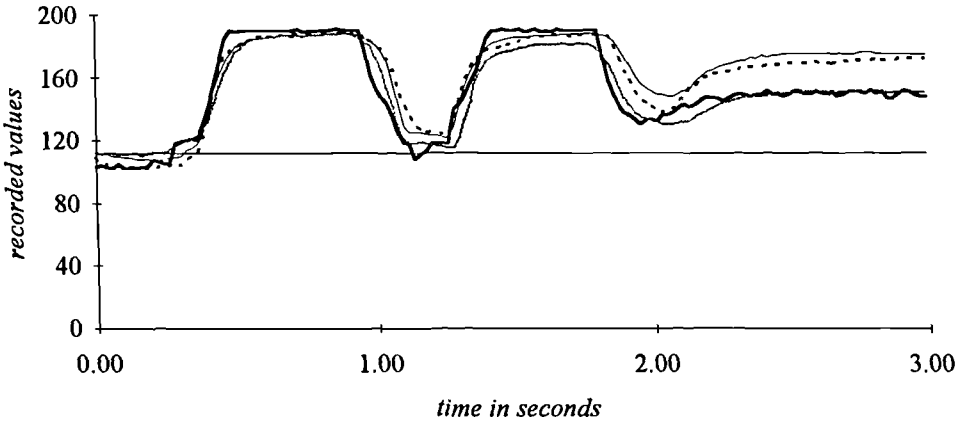
From the results there were a number of examples where this effect was obvious. The effects of the stretch actions of the corner on the centre of the lips were visible in the test "/ar/ /p/ /ar/" (plots d] and e] in Figure 6.28) where the input signals have measured only minimal variations yet the output signal indicated significant changes. Analysis of the corner stretch signals (plots a] and c]) confirmed similar differences in the output variations resulting from the poor conditioning of corner control signals.

6.4.3.3 Application of Conditioning Compensation To Reduce Discrepancies

From the measured characteristics, $\underline{F_{total}}$, in Sections 6.2 and 6.3, it was shown that non-linear regions exist which were likely to produce differences in the present results due to the inaccuracies of the bi-linear conditioning function. From the results, the actual effect was difficult to isolate clearly from the other variants. The example in Figure 6.32 shows the data produced for the test "/oo/ /p/ /oo/" at K_{upper} and K_{lower} for linear, bi-linear and non-linear compensation conditioning functions (c.f. Section 4.3.3). The results showed an improvement in the similarity of the measured displacements from the application of the look-up table. No clear conclusion could be made due to the difficulty in separating the errors of conditioning from the other factors. Time also prevented further investigation into the application of this method to all of the tests.



a) upper lip centre



b) lower lip centre

——— input ····· 2 part ——— non-linear ——— 1 part

Figure 6.32 Plots Of Different Conditioning Of "/OO/ /P/ /OO/" Test Results With Lag Compensation

6.4.3.4 Other Factors Likely To Produce Discrepancies

The following points were also defined during the investigation.

1. Other effects discussed in Section 6.2 and 6.3 included the effect of the jaw action on the K_{upper} and K_{lower} to produce fluctuations. Probable examples of these changes are shown in test "/ee//p//ee/" (plot e] of Figure 6.29) and in test "Surprise" (plot d] Figure 6.25).

2. The variations in the attachment of the jaw piece to the skin, which could move independently of the jaw, produced inconsistencies in the measured signal. These were considered as the main error to be avoided due to the importance of the jaw action in speech production.

3. An example of where it was possible to identify the differences between the perceived idea of signal variation and actual input was the signal at $K_{corner(right)stretch}$ in the test of "Surprise" (plot c] in Figure 6.25). The description of 'Surprise' in Section 3.4 and the researcher's visual assessment of his own action, see Figure 3.12, do not suggest any corner stretch. The input measurement indicated the opposite and this resulted in the reproduction of that action (similar effect is shown in the photograph of Fig 6.20). This was obviously the result of the measurement system but it does pose the question of whether it actually effects the overall meaning of the action.

4. The corner action measurements comprised of a significant number of discrepancies. The significant differences that can occur at the K_{corner} are the combination of some or all of the above effects;

a. variation in the orientation of the reflector due to the effects of the skin or the jaw action;
b. inability of the mechanical linkage to find an identical position each time for reset;
c. conditioning errors due to the displacement of the key points exceeding the range of the sensor;
d. differences in the trajectory of displacement for both faces; and

e. inability of the sensor to provide accurate readings outside its respective region of interest.

All of the above may be additive or act alone to produce the variations but in the final analysis, it is difficult to distinguish them.

It was concluded from the results of the graphical analysis that the majority of the variations in signal magnitude were in the region of 20 to 30 values which equated to displacements of 2 to 3 mm. This poses the question of whether these inconsistencies, evident in the graphical analysis, actually have any effect in the overall perception of the final animation. In summary, the variations of most concern were those which produced incorrect control information, i.e. errors in the sensor system. The proposed method of non-linear conditioning should, however, reduce the fluctuations resulting from poor conditioning and remove the effects of drive interaction.

6.4.4 Subjective Analysis Of Final Animation

The objective analysis of the previous sub-section, drew conclusions on the relationships that exist between the identical key points on the live and replica faces and also on the possible inter-relationships between individual points for specific tests. What was difficult to establish from the graphical analysis was an assessment of the overall facial changes resulting from the individual displacements and the effects of their temporal interactions and overlaps. The evaluation of the overall change was important as **[Petajan88a]** stated "the perception of visual speech and expression is not the recognition of individual point displacements but the overall change in the shape of the oral cavity and the face as a whole".

The subjective analysis used in this research was defined as the viewer's visual perception or understanding from the actions of the replica based on comparisons with either the live face, the accompanying soundtrack or a perceived idea of the action. For visemes, this idea was based on one's experience of visual speech.

Within this thesis, the analysis was restricted purely to the opinions of the researcher. To generate useful data from analysis by a series of independent viewers would

require the development of a thorough and wide-ranging procedure to ensure against biased results. It would also be essential to acquire analysis from a large number of viewers. Due to time restraints and the complexity of producing this type of analysis, the decision was made to restrict the analysis to purely the researcher. This was considered satisfactory given the knowledge and experience of the researcher on visual speech production. During the analysis, the researcher endeavoured to remain impartial in the acknowledgement of success or failure in the animation.

6.4.4.1 Analysis Of Experimental Data

This section considers the subjective analysis of the experimental data shown in video sequence V.5 and in the photographs of Figures 6.13 to 6.16. The video evidence contains examples of the primary action units, emotional expressions and the production of the major visemes without a soundtrack. The subjective analysis was applied to draw conclusions on firstly, the individual point changes and secondly on the overall facial change. With regard to the individual displacements, the researcher sought to assess whether the differences highlighted in the graphical analysis are firstly visible and secondly have any effect on the overall perception of the action.

Without the soundtrack, analysis of the overall actions was based on the researcher's knowledge of what the final animation should produce. The analysis of the primary action units, listed in Table 6.1, suggested that the system was capable of the production of recognisable static actions that were of a sufficient standard to match their description. The viseme analysis revealed that the majority of the actions were visually distinct. They were recognisable as being perceptually similar to the actions expected in the production of the primary visemes without any audible information. This type of analysis is used in lip-reading assessments to evaluate the amount of information perceptible without the audio channel **[Montgomery87]**.

As an example, the test "/ar/ /f/ /ar/" showed the final system's ability to sense and animate key point displacements at differing rates and magnitudes of motion and deal with the complex motion overlaps that occur in facial actions. The final animation was perceptually distinct and recognisable.

The animation of the primary expressive actions resulted in a number of conclusions being drawn. Firstly, the system was capable of the consistent production of the distinct emotions; happiness and surprise, and of their blending to produce "pleasant surprise". This showed that the system was capable of complex combinations of the primary actions. Secondly, the system failed to produce the distinct facial expressions that represent the other emotions such as fear or sadness. The research in Section 3.4 indicated that these expressions were primarily the result of actions in eye region, corner depressor and the nose wrinkler, all of which are not produced in the present system. Consequently, it was unlikely that the system should be capable of their production.

An important investigation was the assessment of any correlation between graphical inconsistencies and those from the perceptual analysis. An example where the graphical analysis was confirmed by the subjective evaluations of the test playback was shown in the test "/ar/ /f/ /ar/", at $K_{corner(left)}$, the stretch action appeared significantly more pronounced than expected yet it did not effect the recognition of the viseme.

Certain distinct fluctuations highlighted in the graphical analysis were found to have minimal effect on the overall visible animation. For example, the differences at K_{lower} for the tests "surprise" or "/ee/ /p/ /ee/" (plot e] in Figures 6.26 and 6.29) were not recognised by the researcher in the final animation. Alternatively, the errors produced at the corners in the tests "surprise" or "random jaw open" (plots a], b] and c] in Figures 6.26 and 6.24) were clearly visible in the final performance. Whether this affects the perceptual meaning by possibly conveying "pleasant surprise" is beyond the realms of this project.

6.4.4.2 Analysis Of The Primary Actions With Speech Signal And Live Face

The photographs of Figures 6.17 to 6.20 provided comparison between the static changes on the live face and on the replica. The video sequences V.7 and V.8 permitted subjective evaluations on the consistency of the system to produce the primary actions and visemes in synchronisation with an audible speech signal and also in direct comparison with the live face. These sequences confirmed the conclusions

derived in the experimental analysis. The majority of the animated actions were produced in synchronisation with the acoustic signal and are comparable with the live face for gross facial actions.

Specific examples of what the researcher would perceive as realistic animation in terms of similarity between visible actions of replica and the speech sound, or the live facial actions or the perceived idea are : "/ar//f//ar/", "/p//ar//p/", "/p//oo//p/", "/ar//p//ar/" and "/oo//p//oo/".

Examples of "poor" animation where it was difficult to derive any meaning or where the actions do not match those of the speaker or those expected to accompany the spoken syllables were "/ee//p/ee/" and "/p//ee//p/". The poor animation is the result of incorrect production of corner stretch actions for the visemes "/ee/", "/s/" and "/r/". This effect was also the result of the inability to see the teeth during production and also due to the inability of the corner system to produce graded actions.

The video sequence V.8 showed the differences that were inherent in the creation of a replica face when compared directly with the live face. The advantage of this type of analysis was the ability to confirm that the "mirror" correspondence exists, to a recognisable level. It also allowed comparison of the synchronous actions to produce an improved assessment of the possible causes of the discrepancies. This comparison was particularly beneficial in visualising the time lag between the input and output actions. A specific example can be seen in the brow actions, Video V.8, where the replica motion appeared almost 90° out of synchronisation from the live input.

The disadvantage of this type of analysis was the fact that the comparison emphasised the inaccuracies and inconsistencies in the construction of the replica which ultimately reduced the overall effect of the final performance. This is discussed further in Chapter 7.

6.4.4.3 Subjective Analysis Of Continuous Speech

The actual analysis of lip synchronisation represents a difficult problem. As **[Lewis91]** stated there is no adequate definition for the production of 'good' or realistic lip synchronisation. When it is produced correctly, it is automatically

perceived and accepted as life-like. As a result, analysis can only recognise the differences between our audible perception of "what is heard" and the visual perception of "what is seen". The following arguments have attempted to produce analysis of the final performance though it is accepted that a certain amount of bias may exist in the conclusions.

Using the test sentence of "the good, the bad and the ugly", the video sequence V.9 illustrates the ability of the system to produce lip synchronisation of a quality such that over 50% of the syllables were considered perceptually correct. The repeated performances of the sentence indicated that the system was capable of not only consistent reproduction of the correct syllables and words but that such reproduction could also be achieved at different rates and rhythms dependent on the performer's actions. The changing lip movements were considered to be of comparable intensity and duration to match the actions of the performer and the audible words.

From this it was concluded that the system has the possibility to automatically produce a wider variety of performance compared with the animation produced by other control techniques.

The delay highlighted in the time series analysis of Section 6.4.2 is difficult to distinguish in most sequences but as the rate of speech production was increased so the discrepancies in the time domain became more apparent and reduced the realism of the final animation. It was concluded that the optimum rate for speech production, for the present system, was at a slower rate than normal speech. This correlates with the question of lip-reading speed and also with the processes of acting. This question is considered in the final chapter.

The sequences V.9 and V.10 show what were considered to be improved performances from the replica. Whilst the experimental data of individual visemes was essential for the graphical assessment of the system, the actual production and hence animation of continuous speech in these clips confirmed that continuous speech is not merely a sequence of isolated segments. The visemes' lip shapes do appear within the animation but their context, rate and intensity vary from each production as a result of the effects of co-articulation. The final sequences therefore, highlight the partial success in the implementation of continuous speech animation. The final system overcomes this problem automatically to produce the correct contextual

changes in the animation. The reason for the lack of highly recognisable co-articulation lies firstly in the inability to animate the more subtle forms of lip shape, such as pursed lips or lips together, and secondly the present practical limitations in the sensor system, highlighted in the previous section, preventing a full description of all the lip changes. This argument is discussed further in Chapter 7.

6.4.4.4 Overall Discussion From Subjective Analysis

From the subjective analysis, the following points were made.

The primary articulatory action was that of jaw rotation. It plays a vital role in the production of most of the visemes and hence on the visual speech perception. The success or failure of the lip synchronisation sequences was largely dependent on the displacement of the jaw. This is confirmed by the theory of speech production in Sections 3.2 and 3.3.

The video sequences V.5, V.7 and V.8 all suggest that the overall design principle of recombination of individual drive displacements to produce the same overall change as the input, was broadly successful given the practical limitations of the present drive system. The animated sequences also indicate that the majority of complex facial actions can be reduced to a limited set of visible displacements and still be visually distinct.

In the perception of visemes or the basic emotions, certain individual actions play the major role in their correct recognition. For example, happy is easily recognisable by corner stretch, surprise by the combination of brows raise and jaw drop and the viseme "oo" by the protrusion action. This was also true for the animated sequences presented in this section. The successful production of the viseme "f" indicates that the system also has the capacity to deal with the multiple action overlaps that occur in its animation.

The final performance of the alphabet, the numbers and the different sentences in video sequence V.10, indicated that the system was also capable of producing other visemes, apart from the primary set. This suggests that the assumption made in section 4.4 that the majority of lip actions are similar to the primary actions but either

to a lesser degree of visible change or through some different combination is true. The video sequence of the alphabet also confirms the visual similarities in the groupings of phonemes.

6.5 Summary

This chapter has described the results and analysis of the final system developed to enhance performance control. It has presented examples from the results produced in an extensive experimental analysis designed to assess all aspects of the system. Conclusions have been drawn on the hypothesis of facial control, the key point design principle, the proposed optical sensing system and the performance of the animatronic face.

The results of Sections 6.2 and 6.3 established, firstly, that the replica was capable of the production of visually distinct actions of comparable intensity and orientation to those produced by the live face. Secondly, the results demonstrated that the proposed sensing system was capable of the production of consistent measurement from the defined key point displacements in experimental conditions.

The final analysis in Section 6.4 presented objective and subjective results from the investigation of human facial control of animatronic performances. The objective analysis was successful in evaluating a number of conclusions. Firstly, the analysis confirmed the existence and measured the time delay inherent within the overall performance system. Secondly, the graphical analysis of the recorded data allowed evaluations to be drawn on the individual key point relationships in terms of magnitude, rate and rhythm. Finally, conclusions were drawn on the probable cause for the discrepancies visible in the graphical data.

From this analysis, the conclusion was drawn that practical realisation of the sensing, control and drive systems, based on the principle of distinct key points, was broadly successful in the production of individual displacements at the replica. These displacements were of comparable intensity and at similar rate and rhythm to the measured input actions from the live face. With reference to the functional model of

Section 4.3, the results suggest that $\underline{S_r} = \underline{S_l}$ and therefore $\underline{v} = \underline{u}$ for certain primary actions.

The subjective analysis of the video sequences V.5 to V.10 and the photographs in Figures 6.13 to 6.20 drew the conclusion that the hypothesis for facial control through a limited set of data was, for the majority of primary actions, successful in the production of realistic facial animation. The evaluation showed that the recombination of the individual displacements on the replica face was satisfactory for the overall animation of the primary actions. This confirmed the validity of the hypothesis of distinct key point relationships as a design principle for all elements of the system. The system was capable of the consistent production of the visually distinct primary action units in isolation and in different contexts within a sequence of spoken words.

The visual analysis of the final performance sequences indicated that the final system displayed the potential to achieve realistic animation of the actions of the lips in synchronisation with an audible speech signal. A sufficient proportion of the lips' actions were considered as perceptually correct in terms of comparable intensity, rate and rhythm.

From all of the evaluations, the conclusion was drawn that the overall design principle of key point reduction of the facial system, had the potential to enhance performance control of Animatronic characters.

The different analysis techniques enabled a comprehensive evaluation to be drawn on all elements of the proposed system that was unobtainable from the individual results. A number of points can be drawn on the techniques of evaluation. The results of the graphical analysis can only produce indications of possible failures in the final animation. The individual discrepancies in the graphical data do not necessarily produce an incorrect or poor animated performance. For example, the test "/oo/ /p/ /oo/" in Figure 6.27, Video sequence V.5 and the photographs in Figures 6.14 and 6.18. The graphical analysis indicates a difference at K_{upper} yet subjectively there is no perceptual difference. Similarly, a perceptually poor animated action does not necessarily yield poor graphical results. For example, the test "/ee/ /p/ /ee/" in Figure 6.29.

Finally, a number of conclusions were drawn on the present limitations of the practical system.

1. The sensor support mask produced a number of errors as a result of its inability to allow, firstly, unrestricted facial movement and, secondly, the consistent placement of the sensors and reflectors in the identical positions to retain the same transfer function between action and control signal ($F_{physical}$). The restriction of the jaw action produced errors at the other sensors through the relative change in the position of the mask. Variations in the measurement of the lower lip are caused by the inability of the mask to position the sensor in the same relative position for every recording. The overall principle for a head mounted system is still considered the preferred option but an improved design should be developed to produce a system that is speaker independent, not restrictive to the speaker, easily removed or replaced and yet still retains the positions of the sensors irrespective of global head actions.

2. The delay factor has an adverse effect on the present system for average to fast rates of speech by the performer. This is related to physical effects present in the drive system and on the insufficient sampling rate used at present.

3. The present sensor system has three specific restrictions: it has a limited range of measurement resulting in the inability to measure all changes; the use of infra-red sensors results in an inability to visualise the exact direction of sensing leading to possible incorrect positioning relative to the facial reflectors; and the limited field of view of the sensor makes it susceptible to fluctuations from motion across and about the axis.

4. Another problem of the present system is the physical attachment of reflectors to the facial surface. The assumption that the point displacement can be transposed to an axis away from the face only holds for certain motion. At maximum displacements, the skin acts in various ways to change the relationship between reflective surface and actual facial point. These attachments also produce physical discomfort to the performer after long periods of wearing. The problem of attachment exists in all techniques based on the analysis of specific points. **[Fromkin64]**, **[Brooke83]**, **[Finn88]** and **[Himer91]** all produced measurement using different types of attachment to the

skin. The only techniques that have the ability to measure facial changes without attachment are based on complex image processing techniques (**[Petajan88b]**, **[Mase89]**). The problem of attachment remains one for which there is no solution at present and further work must be undertaken to minimise its effects if a final system is to be realised.

5. Within the replica system, discrepancies result from limitations in the conformation and appearance of the face and in the mechanical production of the facial displacements. The conformation of the skin and the production of tears in the skin at the corners of the mouth have significant effect on the final perception of the visual actions. The design of the mechanical linkages fails to produce displacements of identical magnitude and along the correct trajectories.

6. The present bi-linear mapping principle has been largely successful in the realisation of the final animation. However as the system is developed, reductions should occur in the practical limitations, resulting in the need for a greater degree of accuracy in the mapping system. The initial results presented in Section 6.4.3 suggest that the non-linear compensation techniques proposed in Section 4.3.3 have the possibility to produce an improved relationship between input and output. At present the compensation technique fails to produce any perceptual improvement as shown in the video sequence V.6. Further investigation would be necessary on its capacity to improve the overall performance.

7. The results in Section 6.2 and 6.3 showed that interactions occur at the key points on the replica. An example of this is the significant effects produced by the jaw to the measured signals at the corner, as shown in Figures 6.9 and 6.10. The following solution is proposed to improve the conditioning of the control signals and hence improve the final animation. At the corner, the control system can be considered as a multi-sensor system used for multi dimensional measurement where every possible displacement of the corner is defined by three signals **[Trankler89]**. In Section 4.5, the control: drive relationship was assumed to be distinct based on the diagonalisation of the overall function as shown in diagram a) of Figure 6.33. In reality, the jaw has significant effect resulting in the representation shown in diagram b) of Figure 6.1. In Section 6.3, the individual key point function characteristics were measured. Similarly,

the effects of the jaw on the corner actions of stretch and protrude were also measured. By extending the theory of non-linear compensation (c.f. Section 4.3.3), it would be possible to define the characteristics for a whole range of displacements at different levels of jaw drop. This could lead to the development of a search strategy to derive the required control signal from the analysis of two measured signals. An example is shown in diagram c) of Figure 6.33.

Diagram a)

sensor \ drive	D_{prot}	D_{stret}	D_{jaw}
S_6	*		
S_7		*	
S_{jaw}			*

Diagram b)

sensor \ drive	D_{prot}	D_{stret}	D_{jaw}
S_6	*		
S_7		*	
S_{jaw}	*	*	*

Diagram c)

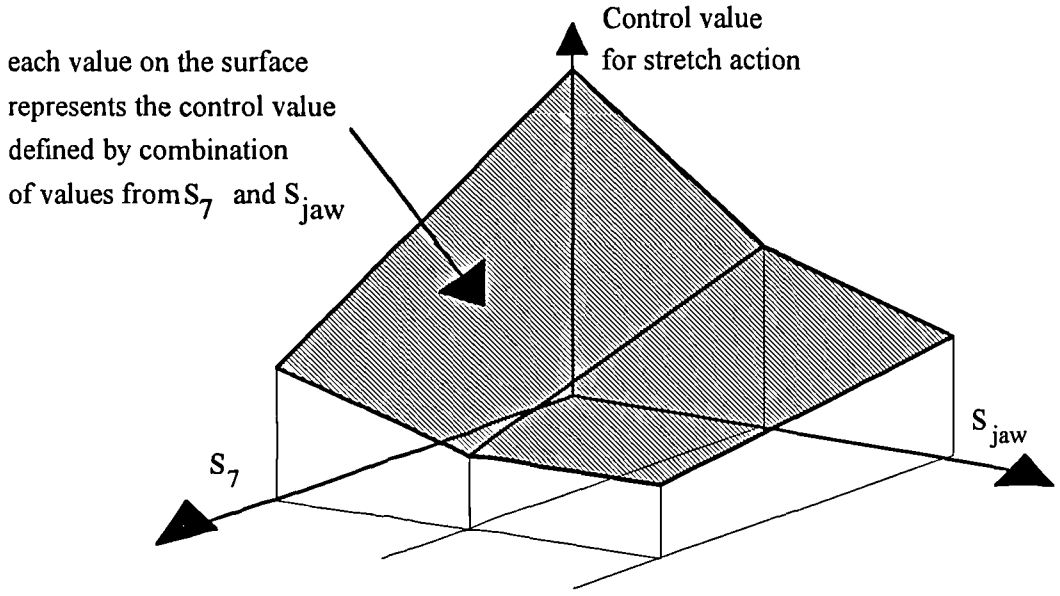


Figure 6.33 Example Of Possible Mapping Compensation For Stretch Actions At The Lip Corner

Chapter 7

Final Discussion And Future Work

Chapter 7

Final Discussion And Future Work

This thesis has described multi-disciplinary research on the development of original techniques to enhance the performance of facial action animation in Animatronics.

The goal of the facial animation is to communicate an idea, story or message to a viewer through distinct actions that are sufficiently similar to their idea or experience of human movements and expressions, that they view them as being the product of a living character.

The movements of the lips during speech provide vital clues about what is being said. The animation of these movements, when synchronised to the audible speech, represents an important part in a performance. The animation must produce a plausible representation of the real articulations that would accompany the spoken words to achieve the illusion of life. For realistic lip synchronisation, the objective is not the actual recognition and animation of the individual speech elements but the production of the continuously changing lip movements in time with the acoustic signal. The animated actions must be of comparable intensity, rate and rhythm and should produce the changes resulting from the blending effects of other visemes or other facial expressions. The control of this type of performance in animatronics and computer generated animation has been attempted in numerous ways.

The review in Chapter 2 considered the present techniques and deduced a number of limitations in their methods of realising lip synchronisation. Acoustic speech driven techniques produce animation that lacks the variety of action associated with natural speech due to the problems of segmentation of the input signal and the exclusion of the important expressive signals associated with speech. Acoustic methods also preclude the important element of performance control; the performer. The hand control systems used in Animatronics are designed to provide the performer with the maximum amount of control of the animation in order to produce life-like performances. Such systems are, however, limited in the production of complex lip movements due to the unnatural cognitive processes involved in the translation of hand action to lip action. Successful performance is achievable only by specialised performers.

In conclusion, this thesis proposed that a more innate form of control would be realised through the automatic extraction of visible information from the performer's facial actions that are directly associated with speech articulation and expression. This led to the development of the main hypothesis of this research that improved results in the final performance would be achieved by the application of an original design principle, based on an optimum set of visible key points on the face, to the production of an animatronic facial model and to its method of control.

In summary, the theoretical and practical work presented in this thesis broadly supports the hypothesis of facial action animation through designs based on the key point principle. Some of the main achievements of this research in realising the final performance system are outlined in the following section.

7.1 Specific Achievements

The hypothesis of improved performances through the automatic control from a performer's face led to the investigation into human facial communication documented in Chapter 3. The examination of speech production highlighted the problems that exist in the recognition of the visual actions of the lips and the effects of interactions between the articluatory actions of corresponding phonemes in continuous speech. The examination of facial expression described the human ability to produce a large variety of visible actions. The work of Ekman, **[Ekman82]**, was

adopted as a comprehensive notation scheme to describe the visible actions of the face. This investigation emphasised the complexity of the human facial system and its related actions.

The aim of the research was, therefore, to reduce to a minimum the amount of control information necessary to produce an output performance which would convey the desired messages. To this end, the research proposed and subsequently produced a more compliant system based on the reduction of the facial image to a minimum number of points. The hypothesis, described in chapter 4, proposed that a primary set of key points displacements exist on the surface of the face, that can fully describe the primary visible actions of the whole face during speech production and facial expression. This derivation was achieved through the investigation of the human face and through the experimental analysis of the researcher's face in static poses, described in Chapter 4. The results derived from that investigation confirmed the existence of a set of key points that were displaced along distinct trajectories during the production of the primary visible actions.

From this principle for description of facial actions in terms of a reduced set of key point displacements, it was clear that this description was intrinsic to the design of both drive and control systems of the overall performance system. This led to the production of a theoretical model, in Section 4.3, to describe the principle of key point design. The overall design was based on the extraction of input signals from distinct points on the performer's face which were mapped using the proposed 'mirror' correspondence to an identical set of points at the animatronic face. The subsequent recombination of the individual displacements should then reproduce the same overall facial actions at an identical rate and rhythm and with comparable intensity.

A novel technique to measure the key point displacements on the face through the application of low cost optical sensors was developed in this research, described in Sections 4.2 and 5.3. Its design was based on the criteria of continuous signal measurement with high sensitivity over a small range of point displacements. Through comprehensive assessment of the sensing system in experimental conditions, the thesis confirmed its capacity to operate, firstly, as a measurement system, described in section 5.3 and, secondly, as a facial point measurement system, in sections 6.2 and 6.3.

To produce a thorough assessment of the proposed hypothesis and the relevant theories, a method was developed to generate both subjective and objective data. This was based on the construction of an animatronic replica face of the exact shape and dimension, as the researcher, and capable of identical facial displacements of the same set of key points.

The realisation of the animated 3-D model represented a significant achievement given the practical complexities involved in construction, as discussed in Chapter 5. The video sequences V.3 and V.4 highlighted the wide variety of possible actions from the drive design and indicated the capacity to produce a number of distinct actions that are comparable with the human face. The successful realisation of the replica face coupled with the practical development of a data acquisition system, in Chapter 5, produced a system capable of the analysis of the research objectives.

The results presented in Chapter 6 show that the derived hypothesis represents a novel and valid principle of design. The results show that the system is capable of extracting and mapping sufficient control information from the key points on the performer's face to those on the replica, to produce animation of the same overall facial action.

This is shown by the graphical similarities of individual point displacements, in Figures 6.21 to 6.31, and the subjective analysis of the final performance, in video sequences V.5 to V.10, in comparison with the visible input actions of the performer and in the synchronisation with the audible speech. The results indicated that the system is capable of producing realistic lip synchronisation and facial expression at varying rates and rhythm and of differing intensities. The final performances show that the system automatically overcomes some of the problems inherent in other techniques, notably co-articulation and expressive blending.

The photographs in Chapter 6 and the video sequences V.5, V.6 and V.8 show the successful animation of certain individual and complex combinations of visually distinct actions. The actions produced were recognisable as either the individual action units, defined by Ekman, or as the combinations of lip movements associated with the production of speech or as the overall complex facial changes associated with the emotions 'surprise' or 'happiness'.

The final performances in V.9 and V.10, show that the overall system is capable of producing certain sequences of realistic animation that are, visually, very striking. The final animation exhibits a range of visemes, emotions, life-like nuances and idiosyncratic actions unlikely to be matched by other existing techniques, specifically those that apply segmentation to the overall performance.

This system, therefore, offers a means to directly incorporate the traditional talents of actors into animatronic performances. The performances presented in chapter 6 were produced at a slower than natural rate and with exaggerated actions to improve the perception of the lip movements. This is related to the views of lip-readers who have found that perception is increased as the rate and rhythm are slowed. The researcher is of the opinion that this is also related to an actor's performance where they produce facial actions which are exaggerated and at a slower rate than normal speech to ensure the viewer can fully perceive the message.

This research represents the development of a new and original method to enhance the performance control of animatronic characters through the automatic sensing of natural facial actions.

7.1 Future Work

The subjective analysis of the final performance in Section 6.4 was restricted to the opinions of the researcher which were unavoidably biased in terms of the recognition of the specific actions. This critical assessment was important in drawing conclusions on the specific actions of the replica. What remains unanswered is the ability of independent viewers, and specifically people with hearing impairments, to perceive the desired words or messages from the actions of the replica. This type of experimental analysis would have to be carefully planned to prevent the spurious results caused by either vague or biased questions or measurement of other variables.

For the present system, further work is necessary to overcome the practical limitations highlighted in the analysis of chapter 6. Specific examples include: the development of an improved design to support the sensors about the face, firstly, to remove the present errors and secondly, to be non specific to allow evaluation of the

system for different speakers; and the improvement to the sensor system to allow measurement over a greater range.

Visual analysis of the final performance draws the conclusion that the addition of actions to produce, say lip corner depressor or nose wrinkler, would improve the performance of the animation of the primary emotions. This addition of a new driving action would have to be accompanied by the addition of a sensor to retain the balance of the system. This leads to the argument that the design of both control and drive systems, and their successful application in the final performance, is inter-twinned. The design of a complex drive system that has the ability to produce a high subtlety of motion will ultimately be redundant if the control system cannot provide information to a comparable level. Similarly, control information will be redundant if the drive system can only achieve a limited number of actions. This desired increase in subtlety of performance appears related to our perceived opinion on the differences between the replica and the actual human face.

Herein lies the problem of facial animation which uses the conform of the human facial shape. A phenomenon exists in the visual perception of the face. As the facial model appears more realistic and specific, it allows greater comparisons to be drawn with the actual human face. This results in the fact that any inaccuracies or differences between them become more prominent and even disturbing **[Parke82]**. This problem may also exist as a result of producing the animated actions with actual speech signals.

This phenomenon is likely to be reduced when a fantasy character is viewed. In this case the face is conformed to contain only the key elements necessary to convey the idea of "likeness" or "similarity" to the human face. The production of lip synchronisation then moves away from the physical displacements toward animation which hints at the overall shape and the rate and rhythm of the desired actions. Consequently, the application of the present sensing system to a face of non-human conformation could improve the present lip synchronisation without the need to increase the complexity of the control system. The application of an improved version of the sensor system to other facial models offers great possibilities, specifically within computer animation, to overcome a number of their present control limitations.

An alternative application for the sensing system is as a supplementary source of information on the production of speech. Other research, **[Yuhas89]** and **[Petajan88a]**, has shown that information on visible speech significantly improves automatic speech recognition compared to acoustic speech alone. The results in chapter 6 indicate that the sensing system developed in this research is capable of the recognition of a limited vocabulary of visual speech segments which may provide useful information in circumstances when the acoustic signal is degraded. The system has also shown its potential, through the ability to produce synchronised lip animation, to be capable of recognising visible speech in continuous speech.

7.2 Summary

This thesis has described the following contributions;

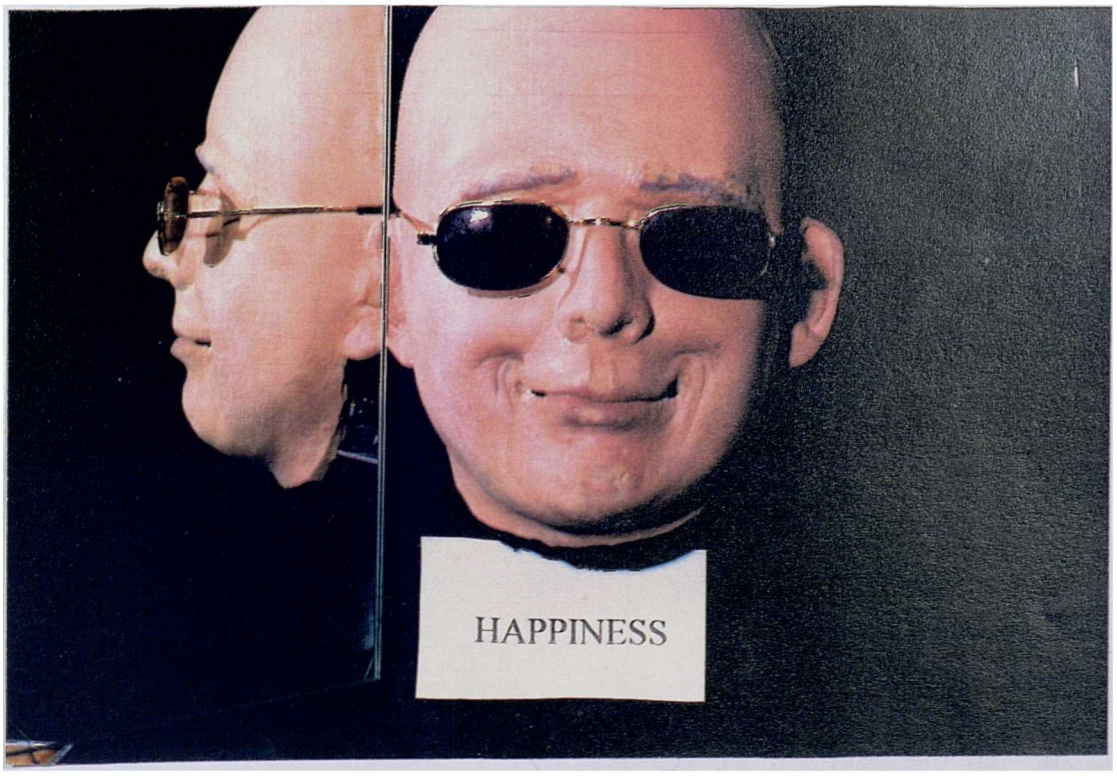
1. The development of an original design principle based on the reduction of the human facial action system to a specific set of key point displacements.

2. The production of results and analysis which confirm the validity of the key point principle and the hypothesis of automatic facial control in the production of realistic lip synchronisation and overall performance.

3. The development of a novel technique to sense the facial actions of the performer through the use of low cost optical sensors.

4. The realisation of a practical system including the construction of the animatronic replica face, based on the key point principle, to produce facial animation.

5. The correlation of a wide range of disciplines other than electronic engineering, to apply a greater scientific base to the field of animatronics which has been, until now, largely based on practical and subjective approaches, with little literature available.



"We do not receive wisdom, we must discover it for ourselves, after a long journey through the wilderness which no one else can make for us, which no one else can spare us, for our wisdom is the point of view from which we come at last to regard the world."

M. Proust 1919 "Remembrance Of Things Past"

Bibliography

Bibliography

[Arby86] ..... Arby, C. Laws For Lips., *Speech Communication.*, vol. 5., 1986., pp97-104.

[Argyle88] ..... Argyle, M. Bodily Communication., *Methuen & Co Ltd.*, 1988., pp.-21.

[Bailey82] ..... Bailey, A. Walt Disney's World Of Fantasy., *Walt Disney Productions,*. 1982., pp. 220-225.

[Bailly87] ..... Bailly, C. Automata: The Golden Age., *Sotherby Publ, U.K..*, 1987., pp. 13-23, 204-216.

[Barry88] ..... Barry, B. Errors in Practical Measurement in Science, Engineering and Technology., *John Wiley & Sons, U.K..*, 1988.

[Bengeraul82] ..... Bengeraul, A. Co-Articulatory Effects In Lip-Reading., *Journal of Speech Research.*, vol. 25., 1982., pp. 600-607.

[Berger72] ..... Berger, K. Speech-Reading: Principles And Methods., *National Educational Press.*, 1972.

[Binnie74] ..... Binnie, C. Auditory And Visual Contribution To The Perception Of Consonants., *Journal Of Speech And Hearing Research.*, vol. 17., 1974., pp. 619-630.

[Birdwhistell70].....Birdwhistell, R. Kinesics & Context., *Philadelphia University of Pennsylvania Press, USA.*, 1970., pp240.

[Blurton71].....Blurton Jones, N. Criteria For Describing Facial Expression In Children., *Human Biology.*, vol. 141., 1971., pp365-413.

[Boucher75].....Boucher, J. Facial Areas of Emotional Information., *Journal of Communication.*, vol. 25., 1975., pp21-29.

[Brooke83].....Brooke, N.M., Summerfield, Q.Analysis, Synthesis and Perception of Visible Articulatory Movements., *Journal of Phonetics.*, vol. 11., 1983., pp. 63-76.

[Brooke86].....Brooke, N.M., Petajan, E. Seeing Speech : Investigations Into The Synthesis And Recognition Of Visible Speech Movements Using Automatic Image Processing And Computer Graphics., *Proceedings IEE Conference on Speech I/O : Techniques and Applications.*, publication no 258., 1986., pp. 104-109.

[Brooke89].....Brooke, N.M. Visible Speech Signals : Investigating Their Analysis, Synthesis And Perception., *Chapter 18 in 'Structure of Multimodal Dialogue' ed. by Taylor,M et al.*, 1989., pp. 249-258.

[Brooke90a].....Brooke, N.M. Classification of Lip-Shapes And Their Association With Acoustic Speech Events., *Proceedings of International Research Workshop on Speech Synthesis.*, 1990., pp. 245-248.

[Brooke90b] .....Brooke, N.M. Visual Speech Intelligibility of Digitally Processed Facial Images., *Proceedings of Institute of Acoustics.*, vol. 12 part 10., 1990., pp. 483-490.

[Brooke90c].....Brooke, N.M. Computer Graphics Synthesis Of Talking Faces., *Proceedings of Tutorial on Speech Synthesis.*, 1990., pp. 65-71.

[Campbell86]..... Campbell, R. The Lateralization Of Lip-Read Sounds: A First Look., *Brain And Cognition.*, vol. 5 part 1., 1986., pp. 1-21.

[Chambers93]..... Chambers Paperback dictionary., *W.R.Chambers Ltd.*, 1993.

[Chatfield83] ..... Chatfield, C. Statistics for Technology., *Chapman and Hall, London; New York.*, 1983., pp. 224-240.

[Choi90]..... Choi, C. 3-D Facial Model-Based Description and Synthesis Of Facial Expressions., *Electronics and Communications In Japan.*, vol. 74 part 7., 1990., pp. 1270-1280.

[Choi91a]..... Choi, C. System Of Analysis And Synthesising Facial Images., *IEEE International Symposium On Circuits And Systems.*, vol. 5 part 4., 1991., pp. 2665-2668.

[Choi91b] ..... Choi, C. Analysis And Synthesis Of Facial Images., *Proceedings of ICASSP 91.*, vol. 4., 1991., pp. 2737-2740.

[Cinefex16]..... *Cinefex Magazine no.16.*, Rick Baker: Maker Of Monsters., Article by J. Fox., Publ. by D. Shay (USA)., Ed. by J. Duncan., April 1984., pp 4-71.

[Cinefex43]..... *Cinefex Magazine no.43.*, Ego Trip (Total Recall)., Article by P. Roberts., Publ. by D. Shay (USA)., Ed. by J. Duncan., August 1990., pp 4-33.

[Cinefex46]..... *Cinefex Magazine no.46.*, Rick Baker Revisited., Article by R. Magid., Publ. by D. Shay (USA)., Ed. by J. Duncan., August 1991., pp 4-29.

[Cinefex47]..... *Cinefex Magazine no.47.*, A Once And Future War (Terminator II)., Article by J. Duncan., Publ. by D. Shay (USA)., Ed. by J. Duncan., November 1991., pp 4-59.

[Cinefex50a]..... *Cinefex Magazine no.50.*, Zealots And Xenomorphs (Alien^3)., Article by B. Norton., Publ. by D. Shay (USA)., Ed. by J. Duncan., May 1992., pp 26-53.

[Cinefex50b]..... *Cinefex Magazine no.50.*, Zealots And Xenomorphs (Alien^3)., Article by P. Sorenson., Publ. by D. Shay (USA)., Ed. by J. Duncan., May 1992., pp 54-71.

[Cinefex52]..... *Cinefex Magazine no.52.*, Life Neverlasting (Death Becomes Her)., Article by K. Martin., Publ. by D. Shay (USA)., Ed. by J. Duncan., November 1992., pp 54-78.

[Cohen90]..... Cohen, M. Synthesis Of Visible Speech., *Behaviour Research Methods, Instruments And Computers.*, vol. 22 part 2., 1990., pp. 260-263.

[Coiffet83]..... Coiffet, P. Introduction to Robot Technology., *Kogan Page Ltd.*, 1983., pp33-37, 121-129.

[D'Azzo81]..... D'Azzo Houpis, T. Linear Control System Analysis And Design., *McGraw-Hill Inc.*, 1981., pp141-183.

[Eastman83]..... Eastman Kodak Company Complete Kodak Animation Book., 1983., pp42-46.

[Ekman73]..... Ekman, P. Darwin and Facial Expression., *Academic Press, USA.*, 1973., p7.

[Ekman78]..... Ekman, P. The Facial Action Coding System; A technique for Measurement of facial movement., *Consulting Psychologists Press Palo Alto, Ca(USA).*., 1978.

[Ekman79]..... Ekman, P. About Brows: Emotional And Conversational Signals., Human Ethology Ed. by M. Von Cranuch., Cambridge University Press (UK)., pp. 169-249., 1979.

[Ekman82]..... Ekman, P. Methods for Measuring Facial Action ., *Handbook of Methods of Non-Verbal Research Ed. By Scherer, K.*, Cambridge University Press., 1982., pp45-90.

[Ekman92]..... Ekman, P. Facial Expression Of Emotion., *Psychological Science.*, vol. 3 part 1., 1992., pp. 34-38.

[Engler73]..... Engler, L. Making Puppets Come Alive., *D & Charles Newton Abbot, U.K.*., 1973., pp1-9.

[Faigin90]..... Faigin, G. The artists complete guide to facial expression., *Phaidon Press, U.K.*, 1990., pp32-72.

[Finch82]..... Finch, C. Of Muppets & Men., *Muppet Press/Joseph, U.K.*., 1982.

[Finn88]..... Finn, K. Automatic Optically Based Recognition Of Speech., *Pattern Recognition Letters.*, vol. 8 part 3., 1988., pp. 159-164.

[Fisher68]..... Fisher, C. Confusion Among Visually Perceived Consonants., *Journal of Speech & Hearing Research.*, vol. 11., 1968., pp796-804.

[Fromkin64]..... Fromkin, V. Lip Positions In American English Vowels., *Language & Speech.*, vol. 7., 1964., pp. 215-225.

[Fry79]..... Fry, D. The Physics Of Speech., 1979., pp. 97-125.

[Goleman81]..... Goleman, D. The 7000 faces of DR. Ekman., *Psychology Today.*, vol. 15., 1981., pp42-49.

[Gopel89]..... Gopel, W. Sonsors: A Comprehensive Survey., *VCH Publishers, USA.*., 1989., pp299-311.

[Guenter89]..... Guenter, B. System For Simulating Human Facial Expression., *Proceedings Of Computer Animation* ., 1989., pp. 191-202.

[Hardcastle78]..... Hardcastle, W. Physiology of Speech Production., *Academic Press, USA.*, 1978., pp. 88-121.

[Herbison-Evans84] Herbison-Evans, D. Dance, Video, Notation And Computers., *Leonardo.*, vol. 21 part 1., 1984., pp. 45-50.

[Hill88]..... Hill, D. Animating Speech; Automated Approach using Speech Synthsis By Rules., *The Visual Computer.*, vol. 13, No. 5., 1988., pp277-287.

[Hilton87]..... Hilton, J. Performance., *MacMillan Publishers Ltd, U.K.*, 1987., pp40-44.

[Hitchcox89]..... Hitchcox, A. Pneumatics Brings Dinosaurs To Life., *Hydraulics and Pneumatics.*, 1989., pp. 45-48.

[Hitchcox90a]..... Hitchcox, A. Disney Makes Kid Stuff Serious Business., *Hydraulics and Pneumatics.*, 1990., pp. 31-33.

[Hitchcox90b]..... Hitchcox, A. Special Effects, Hydraulics Style., *Hydraulics and Pneumatics.*, 1990., pp. 34-40.

[Housman90]..... Housman, D. Personal Communication ., *Jim Henson Creature Shop.*, 1990.

[Hulton75]..... Hulton, K. Jean Tinguely., *Thomas & Hudson.*, 1975.

[Hutchinson87] ..... Hutchinson, D. Film Magic: The art & secience of special effects., *Simon &Shuster Ltd, U.K.*., 1987., pp18-65.

[Jackson88] ..... Jackson, P. Theoretical Minimal Unit For Visual Speech Perception : Visemes And Co-articulation., *Volta Review* ., vol. 90 part 5., 1988pp. 99-115.

[Jeffers71]..... Jeffers, J. Speech Reading., *Charles C Thomas Publ.*, Springfield, USA.., 1971.

[Kalra91]..... Kalra, P., Magnenat-Thalmann, N. SMILE: A Multi-Layered Facial Animation System., *Proceedings Of IFIP WG 5.1 Working Conference Modelling in Computer Graphics* ., vol. 92/00502., 1991., pp. 189-198.

[Kehoe85]..... Kehoe, V. The technique of the professional Make-up artists for filem & television., *Focal Press, Ltd, U.K.*., 1985., pp141-232.

[Knapp72]..... Knapp, M. Nonverbal Communication in Human Interaction., *New York, Holt, Rinehar & Winston Inc.*, 1972.

[Kopra86]..... Kopra, L. Development Of Sentences Graded In Difficulty For Lipreading Practice., *Journal Of The Academy Of Rehabilitative Audiology.*, vol. 19., 1986., pp. 71-86.

[Korane90]..... Korane, K. Magic Engineering at Disney., *Machine Design.*, 1990., pp. 26-28.

[Kricos82]..... Kricos, P. Differences In Visual Intelligibility Across Talkers., *Volta Review.*, vol. 84 part 4., 1982., pp. 219-225.

[Ladefoged78]..... Ladefoged, P. The Phonetic Basis For Computer Speech Processing., *A Course In Phonetics.*, 1978., pp5-9.

[Laver80]..... Laver, J. Phonetic Description of Voice Quality., *Cambridge University Press, U.K.*, ., 1980., pp12-36.

[Lesner87]..... Lesner, S. Training Influences On Visual Consonant And Sentence Recognition., *Journal Of Ear And Hearing.*, vol 8., 1987., pp283-287.

[Lesner88]..... Lesner, S. The Talker., *Volta Review.*, vol. 90 part 5., 1988., pp. 89-98.

[Levitan79]..... Levitan, E. Handbook of Animation Techniques., *Van Nostrand Reinhold Co, U.K..*, 1979., pp42-46.

[Lewis87]..... Lewis, J. Human Factors In Computing Systems And Graphics Interface. Automated Lipsync And Speech Synthesis ., *Proceedings Of Conference CHI and GI .*, 1987., pp. 143-147.

[Lewis91]..... Lewis, J. Automated Lip Synchronization: Background And Technique., *Journal Of Visualization And Computer Animation.*, vol. 2 part 4., 1991., pp. 118-122.

[Luzzadder89]..... Luzzadder, W. Fundamentals of Engineering Drawing., *Prentice Hall, USA.*, 1989.

[MacKay87]..... MacKay, I. Phonetics : The Science Of Speech Production., *College Hill Productions, USA.*, 1987., .

[Magnenat-Thalmann89a] Magnenat-Thalmann, N. The Problematics Of Facial Animation., *Proceedings Of Computer Animation* ., .1989., pp. 47-56.

[Magnenat-Thalmann89b] Magnenat-Thalmann, N. Design, Transformation And Animation Of Human Faces., *Vision Computer (West Ger).*, vol. 5 parts 1-2 ., 1989., pp. 32-9 .

[Marston88] ..... Maston, R. Optoelectronic Circuits Manual., *Heinemann Prot Publ, U.K.*., 1988., pp105-125.

[Mase89]..... Mase, K. Lipreading: Automatic Visual Recognition., *Image Understanding And Machine Vision.*, vol. 14., 1989., pp. 124-127.

[Massaro87] ..... Massaro, D. Speech Perception by Ear & Eye., *Lawerance Erlbaum Publ, USA.*, 1987., pp16-36.

[McClure87]..... McClure, E. New Tricks For Film Fantasy., *Computer Graphics World.*, vol. 10 part 2 ., 1987., pp. 77-80 .

[McGrath84]..... McGrath, M., Brooke, N.M., Summerfield, Q. Roles Of Lips and Teeth in Lipreading Vowels., *Proceedings of Institute Of Acoustics.*, vol. 6 part 4., 1984., pp. 401-408.

[McGurk76] ..... McGurk, H. Hearing Lips, Seeing Voices., *Nature.*, vol. 264., 1976., pp. 746-748.

[Minson89]..... Minson, J.The Prime Mover ., *The Guardian Newspaper* ., Apr 20, 1989.

[Montgomery83] ... Montgomery, A.Physical Characteistics Of Lips Underlying Vowel Lipreading Performance., *Journal Of Acoustical Society Of America.*, vol. 73., 1983., pp. 2134-2144.

[Montgomery87] ... Montgomery, A.Effects Of Consonantal Context on Vowel Lipreading., *Journal of Speech and Hearing Research.*, vol. 30., 1987., pp. 50-59.

[Morishima91a] ....Morishima, S. Real Time Model Based Facial Images., *Systems And Computers In Japan.*, vol. 22 part 13., 1991., pp. 59-69.

[Morishima91b] ....Morishima, S. Real Time Facial Action Image Synthesis System Driven By Speech And Text., *Proceedings Of SPIE* ., vol. 1360 part 2., 1991., pp. 1151-1158.

[Morishima91c].....Morishima, S. Facial Motion Synthesis for Intelligent Man-Machine Interface., *Systems And Computers In Japan.*, vol. 22 part 5., 1991., pp. 50-59.

[Morishima91d] ....Morishima, S. A Media Conversion From Speech To Facial Image for Intelligent Man-Machine Interface., *IEEE Journal on Selected Areas In Communications.*, vol. 9 part 4., 1991., pp. 594-99.

[Morishima92] .....Morishima, S. Image Synthesis And Editing System For A Multi-Media Human Interface With Speaking Head., *Proceedings of IEE International Conference On Image Processing And Its Applications.*, Conf No. 354., 1992., pp. 270-273.

[Morris87] .....Morris, D. Manwatching: A Field Guide To Human Behaviour., *Grafton Books(U.K.).*, 1987.

[Nitchie12].....Nitchie, E. Lip-reading principles: an art., *Volta Review.*, vol 15., 1912., pp276-278.

[Nitchie50].....Nitchie, E. New Lessons In Lip-Reading., *J.B.Lippincott Publ. (U.S.A).*, 1950.

[Ohba92] .....Ohba, R. Intelligent Sensor Technology., *John Wiley & Sons, U.K.*., 1992., pp63-71.

[Oliver87] .....Oliver, T. The Brody Family Gets Jawed Again., *Hydraulics and Pneumatics.*, 1987., pp. 60-62.

[Pallas91].....Pallas-Areny, R. Sensors &Signal Conditioning., *John Wiley & Sons, U.K.*., 1991., pp. 1-26.

[Panis92]..... Panis, S. Control Of Movement Of A Synthesized Mouth From A Text Stream., *Proceedings of IEE International Conference on Image Processing & its Applications.*, Conf No. 354., 1992., pp. 266-269.

[Parke82]..... Parke, F. Parametric Models for Facial Animation., *IEEE Computer Graphics & Applications.*, 1982.

[Patel91]..... Patel, M. FACES: Facial Animation, Construction And Editing System., *Proceedings of EUROGRAPHICS '91.*, vol. 1., 1991., pp. 33-45.

[Pearce86]..... Pearce, A. Speech And Expression: A Computer Solution., *Proceedings Of Conference On Graphics Interface.*, 1986., pp. 136-140.

[Pelachaud92]..... Pelachaud, C., Balder, N. Correlation of Facial and Vocal Expressions for Facial Animation., *Informatique '92 - International Conference On Interface to Real and Virtual Worlds* ., 1992., pp. 95-110.

[Petajan88a]..... Petajan, E., Brooke, N.M. An Improved Automatic Lip-Reading System to Enhance Speech Recognition., *Proceedings Of Conference On Human Factors In Computing Systems.*, 1988., pp. 19-25.

[Petajan88b]..... Petajan, E., Brooke, N.M. Experiments In Automatic Visual Speech Recognition., *Proceedings 7th Symposium of Federation of Acoustical Societies of Europe.*, 1988., pp. 1163-1170.

[Pichora-Fuller91]. Pichora-Fuller, M. The Design Of CAST: Computer Aided Speechreading Training., *Journal Of Speech and Hearing Research.*, vol. 34., 1991., pp. 202-212.

[Pisoni74]..... Pisoni, D. Cateogorical And Non-Categorical Modes Of Speech perception., *Journal Of The Acoustical Society Of America.*, vol 61., 1974., pp1352-1361.

[Platt81]..... Platt, S. Animating Facial Expressions., *Computer Graphics.*, vol. 15, No. 3., 1981., pp245-252.

[Reynolds82] ..... Reynolds, C. Computer Animation with Scripts And Actors., *Proceedings Of SIGGRAPH '82 Computer Graphics.*, vol. 16 part 3., 1982., pp. 289-296.

[Robertson88] ..... Robertson, B. Mike, The Talking Head Interactive Animation., *Computer Graphics World.*, vol. 11 part 7., 1988., pp. 57-59.

[Rovin77]..... Rovin, J. Movie Special Effects., *Thomas Yoseloff, U.K..*, 1977.

[RSData83]..... R.S. Data Sheet 4276 "Reflective And Slotted Opto Switches., *R.S. Components, U.K..*, 1983.

[Scherer82]..... Scherer, K. Methods of Research on Vocal Communication., *Handbook of Methods in Nonverbal Research, Ed. by Ekman, P. & Scherer, K.*, 1982., pp. 182-186.

[Seely58]..... Seely, F. Analytical Mechanics for Engineers., *John Wiley & Sons Inc, U.K..*, 1958., pp175-308.

[Sheridan92]..... Sheridan, T. Telerobotics, Automation & Human Supervisory Control ., *MIT Press.*, 1992., pp46-49.

[Smeele91]..... Smeele, P. The Contribution Of Vision To Speech Perception., *Proceedings of EUROGRAPHICS '91.*, vol. 3., 1991., pp. 1495-7.

[Smith86]..... Smith, T. Industrial Light & Magic; The Arts of Special Effects., *Columbus Books Ltd, U.K..*, 1986., pp4-14, 66-77.

[So91] ..... So, I. Model Based Coding Of Facial Images., *Visual Communications And Image Processing.*, vol. 1605 part 1., 1991., pp. 263-272

[Storey88]..... Storey, D. Reading the Speech of Digital Lips: Motives and Methods for Audio-Visual Speech Synthesis., *Visible Language.*, vol. 22 part 1., 1988., pp. 112-127.

[Sumby54]..... Sumby, J. Visual contribution to Speech., *Journal Of The Acoustical Society Of America.*, vol. 26., 1954., pp212-215.

[Summerfield82] ... Summerfield, Q. Analysis Perception of Visible Articulatory Movements., *Journal of Phonetics.*, vol.11., pp63-76.

[Swain92] ..... Swain, B. Pixel Tricks Enrich Flicks., *New Scientist.*, 1992., pp. 23-27.

[Terzopuolos90] .... Terzopoulos, D. Analysis of Facial Images Using Physical And Anatomical Models., *Proceedings Of 3rd International Conference On Computer Vision Publ. by IEEE Computer Society (USA).*, 1990., pp. 727-732.

[Thomas81a] ..... Thomas, F. Disney Animation: The Illusion Of Life., *Walt Disney Productions, USA..*, 1981., pp441-507.

[Thomas81b] ..... Thomas, F. Disney Animation: The Illusion Of Life., *Walt Disney Productions, USA..*, 1981., pp528-530.

[Thring83]..... Thring, M. Robots & Telechirs., *John Wiley & Sons Inc, U.K..*, 1983., pp49-59, 82-131.

[Tost88] ..... Tost, D. Human Body Animation; A Survey., *The Visual Computer.*, vol. 3., 1988., pp254-264.

[Trankler89]..... Trankler, H. Signal Processing., *Sensors: A Comprehensive Survey Ed. By Gopel, W..*, VCH Publishers, USA.., 1989., pp299-311.

[Vila90]..... Vila, J. A Computerized Phonetics Instructor: BABEL., *CALICO Journal.*, vol. 7 part 3., 1990., pp. 3-29.

[Walden87]..... Walden, B. Perception of Synthetic Visual Consonant-Vowel Articulations., *Journal of Speech and Hearing Research.*, vol. 30 part 3., 1987., pp. 418-424.

[Warwick74] ..... Warwick, R. Grays Anatomy., *Longman, UK.*, 25th Ed., 1974., pp. 490-503.

[Waters87a]..... Waters, K. Muscle Model For Animating 3-D Facial Expression., *Computer Graphics.*, vol. 21 part 4., 1987., pp. 17-24.

[Waters87b] ..... Waters, K. Animating Human Heads., P*roceedings Of Computer Graphics.*, 1987., pp. 89-97.

[Waters89]..... Waters, K. Computer Synthesis of expressive 3-D facial character animation., *Ph.D Thesis, Middlessex Poly, U.K.*., 1989., pp204-239.

[Wiggers82]..... Wiggers, M. Judgements of Facial Expressions of Emotion Predicted from Facial Behaviour., *Journal of Non Verbal Behaviour.*, vol. 7., 1982., pp101-116.

[Williams90]..... Williams, L. Performance Driven Facial Animation., *Computer Graphics.*, vol. 24 part 4., 1990., pp. 235-242.

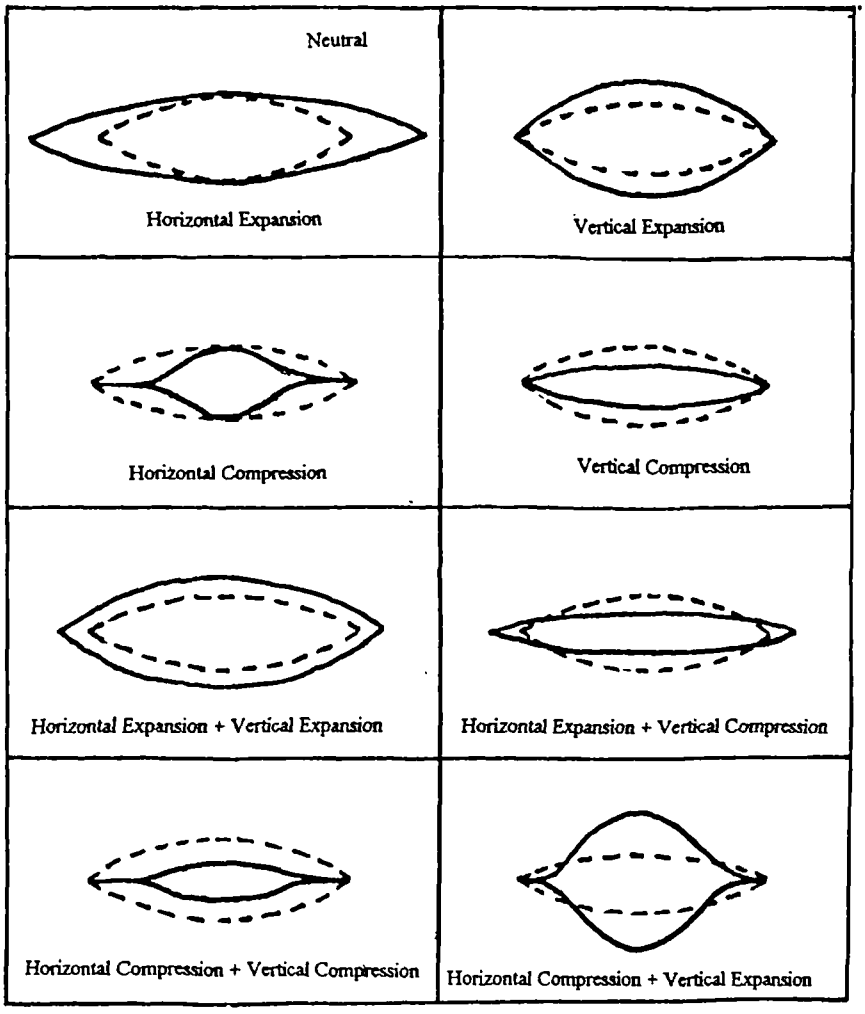
[Yuhas89]..... Yuhas, B.Integration of Acoustic and Visual Speech Signals Using Neural Networks., *IEEE Communications Magazine.*, vol. 27 part 11., 1989., pp. 65-71.

Appendix

Appendix A

Physiology of Speech Production

[Hardcastle78] defined the minimal set of lip shapes as shown in Figure A.1 based on reduction of the complex lip shapes to a description comprising of horizontal and vertical expansion or compression, with or without lip protrusion.



@ copy of diagram p37 The Phonetic Description Of Voice Quality Laver, J.

Figure A.1 Hardcastle's Description Of Lip Shapes

A.1 The Muscle That Closes The Lips

A.1.1 ORBICULARIS ORIS [15]

General description: large sphincter muscle encircling lips and forming majority of muscle in lips. The upper and lower fibres of the muscle meet at the angles of the mouth.

Course : muscle encircles the lips joining at the corners.

Main function is to adduct (close tightly) the upper and lower lips against the teeth. This is important for any sound involving the closing of the lips; /b/,/p/,/m/.

For any labiodental frictive, /f/ and /v/, the inferior part of the muscle presses the lower lip against the teeth. The action of drawing the lower lip upwards is assisted the elevators of the **MANDIBLE** (jaw), notably the **MASSETER [21]**, **TEMPORALI [20]** as well as the inferior fibres of the **LEVATOR ANGULI ORIS [8]**.

A.1 The Muscles That Raise The Upper Lip

A.2.1 ZYGOMATICUS MINOR [10]

Course : downwards and slightly medially.

Main function to raise the upper lip. Of importance during articulation of labiodental frictives, as it elevates the upper lip slightly to allow air to flow through the narrow oral cavity caused by the lower lip against the teeth.

A.2.2 LEVATOR LABII SUPERIORIS [9]

Course : downwards.

Main function to raise the upper lip primarily the medial region.

A.2.3 LEVATOR LABII SUPERIORIS ALAEQUAE NASI [6]

Course : downwards and laterally along sides of the nose.

Main function to raise the medial part of upper lip and elevate wings of the nose.

A.3 Muscles that lower the lower lip.

A.3.1 DEPRESSOR LABII INFERIORIS [14]

Course : upwards and medially.

Function is the main depressor of the lower lip. It is important for release of bilabial stops; /b/,/p/,/m/. This is assisted by muscles which depress the **MANDIBLE**.

A.4 Muscles that round the lips.

A.4.1 ORBICULARIS ORIS [15]

Course : muscle encircles the lips.
In its sphincter activity, it can round the lips for the production of rounded phonemes such as /u/,/y/.
The different degrees of rounding are caused by different tensions in the muscle and by different elevations or depressions of the **MANDIBLE**. The lips may not be necessarily be protruded during rounding, through the retractive nature of the muscles **BUCCINATOR [17]**, **ZYGOMATICUS MAJOR [11]** and the **RISORIUS [12]**.

A.5 Muscles that protrude the lips.

A.5.1 MENTALIS [16]

Course : downwards.
Its activity is one of drawing skin of chin and lower lip upwards which has the effect of everting the lower lip and protruding it slightly.

A.5.2 ORBICULARIS ORIS [15]

Course : muscle encircles the lips.
When contracted simultaneously with **MENTALIS [16]** causes protruded, rounded lips for production of /u/,/o/.

A.6 Muscles that retract the angles of the mouth.

A.6.1 BUCCINATOR [17]

Course : forward towards angles of the mouth, parallel to **RISORIUS [12]**.
It retracts the angles of the mouth perhaps compressing the lips against the teeth and is antagonistic to the protruding of the lips by the **MENTALIS [16]**, **ORBICULARIS ORIS [15]**.
It is used in the articulation of labiodental frictives and spread vowels; /i/,/e/, and is probably a protagonist in the production of bilabial frictives.

A.6.2 ZYGOMATICUS MAJOR [11]

Course : downwards and medially.
Its function is to draw the corners of the mouth laterally and slightly upwards and is important in the production of labiodental and spread vowel articulations.
It may also function with the **RISORIUS [12]** in the spread of lips for the production of /s/.

A.6.3 RISORIUS [12]

Course : horizontal and forwards toward angle of mouth.

Its main action is one of drawing the angles of the mouth laterally which helps in the spread of lips for /s/ and vowels /i/,/e/ production.
It also helps in the production of labiodental frictives.

A.7 Muscles that raise the angles of the mouth.

A.7.1 LEVATOR ANGULI ORIS [8]

Course : downwards and slightly laterally.
Main function to elevate the angles of the mouth and the upper lip. It may also help in the raising of the lower lip in bilabial closures.

A.7.2 ZYGOMATICUS MAJOR [11]

Course : downwards and medially .
Main function is to draw the angles of the mouth upwards as well as laterally. It is also important in the articulation of labiodental phonemes.

A.8 Muscles that lower the angles of the mouth.

A.8.1 DEPRESSOR ANGULI ORIS [13]

Course : vertically upwards.
Main function is to depress the angles of the mouth. It also acts in lowering lower lip for release in bilabial stops.

A.8.2 PLATYSMA [18]

Course : upwards and forwards towards the **MANDIBLE**.
Main function is to have an effect in drawing down the angles of the mouth and also helps in protrusion of lips with **MENTALIS [16]** and **ORBICULARIS ORIS [15]**.

A.9 Muscles that elevate the jaw.

A.9.1 PTERYGOIDEUS INTERNUS [22]

Course : runs posteriorly and downward.
Main function is to raise the **MANDIBLE** along with the **MASSETER [21]**, **TEMPORALIS [20]**. Its action opposes the natural tendency to move backward when raised because of the nature of the temporomandibular joint.

A.9.2 MASSETER [21]

Course : downward.
Main function is to elevate the **MANDIBLE** and hence the tongue. It is the most powerful of the mandibular muscles and plays important role in production of alevolar articulations and in mastication.

A.9.3 TEMPORALIS [20]

Course : downwards and anteriorly.

Main function to elevate the **MANDIBLE**.

Appendix B List Of Phonetic Groupings And Examples

The following tables show examples of contrasting consonants and vowels with examples of their application in the English language.

Consonants	
phoneme description	example
p	*p*ie
t	*t*ie
k	*k*ye
b	*b*y
d	*d*ye
g	*g*uy
m	*m*y
n	*n*igh
ŋ	ra*ng*
f	*f*ie
v	*v*ie
θ	*th*igh
δ	*th*ee
s	*s*igh
z	*Z*
(sh)$_1$	*sh*y

Vowels	
phoneme description	example
ee	h*ee*d
i	h*i*d
ay	h*ay*ed
ea	h*ea*d
a	h*a*d
ar	h*ar*d
u	c*u*t
er	h*er*d
o	h*o*d
aw	h*aw*ed
oe	h*oe*d
ou	c*ou*ld
oo	c*oo*ed

$(sh)_2$	vi*s*ion
l	*l*ie
w	*w*hy
r	*r*ye
h	*h*igh
ch	*ch*ime
ji	*ji*ve

Appendix C

Software Listing For Data Acquisition System

This appendix contains examples from the final listing of the software developed to realise data play and record and the conditioning of the control data.

C.1 Function To Condition The Playback Data

```
/*********************************************************************/
/*                          function : cond                          */
/*      objective : to condition play files according to desired method */
/*                          arguments : no                           */
/*********************************************************************/
void cond( int no)
{
long loop;
int port;
int u,v,w;                      /* pointers to values in condvalue array */
int value,upper,neutral,lower;  /* variables to hold conditioning parameters */

   openreadfile(playname,playdir);          /* open text file */
   openwritefile(playtempfile,playdir);     /* open temp file */
   loop=0;
   while(loop<sample)
         {
         for(port=0;port<chan;port++)
               {
               if (port == 0)               /* if channel number is 0 or 11 to 15 */
                     {                              /* then arbitary value of 128 given */
                     fscanf(fin,"%d\t",&value);
                     tempone.reading[port]=128;
```

```
                }
            else if ((port >10)&&(port<chan))
                {
                fscanf(fin,"%d\t",&value);
                tempone.reading[port]=128;
                }
            else
                {
                fscanf(fin,"%d\t",&value);          /* read original value in */
                if (no == 3)                        /* if nonlinear conditioning..*/
                    tempone.reading[port]= nlinout.table[port][value];
                else                                /*if linear conditioning..*/
                    {
                    v = (port)*3+2;
                    w = (port)*3+1;
                    u = (port)*3;
                    upper = condlimits.reading[u];
                    neutral = condlimits.reading[w];
                    lower = condlimits.reading[v];
                   if (no == 2)
                    tempone.reading[port]= conditioning2(upper,neutral,lower,value);
                    else
                        tempone.reading[port]  = conditioning1(upper,lower,value);
                    }
                }
            }
        fwrite(&tempone,sizeof(tempone),1,fout);    /*write out row of new values */
        fscanf(fin,"\n");                           /* skips end of line in text file */
        loop++;
        }
    fclose(fin);                        /* close both files */
    fclose(fout);
}
```

C.2 Function To Produce Linear One Part Conditioning

```
/*********************************************************************/
/*                    function : conditioning1                       */
```

```
/*          objective : to perform one part linear conditioning                */
/*      arguments : input conditionig parameters,original data value           */
/*                  result : new conditioned play value                        */
/*************************************************************************/
int conditioning1(int uppin,int lowin,int xvalue)
{
int uppout = 255;                       /* output parameters */
int lowout =0;
int y;
double temp;
float scale1, z;
float one = 1.000;

   scale1 = (one *(uppout - lowout)) / (uppin - lowin);      /* calculate function scaling */
   if (xvalue>uppin)                    /* if original value outside range, arbitary value given*/
         y = uppout;
   else if (xvalue<lowin)
         y = lowout;
   else   {
         z = ((xvalue - lowin) * scale1); /* caluculate output conditioned value */
         temp = (z + lowout);
         y = round(temp);
         }
   return(y);
}
```

C.3 Function To Produce Linear Two Part Conditioning

```
/*************************************************************************/
/*                      function : conditioning2                           */
/*          objective : to perform two part linear conditioning            */
/*      arguments : input conditionig parameters,original data value       */
/*                  result : new conditioned play value                    */
/*************************************************************************/
int conditioning2(int uppin,int neutin,int lowin,int xvalue)
{
double temp;
int y;
int uppout = 255;          /* output parameters */
```

```
int neutout = 128;
int lowout = 0;
float scale1, scale2, z;
float one = 1.000;

    if (neutin==lowin)
            {
            scale1 =0.0;
            scale2 = (one *(uppout - neutout)) / (uppin - neutin);
            }
    else if (neutin==uppin)
            {
            scale1 = (one*(neutout - lowout)) / (neutin - lowin);
            scale2 = 0.0;
            }
    else
            {
            scale1 = (one *(neutout - lowout)) / (neutin - lowin);
            scale2 = (one*(uppout - neutout)) / (uppin - neutin);
            }
    if ((xvalue >= neutin)&&(xvalue>=uppin))
            y = uppout;
    else if ((xvalue<neutin)&&(xvalue<=lowin))
            y = lowout;
    else if (xvalue >= neutin)
            {
            z=((xvalue - neutin) * scale2);
            temp = (neutout + z);
            y = round(temp);
            }
    else
            {
            z = ((xvalue - lowin) * scale1);
            temp = (z + lowout);
            y = round(temp);
            }
    return(y);
}
```

C.4 Function To Produce Look Up Tables For Non-Linear Compensation Conditioning

```
/*************************************************************************/
/*                    function : createlookup                            */
/*                    objective : to produce lookup table for nonlinear conditioning */
/*                    arguments : name and dir of file storing control variation */
/*************************************************************************/
void createlookup(char *name, char *dir)
{
int port,temp;
int in,row,x;
int value;
int conslow,consup;
struct values temparray;

   openreadfile(name,dir);
   row =0;
   for (port=0;port<chan;port++)
         {
         fscanf(fin,"%d\t",&value);
         maxin.reading[port] = minin.reading[port] = nlininp.table[port][row] = value;
         }
   for (row=1;row<range;row++)
         {
         for (port=0;port<chan;port++)
               {
               fscanf(fin,"%d\t",&value);
               nlininp.table[port][row] = value;
               if(value > maxin.reading[port])
                     maxin.reading[port] = value;
               else if (value < minin.reading[port])
                     minin.reading[port] = value;
               }
         }
   rewind(fin);
```

```
    for (port=0;port<chan;port++)
            {
            if (nlininp.table[port][0]-nlininp.table[port][(range-1)] > 0)
                    {
                    for (row=0;row<range;row++)
                            temparray.table[port][row] = nlininp.table[port][row];
                    x = range-1;
                    for (row=0;row<range;row++)
                            nlininp.table[port][row] = temparray.table[port][x-row];
                    }
    conslow = 0;
    consup = 255;
    for (in=0;in<range;in++)
            {
            if (in <= minin.reading[port])
                    nlinout.table[port][in] = conslow;
            else if (in >= maxin.reading[port])
                    nlinout.table[port][in] = consup;
            else
                    {
                    row=0;
                    while ((nlininp.table[port][row] != in)&&(row<range))
                            row ++;
                    if (nlininp.table[port][row] == in)
                            nlinout.table[port][in] = row;
                    else nlinout.table[port][in] = nlinout.table[port][(in-1)];
                    }
            }
    }
    fclose(fin);
}
```

C.5 Final Listing Of The Program " PLAYRECORD "

This function performs the final output of the conditioned playback data and the input of recorded data from the sensor system.

```
/*********************************************************************/
```

```
/*                          function : playout                                  */
/*objective : to play out conditioned data and record input data at same time.  */
/*                          arguments : none                                    */
/*******************************************************************************/
void playout()
{
int del[] = { 20, 10, 100, 2, 1, 0 };          /* array of possible msec time delays */
long loop;
int port;
struct datavalues temprec;                     /* structures for initialization of */
struct datavalues tempplay;                    /* record & play data arrays      */
char rubbish[2];
Timer total;                                   /* constructs a stopwatch */

   openwritefile(rectempfile,recdir);          /* reopen file ready for record data*/
   openreadfile(playtempfile,playdir);         /* reopen open play file */

/* actual recording */

   printf("press return to start recording..");
   gets(rubbish);

/* for each reading, all 16 ports are scanned into the array at the            */
/* same instant using the C command inp(port addr.). The array is then         */
/* written to the temp file where it is stored in binary form. Before the      */
/* next reading is taken, the C library function delay(x) is called which      */
/* suspends execution for x msec. This is repeated until the number of         */
/*  desired samples have been taken.                */

/* start recording */

   loop=0;
   total.reset();              /* reset stop watch */
   total.start();              /* start stop watch */

   while(loop<sample)
         {
         fread(&tempplay,sizeof(tempplay),1,fin);
         for(port=0;port<chan;port++)
               {                                                  /*read one set*/
               outp((0x310 + port),tempplay.reading[port]);       /*of values*/
```

```
                temprec.reading[port] =inp(0x300 + port);
                }
        fwrite(&temprec, sizeof(temprec),1,fout);
        delay(del[SETTING]);                                    /*wait for x millisec*/
        loop++;
        }

    /* stop the watch and print out the time as check */
    total.stop();
    printf("length :%d, actual time:%.3f\n",length,total.time());

    fclose(fout);           /* close the temp files */
    fclose(fin);
}
```

Appendix D

Principles Of Error Analysis

"Because no measurement is free from error, steps must be taken to evaluate the accuracy and precision of the measurement. Precision is descriptive of the degree of care and refinement employed in making a measurement and accuracy is defined as the conformity of the measured value with the true value" **[Barry78]**.

The following methods were developed to ensure that the precision of the experimental procedure had minimal effect on the accuracy of the recorded data, allowing the results to be considered a true representation of the system characteristics. For full discussion on error analysis in practical measurements refer to **[Barry78]**.

In any experiment, it is important to separate the results from possible errors and the most obvious technique is to produce repeated tests. These also allow evaluation of the reliability of the data recorded and provide indications of the consistency of a system. This number should be sufficient to produce a distribution of the data sufficient to be considered as indictative of the actual function. For a series of repeated measurements, the true value can be defined as

$$x_{true} = x_{mean} \pm \frac{\sqrt{\Sigma v^2}}{n}. \qquad \text{Equation [D.1].}$$

or in expanded form,

$$x_{true} = \frac{(x_1 + x_2 + \ldots + x_n)}{n} \pm \frac{\sqrt{v_1^2 + v_1^2 + \ldots + v_n^2}}{n} \qquad \text{Equation [D.2].}$$

where x_n is the sampled value and v_n is the residual value equal to $x_n - x_{mean}$.

The standard error, σ_m, can be classified as a measure of the error of the mean of a series, if n is sufficiently large. Therefore

$$\sigma_m \cong \frac{\sigma_s}{\sqrt{n}} \cong \frac{\sqrt{\sum v^2}}{n} \cong \sqrt{\frac{\sum v^2}{(n-1)}} \qquad \text{Equation [D.3].}$$

which leads to the equation $x_{true} = x_{mean} \pm \sigma_m$. Equation [D.4].

In this equation x_{mean} represents the most likely value for x and $\pm\sigma_m$ indicates that there is a 68.3% certainty that the true value lies between $x_{mean} \pm \sigma_m$ **[Barry78]**. $\pm 2\sigma_m$ indicates that there is a 95.5% certainty that the true value lies between $x_{mean} \pm 2\sigma_m$ and $\pm 3.29\sigma_m$ indicates a 100% certainty that the true value lies in the defined range. These ranges represent the tolerance for the data recorded and indicate the precision of the measurement technique.

[Barry78] stated that the standard error of the mean of several means can be defined as $\sigma_{mT} = \sqrt{\frac{(\sigma_{m1})^2 + (\sigma_{m2})^2 + \ldots (\sigma_{mn})^2}{n}}$. Equation [D.5].

This represents a measure of the reliability of the overall test procedure.

Appendix E Design Of Multiple Analogue Interface Board

This board was designed by the researcher to assist in the data acquisition from different sources of control input whilst allowing different types of control output. The motivation of this design was to remove the need to constantly change between connectors. The overall system design is shown in Figure E.1. The board allowed for different input and output inter-connections between the various control inputs, the data acquisition board in the PC and the Henson Performance Control System.

The different inputs were as follows;

a. Sensor System on LIVE face ($\underline{S_l}$) and Sensor System on REPLICA face ($\underline{S_r}$) each with up to 8 input channels and common ground connection, suitable for upper or lower face only; or

b. Multiple Hand Control Input System ($\underline{C_{hand}}$) with up to 16 input channels and common ground connection. A positive supply rail, $+V_{cc}$ (5 volts) was connected for linear potentiometer scale.

The HPC system has up to 24 input channels allowing the following connections;

a. Multiple Hand Control Input System ($\underline{C_{hand}}$) at channels 0 to 15; or

b. Direct Software Control from the PC ($\underline{c}$) at channels 0 to 15; and

c. Sensor System on LIVE face ($\underline{S_l}$) at channels 16 to 23.

Depending on the type of control required either $\underline{C_{hand}}$ or $\underline{c}$ were physically connected between the interface board and the HPC system input.

The interface board therefore allowed the transfer of different combinations of input and output data dependent upon the required operation and the software used.

In the final system, the majority of data input was in the form of control from both sensor systems, $\underline{s}$, and the data output was in the form of direct software control, $\underline{c}$.

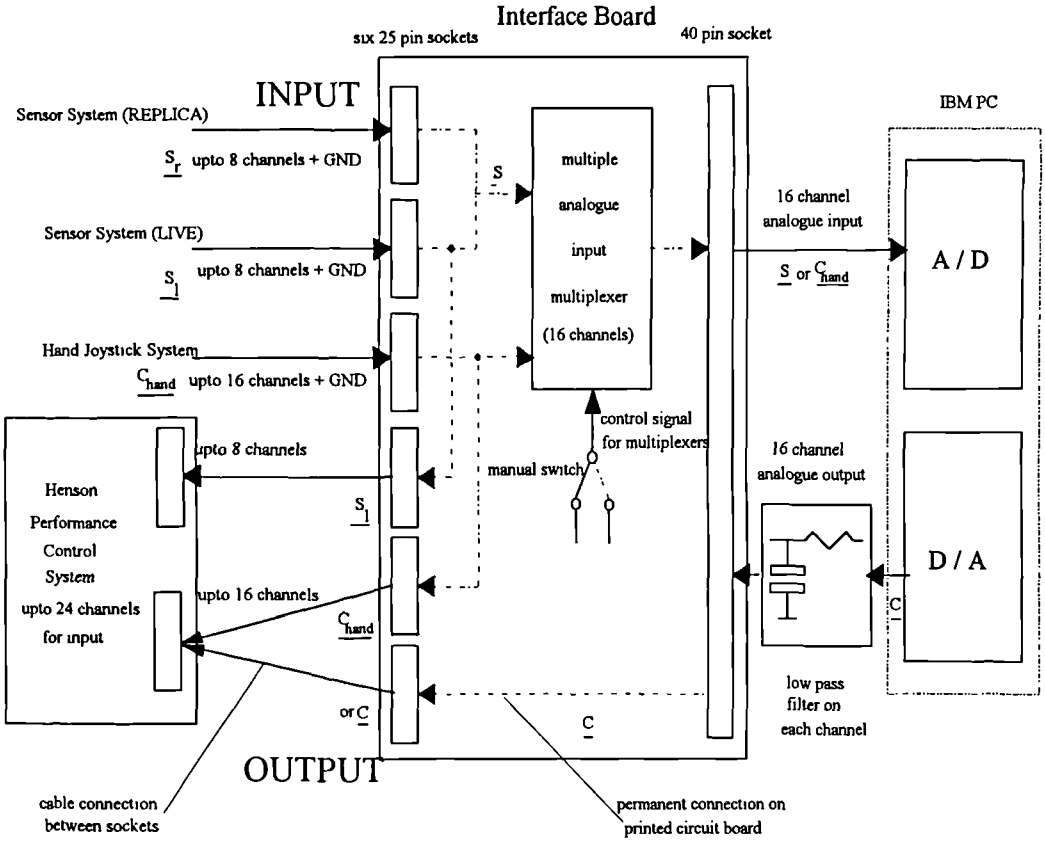


Figure E.1 Block Diagram Of Data Interface Board and Other Inter-Connections

E.1 Operation Of Triple Two-Channel Analogue Multiplexer

The device used, MM74HC4053, was an analogue multiplexer incorporating built-in level shifting. The level shifting allows a control input between GND and V_{cc} (+5V) for an analogue signal range of V_{ee} (0V) to V_{cc} (+5V). The device used has three separate control inputs and a common enable / inhibit input to allow independent selection of one of two channels in one of three sections. Within this specific design, all sections use common input signals. The logic table is shown in Table E.1 and the logic diagram is shown in Figure E.2.

enable / inhibit input	Control input A	Control input B	Control input C	Output Channels Turned On
1	*	*	*	none
0	0	0	0	A_x, B_x, C_x
0	1	1	1	A_y, B_y, C_y

x represents the inputs from the Sensor Systems ($\underline{S_l}$ and $\underline{S_r}$)

y represents the inputs from the Hand Control System $\underline{C_{hand}}$.

Table E.1 Logic Table For Triple Two-Channel Analogue Multiplexer

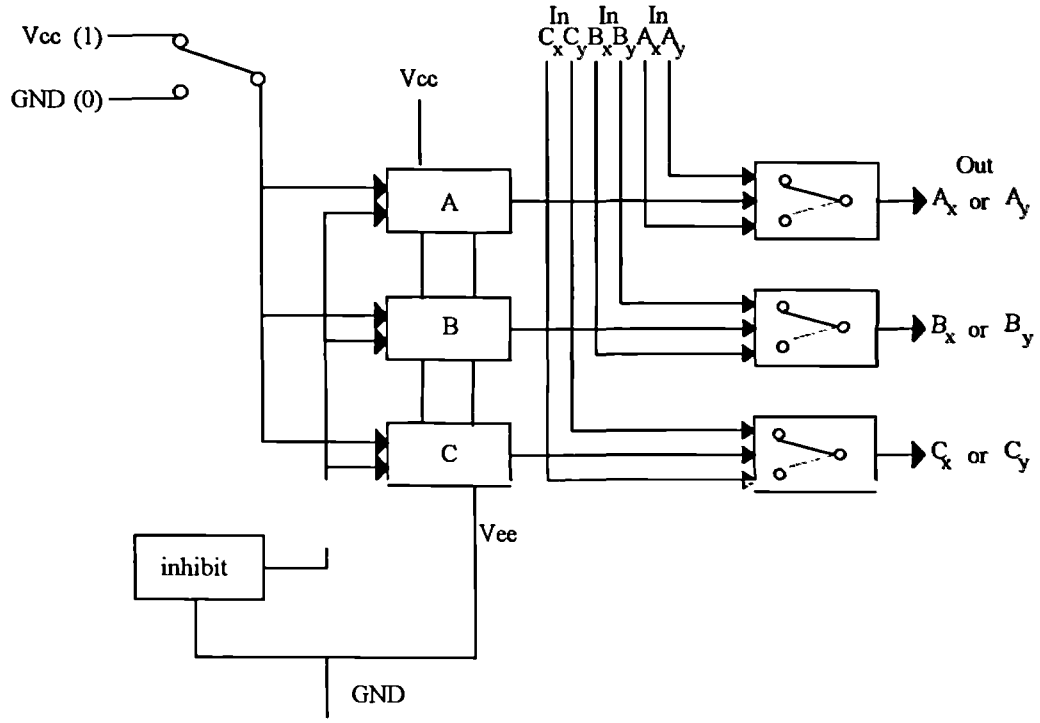


Figure E.2 Logic Diagram For Triple Two-Channel Analogue Multiplexer