The AMODEUS Project
ESPRIT Basic Research Action 7040

Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge

Phil Barnard and Jon May

21st November 1994

BC 13

published as:

Barnard, P. & May, J. (1995) Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge. In F. Paterno' (ed) *Eurographics Workshop on the Design, Specification and Verification of Interactive Systems*, pp. 15-48. Berlin: Springer Verlag.

2 Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge

Phil Barnard
Jon May

ABSTRACT
Advanced graphical interfaces are increasingly dynamic, multimodal and involve multi-threaded dialogues. This paper provides a theoretical perspective that can support an analysis of the issues involved in their use: the Interacting Cognitive Subsystems (ICS) framework. This framework is used to examine alternative ways in which information from different data streams can be blended within perception, thought and the control of action. The potential applicability of the core constructs to interface design is considered. The paper concludes by outlining a specific strategy for bringing this form of understanding into closer harmony with the formal methods community in computer science.

2.1 Introduction

2.1.1 The design problem

As human interfaces to information technology become increasingly advanced, the representational and communicative capabilities they embody have broadened considerably. We can no longer consider interfacing to be a simple matter of issuing commands – by lexical or graphical means – and ensuring that a user understands the consequent change in display state. Advanced graphical interfaces are increasingly dynamic, multimodal and involve multi-threaded dialogues. These interfaces may incorporate video communications technologies, computer controlled films or dynamic animations, intelligent agents, voice input and so on. The end users of such technologies must actively interpret what their full range of senses tell them, remember what a range of computer and human agents are doing, and carefully craft rather complex response patterns – often 'social' in nature. As designers, software engineers must now develop complex communicative environments. In doing so they must pose and answer a full range of questions about the complete interactive system, incorporating devices and users, operating in one or another domains of application – such as air traffic control, computer-aided design, computer supported co-operative work, or computer games.

In thinking about the issues raised, it is natural first to consider the domain tasks and explicit knowledge that a user must already have or needs to acquire in order to use an interface effectively. What is the relevant domain and task knowledge? How is the deployment of such knowledge constrained by limitations on human memory or upon human abilities to do more than one thing at a time? Providing answers to such questions has been the traditional contribution of human factors and cognitive engineering approaches to the design, development and evaluation of computer interfaces.

In the context of advanced systems it is also necessary to consider how the human mental mechanism integrates information over modalities (voice and vision) or over sources and locations (as when a pointing gesture is used to resolve reference to an item). Tone of voice, facial expression or accompanying gestures may radically alter the appropriate interpretation of a message and other factors such as the attribution of agency. The perceived emotional or social status (yes – even of computers!) can radically change the properties of interactions.

Much of the relevant human knowledge that governs these considerations is ‘latent’ and its analysis lies outside the capabilities of most current approaches to cognitive engineering.

2.1.2 Latent knowledge and its relevance to the design of advanced graphical interfaces

While our overt knowledge of the procedures and properties of interfaces helps to determine what we can achieve with them and what we can explain to, or teach others, latent knowledge plays a powerful role in user performance. It acts throughout the human mental mechanism systematically to constrain perception, thought and the control of action. For example, when things happen in the world, they are typically perceived as unified ‘events’. We see a flock of birds fly across our field of view; we hear someone shout our name; or we may feel drops of rain on our face. We can interpret these events as unitary even though the information that we receive about them is clearly not unitary: constituent information can be separated in space and in time, or it can originate in different sensory ‘modalities’. In a tennis game, we may see a ball approaching and, at the point of contact, both hear and feel its impact on the racquet. We are able to bring constituents together by learning to recognise the invariant structural patterns that typify events ‘in the world’ [28].

There is a substantial literature on the unification of information within individual modalities like vision [18] or audition [11, 17]. Although rather less is known about it, there is more than ample evidence that multimodal integration supports a whole range of human understanding. When we watch an accomplished ventriloquist, there is a spatial separation between the source of the voice and the position of the dummy. The audience will nevertheless understand the action as a single fused event sequence focused on the dummy. For decades billions of people have been happy to watch events on television while the sound accompanying them has come from a small speaker to one side of the screen. A thunderclap follows a flash of lightning by an unpredictable number of seconds, but we can understand them as a consequence of the same event, and can even make productive use of information about the temporal disparity. Although we can be aware that multimodal integration is taking place, the latent knowledge that determines its real time operation is often not readily accessible. Some explanations rely upon very general abstractions like the Gestalt principles of common fate and proximity. So, for example, Radeau [41] argues that “the rules underlying auditory-visual pairing could state that if elements from two different sensory modalities have the same temporal patterning, are asynchronous and come from locations which are not too far apart in space, then they can probably be assigned to the same event”.

The consequences of systematic departures from normality may be all too accessible. The misalignment of voice and lip movements that occurs when the soundtrack of a film goes out of synchrony is highly disruptive, because we ‘see’ one speech stream and ‘hear’ another. Other instances may be equally inaccessible to conscious awareness. When the speech information we hear is at variance with the lip movements of the speaker, a listener may ‘hear’ what was seen, or may blend the two sources to perceive a word that was neither seen nor heard (e.g. [30]). If we place a finger on one hand in a glass of hot water and a finger on the other hand in a glass of cold water until temperature adaptation occurs, then place both fingers into a third glass of warm water, the resulting sensations conflict. The *same* water feels cold to one finger and hot to the other (see [37]).

Although some of these examples may appear somewhat esoteric, interest in the fusion of information within and between human sensory modalities is not simply of academic interest. Multisensory fusion is vital in advanced robotics and, of course, in virtual reality and telepresence systems. Current computer system design is preoccupied with developing multimedia and multimodal forms of interaction. Video, dynamic animation, speech modes, head ‘mice’, data gloves, force feedback and other forms of haptic information exchange are increasingly being bolted on to traditional text and gesture based interactions. Designers keen on illustrating how such technology might play a part in a tourist information system might,

for example, show how two, very small video windows of different parts of the town can be simultaneously running on the screen, accompanied, of course, by a single informative voice over. At the same time, the user may actually be accessing a database of train times. All of these 'new' classes of system involve the management of multiple 'streams' of information, either within the same or in different modalities. At the heart of the vast majority of the systems are advanced graphical interfaces. Observation of existing 'demonstrator' applications suggests that computerised 'events' may not always be accompanied by an appropriate fusion of related information or by an appropriate separation of unrelated streams. If such systems are to be effective, they must be kept within the limits of normal human capabilities for handling multiple sources of information.

2.1.3 Structure of the paper

This paper will be broadly concerned with the capabilities, and limitations of the deployment of latent human knowledge in interface usage. The paper will not provide direct advice to designers on specific issues. Rather, the intention here is to provide a broader view of how the human information processor may function to deploy its latent knowledge in perception, thought and the control of action. Our wider objective is to develop an explicit account of the uni-modal and cross-modal blending of information. In the specific context of HCI, this endeavour contributes directly to the development of techniques for the approximate modelling of cognitive activity based upon Interacting Cognitive Subsystems – ICS (e.g. [5, 34]).

The remainder of this paper is divided into three main sections. The next section is theoretical and reviews those features of the ICS framework that are relevant to the blending of information sources to form unified streams of data. The following section (2.3) focuses on the four types of representational blending that occur within the ICS architecture. In each case potential applicability to interface design is discussed. A summary of the main points to emerge from our preliminary application of the ICS framework to these issues is presented in section 2.4, which also outlines an agenda for bringing this form of understanding into closer harmony with the design and formal methods communities in computer science.

2.2 Interacting Cognitive Subsystems

Interacting Cognitive Subsystems (ICS) is a part of a theoretical movement within cognitive psychology (e.g., [47]) that represents the human information processing mechanism as a highly parallel organisation with a modular structure. The ICS architecture contains a set of functionally distinct subsystems, each with equivalent capabilities, yet each specialised to deal with a different class of representation. These subsystems exchange representations of information directly, with no role for a 'central processor' or 'limited capacity working memory'. The assumption is that we are dealing with a *system* of distributed cognitive resources, in which behaviour arises out of the co-ordinated operation of the constituent parts. In order to be able to discuss specific issues concerned with the different forms of representational blending, it is essential to have a general overview of the architecture and its operation.

2.2.1 The systemic organisation

As a fundamentally systemic approach to mental processing, ICS is comprehensive: it encompasses all aspects of perception, cognition, and emotion, as well as the control of action and internal bodily reactions. Acting together, nine component subsystems deal with incoming sensory information, structural regularities in that information, the meanings that can be abstracted from it, and the creation of instructions for the body to respond and act both

externally, 'in the real world', and internally, in terms of physiological effects. Figure 2.1 outlines the overall architecture whilst Table 1 lists the nature of the mental representations that each subsystem processes. The subsystems can be classed as *peripheral* if they exchange information with the world via the senses or the body (the Sensory subsystems AC, VIS, and BS; and the Effector subsystems ART and LIM), or *central* if they only exchange information with other subsystems (the Structural subsystems MPL and OBJ; and the Meaning subsystems PROP and IMPLIC).

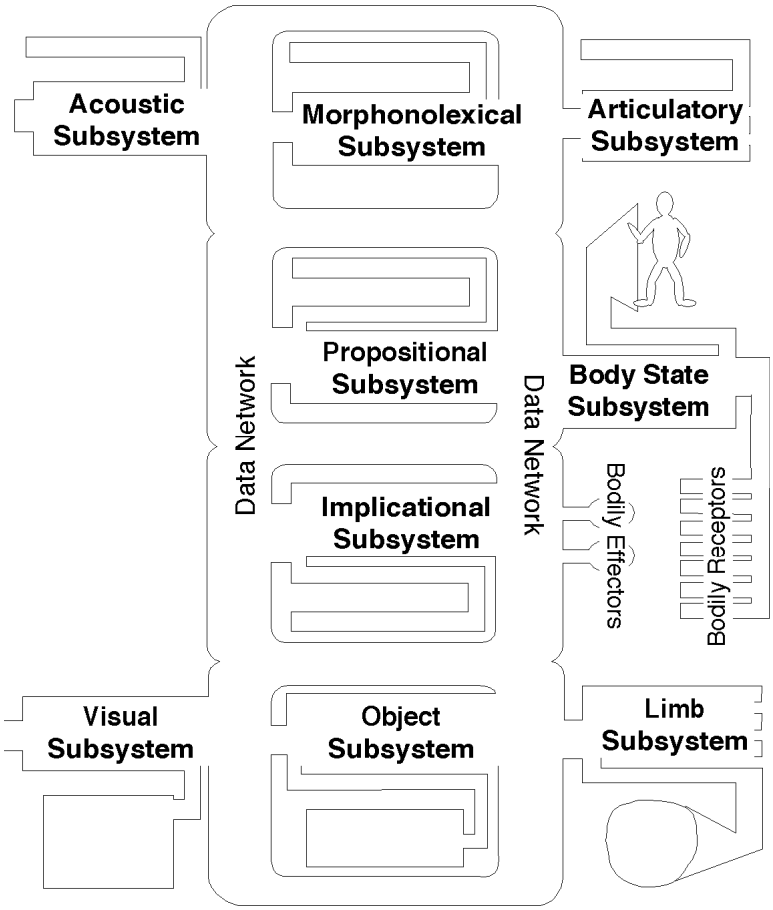

Figure 2.1: The systemic organisation of ICS

2.2.2 The internal structure of each subsystem

Although each of these subsystems deals with a different class of information, they have the same basic internal structure (Figure 2.2). The representations they receive arrive at an *input array*, where they are *copied* into an *image record*, while simultaneously being operated on by a number of *transformation processes*. The input array brings together all the information currently represented in the form appropriate for its subsystem, and so its content is dynamically changing from moment to moment. The Copy process continually transfers this information, without changing it, to the Image Record, which acts as a memory 'local' to the subsystem. Any representation that has ever been received at the input array of the subsystem is registered in the record, and in the long-term any communalities and regularities of the varied representations can be abstracted from this stock of past experience. In the shorter-term, it presents an 'extended representation' of the transient information on the input array, making available derivatives such as a rate of change of a representation or repeated occurrences of the same patterning within the constituent structure of those representations.

PERIPHERAL SUBSYSTEMS		
a) Sensory		
(1)	Acoustic (AC):	Sound frequency (pitch), timbre, intensity etc. Subjectively, what we 'hear in the world'.
(2)	Visual (VIS):	Light wavelength (hue), brightness over visual space etc. Subjectively, what we 'see in the world' as patterns of shapes and colours.
(3)	Body State (BS):	Type of stimulation (e.g., cutaneous pressure, temperature, olfactory, muscle tension), its location, intensity etc. Subjectively, bodily sensations of pressure, pain, positions of parts of the body, as well as tastes and smells etc.
b) Effector		
(4)	Articulatory (ART):	Force, target positions and timing of articulatory musculatures (e.g., place of articulation). Subjectively, our experience of subvocal speech output.
(5)	Limb (LIM):	Force, target positions and timing of skeletal musculatures. Subjectively, 'mental' physical movement.
CENTRAL SUBSYSTEMS		
c) Structural		
(6)	Morphonolexical (MPL):	An abstract structural description of entities and relationships in sound space. Dominated by speech forms, where it conveys a surface structure description of the identity of words, their status, order and the form of boundaries between them. Subjectively, what we 'hear in the head', our mental 'voice'.
(7)	Object (OBJ):	An abstract structural description of entities and relationships in visual space, conveying the attributes and identity of structurally integrated visual objects, their relative positions and dynamic characteristics. Subjectively, our 'visual imagery.'
d) Meaning		
(8)	Propositional (PROP):	A description of entities and relationships in semantic space conveying the attributes and identities of underlying referents and the nature of relationships among them. Subjectively, specific semantic relationships ('knowing that').
(9)	Implicational (IMPLIC):	An abstract description of human existential space, abstracted over both sensory and propositional input, and conveying ideational and affective content: *schematic models* of experience. Subjectively, 'senses' of knowing (e.g., 'familiarity' or 'causal relatedness' of ideas), or of affect (e.g., apprehension, desire).

Table 1. The subsystems within ICS and the type of information with which they deal (based on [6]).

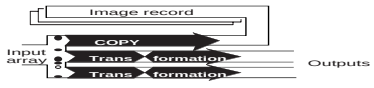

Figure 2.2: The internal structure of each subsystem

While the Input Array, the Copy process and the Image Record are crucial in defining the internal function of each subsystem, the Transformation processes are the key to the function of the overall, systemic organisation. In normal operation, these transform the information represented on the input array into a different representation, for use by another subsystem. The VIS subsystem, for example, contains a VIS→OBJ transformation process that uses the information contained within the sensory, Visual representation to derive the more abstract,

Object representation. The transformation processes within a subsystem are independent, and so can act simultaneously, on the same part of the input array, or on a different part. A consequence of this is that each subsystem can produce multiple, different output representations at the same time, each from a different transformation process. A single transformation process can only produce one output at a time, though, since it can only process a single coherent stream of data. This constraint is an important one, and the definition of a 'coherent' stream will be discussed later.

2.2.3 Configurations of processing

Since the subsystems are specialised to receive information represented in a particular way, none of these subsystems can do much on its own. In practice, cognition is the consequence of several subsystems functioning in a chain, or configuration, each taking its input representation and producing an output representation for use by a subsequent subsystem. The parallel nature of the architecture means that information 'flows' through this configuration, rather than pulsing in steps. The flows that are possible are defined by the outputs that each subsystem can produce, a crucial constraint of the architecture. The particular transformation processes 'allowed' within ICS have been systematically derived. Only those processes that are both logically plausible (given the nature of the information held within each representation) and empirically justified (given the experimental and phenomenological evidence about human cognition) are contained in the architecture. The internal structures, transformation processes and systemic organisation of ICS are brought together in Figure 2.3.

The subsystems dealing with vision (VIS) and hearing (AC) produce one output reflecting the structural regularities of their native representations (for OBJ and MPL respectively), and another output reflecting its global patterning (for IMPLIC). The latter outputs provide information of both 'cognitive' and 'affective' significance. The structural subsystems (OBJ and MPL) produce effector representations (for LIM and ART, respectively), and derive the referential, or semantic, level of meaning (for PROP) from their native representations. Note that these two subsystems do not produce Implicational representations. To enable skilled reading and naming of objects, the OBJ subsystem is able to recognise the shape of words and letters and to produce their corresponding MPL representation. However, it is assumed that there is no corresponding inverse transformation (from a word to its shape, MPL→OBJ). The third sensory subsystem is the Body State subsystem (BS). This subsystem can produce Implicational representations that reflect bodily and skeletal states. It can also produce effector representations (ART and LIM) to provide proprioceptive feedback during motor action. Unlike the processing of visual and acoustic information, there is *no* subsystem corresponding to the structural level (OBJ & MPL) for body state information.

Of the two Meaning subsystems, IMPLIC uses its schematic representations to derive a more detailed referential meaning appropriate to the current situation and state (for PROP) and to produce affective responses in the body via the SOM (somatic) and VISC (visceral) representations. These are not directly received by any of the other subsystems, but the consequences of the action of these processes are, of course, indirectly sensed by BS. The referential level of meaning, PROP, is unique in neither receiving representations from sensory subsystems, nor producing effector representations. Instead, it receives representations from each of the other central subsystems, and produces output for each of them. This gives it a central role in many of information flows underpinning thought. The remaining two subsystems, ART and LIM, directly control the articulatory and skeletal musculatures, and do not produce representations for other subsystems. However, as with somatic and visceral states, the bodily consequences of their outputs may be sensed by BS and fed back both locally and to the Implicational subsystem.

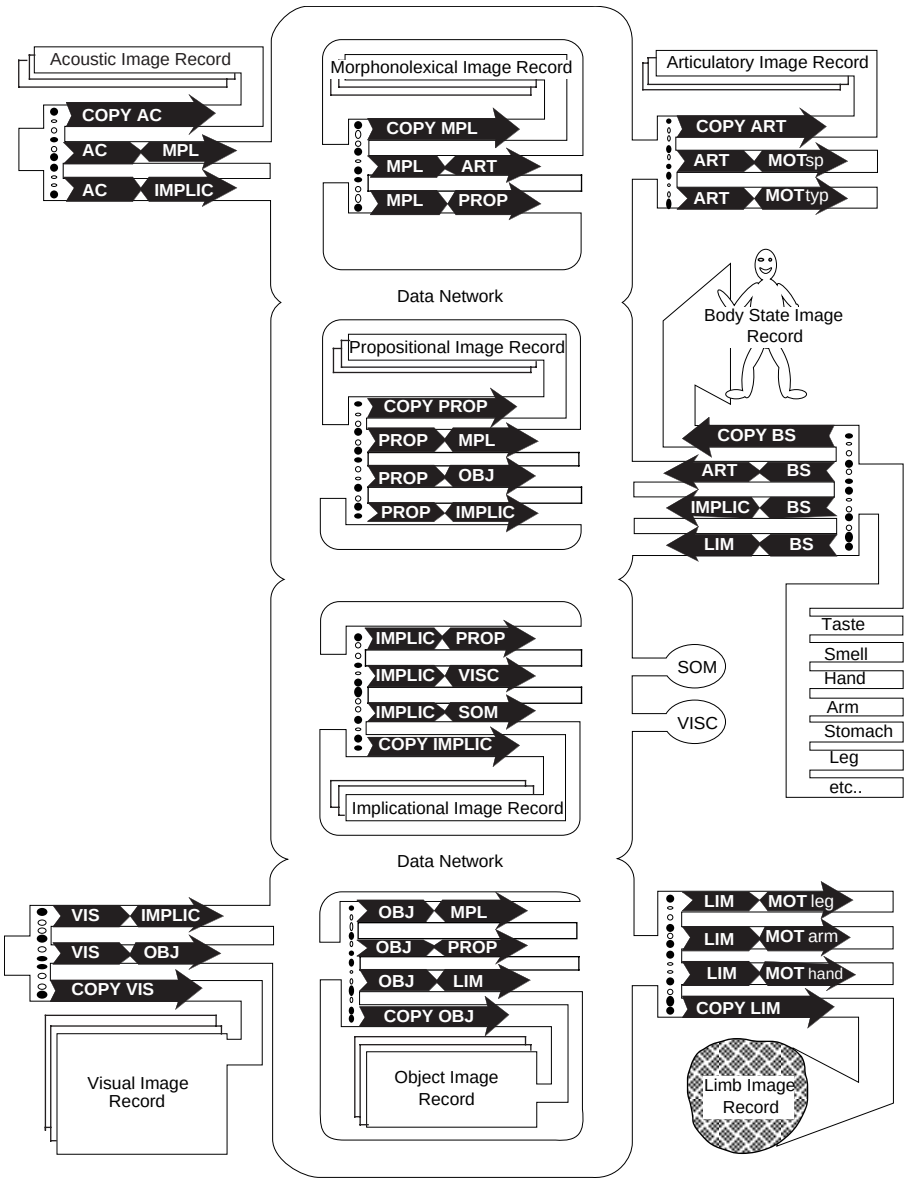

Figure 2.3: An overall view of the ICS architecture, showing the transformation processes.

2.2.4 Integrating information over sources and over time

The transformation processes shown in Figure 2.3 can be organised into a significant range of simple and complex configurations. The more complex configurations can include cyclical exchanges of representations between pairs and even triplets of central subsystems (e.g., PROP→IMPLIC & IMPLIC→PROP, and OBJ→MPL, MPL→PROP & PROP→OBJ). The configurations that are possible are systematically constrained by the availability of resources, since any given process can only be doing one thing at a time [3]. It should be also be clear from the figure that all central and effector subsystems receive inputs from at least two others. The sensory subsystems can also receive their inputs from multiple sensory 'transducers', in that we have two eyes and ears, and many sources of proprioceptive information. Consequently, any one process can receive input about the same event from different sources, each source may provide different information about that event. These are available to be blended together, or fused, to form a richer representation than any single source can provide. The potential for multiple inputs to subsystems provides the first way in which the integration

of information can be dealt with in the ICS architecture, and so allows us to reason about the types of multimodal information that can be cognitively useful.

Integration of information also occurs as a consequence of the internal structure of subsystems. In Figure 2.2, the operation of transformation processes was illustrated through the direct transformation of information on the input array into an output code. In addition, the transformation processes can also make use of information held in the image record and produce an output representation on the basis of these stored representations. This has several implications. Most obviously, it enables a form of episodic long-term memory, and supports the abstraction of regularities in the history of input to a subsystem. In the very short term, it also allows the transformation process to operate upon an 'extended representation' rather than upon the moment to moment contents of the input array. Since the Copy process is continually transferring information from the input array to the image record, this enables the Copy process and a transformation process to operate serially, with the image record acting as a short-term 'buffer' between the input array and the transformation process. The buffered mode of operation is mainly useful where the representation on the input array is changing too quickly for the transformation process to use, or where it is the nature of the change that is relevant to the transformation rather than the information itself. In this sense, buffered processing supports the integration, or blending of information, over time within a single source as well as over sources at one time.

One effect of the buffered mode of operation is to allow the transformation process to produce output at its natural processing rate, and to cope with variation in the timing of input data flow. A further constraint arising from the architecture of ICS is that, within a subsystem, only one transformation process can access the image record at a time, whether for revival of long-term records or for operation in buffered mode. It is also assumed that buffering helps co-ordinate the rate of flow throughout a configuration.

The differences between the direct and buffered modes of operation have subjective consequences. The Copy process gives rise to a general sense of conscious awareness of arriving information, but the operation of transformation processes is assumed not to be available to our conscious experience. Since there are normally Copy processes active in each of the subsystems, we can be diffusely aware of information at several different levels of representation, and we can even be aware of uncorrelated streams of information, depending upon the particular configural flows of information that are active at the time. In contrast, the buffered mode of operation corresponds to *focal awareness* of the extended representation: if any process within a configuration is operating in buffered mode, then the extended representation that it is operating upon will be the subjective focus of attention. Since only one process within a configuration can be buffered at a time, it follows that our focal awareness, will be restricted. We can nevertheless remain diffusely aware of the activity of the Copy processes throughout the system.

In summary, the architectural principles underlying ICS allow integration of information in two ways: in direct processing by the blending of information from different sources, and in buffered processing by transformation processes using an extended representation from the image record, hence supporting the integration of information over time. The consequences of the integration depend, of course, upon the nature of the information being integrated, and this corresponds to the subsystem at which integration is occurring. The differentiation between forms of integration consequently mirrors the differentiation between the types of subsystem, being divisible at a gross level between peripheral and central integration.

2.3 Peripheral integration

The peripheral subsystems (see Table 1) are those that exchange information with the physical world (both internal and external to the body). The Sensory subsystems (AC, VIS and BS) do

not receive representations from other subsystems, and the Effector subsystems (LIM and ART) do not produce representations for other subsystems to use directly. These two groups of subsystems are consequently assumed to have different roles in the integration of information. In brief, Sensory integration cannot be cross-modal, since each Sensory subsystem only receives information from its own set of transducers, while Effector integration is limited to the blending of intended actions with proprioceptive feedback.

2.3.1 Sensory Integration

As described above, the Sensory subsystems do not receive representations from other subsystems, but only from their sensory transducers. In consequence, these subsystems cannot integrate information from different modalities, although they can integrate multiple streams of information produced from within the capabilities of their respective transducers.

When an orchestra is playing, for example, the Acoustic subsystem can produce output that is a composite of the sound of all of the instruments, or it can 'focus in' on one particular sound, perhaps the strings or the percussion. Within these streams of information, it is not usually easy for a novice listener to focus in to attend to a particular violin, although it is usually possible to distinguish different percussive instruments. At a different level, a similar effect occurs in our processing of speech information. Different formants within a stream from a single source blend to form a percept that corresponds to a vowel sound. The knowledge deployed in these forms of integration is characteristically not available to conscious introspection, but is implicit or latent within the processing mechanism itself.

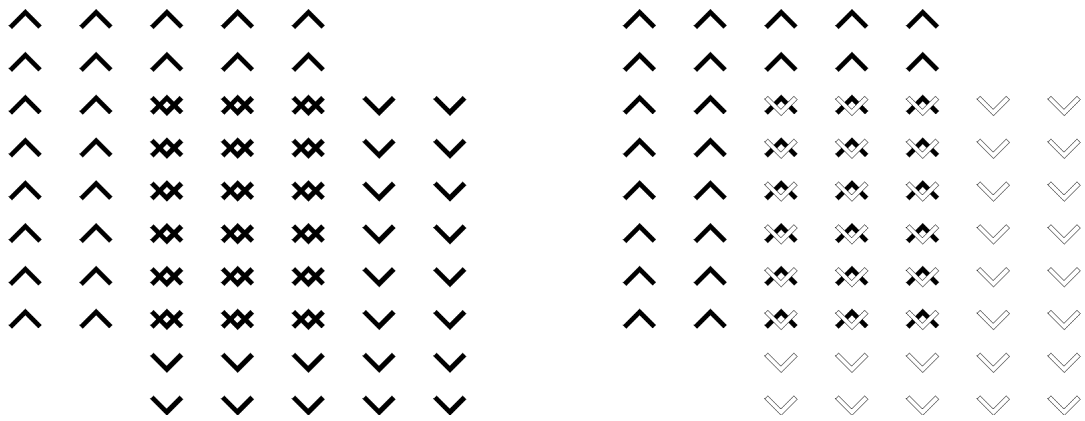

Figure 2.4: Visual integration (left) and differentiation (right)

Similarly, the Visual subsystem is able to operate upon representations of a flock of birds flying through the sky, and to attend to one particular bird. There are also situations where the Visual subsystem is unable to differentiate between different constituent elements of a visual scene. The left of Figure 2.4 shows an example of this: two overlapping rectangular areas forms, each composed of arrow-heads of different orientations. The Visual subsystem is easily able to resolve each constituent element of the areas where they do not overlap, but in the overlapping region, where their constituent elements intersect, a new 'xx' form is created which cannot be resolved visually into its two component parts (although it can be resolved 'conceptually', which will be discussed later). What the visual system 'sees' here is the integrated, or blended, form. The rectangular areas on right of Figure 2.4, where the arrowheads have different form, remain differentiated. Were the regions to be placed on separate overhead projector foils and one passed over the other, more dynamic forms of integration can be experienced with, for example, the precise perceptual 'blend' achieved from the pair on the left changing over time to resemble zig-zags and squares. The kinds of

blends that can be achieved within the sensory subsystems are constrained by the information provided by the receptors in terms of range and constituent form. So, for example, individuals can only hear or see things within well defined wavelength limits and internal resolution.

2.3.2 Effector Integration

Like the sensory subsystems, the effector subsystems (ART and LIM) communicate with the external world, in that they produce representations that cause the articulatory and skeletal musculatures to carry out actions. Unlike the sensory subsystems, their input arrays receive representations from other subsystems. Each receives one stream of representations from the appropriate structural subsystem (MPL→ART and OBJ→LIM) and another from the Body State subsystem (BS→ART and BS→LIM). The output of the structural subsystems defines the intended speech or action, while the output of the Body State subsystem represents proprioceptive feedback about the current state of the musculatures. Clearly, if the LIM subsystem is receiving representations from OBJ about a hand movement, it needs to have the information about the current position of the hand to be able to construct the appropriate motor representation. Similarly, the ART subsystem needs to know where the tongue is in the mouth and what the lips are doing before it can construct a representation for the production of the sounds specified by MPL.

Unlike the integration of representations by the sensory subsystems, where the same event, or consequences of the same event, were being brought to the sensory input arrays by different transducers, effector integration brings together two representations of different origin, and of potentially different content and structure. One will be internally derived, representing intended action, and one will be externally derived, representing the actual position and motion of the bodily parts. This is not a problem most of the time, for in normal co-ordinated performance the proprioceptive feedback reflects the actions that have just been carried out. As long as the structural subsystems are producing a coherent output data stream, and actions are being carried out roughly as intended, the proprioceptive feedback will be compatible with it, and so will be integrable.

In normal motor action or in speech, we are rarely aware of the many little compensatory adjustments that we continually make to our posture and movements, or of the many different ways in which we make a speech sound, depending upon the sound that has preceded it and the sound that is to follow it. Corrections to motor output can occur extremely rapidly, driven by the proprioceptive feedback from BS, without disturbing the ‘planning’ stream arriving from the structural subsystem. For example, if the lower lip is perturbed during speech, adjustments occur not only in that lip, but also in the upper lip, within 36-72 ms of the perturbation [1]. The adjustment occurs on the very first trial, showing that it is not learnt during the experiment, and its nature is determined by the speech that is being produced, showing that it is a functional response related to the planned speech, not a reflex. Monster, Herman & Altland examined the effects of adding a force to the ankle joint, finding that a load torque of 1.4 kg-m in either direction produced a 4° error in perceived position, with the sole of the foot feeling more flexed than it actually was [36]. Under ischaemic anaesthesia of the arm (cutting off blood flow to remove sensation), people report its perceived position to be closer to the body than it actually is, with the degree of error increasing as duration of ischaemia increases [25]. Muscular exertion can also produce distorting after-effects, as in the Kohnstamm effect. To experience this, stand in a doorway, with your arms at your sides. Keeping your arms straight, push against the door frame with the backs of your hands for 10 to 20 seconds. When you step out of the doorway, your arms will ‘feel’ light, and will seem to rise almost effortlessly. Once again much of the ‘key knowledge’ deployed by the human information processing mechanism is latent.

As with sensory integration, the precise forms of effector integration will be constrained by the nature of the human skeleton and musculatures – there are only two arms and two legs

with well defined degrees of freedom in the movements that can be achieved. Many of the points concerning the integration of information in the control of action may appear obvious as, indeed, may many of those points covered in the previous subsection on sensory integration. Nonetheless, achieving an understanding of the operation of these processes is of key importance to design. The dot matrix printer illustrates how technology can be developed to support the blending of constituents of characters into increasingly well defined and fully integrated form. In contrast, variability in peoples' sensory abilities frequently goes unrecognised in design. The capabilities of the visual and auditory coding space decline with age and other factors. The genetics of colour vision radically alter how colours are resolved by a particular individual. Many systems and information displays, like graphs, are produced in which it is very difficult for a significant proportion of users to distinguish colours that mark vital contrasts. Similarly, the environment in which perception and action occur may move beyond normal limits, as when a pilot is flying on instruments in altered gravitational conditions. In all these and other instances, it is important to understand the properties of the underlying mechanism.

As each new advance occurs, new challenges are posed with renewed opportunities for problematic design. In one relatively recent example, in order to get a large map onto a small screen, compression algorithms were proposed to enable a large area to be shown concurrently with a high resolution representation of a focussed part of that area. A user controllable fish-eye lens solved this "representational" part of the problem. However, the dynamic distortion of the objects and scales represented introduced substantial difficulties with the basic perceptual processing of the visual information [33]. In one sense, virtual reality and telepresence systems share an important property with the early dot matrix printer: a requirement to increase resolution in the relevant human coding spaces. A designer may have a choice between updating the constituent graphic elements of a basic visual unit individually as they are computed, or to wait and update the constituents of the whole unit simultaneously. The choice may radically affect how features are blended in both space and time.

2.4 The structure of representations

To explain why certain representations can be integrated but not differentiated, and others can be differentiated but not integrated, we need to consider their structure. The subsystems in ICS all share the same functional architecture. While the contents of their representations differ, all can be treated as following a common set of structural rules. ICS treats a representation as consisting of a number of basic units, which may have constituent elements (a substructure) and be grouped together in some way (a superstructure). One of the basic units will be the 'psychological subject' of the representation, and the others will form its predicate structure. The psychological subject is the 'theme' of the representation: the element of the scene to which attention is directed.

This applies for representations received by each of the subsystems. In the Visual example shown in Figure 2.4, the psychological subject of the representation may be one of the arrowheads, one of the double-cross forms, one of the larger rectangular areas, or the overlapping area (and there are other possibilities). If an arrowhead or a double-cross is the psychological subject, then the changing hues and brightness levels within it and around its edges will be its substructure; the rows, diagonals and columns it belongs to will form its superstructure; and the other 'pools' of brightness and darkness within these superstructural groups will form the basic units within its predicate structure. If one of the larger regions is the psychological subject, then its textural constituents will form its substructure; the other regions with differing constituent elements will form the predicate structure; and together the figure as a whole will form their superstructure.

Since the Visual subsystem deals with low-level information about brightness, hue, and so on, its representation cannot ‘split up’ the intersecting black lines of the double-crosses to ‘see’ the original arrowheads: to do this, we need the information that would be obtained by transforming the Visual representation into an Object representation, where lines or differing orientation can be represented. If the arrowheads were of different hues, however, the Visual representation of a double-cross would include substructural detail sufficient to allow only the parts in one hue to form the psychological subject: in the right of Figure 2.4, the arrowheads can be ‘seen’ within the double-crosses. The structure of the Object representations that can be produced from the figure are shown in Figure 2.5.

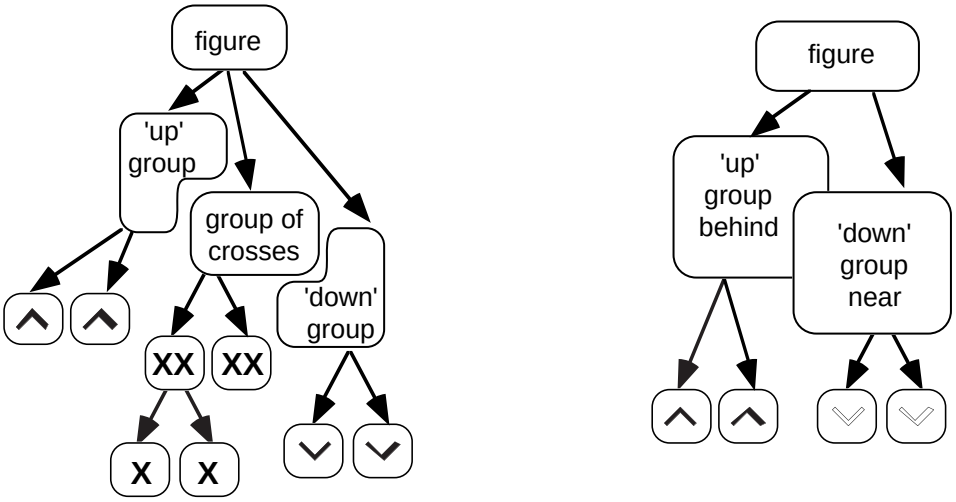

Figure 2.5: The structures of object representations resulting from Figure 2.4

In the Acoustic orchestral example, a psychological subject may be formed from the overall sound of the music, in which case the constituent ‘voices’ within the music would be the substructure. If one of these voices were taken as the psychological subject, the other voices would form the predicate structure, and the music as a whole the superstructure. The nature of the constituent elements of the voice would then determine whether or not a further ‘thematic transition’ could be made to bring one of them to the fore as a new psychological subject. In the case of the strings, their acoustic forms would probably not present enough discriminating detail to allow this transition to take place, and like the different arrowheads in Figure 2.4, they would ‘overlap’ to form an integrated whole. The acoustic forms of different percussive instruments might be more discriminable, and if so would be differentiated as basic units.

These two sensory examples have shown how differences in the external data can affect the theme or subject of the representation. The representations of the central subsystems can also be influenced by internally generated streams of information, and so thematic transitions can be ‘willed’, allowing attention to be directed. Even in the left of Figure 2.4, the double-cross forms can be ‘taken apart’ at a structural level to focus on the ‘up’ and ‘down’ parts separately. This ‘central integration’ is discussed further in section 2.5.

2.4.1 Coherent data streams

The notion of structure is the key to the definition of a coherent data stream. A crucial point in the arguments about integration of information is the constraint that a single transformation process can only operate upon a single coherent data stream at any given time. If two data streams cannot be integrated into a common representation, a transformation process will not be able to produce an output representation that is based upon both of them, and only one will pass onwards for subsequent processing. So in the classic case of the cocktail party

phenomenon [14], there may be many people speaking in the same room concurrently. However, at any given point in time, only one of these can be recoded through to the MPL subsystem and from there to subsequent processing of its meaning.

At the level of the sensory representations, the nature of the representations is largely a property of the neuroanatomy of the sensory transducers. Events in the 'real world' that cannot be discriminated by our senses necessarily form integrated basic units, and hence a single data stream, while discriminable events form differentiated basic units, and separable data streams. In Visual representations, brightness, hue and spatial organisation are the delimiting dimensions; in Acoustic representations, the frequency, timbre, intensity, and onset; in Body State representations, the type of stimulation, its location (and hence the density of receptors at its site), intensity, onset, and so on (see Table 1). At the central representations, the structure depends more upon the nature of the information flow, a factor to which the discussion will return later.

This should not be taken to mean that the separation or integration of data streams within the sensory subsystems is largely a matter of the underlying 'neural wiring'. Integration will occur because there are topographic and dynamic patterns that recur in the external world or the internal world of the body [28]. Within the ICS framework, a transformation process 'learns' to map inputs onto outputs. So for example, the Acoustic to Morphonolexical transformation (AC→MPL) extracts regularities in the underlying sound stream which, in this case, come to form key contrasts in the structural description of the 'sounds' used by the linguistic community in which the individual lives. These regularities are captured by distinguishing substructure, basic units and superstructures at each level of representation. For example, frequency patterns in sound over time represent the formant structures of speech sounds. While aspects of the substructures of these sounds will reflect properties of the human vocal tract, the way in which they are put together to form basic units of speech sounds and their superstructure depends upon experience.

As it moves through successive transformations, the nature of the information changes. When transforming a representation of sound (AC) to a structural description of sounds (MPL) the elemental acoustic information is discarded – the basic units of the input representation become the substructure of the output representation, and the superstructure of the sound stream indicates the basic units. Conversely, a transformation to an effector representation (e.g. MPL→ART) involves the reverse process, of taking an abstract structural description of the sound and computing more detailed articulatory representation necessary for the subsequent motor control of speech. This all depends upon experience with the appropriate sounds, and very young pre-speech children babble in the full range of human vocalisations, attempting to imitate the speech they hear. Eventually they hear themselves producing the appropriate set of sounds, and the inappropriate sounds cease to be part of their articulatory repertoire. Still later, we cease even to be able to differentiate between speech forms that are not used in our native languages.

The integration and differentiation of sensory representations according to these structural invariants necessarily constrains the structure of the output representations. In this respect, the products of processes constrain the higher level invariances that can be abstracted in exactly the same way as sensory transducers. Although these interdependencies are potentially rich, the architecture of ICS and its associated view of information representations enable us to frame theoretically motivated distinctions about the coherency of data streams. They also enable us to understand in more precise terms the part that might be played by the classic Gestalt principles of 'common fate' and 'proximity' that were mentioned earlier in section 2.1.2. Within the ICS framework, the principle of common fate can be defined more precisely as a property of sets of basic units within a representation (not necessarily a sensory representation). When basic units behave with highly similar characteristics over time, or are part of some invariant within the representational space, they can be interpreted as single superstructural element. Even though proximity as a concept is most obvious when applied to

spatial location, it can equally well be applied to auditory representations, as well as to central representations. To be useful, the dimensions on which units are 'proximal' must be thought of as those of the 'space' within which they are represented.

In summary, at any level of representation constituents represented on the input data array can be blended into a single data stream by a transformation process if they can be interpreted as a basic unit, given the learning history of the subsystem. The blending of units into a coherent data stream occurs when a transformation process imposes a superordinate structure upon the basic units. The imposition of this structure will again depend upon the learning history of the process and the ways in which basic units have come together in those structures in the past.

2.4.2 Integration of coherent data streams

Given the structural organisation of representations, we can now start to account for phenomena, of the sort introduced earlier, which involve the integration of information to form a single data stream. The birds flying in a flock across the sky can be interpreted in sensory representations (VIS→OBJ) as a unified superstructure of basic units with shared dynamic properties, as can points of light moving in a co-ordinated way [18]. Similarly, the formants in speech sounds can be represented as basic units, made up of energy distributions in sound space, with a superstructural organisation representing phonemes. Some blends occur because the information is not distinguished on the data array (e.g. low level localisation; psycho-acoustic fusion; critical flicker frequency), while other blends occur because information on the data array systematically co-occurs, giving rise to phenomena associated specifically with speech [17].

At the other extreme, the ICS architecture is capable of understanding thunder and lightning as constituents of the same event at a deeper, propositional level, in terms of their onset asynchrony. Features such as the intensity and unexpectedness of the event may have affective constituents that blend at an affective, Implicational level (see Table 1). The blending of information within the central subsystems is capable of occurring over a very much longer duration and is also likely to be 'interactive', incorporating the effects of the individual's own actions. Again, such 'interactivity' is not purely a property of the processing of information by central subsystems. Systematic effects of integration and differentiation occur at shorter durations. In between examples of psychoacoustic forms of integration and thunder and lightning, forms of integration that depend upon morphonolexical and articulatory representations would be expected.

Delaying auditory feedback of our own voices can have extremely disruptive effects on normal articulation of speech [29], maximally so at delays of around 200msec. At very small delays, or those exceeding around 300msec, the effects can be negligible. If the delay is small, buffered representations can be used to overcome the disparity, and to treat the acoustic feedback and the intended speech as part of the same data stream. At the longer delays, the acoustic feedback is so different to the intended speech that the two streams are clearly differentiated, and do not interfere. Between these limits, conditions are presumably created where the superstructure of the acoustic representation is similar enough to that of the intended speech to allow integration, yet the content of the basic units (intensity, for example) is far enough from the intended speech to suggest that there has been some problem in articulation (that the onset of the sounds was incorrect, for example). Interestingly, non-speech sounds can have equivalent effects when the overall rhythm of the feedback is driven by the speech output [27] – perhaps illustrating here how 'common fate' in the superstructure of acoustic representations can itself contribute to blending.

There can be additional problems in identifying data streams when there are multiple sources of feedback. Within the architecture of ICS, feedback about voice also occurs internally via the products of the BS→IMPLIC and BS→ART processes. The model therefore

suggests that blending of articulatory constituents can be influenced by affective and oral routes, in addition to auditory feedback. This helps us to understand why stutterers are often strongly influenced by anxiety [16] as well as oral sensory feedback. With instruction and practice, people can learn to cope with delayed auditory feedback, relying on the proprioceptive information from the mouth and tongue instead of auditory information to co-ordinate their speech [2].

2.5 Central Integration

As with the peripheral subsystems, integration of information by the central subsystems corresponds to the nature of the representations dealt with by the subsystems in question. The structural subsystems, MPL and OBJ, integrate information reflecting the structure of sound and visual space respectively, while the Meaning subsystems, PROP and IMPLIC, integrate referential and schematic information. Unlike the peripheral subsystems, the central subsystems can exchange information with each other, and can form cyclical configurations. The representations that they receive are highly likely to have been influenced in part by their own outputs. The potential for integration is correspondingly much greater, and so we shall describe each of the four central forms of integration in turn.

The OBJ subsystem receives input from visual processing (VIS→OBJ) and propositional processing (PROP→OBJ). It also acts as an interchange between information flow interpreting the visual environment (OBJ→PROP) and that involved in structuring output for the subsequent control of skeletal movement by the LIM subsystem (OBJ→LIM). No direct integration of multimodal sources occurs at this subsystem, since there are no sensory BS→OBJ or AC→OBJ processes, and no central MPL→OBJ process. Any effects of multimodal origin of information at this subsystem must therefore be indirect, with the configuration of information flow passing through the PROP subsystem.

In many ways, the MPL subsystem is an analogue of the OBJ subsystem, but in the domain of sound. It acts as an interchange point between the comprehension (MPL→PROP) and production of language (MPL→ART). In another important respect it differs quite markedly. The MPL subsystem takes crossmodal input (from OBJ→MPL) but produces no reciprocal output back to the OBJ subsystem. Unlike OBJ, it is possible for MPL to integrate multimodal information, although apart from Acoustic representations it is limited to receiving representations about word forms and the recognition of well-known objects or scenes.

Alone among the four central subsystems, the PROP subsystem neither receives information directly from sensory subsystems nor produces output for effector systems. Just as the OBJ and MPL subsystems act as a mediating point between sensory ‘input flows’ and effector ‘output flows’, the PROP subsystem mediates the flow of information between the structural subsystems and the higher level schematic representations of the implicational subsystem. It is able to do this in both directions, both receiving representations from the structural subsystems to produce IMPLIC representations, and receiving IMPLIC representations to produce structural representations. This gives it a pivotal role in most ‘cerebral’ configurations, and in all that include a cyclical loop. Although as far as multimodal integration is concerned, PROP is only able to blend the structurally organised output of MPL and OBJ rather than the lower-level output of AC and VIS, its ability to bring them together with representations derived from the schematic content of IMPLIC means that it is the key to giving implicit, latent knowledge a more explicit form.

In marked contrast, the Implicational subsystem is deeply multimodal. It receives input from sound (AC→IMPLIC), vision (VIS→IMPLIC), and proprioception (BS→IMPLIC), as well as the products of referential, semantic meaning (PROP→IMPLIC). The information received directly from the sensory subsystems is quite distinct from that which comes through the sequential analysis of sentences (via MPL and PROP) or the spatial analysis of visual

scenes (via OBJ and PROP). While these indirect representations convey considered, or inferred, affective content, the direct inputs from sensory subsystems concern broad 'gut reactions' to the general tone of voice, facial expression, arm gestures, and bodily arousal. In fact, as discussed below, the directness of these sensory inputs to IMPLIC means that the IMPLIC→PROP process may provide an output that can be integrated with the structural subsystems' output to PROP, before PROP is able to produce an IMPLIC representation that would give a more 'rational' schematic view of the representations' referential content.

2.5.1 Object Integration

As described in the previous section, inputs of direct multimodal origin do not arrive at the object subsystem. In order to understand many effects of intermodal conflict, such as the ventriloquism effect and its analogues [41], it is necessary to look elsewhere within the architecture. The object subsystem is able to produce different output representations, and since it is a central subsystem, it can operate in cyclical configurations with the propositional subsystem and this does have some consequences for integration.

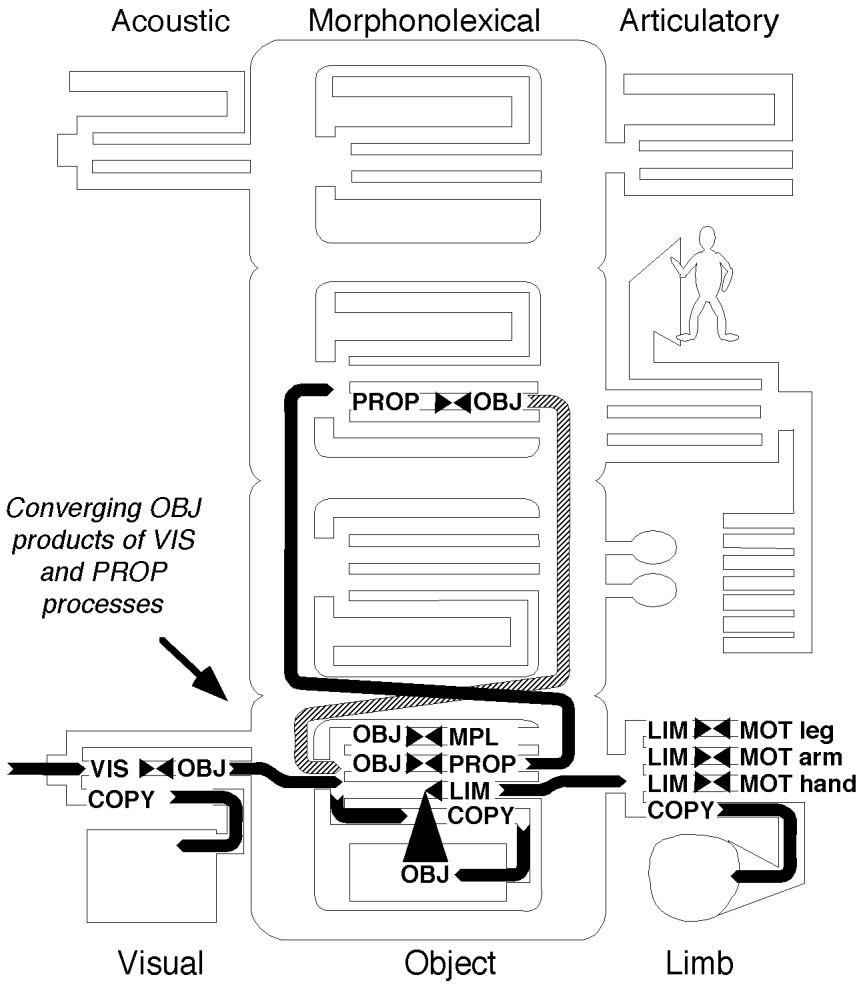

Figure 2.6: Converging object products of visual and propositional processing activity

Figure 2.6 shows a configural flow for motor action under the direct control of a visual source of information. This would be the sort of configuration controlling the hitting of a tennis ball on a particular trajectory in the real 3-d world or even in 2-d world of a computer game. The Visual representation of the ball, moving through space is transformed by the VIS→OBJ process, forming the primary input to the OBJ input array. The ball is almost

certainly moving too fast for the OBJ→LIM transformation to use this representation directly, and in any case the actual position of the ball is not as relevant as its velocity and acceleration, which are derivatives of these moment to moment representations. To have access to this information, and enable the individual's motor actions to begin in time for their racket to intercept with the tennis ball, the OBJ→LIM process must act in buffered mode, processing the extended representation in the image record.

At the same time as the OBJ→LIM process is producing its representations of intended action, the OBJ→PROP process is able to produce a referential interpretation of the structural representation. Since the OBJ→LIM process is accessing the image record, the OBJ→PROP process can only operate upon the information on the input array. This may mean that it is unable to keep up with the speed at which the representations arrive, and that the referential output of the process is not as smooth as it would be if it were buffered, having a more 'snapshot' nature. It also means that the referential representations it produces are not based upon the derivative information available from the extended representation, and so are different in content as well as nature to the products of the OBJ→LIM process. As the solid arrow in Figure 2.6 indicates, once they have been produced, they enter the input array of the PROP subsystem, where they can be operated on by the PROP→OBJ process. These centrally generated OBJ representations will now be based upon a referential interpretation of the visual scene, rather than the raw visual information. Since the OBJ→PROP process can operate on any part of the representations arriving at the OBJ input array, the products of this reciprocal PROP→OBJ process may now include information derived from the position of the opponent's arm as they hit the ball, or of their position on the court, instead of being based on the velocity of the ball. Whatever the content of the PROP representations, the PROP→OBJ process will use the individual's referential, rule-like experience of tennis playing to feed back new OBJ representations via the downward, hatched arrow in Figure 2.6.

Like the output of the VIS→OBJ process, the products of the PROP→OBJ process also arrive at the input array, and so provide an opportunity for information integration to take place. Just as with sensory and effector integration, central integration can only occur if the data streams are compatible: if they provide conflicting or irreconcilable information, the transformation processes will simply not be able to produce any useful output. In this example, the referentially mediated OBJ representations (of, say, the opponent's actions) are highly likely to be closely correlated with can be integrated with the representations derived from the raw visual information (of the consequences of their actions), and so they should form a single coherent data stream with few conflicts. The extended representation in the image record contains these integrable representations, not solely the products of the VIS→OBJ process. In consequence, the OBJ→LIM process of a skilled tennis player is able to operate on the basis of more than basic, visuomotor reactions. It can blend in the referential knowledge derived from their experience of playing tennis – for example, to recognise from their opponents' posture, action and position that a ball is likely to have spin, and so to predict its trajectory more accurately.

This is not the whole story, of course, for vision is not the only sense active during a game of tennis. As noted earlier, the impact of the tennis ball is also heard and felt. Bodily effects and the sound of the impact individually influence Implicational understanding (via BS→IMPLIC and AC→IMPLIC), and bodily effects also provide feedback to movement control (BS→LIM), but these sources cannot form an integrated data stream at the *object* level of representation, since there are neither direct paths from AC or BS to OBJ, nor from IMPLIC to OBJ. For these sources of information to be used by OBJ→LIM, they will have to pass through PROP, as we shall describe in the section on Propositional Integration.

In summary, the OBJ subsystem can integrate over bottom up, visual information (from VIS→OBJ) and top-down, referential information (from PROP→OBJ). This provides a way that propositional mappings and image records can come to influence the processes of visual interpretation. Standard texts (e.g. [24]) often make use of simple but powerful

demonstrations. Presenting scenes containing a Jersey cow or a Dalmatian dog as a pattern of dots or blobs against a similar background can make the 'object' percept hard to establish. In Figure 2.7, for example, the Visual information is inadequate for an Object representation to be formed easily. However, once the 'object' is known, it is subsequently impossible not to see the pattern involving that 'object'. In this case, where the picture shows two ducks swimming on the River Cam, stored referential knowledge about the appearance of ducks can be automatically mapped to the OBJ input array (PROP→OBJ) to constrain the interpretation of input.

Figure 2.7: Propositional knowledge about this scene can make it easier to form an object representation.

Similarly, in Figure 2.4, the double-cross forms in the left-hand overlapping region cannot be resolved at a visual level into their component arrowheads, but once we 'know' what they consist of, the PROP→OBJ input allows the integrated OBJ representation to 'attend' to either the upward pointing part of the form, or the downward pointing part (note, though, that it is harder still to attend to both at once). The same route allows the overlapping arrowheads in the right-hand region to be 'seen' as double-cross forms, albeit multicoloured. This is what was meant earlier by the word 'conceptually'. This form of propositional, top-down control over perception is a pervasive part of our everyday cognitive activity, normally helping us to impose order on what may be an incomplete or ambiguous sensory data stream. The many visual illusions and ambiguous figures (e.g., the Necker Cube) that fill textbooks on perception testify to its role.

The tennis example was useful in emphasising how the dynamic nature of cognition is central to the integration of information, since it results from the concurrent activity of independent processes transforming a flow of representations. The particular streams of representations being integrated in that example make it a little harder to discuss the role of structure, since there are many different tennis playing situations, and the role of PROP→OBJ depends on the skill and experience of the individual tennis player. The structure of the representations being provided to OBJ by VIS and PROP is still crucial, though, as the following, more static, example indicates.

In the left hand part of Figure 2.8, it is easy to see one of the elements 'popping-out' from its companions, yet in the right hand part, it is quite hard on first sight to find the unique element. The two arrays have been carefully controlled to balance their visual characteristics, and the substructural features of each element, but the superstructures are clearly different. Taking the left-hand side first, VIS→OBJ is able represent each element as a three dimensional cube, with the majority of them having the same orientation. The array as a whole

can consequently be seen as having the orientation of its constituent elements: the superstructure of the OBJ representation takes on the common attributes of its basic units. One of the elements does not have the same attributes however, and so VIS→OBJ cannot represent it as part of this superstructural group, even though it is spatially 'within' it. In consequence, it 'pops-out'. The bias of the VIS→OBJ process is to represent disparities within the visual scene as the theme of the representations it constructs for OBJ, since these are in practice likely to be the 'figure' against the 'background'. In this example the odd element out will typically become the psychological subject of the OBJ representation, and the rest of the array, with its common superstructural description, its predicate structure.

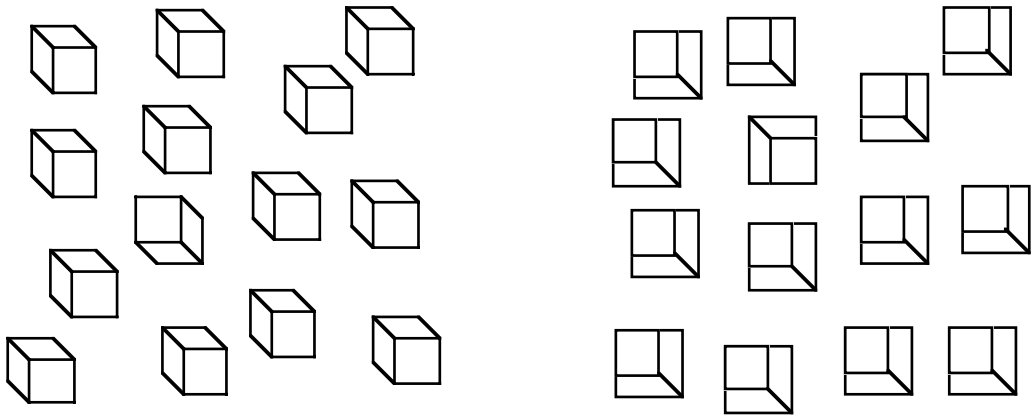

Figure 2.8: Examples of visual 'pop-out' that vary according to the visual and propositional information about the superstructure of the object representation

In the right hand part of Figure 2.8 the visual information is different. The basic units of the visual representation now no longer have the characteristics that allow VIS→OBJ readily to produce three-dimensional representations, and so it is able to produce a common, two-dimensional superstructure for the OBJ representation that adequately fits all of its basic units. None of them 'pop-out'; the incongruous element has to be found by searching the array. Here we can show the role of PROP→OBJ upon the formation of the structure of the integrated OBJ representation. As soon as you 'know' referentially that the elements in the right hand array are 'square holes in a flat surface', the OBJ→PROP & PROP→OBJ loop is able to add three dimensional information to the OBJ input array: the superstructure of the representation now becomes a three dimensional form with the basic units 'descending' into it. Now one of them can immediately be 'seen' to descend in an incompatible way, and it 'pops-out' of the array. In this case 'pop-out' can be seen to be a property of the OBJ representation, depending upon the integration of PROP→OBJ and VIS→OBJ representations.

The exact way in which a scene is structured obviously exerts a major influence on the way in which the constituent elements appear when displayed upon a computer screen. They also constrain the way in which people search for a specific target. In a very simple example, a typical task for a text based display is to present a listing of filenames in a directory. In many systems an attempt is made to make use of the full screen display by writing names alphabetically across lines with a tab separating each item on the line. With a large list, many lines are presented, and the overall appearance created is of several *columns* of items. Under these circumstances users tend to direct their search down the columns rather than use the normal reading strategy of moving across the lines. Since the column strategy is at variance with the alphabetic listing across the page, this makes it hard to locate specific targets accurately and rapidly. This is not simply a problem for graphically challenged display technologies. Similar effects have been discussed in the context of button organisations within hypertext systems, and have been identified in conjunction with subject-predicate analyses

[34, 35]. Indeed, as graphical interfaces become more advanced the number of ways in which these problems arise will undoubtedly increase.

A transformation process in OBJ that has not been mentioned in this discussion is OBJ→MPL, which produces representations of the names of objects, and the sound of words or word parts in reading. The absence of a direct route from MPL→OBJ means that the structural representations of sounds do not play a role in the integration of object representations. The converse is not true. When people read words, for example, the transformation from an OBJ to an MPL representation is not subject to the same temporal dynamics as spoken speech, which produces an MPL representation from the raw AC representations. This has very strong implications for multimodal integration, particularly in the context of advanced graphical systems.

2.5.2 Morphonolexical Integration

Figure 2.9 shows how information flows come together at the Morphonolexical system. Like the other structural subsystem, OBJ, the MPL subsystem receives information flow from sensory (AC→MPL) and referential sources (PROP→MPL). Unlike OBJ, MPL also receives input from a crossmodal source, OBJ→MPL. This asymmetry has its origins in language processing and supports the skill of reading.

Just as OBJ plays a crucial mediating role in visuomotor action, co-ordinating the intended motor actions with the visual scene, MPL is central to the production of speech, co-ordinating intended speech with the acoustic scene: the general noise level, the individual's immediately recent speech acts, and the speech acts of others. There is a direct analogy between the buffered OBJ→LIM process in the tennis example of Figure 2.6, and a buffered MPL→ART process in the production of speech. There is also an analogy between the OBJ→PROP & PROP→OBJ cycle integrating with the VIS→OBJ flow to influence object perception, and the MPL→PROP & PROP→MPL cycle integrating with the AC→MPL flow to influence our perception of the acoustic world. There are very strong effects of context on the recognition of spoken words, particularly when the speech is degraded on a noisy channel. Just as what we see depends upon what we expect to be seeing (cf. Figure 2.7), what we 'hear' depends upon what we referentially 'expect' to be hearing.

This key role of the MPL subsystem in controlling speech results in its transformation processes operating at a rate compatible with the flow of transient, acoustically derived representations. The structural nature of these representations is dominated by temporal relationships. In contrast, in reading, the objects are written words, which are less transient entities, in two dimensional space. The products of OBJ→MPL processing generated in the course of reading cannot be handled directly by MPL processes, since they are on the wrong time base: the sequential patterning of the MPL basic units is not equivalent to that of normal speech processing. The mechanism copes with this by using buffered processing, either of MPL→ART, if the words are to be read aloud directly, or of MPL→PROP if referential comprehension is intended.

This model of reading supports the theoretical analysis of classic modality differences in short term memory [3], where the last item of aurally presented lists of words is usually recalled with greater accuracy in serial recall tasks than the last item in visually presented lists. An auditory list allows a direct flow of Acoustic, through MPL to Articulatory representations, resulting in consistent, temporally ordered representations in the proximal part of the Articulatory image record that can be revived to support short-term memory retrieval tasks. The buffering of the MPL transformation provides a basis for explaining the effects obtained on the last item of a visually presented list of written words.

However, there are circumstances under which a full 'auditory' recency effect occurs with visual material [13, 48]. These circumstances involve lip-read material. Barnard [3] argues that these effects are obtained because lip-read information, although visual in origin and

hence processed by VIS→OBJ and OBJ→MPL, is on the same time-base as speech, and so can be transformed directly by the relevant processes in the MPL subsystem. Crossmodal effects occur when a redundant item (a suffix) is presented to be lip-read at the end of an aurally presented list [48], but only if the inter-item timing allows them to be dynamically interpretable as constituents of the same data stream, and so integrable by MPL transformations.

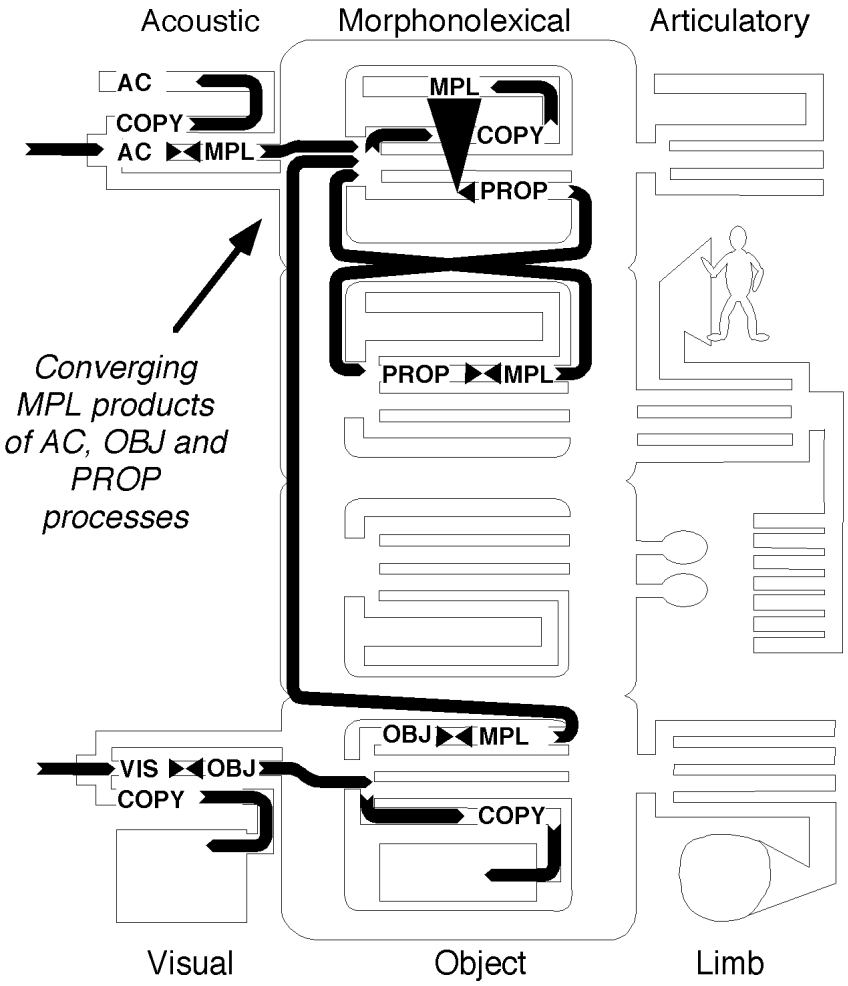

Figure 2.9: The converging MPL products of acoustic, propositional and object-based processing. Here the configuration is for the comprehension of speech, where the individual is listening to the form of the information in the speech stream (buffered at MPL).

This general analysis of information flow receives considerable support from another classic phenomenon, this time explicitly dealing with multimodal integration: The McGurk effect [30, 31]. In the basic McGurk effect video recording techniques are used to alter the relationship between lip movement and heard speech. Acoustic information for one utterance is presented with the visual information for another. Initially, subjects were instructed to watch the speaker and repeat what had been said. The subjects in this study [31] experienced neither intermodal conflict nor domination of information in one modality by information in another. The two sources 'fused' to yield a unified percept. So, for example, when the sound of /ba-ba/ was dubbed onto lip movements for the utterance /ga-ga/, 80% of pre-school children and 98% of adults reported hearing /da-da/. Dubbing /ga-ga/ onto /ba-ba/ also produced illusions, with subjects reporting that they heard /gab-ga/ or /bag-ba/. Similarly,

presenting the sound of /pa-pa/ with the lip movements for /ka-ka/ elicited /ta-ta/ as the dominant response; here the reverse dub resulted in such responses as /pak-pa/ and /kap-ka/.

Visual dominance theories or an acoustic averaging hypothesis for similar dichotic fusions [17] cannot explain this effect. Information flow within the ICS architecture provides an explanation: crossmodal integration occurs through the convergence of multimodal sources at the MPL input array. Since the arriving representations can fit a common structural description, and are on the same time base, integration into a coherent data stream is possible. The integrated structure that results will be dependent upon the content of the representations that are to be integrated, and the dynamic status of the processing activity that is to operate on the integrated structure. The McGurk effect has now been widely investigated and the precise character of the effect does appear to vary, being dependent on detailed linguistic influences in some circumstances [30], and on the actual language involved. There are reports for example that the effect is hard to obtain with Japanese [45] due to its syllabic structure; but that it does occur in English when voices of one gender are dubbed onto lip movements of the opposite gender [23].

As advanced graphical interfaces develop, multimodal characteristics assume massive importance. Superficially, the synchronisation of voice in video film, in videophone conversation, in the use of animated characters who ‘speak’, and even in ventriloquism, are all cases within the ICS framework where multimodal integration could be occurring at the MPL level. However, the bulk of the evidence suggests that multimodal integration at the MPL level must be subject to very tight temporal constraints, and also depends upon detail being abstracted from lip-read information about the linguistic forms being articulated.

While voice asynchrony in high definition films may well cause considerable discomfort, the visual information in very small video images presented on computer screens (e.g. see Figure 2.12) may be quite inadequate to convey the kind of articulatory contrasts that lead to real time blending on the input array of the MPL subsystem. Animated faces, videoconferencing images [46], or the mouth movements of a ventriloquist’s dummy would not be detailed enough to let OBJ→MPL produce the detailed MPL representations that are necessary for blending of individual speech sounds to occur. Nevertheless, the overall superstructure and rhythm of the visual stream might be sufficiently well correlated with the Acoustically derived MPL representations for blending of larger units of speech to occur, allowing the speech stream as a whole to be propositionally attributed to a particular visual source, without causing crossmodal conflict at the level of the individual sounds.

An exact synchronisation of video and audio outputs dealing with speech may not be necessary, unless both sources are of sufficiently high fidelity that any offset would cause differentiation of data streams at the MPL level. Their superstructure and the temporal dynamics of the data streams may be sufficient to ensure an attribution of voice and moving image to a single source. However, rapidly improving technology means that larger images of ‘speaking faces’ are being incorporated into new systems. For these graphic displays, detailed lip movements may be readily discernible, forming a basis for direct blending, conflict or temporal offset on the input array of the MPL subsystems. Along with increased size may come greater requirements for achieving close synchrony between sound and vision.

Even in situations where the visual and acoustic data cannot be reconciled at a superstructural level, it may be possible cognitively to attribute a speech stream to a visual object. In the same way that the VIS representations could not integrate the differently coloured arrowheads, and a higher-level OBJ representation was required, the MPL representation will not be sufficient, and the higher-level PROP representation will be involved.

2.5.3 Propositional Integration

As noted earlier the propositional subsystem takes no direct input from sensory sources, but the role it plays in the indirect integration of information derived from multimodal sources should not be underestimated. This level of representation deals with abstract referents, their properties and semantic inter-relationships. We have already described the role of the PROP subsystem in cycles with OBJ and MPL, providing a referential stream of information to be integrated with and inform our structural interpretation of sensory data. It also plays a directly integrative role in its own right.

The PROP subsystem receives input from OBJ, MPL and IMPLIC sources. Each of these deals with information originating in sensory subsystems. The relevant data flows are highlighted in Figure 2.10. If there are no conflicts between the data flows from OBJ→PROP and MPL→PROP, and if they are compatible with the schematic models active at the implicational level, then multimodal events can be understood indirectly in terms of their integrated PROP representations.

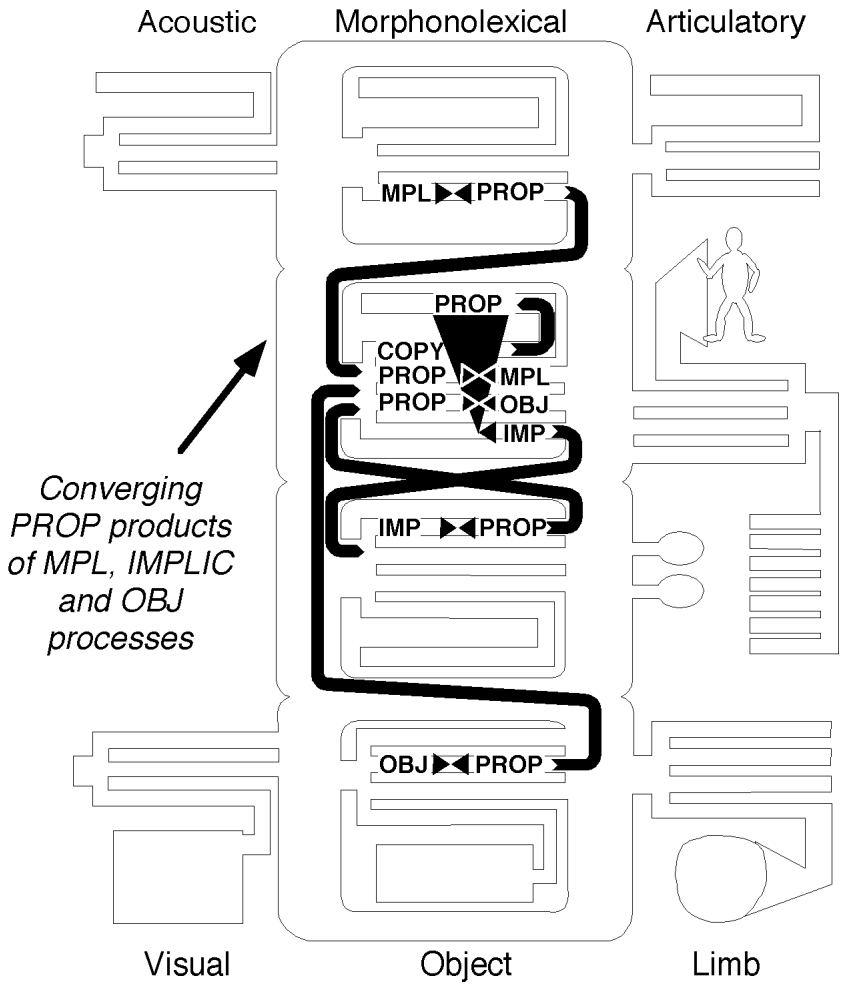

Figure 2.10: Converging propositional products of MPL, OBJ and IMPLIC processes.

Our ‘perception’ of thunder and lightning being related to the same environmental event really comes down to a propositional understanding in which the structurally interpreted sensory information is semantically integrable as consequences of a single environmental referent. Likewise, a language stream can be attributed to an animated non-human referential agent so long as they combined agent/speech representation is within the bounds of schematic models of what is reasonable. It is perfectly acceptable behaviour to see and understand a

cartoon tree as capable of speaking to you, but it would take very special circumstances for it to be acceptable for you to be seen talking to a real tree in a real park.

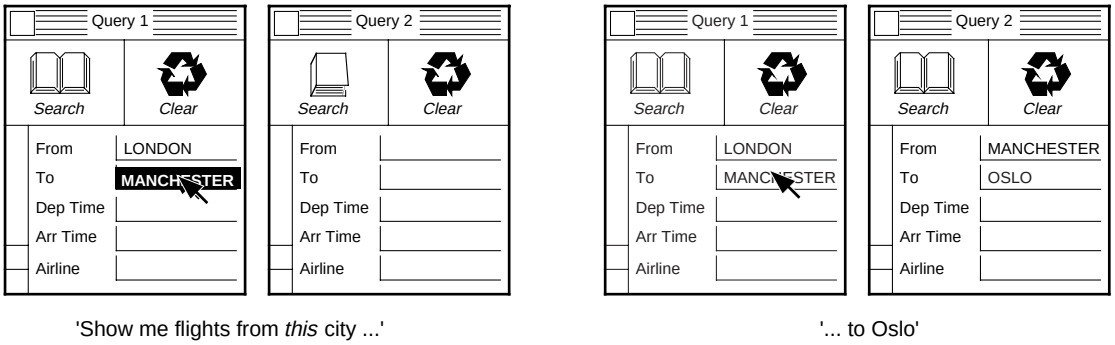

Figure 2.11: A flight information system that allows deictic reference (Nigay, 1994)

In normal discourse, we understand what is said in relation to some frame of reference – and more often than not, the frame of reference involves the external world. Reference to things in the world is an integral part of communication – deixis. We refer to things ‘out there’ by pointing or by sharing very particular forms of semantic common ground. In terms of the ICS architecture, deictic reference is resolved through the integration and blending of propositional information from different sources (OBJ→PROP; MPL→PROP). Some prototype computer interfaces explicitly support deictic expression. For example, the flight information system illustrated in Figure 2.11 might allow a user to ask for ‘flights from this city to Oslo’ verbally, co-ordinated with mouse motions that point at the name of the origin city elsewhere on the screen [38, 39].

This kind of application is interesting to analyse for a number of reasons. Pointing with a particular interface device while simultaneously talking is a novel skill combination for most people, even though we can point with our own hands and talk, and would require some experience and learning to acquire. More important are the technological constraints imposed by current speech recognition systems. These systems typically take a significant amount of time to recognise a word and provide feedback about its identity. Even when trained they are also subject to significant error rates. The delay may well be crucial for appropriate deictic fusion at the propositional level. With significant delays, users are likely to become impatient and, where possible, achieve the same ends by other means.

For such systems to be effective the recognition technologies may have to attain a level of performance commensurate with the ‘normal’ requirements for propositional integration. With advanced graphical interfaces, and particularly those developed in the context of computer supported work, pointing may have far less deictic precision than that normally attained with a mouse. Figure 2.12 combines a graphical workspace in which a user is pointing with arm and finger to an item in the display. In such circumstances, the referent of a verbal statement “this one” can be quite ambiguous and understanding depends on the extent to which a range of cues intersect to minimise the propositional ambiguity [7].

In the broader experimental literature on multimodal events, a great deal of effort has been devoted to understanding what happens with inter-modal conflicts in attributes of sound sources and light sources, such as their spatial or temporal separation [41]. Very often, the tasks require people to exercise considerable judgement. So, for example Rock & Victor [43] asked people to grasp a square whilst simultaneously viewing it through a lens that contracted its image to half size. Subjects were then asked to match what they had experienced to one of several alternatives. Their judgements were more biased by what they had seen than by what they had felt. Indeed, in many cases ‘visual dominance’ is often assumed in the resolution of multimodal conflicts. This is by no means universally the case. In other circumstances,

Walker & Scott [50] showed that people judge a light as of shorter duration than an tone of identical duration. The form of analysis evolving here would suggest that many of the cases of inter-modal conflict are best understood not in terms of the information processing of sensory or structural representations – but rather as a consequence of cyclical central processing in which propositional attributions may be derived from various integrations of information originating in OBJ, MPL and IMPLIC processing activity.

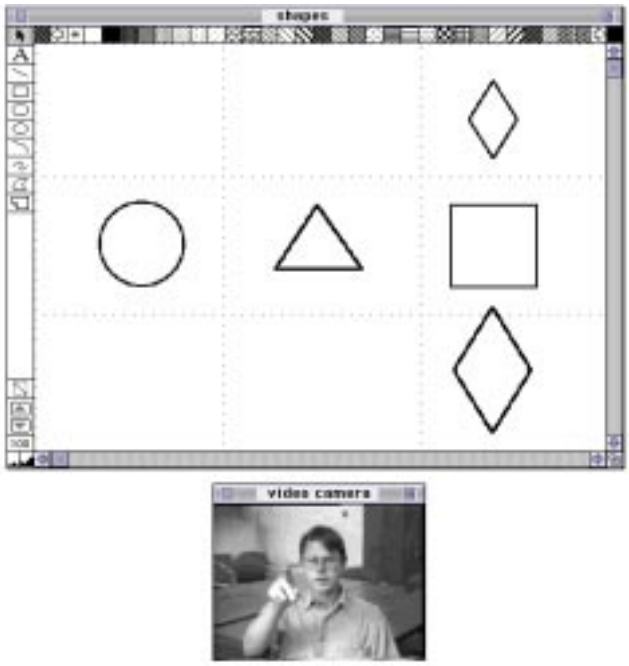

Figure 2.12: A mock-up of a 'shared drawing' package that encourages its users to make deictic references.

The cycles that PROP can make with OBJ and MPL have already been discussed, but it should now be clear that PROP is able to bring these two cycles of activity together. In fact the two cycles can in principle act independently, with their representations remaining differentiated at the PROP input array, and only one (or even neither) of PROP→OBJ or PROP→MPL being buffered. This means that it is quite possible for someone to play tennis, with an OBJ⇔PROP cycle helping them hit the ball, while engaging in a conversation with their opponent and using an MPL⇔PROP cycle to co-ordinate the referential meaning of their speech. What will not be possible, since only one of the processes within PROP can access the image record at a time, is for image record access to occur in both cycles simultaneously. In practice, when the PROP⇔OBJ cycle needs to access experientially derived referential information about tennis, the player will have to momentarily disengage their PROP→MPL process from buffered mode, and so pause the production of intended speech. Similarly, a need to access referential information to drive the speech stream will prevent PROP→OBJ from using the image record. The MPL→ART process will continue to drive the articulatory mechanisms, and OBJ→LIM the motor mechanisms, and so this interchange of buffering will only cause delays in either the speech or play if critical referential activity is required in both streams at the same time. Then the tennis player will have to pause their speech, or risk missing their shot.

In addition to the cycles with the structural subsystems, PROP can both receive representations from IMPLIC and produce IMPLIC representations in return, allowing it to form a third cyclical configuration, PROP⇔IMPLIC. Indeed, propositional representations can rarely, if ever, be sensibly considered in isolation from the processing of the schematic representations of IMPLIC. The cycle between these two levels of meaning is so pervasive

within human cognition that it is known as the 'central engine' of cognition. This cycle is also important in bringing latent schematic knowledge into referential form, which is required for it to be verbally expressed (since it is assumed that there is no direct IMPLIC→MPL process).

2.5.4 Implicational Integration

A sweeping, rather dismissive gesture of the hand is over in a moment and almost certainly contributes to a propositional understanding of a speech stream, not by detailed parsing of its spatio-temporal attributes, but by direct apprehension of some quality of the movement (VIS→IMPLIC followed by IMPLIC→PROP). In the same way, facial expressions indicate the mood of the speaker, or provide clues as to how a listener is reacting to what you are saying. Similar representations may be abstracted from tone of voice (AC→IMPLIC) or even from feelings in the body, such as running vigorously on the spot whilst being persuaded to try harder (BS→IMPLIC). All of these sorts of information can be integrated into a unified data stream at this most abstract level of representation within the ICS architecture (Figure 2.13).

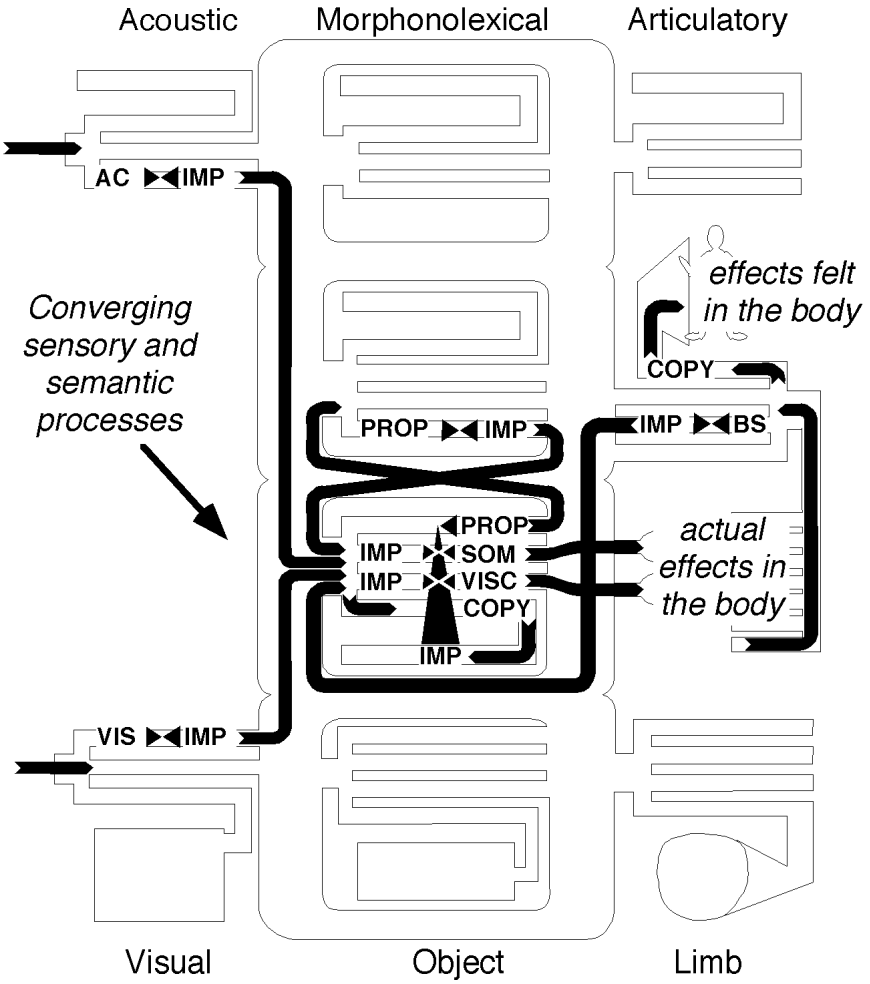

Figure 2.13 Sensory and semantic inputs converging at the Implicational level.

It is at the implicational level that cognition and affect inter-relate most clearly. So, for example, affective judgements about the sound quality of walkman tape recorders tend to be more positive when the subjects are doing arousing physical exercise than when they are at rest [15]. Following the integration of this sensory input to IMPLIC, consequences can then flow back to the PROP subsystem. It is important to note that the other output from IMPLIC only maps to the somatic (SOM) and visceral (VISC) representations to produce responses in

the body. Its influence on OBJ and MPL representations, and so their integration with affective information, can only occur indirectly, through the mediation of PROP. Similarly, structural interpretations of the sensory data cannot directly influence its affective representation, but must be channelled through PROP, hence the importance of the PROP⇔IMPLIC central engine.

The basic units of the implicational level of representation are holistic concepts and properties, which come together to form schematic models of a rather generic kind (see Table 1). It also the level of representation at which cognitive and affective concerns come together [49]. Like any other type of representation within ICS, incoming representations can be integrated by transformation processes to form a coherent stream of data. This depends on the way an individual's learning experience has led to the formation of basic units at this level and their combination into higher level superstructures. The form and content of the schematic models constrain our judgement, decision making and overt behaviour.

The effects of implicational integration are general. Information about relative visual positioning and sound localisation contribute to the overall generic specification of the physical environment within which we are operating. Properties of facial expression, tone of voice and bodily context will contribute to setting the interpretative context for incoming utterances. The influences of pattern based factors like facial expression and tone of voice can be illustrated by reference to a simple example of how sensory information can contribute directly, via the VIS→IMPLIC and AC→IMPLIC processes, to the kind of holistic meaning represented at the implicational subsystem. Figure 2.14 shows two shapes. When asked 'which one is Uloomo and which one is Takete?', the majority of people point to the form composed of round elements as Uloomo, and the form composed of angular lines as Takete [19]. The visual characteristics of the shapes and the acoustic and articulatory characteristics of the sounds seem to match in this particular way, and not the other way around: 'Uloomo' connotes a generic quality of roundness, and 'Takete' conveys a similar generic quality of sharpness. The very difficulty people have in justifying their assessments is characteristic of the involvement of IMPLIC representations, set apart as they are from verbalisation without the mediation of PROP.

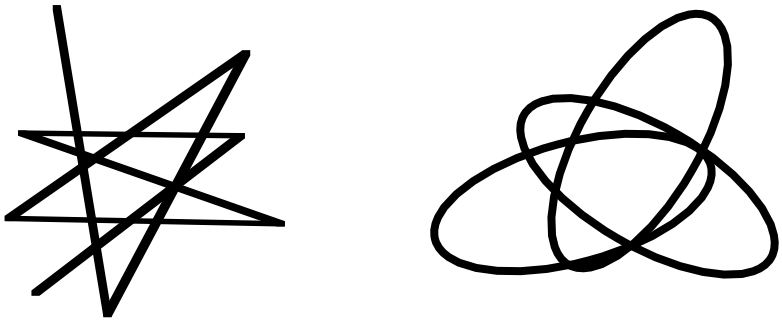

Figure 2.14: Two 'connotative' shapes (after Davis, 1961)

Implicational information derived from visual sources may also inter-relate and blend with Body-State and Propositional considerations. There is, for example, a massive literature on mood, memory and judgmental effects . To take but one case, asking people 'how do you feel' when the weather is sunny generally produces more positive responses than when it is raining [44]. The pleasantness of an individual's physical environment is assumed to provide a continual affective input via VIS→IMPLIC and BS→IMPLIC to the 'central engine' of propositional and implicational processing. The basic judgemental effect can be radically altered through the creation of specific propositional representations. If *before* asking them how they feel, their attention is directed to their environment by being asked 'what's the

weather like?', the effects of weather on their subsequent judgement of how they feel can be removed [44]. Answering the question about the weather requires PROP→IMPLIC and IMPLIC→PROP processes to assess the status of the weather, and what it means qualitatively. When they answer the later question, any weather related affective representations at IMPLIC will be schematically linked with the earlier question, and will be less likely to produce latent or implicit effects on the a judgement of their own 'overall state' [49].

At first glance, the relevance to design issues in human-computer interaction of somewhat esoteric aspects of the connotative meaning of visual form, facial expression and the weather may appear somewhat remote. However, with both simple technologies [4] and advanced interfaces, this is far from the case. For example, Figure 2.15 shows a schematic view of an hypothetical military radar-screen. Here it is easy to recognise which shapes are intended to represent formations of friendly aircraft or ships and to recognise those intended to represent the enemy. In this context, connotation via the sensory-schematic-referential chain can be extremely useful in communicating abstract information quickly, a factor recognised by the earlier investigators of such displays [40].

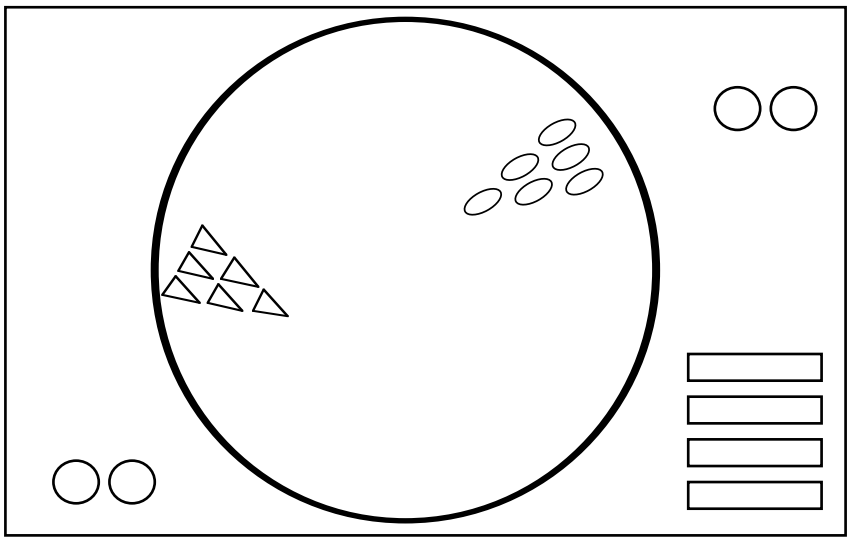

Figure 2.15: A schematic view of a military radar screen, showing friend and enemy formations using connotative forms

Processing in the Implicational subsystem also plays a vital role in the overall co-ordination of communicative behaviour, and so is of considerable significance for the development of computerised systems that are intended to support collaborative work. Since the sensory subsystems send their output directly to IMPLIC, it can arrive before the structural and referential subsystems have interpreted its explicit content. The affective tone of a message may be interpreted and fed back to PROP in advance of the structural subsystems' contribution. Gesture, gaze and intonation, for example, have long been regarded as playing a key role with respect to the management of dialogue. They are also associated with the expression of meaning and with more generic social signalling. Both content and timing are important. Beattie [9], for example discusses the differential distribution of speech focused movements and gestures. Lexically related gestures tend to occur an average of 800msec in advance of the word to which they are related [12]. Although such advance gestures could be related to planning, it is equally plausible that they provide the listener with a context for interpreting the speech.

The use of facial cues to add information about a speaker's affective meaning to the content of their speech is functionally useful when dealing with a person whose facial appearance

does correlate with the meaning of their speech. When this correlation is missing, or inappropriate, as is possible in computer generated facial expressions, it can cause problems. Walker et al [51], for example, compared two versions of a synthesised on-screen face accompanying auditory questions (Figure 2.16) with a plain text version of the same questions. Users who were asked questions by the talking faces spent longer writing, wrote more comments, and made fewer mistakes than those responding to textual questions. In addition, users who saw a 'stern' face (the right-hand face in Figure 2.16) spent longer, wrote more, and made fewer mistakes than those who saw the 'neutral' face (left hand side), but they liked the experience and the face less. Here the affective content of the face changed the way that subjects interpreted the situation as a whole, with the vaguely negative tone of the stern face giving a neutral question and answer session the characteristics of an inquisition.

Figure 2.16: a neutral (left) and stern (right) synthesised face, used to accompany auditory questions to computer users (from [15]; reproduced with the permission of the Association for Computing Machinery).

These direct sensory-affective routes are not the only way that affective content can be generated. The PROP→IMPLIC route is just as important, particularly given the possibility for direct feedback via the central engine cycle which can iteratively reinforce its contribution. Referential representations at PROP, derived from structural interpretations of the environment (from OBJ→PROP and MPL→PROP), can provide the sources for the PROP→IMPLIC transformation. In this case, implicit or latent knowledge about the world can be elicited, and unless their attention is drawn to the source of the information, as in the case of the weather question discussed earlier, people remain unaware of its origin, and even its effects upon their subsequent behaviour. For example, when people of different status talk to each other, there is some evidence that the higher status person will tend to use a different proportion of nouns to verbs in their speech than will the lower status person. Change their respective status and they will adjust their speech style. Neither of them will be aware of any change [22]. Similarly, adults when talking to children will automatically adjust their speech style to use shorter, high frequency words. There are many examples of this form [42]. In the context of a meeting, a manager will tend to take up a dominant physical location relative to other participants. Speech cues, location and even dress sense may all contribute to the blending of information to form a particular schematic model.

Within the ICS framework, implicational representations express generic meanings. Their broader function is to capture high level communalities about the world and the self, rather than the specific details. Under circumstances that hinge upon to the integration of multimodal information at the implicational level, the key issue is not whether a specific implicational 'cue' is available, but whether the integrated representation remains adequate to deliver appropriate implicational basic units, and the schematic models to which they contribute. As the applications of advanced graphical systems extend to encompass all domains of human endeavour, it will be increasingly important to supplement our understanding of the mental processing of perceptual form and propositional meaning with an understanding of the role of the wider, holistic meaning that is captured in implicational concepts and their often personal significance. It has long been known that successful computer games are highly engaging [32]. Excitement, curiosity, fantasy and enjoyment are constructs that resonate more closely with implicational rather than propositional meaning.

2.6 Conclusion

2.6.1 Summary

This paper has been concerned with a range of cognitive issues that may arise in the design of advanced graphical interfaces. The concept of advanced graphical interfaces has been taken in its widest sense, encompassing links between graphical presentations and the full range of input and output modalities that can be associated with their use. At the outset, the intention was to convey an understanding of the issues raised rather than to provide detailed design advice. To support that understanding a conceptual framework was presented. This framework is a unified theory of human information processing incorporating two fundamental aspects. The first aspect was a theory of information *flow*. The elements of this theory were basic mental processes organised into subsystems, each concerned with a specific domain of mental life, from sensation through central processing to the effector control of action. Individual subsystems were in turn organised within a wider, superordinate architecture.

The second aspect of the framework involved a theory of information and its *representation*. As with the internal organisation of processes within subsystems, it is assumed that the information represented within all mental codes is organised according to a common set of principles. Representations in each code are formed out of basic units. These basic units are themselves built from constituent elements whose properties reflect the nature of the underlying coding space – be it in a sensory, central or effector domain. The basic units are themselves constituents of a superordinate structuring of the information. This superordinate structuring is governed in part by thematic considerations. Each representation is assumed to be about something (the psychological subject) that is linked to other constituents (the predicate).

The theory of information flow is also closely linked to the theory of representation. First, the flow of information depends upon processes which transform information from one code to another. As such the coding systems are related. With a change from sensory to structural codes information about the sensory constituents is discarded and the sensory superstructure signals the basic units of the structural code. Exactly the same thing happens in the transformation of structural to propositional and in the transformation from propositional to implicational codes. In moving from implicational through propositional and structural codes to effector output, the respective transformation processes accomplish the inverse. The properties of the flow also link quite directly with the thematic structure of representations. The patterns of flow within the architecture are a crucial determinant of the inputs received by a process. This constrains the properties of the central and effector mental codes, since flow patterns determine how information from different sources comes together and blends.

It is upon this last point that the current paper has been focussed. Within the ICS framework, it is assumed that conscious experience is associated with the operation of the copy processes. The actual action of the underlying transformation processes is 'unconscious'. Of course, since information produced by one process will be copied into the image record of a subsequent subsystem, the products of that process will be available to conscious experience, but mediated by a copy process in a different subsystem. Since the elements of an input code are discarded when producing an output code, the *products experienced* at the subsequent level cannot directly represent how those elements came together. The constraints governing combination, fusion or blending of constituent elements are inherent, and consequently the knowledge deployed is 'latent' within the process. It will not directly be reflected in the products of that processing activity. If related streams fail to fuse, the separate products will, of course, be available to conscious experience at the next level of representation.

The arguments presented here have covered a very broad range. The particular illustrations were selected to demonstrate the relevance of the material to the resolution of issues in the design of advanced graphical interfaces. Hopefully, they went further and illustrated the potential utility in design of the kind of deeper understanding offered by cognitive theory. However, for such utility to be realised, means must be available to make the theory useful in the context of systems design, software engineering processes and human factors evaluations.

2.6.2 Relating cognitive theory to evaluation, design processes, and software engineering.

The general idea of utilising cognitive theory in design has long been an objective of research in HCI. Theories have often proved to be of limited scope and difficult to apply in a manner that guides design processes effectively [33]. The concluding section of a long paper is no place to pursue the relevant arguments in detail. It is nonetheless appropriate to provide some pointers. As a part of a long term inter-disciplinary collaboration, three strategies have been pursued for connecting the form of theory presented here to practical processes of design.

The first strategy is to develop heuristics for analysing practical design scenarios based upon the theoretical techniques. It is then necessary to convey those heuristics and the products of the application into design processes. So, for example, heuristics can be described in a technical manual for decomposing the structure of information on displays and principles described that support inferences about user performance with those structures. The products of analyses based upon these sorts of heuristics can them be incorporated into representations that support design. Design Rationales can, for example, be specified in the form of a notation based upon Questions Options and Criteria. Work done as a part of the AMODEUS project has illustrated how the heuristic analysis of design scenarios based upon ICS techniques can come to form criteria in a design space [10].

This first kind of strategy clearly relies upon the training and availability of people who are skilled in the relevant techniques and it is unlikely that this would be a real option for many design teams. An alternative strategy is to build software tools to support design which effectively embody those theoretical skills. This second strategy has also been pursued for some years [8, 34]. Production system techniques are used to automate the process of collecting information about a design space. The same techniques are used to represent theoretical heuristics. As with the skilled expert, these production rules can then build explicit models of the kinds of cognitive activity that will occur during the different mental phases in the performance of an interactive task. This can then be used to provide a predictive evaluation of user performance and theoretically motivated advice about design solutions.

A third strategy is to seek more direct means for interlinking the techniques of cognitive science with those of computer science. In one tactic, methods have been sought to represent interactions using formal methods [26]. In this approach, particular interactional requirements are precisely formulated. Their consequences can then be explored, on the one hand, by

reference to an appropriate models of the interactive behaviour of the computer system, and on the other hand to a model of user cognition. Another more direct tactic has also been explored. ICS is constrained by principles of information flow and information representation, with sufficiently adequate precision to be modelled using formal methods of software engineering. Interactor theory in computer science [21] can be used as a basis for this formal specification. In fact, using the MATIS system of Figure 2.11 as an example, Duke et al. [20] have shown how cognitive and system theory can be directly combined. Using a deontic extension of modal action logic, key constraints on the information flow imposed by the ICS model of user cognition were represented axiomatically. Key aspects of the behaviour of the MATIS computer system were formally represented in equivalent axiomatic terms. With both user and system models represented in the same language, they could be directly combined into a third, syndetic model. This third model represents the axioms governing the conjoint behaviour of the user and the system.

Naturally, much remains to be done to convert these strategies into everyday design methods. However, the opportunities to relate cognitive theory to the kinds of software engineering techniques described throughout this volume are certainly there to be grasped.

Acknowledgement

The research reported in this paper was carried out within the AMODEUS project, ESPRIT BRA 7040, funded by the European Union. We are grateful both for their financial support and for the opportunity to work in a strongly interdisciplinary context. Information about the project, and access to many project documents, is available electronically:

`http://www.mrc-apu.cam.ac.uk/amodeus/`
`ftp://ftp.mrc-apu.cam.ac.uk/pub/amodeus/`

References

[1] Abbs, J.H. & Gracco, V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology, 51*, 705-723.

[2] Attanasio, J. S. (1987) Relationships between oral sensory feedback skills and adaptation to delayed auditory feedback. *Journal of Communication Disorders*, 20, 391-402.

[3] Barnard, P.J. (1985) Interacting Cognitive Subsystems: A psycholinguistic approach to short term memory. In A. Ellis (Ed.) *Progress in the Psychology of Language*, (Vol. 2), Chapter 6, London: Lawrence Erlbaum Associates, 197-258.

[4] Barnard, P. and Marcel, A.J. (1984) Representation and understanding in the use of symbols and pictograms. In R. Easterby and H. Zwaga (Eds.), *Information Design: The Design and Evaluation of Signs and Printed Material*, 37-75. John Wiley: Chichester.

[5] Barnard, P. & May, J. (1993) Cognitive Modelling for User Requirements. In Byerley, P., Barnard, P. & May, J. (Eds.) *Computers, Communication and Usability: Design Issues, Research and Methods for Integrated Services*. Chapter 2.1, pp 101-146, Amsterdam: North Holland, Studies in Telecommunications.

[6] Barnard, P.J. and Teasdale, J.D. (1991) Interacting cognitive subsystems: A systemic approach to cognitive-affective interaction and change. *Cognition and Emotion, 5*, 1-39.

[7] Barnard, P., May, J. & Salber, D. (1994) Deixis and points of view in Media Spaces: an Empirical Gesture. AMODEUS project document UM/WP19;submitted for publication.

[8] Barnard, P., Wilson, M. and MacLean, A. (1988) Approximate modelling of cognitive activity with an Expert system: A theory based strategy for developing an interactive design tool. *The Computer Journal, 31*, 445-456.

[9] Beattie, G. (1983) *Talk*. Milton Keynes: Open University Press.

[10] Bellotti, V., Buckingham-Shum, S., MacLean, A., & Hammond, N. (1994) Multidisciplinary Modelling In HCI Design … In Theory and In Practice, *Amodeus Project Document ID/WP 34;* submitted for Publication

[11] Bregman A. S. & Rudnicky, A.I. (1975) Auditory Segregation: Stream or Streams? *Journal of Experimental Psychology: Human Perception and Performance,* 1, 263-267.

[12] Butterworth, B & Beattie, G. (1978) Gesture and silence as indicators of planning in speech. In: Campbell, R. N. & Smith. P.T., *Recent Advances in the Psychology of Language: Formal and Experimental Approaches.* New York, Plenum.

[13] Campbell, R. & Dodd, B. (1980) Hearing by Eye. *Quarterly Journal of Experimental Psychology*, 32, 85-99.

[14] Cherry, E. C. (1953) Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America,* 25, 975-979.

[15] Clark, M., Millberg, S.,& Ross, J. (1983) Arousal cues arousal-related material in memory: Implications for understanding effects of mood on memory. *Journal of Verbal Learning and verbal Behaviour,* 22 633-649.

[16] Craig, A. (1990) An investigation into the relationship between anxiety and stuttering. *Journal of speech and hearing disorders, 55,* pp. 290-294.

[17] Cutting, J. E. (1976) Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening. *Psychological Review,* 83, 114-140.

[18] Cutting, J. E. (1981) Six Tenets for Event Perception. *Cognition,* 10, 71-78.

[19] Davis, R. (1961) The fitness of names to drawings. *British Journal of Psychology*, 52, 259-268.

[20] Duke D.J. , Barnard P.J. Duce D.A. and May, J. (1994) Syndetic Models For Human-Computer Interaction. *Amodeus Project Document* ID/WP35; submitted for publication.

[21] Duke, D. J. & Harrison, M. D. (1993). Abstract Interaction Objects. *Computer Graphics Forum*, 12 (3).

[22] Fielding, G. & Fraser, C. (1978) The Language of Interpersonal relationships. In Markova, I (ed) The social context of language, Chichester: Wiley.

[23] Green, K., Kuhl, P. Meltzoff, A & Stevens, B. (1991) Integrating speech information across talkers, gender, and sensory modality. *Perception and Psychophysics,* 50, 524-536.

[24] Gregory, R. (1971) *The Intelligent Eye*. London: Weidenfeld & Nicolson.

[25] Gross, Y. & Melzack, R. (1978) Body Images: dissociation of real and perceived limbs by pressure cuff ischemia. *Experimental Neurology, 61,* 680-688.

[26] Harrison, M.D. and Barnard, P.J. (1993) On defining requirements for interactions. In A. Finkelstein (Ed.), *Proceedings of IEEE International Workshop on Requirements Engineering* New York: IEEE, 50-54, 1993.

[27] Howell, P. & Powell, D. J. (1987) Delayed auditory feedback with delayed sounds varying in duration. *Perception and Psychophysics*, 42, 166-172.

[28] Köhler, W. (1947) *Gestalt Psychology.* New York: Liveright.

[29] Lee, B. (1950) Effects of delayed speech feedback. *Journal of the Acoustical Society of America, 22*, pp 824-826.

[30] MacDonald, J. & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics*, 24, 253-257.

[31] McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, 264, 746-748.

[32] Malone, T (1982) Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Human Factors in Computing Systems.* ACM: Washington, 63-68.

[33] May, J. & Barnard, P. (1995) The case for supportive evaluation in design. *Interacting With Computers, 7* (in press).

[34] May, J., Barnard, P.J., and Blandford, A. (1993) Using Structural Descriptions of Interfaces to Automate the Modelling of User Cognition. User *Modelling and Adaptive User Interfaces, 3(1)*

[35] May, J., Barnard, P.J., Boecker, M. and Green, A.J. (1990) Characterising structural and dynamic aspects of the interpretation of visual interface objects. In *ESPRIT '90 Conference Proceedings* (pp.819-834). Brussels (November 1990), Dordrecht: Kluwer Academic Publishers.

[36] Monster, A.W., Herman, R. & Altman, N.R. (1973). Effects of the peripheral and central 'sensory' component in the calibration of position. In J.E. Desmedt (Ed.) *New developments in electromyography and clinical neurophysiology* (vol. 3). Basel: Karger.

[37] Munn, N. (1961). *Psychology: The fundamentals of Human Adjustment*, London: George Harrap.

[38] Nigay, L. (1994). *Conception et Modélisation Logicielles des Systèmes Interactifs.* Ph.D. Thèse de l'Université Joseph Fourier, Grenoble. 350 pages.

[39] Nigay, L., Coutaz, J. & Salber, D. (1993) MATIS: a Multimodal Airline Travel Information System. *AMODEUS Project Document* SM/WP10.

[40] Provins, K, Stockbridge H, Forrest, D. & Anderson, D. (1957) The representation of aircraft by pictorial signs. *Occupational Psychology*, 31, 21-32.

[41] Radeau, M. (1992). Cognitive Impenetrability in Auditory Visual Interaction. In: *Analytic Approaches to Human Cognition.* Alegria, J., Holender, D., Junça de Morais, J. and Radeau, M. (eds.). Amsterdam: Elsevier Science Publishers, BV. 41-55.

[42] Robinson, W. P. (1972) *Language and Social Behaviour* Harmondsworth: Penguin Books.

[43] Rock, I & Victor, J. (1964) Vision and Touch: an experimentally created conflict between the two senses. Science, 143, 594.

[44] Schwarz, N. & Clore, G.L. (1983) Mood, misattribution, and judgements of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 512-523.

[45] Seikiyama, K. & Tokura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805.

[46] Sellen, A. (1992) Speech patterns in video-mediated conversations. In Proceedings of CHI '92 ACM: New York, 49-59.

[47] Shallice, T. (1988) *From Neuropsychology to mental structure.* CUP: Cambridge.

[48] Spoehr, K. & Corin, W. (1978) The stimulus suffix effect as a memory coding phenomenon. *Memory and Cognition*, 6, 583-589.

[49] Teasdale, J. & Barnard, P.J. (1993) Affect, Cognition and Change: Re-modelling depressive thought. Hove: Lawrence Erlbaum Associates.

[50] Walker, J & Scott, K. (1981) Auditory-visual conflicts in the duration of lights tones, and gaps. Journal of Experimental Psychology, Human Perception and Performance, 7, 1327-1339.

[51] Walker, J.H., Sproull, L. & Subramani, R. (1994) Using a Human Face in an Interface' In Proceedings of CHI'94, pp. 85-91. ACM: New York.