Multisensory processing in perception

13

Contents

INTRODUCTION

Natural objects and events are usually multisensory. For example, during a face-to-face conversation with another human, or an encounter with an animal, both vocalizations and visual information (facial expression, lip movement, body attitude) are important. When a motorcyclist twists the throttle and the vehicle accelerates, there is an audible change in engine note, a change in visual movement seen through the helmet visor, and vestibular signals providing a sensation of bodily tilt and acceleration. It would be surprising if the sensory systems did not exploit such correlations between different modalities. Indeed the nausea experienced in virtual reality simulators (Nichols & Patel, 2002) suggests that the brain expects correlated signals to arrive, and is disturbed when they do not.

? *Why are modern flight simulators equipped with electric-pneumatic jacks?*

However the processing framework presented so far includes only limited scope for multisensory integration. Chapter 1 introduced a fundamental theoretical concept in sensory processing, the information processing device or module. By and large the intervening chapters have treated each sensory modality as an independent, autonomous information processing module. Figure 13.1 illustrates the general neural architecture of the system, for three senses. Separate modules process visual (brown), auditory (blue), and somatosensory (yellow) stimuli. Within each modality, dual processing streams specialize in "what" and "where" computations. Information

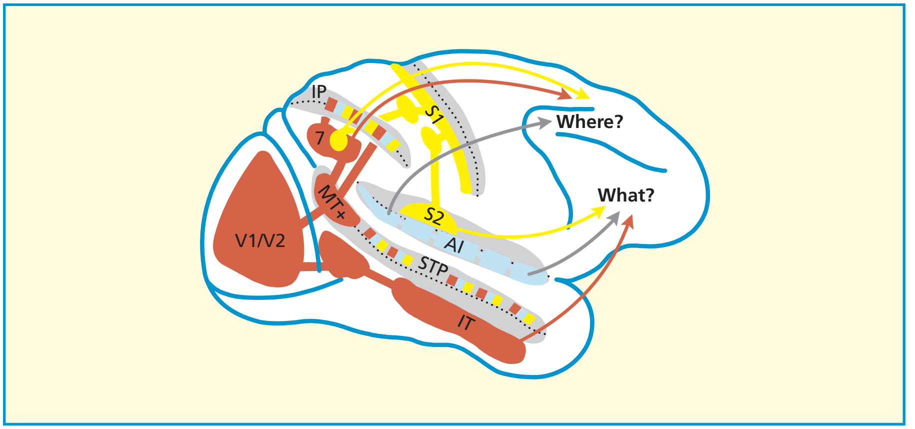

FIGURE 13.1 The general cortical architecture of sensory processing, according to the modularity theory. Specialized processing systems are dedicated to different sensory modalities: Vision (brown), audition (light blue), and somatosensation (yellow). In each system information flow divides into two streams, one carrying "what" information and the other carrying "where" information. Multisensory processing occurs relatively late in the processing hierarchy, in the intraparietal sulcus (IP) and the superior temporal polysensory area (STP), both shown in multiple colors. Redrawn from Schroeder et al. (2003). Copyright © 2003 Elsevier. Reproduced with permission.

? *Analyse your experience of eating a particular kind of food, such as yoghurt or cheese, in terms of its multisensory components.*

KEY TERM

Modular processing
A theory of cognitive processing in which different functions such as vision, hearing, and memory are implemented in separate and independent processing modules.

from the different modalities is integrated only in later association or polysensory areas of the cortex (Driver & Noesselt, 2008), namely the intraparietal sulcus (IP) and the superior temporal polysensory area (STP).

The scheme in Figure 13.1 is consistent with Fodor's (1983) influential theory of cognition as a collection of independent information processing subsystems, often call "modules." **Modular processing** architecture has a number of virtues. Each subsystem can be optimized for its specific computational task, operating with maximum speed and efficiency rather than compromised to serve several functions at once. Any errors or inaccuracies remain confined to one task, rather than propagated widely. New sensory functions or modalities can be created by adding new subsystems as and when they are required. On the other hand completely independent sensory subsystems can pose problems. There is an issue about how time-sensitive processing can be coordinated across independent subsystems so that, for example, our visual perception of lip movements is synchronized with auditory perception of individual speech sounds. Furthermore independent subsystems cannot take advantage of correlated information in different sensory modalities, such as when we both see and hear someone speaking. Finally, as indicated at the beginning of Chapter 1 in Table 1.1, and in the intervening chapters, there is a great deal of overlap in the computational tasks performed in the different modalities, which independent subsystems cannot exploit:

Orienting Vision, audition and somatosensation all orient our attention to external events.

Object identification All the sensory systems supply information that can be used to identify objects, ultimately accessing the same stored representations.

Localization Both vision and audition have high-resolution processes for computing the location of objects and events.

Body motion Visual and vestibular responses contribute to estimates of body motion.

Flavor Information from olfaction, gustation and other senses is combined to mediate perception of food edibility and flavor.

A growing body of research demonstrates that, contrary to the notion that modality-specific subsystems are completely independent, information *is* shared across sensory modalities earlier in processing than previously thought. Integration of cues for body motion was discussed in Chapter 3, and flavor was discussed in Chapter 2, so the next section focuses on multisensory integration during orienting, object identification, and localization. A later section in this chapter will discuss how recent studies of synesthesia are also throwing new light on multisensory integration.

? *How does your local supermarket exploit multisensory processing effects in perception?*

Modularity is a key principle in modern computer hardware and software system design, and offers the same advantages as it does for neural systems.

MULTISENSORY PROCESSING

EVIDENCE FOR MULTISENSORY PROCESSING

Orienting

Orienting to novel or unpredictable stimuli is often studied by measuring reaction times (RTs) to their appearance. Diederich and Colonius (2004a) measured simple RTs to the presentation of stimuli in three different modalities in separate experimental sessions (unimodal stimulation): Flashes of light, auditory tones, or tactile vibrations applied to a toe. Auditory RTs were fastest (132 msec), followed by visual RTs (163 msec) and tactile RTs (177 msec). Using bimodal stimuli (two modalities presented at the same time) RTs speeded up by about 20–30 msec on average relative to unimodal RTs, and trimodal stimulation speeded up reactions by a further 10–20 msec. Cueing experiments produce similar results. In a typical experiment, the subject is required to respond as quickly as possible to a faint visual target stimulus presented either to the left or to the right of fixation, and an auditory or tactile cue is presented to one side of fixation or to the other side. The cue is "valid" when it appears on the same side as the target, and "invalid" when it appears on the opposite side to the target, as also described in the next chapter on attention. Reaction times are faster in valid trials than in invalid trials, indicating integration of information across the different modalities used for the cue and target stimuli (Butter, Buchtel, & Santucci, 1989; Spence et al., 1998).

The faster reaction times using multimodal stimuli may be due to changes in decision processing, such as reduced decision uncertainty, rather than changes in sensory response. Signal detection theory (SDT) offers tools that are used to distinguish between decision and sensory influences on cueing, because it provides separate measures of perceptual detectability (called d′) and decision criterion (called bias or β; see the tutorials section at the end of Chapter 1). McDonald, Teder-Sälejärvi, and Hillyard (2000) and Frassinetti, Bolognini, and Ladavas (2002) used SDT measures in **cross-modal cueing** experiments. Subjects were required to detect weak visual stimuli in the presence of irrelevant auditory stimuli presented either on the same side or the opposite side of fixation, as described earlier. They found that detectability (d′) was significantly higher when auditory stimuli were presented on the same side as visual targets than when they were presented on the opposite side. There were inconsistent effects on bias.

The graphs in Figure 13.2 show the improvement in reaction time and detectability when cue and target were presented on the same side (valid: Dark bars) rather than the opposite side (light bars). Reaction times were approximately 6% shorter, and detectability was 8% higher, when a valid cue was presented. These

KEY TERM

Cross-modal cueing
Occurs when a cue stimulus presented in one sensory modality facilitates the detection or discrimination of a target stimulus presented in a different sensory modality.

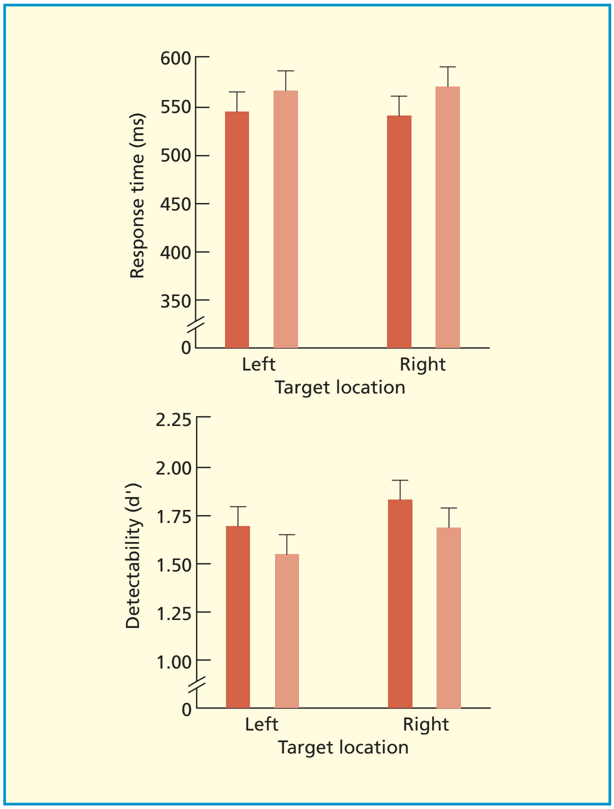

FIGURE 13.2
Reaction time (top) and detectability (bottom) in a cross-modal cueing experiment. Visual targets were presented to the left or right of fixation (horizontal axis). An auditory cue was presented either on the same side as the target (dark bars) or on the opposite side (light bars). Reaction times were faster, and detectability was higher, when the cue and target were presented on the same side. Adapted from McDonald, Teder-Sälejärvi, and Hillyard (2000), figure 3.

results indicate that cross-modal cueing does involve some modulation of relatively low-level sensory processes.

Two explanations have been proposed for enhanced performance using multimodal stimuli:

Parallel activation A bi- or trimodal stimulus generates activity in two or three modality-specific processing modules in parallel. Each is subject to some degree of random variation in its response (see "noise" in Chapter 1), but when activation level in any one module reaches its threshold level, the signal is detected. Processes in the different modalities are in a race to determine which one can reach threshold first. So RTs in response to multimodal stimulation should on average be faster than those to unimodal stimulation on statistical grounds, because there are multiple chances to detect the stimulus rather than just the one. This enhanced response is called "statistical facilitation."

Coactivation Modality-specific responses to multimodal stimuli come together at polysensory brain areas. The combined response is stronger and more reliable than responses generated by unimodal stimuli, so RTs and detectability are both better.

Diederich and Colonius (2004b) estimated the amount of statistical facilitation one would expect from parallel activation, under reasonable assumptions about how the different processes operate. They concluded that multimodal effects are too large to be explained by parallel activation, and must therefore be due to coactivation in polysensory neurons.

Object identification

We can identify a wide range of objects, including concrete entities such as animals and people, and abstract entities such as spoken words. Many objects provide multimodal information. For example, spoken words can be identified from a combination of sound and vision (lip movements). In a noisy environment, such as when holding a conversation or watching television in a crowded room, we have to understand what someone is saying against interfering background sounds. In this situation, visual cues from lip movements, facial expressions, and gestures have a significant impact on speech comprehension. The improvement in comprehension with visual cues is equivalent to an improvement in signal-to-noise ratio of up to 15–20 dB (Spence, 2002). In the **McGurk effect** (McGurk & MacDonald, 1976), the speech sound actually heard by a listener is altered by observation of the speaker's lip movements. For example, if the sound specifies /b/, but the lip movements are consistent with /g/, the subject may report hearing /d/. When the subject closes their eyes, /b/ is heard correctly. A number of other effects have been reported in which the different senses combine to alter perception of objects and their properties:

- *Sound and touch* Manipulating the sound made as the hands are rubbed together can alter the perception of skin texture (Jousmaki & Hari, 1998; Guest et al., 2002).
- *Smell and color* The strawberry smell of a liquid appears stronger when the liquid is colored red (Zellner & Kautz, 1990).
- *Sound and light* Auditory noise presented with light tends to be perceived as louder than noise presented alone (Odgaard, Arieh, & Marks, 2004).
- *Light and touch* Tactile discrimination thresholds are lower during visual observation (Kennett, Taylor-Clarke, and Haggard, 2001).

In all these experiments the researchers took steps to ensure that the change in stimulus quality reflected perceptual processes rather than response bias.

KEY TERM

McGurk effect
A perceptual interaction between vision and hearing in which visual observation of a speaker's lips alters perception of speech sounds.

Localization

The apparent location of a sensory event in one modality can be influenced by information in other modalities. Ventriloquism is a long-established form of entertainment that relies on the audience localizing speech sounds at the moving lips of the ventriloquist's dummy rather than at the stationary lips of the ventriloquist. We experience this effect regularly when watching television, since the audio speaker is always located at the side or even at the rear of the television, yet we perceive the sounds to emanate from within the televisual scene. The illusion is actually due to biases introduced during multisensory interactions. Experimental evidence shows that the apparent location of a sound really is influenced by visual events: A flash of light influences the apparent location of a simultaneous sound (Radeau, 1994). Botvinick and Cohen (1998) describe a dramatic effect involving mislocalization of tactile stimulation. The experimental subject places an arm underneath a table, and a false arm fitted with a rubber glove is placed in view on the table directly above. The experimenter then strokes the subject's arm while simultaneously showing a brush stroking the false arm. Subjects report that the stroking sensation appears to be localized on the false arm—the arm appears to be their own arm despite obvious indications to the contrary.

THE NEURAL BASIS OF MULTISENSORY PROCESSING

Individual multisensory neurons respond to both unimodal and multimodal stimulation. The effect of the multimodal stimulus on a cell's response obeys three rules (Kayser & Logothetis, 2007):

Spatial coincidence A neuron sensitive to both visual and auditory stimuli may have receptive fields in both modalities; visual stimuli must fall in a specific region of space, and auditory stimuli must emanate for a restricted range of directions (these receptive fields were described in Chapters 7 and 4 respectively). The two receptive fields usually overlap in space. Multisensory response enhancement only occurs when the respective stimuli fall inside the area of overlap.

Temporal coincidence Response enhancement only occurs when stimuli in different modalities arrive at about the same time.

Inverse effectiveness If a neuron responds very strongly to a unimodal stimulus, its response changes relatively little when another stimulus is added. But if the unimodal response is weak, there are stronger interactions with other stimuli.

The importance of spatial and temporal coincidence suggests that strong multimodal responses are associated with stimuli that are likely to originate from the same object or event.

Turning to the question of where in the brain these multimodal neurons are found, the established view shown in Figure 13.1 is that multimodal processing occurs in high-level multisensory convergence zones in the brain (reviewed in Driver and Noesselt, 2008), but a number of different routes actually seem to be involved.

Superior colliculus

As mentioned in Chapter 7, this nucleus receives projections from multiple modalities, as well as the cortex, and is thought to be involved in multimodal orienting (Jiang, Jiang, & Stein, 2002). Some neurons in the superior colliculus, for example, show "**superadditive**" responses to multimodal stimulation; their response to stimulation from multiple senses is much greater than the sum of their response to each sense in isolation (see Alvarado et al., 2007).

Primary sensory cortex

Anatomical studies in primate cortex have found direct connections between early unimodal cortical areas, as well as connections between unimodal areas and polysensory cortex. Falchier et al. (2002) used an anatomical tracing technique to investigate connections between auditory cortex, visual cortex, and polysensory cortex (area STP). They found that primary visual cortex receives projections both from auditory cortex and from STP. Schroeder & Foxe (2002) found that auditory association cortex receives converging auditory, somatosensory, and visual inputs. Auditory and somatosensory inputs arrived in feedforward (bottom-up) layers of cortex, and visual input arrived in feedback layers (top-down). Rockland & Ojima's (2003) anatomical tracing study found that visual areas V1 and V2 receive projections both from auditory association cortex and from polysensory parietal cortex. It is also possible that projections from koniocellular neurons in the LGN provide feedforward cross-modal input to early unimodal processing areas (konicellular pathways were described in Chapter 7; Schroeder et al., 2003).

Research on the source of cross-modal effects in humans has relied on **event-related potentials (ERPs)**, which are minute but measurable fluctuations in electrical potential in the brain produced by changes in sensory stimulation. Specialized equipment can detect these fluctuations with millisecond accuracy. Several ERP studies have found short latency audiovisual interactions consistent with responses in early cortical areas (Giard & Peronnet, 1999; Foxe et al., 2000). Also, McDonald et al. (2003) claimed to find evidence for top-down feedback. They recorded ERPs

KEY TERMS

Superadditive response
A response to a combined stimulus which is greater than the sum of the responses to each stimulus presented separately.

Event-related potentials (ERPs)
Minute fluctuations in electrical potential in the brain caused by changes in sensory stimulation.

during a standard cross-modal cueing experiment. When the auditory cue was presented at least 100 milliseconds before a visual target, ERP activity occurred in multimodal cortical areas slightly before activity in visual cortical areas. McDonald et al. (2003) state that the result "strongly suggests that feedback from multimodal to unimodal areas" is involved in cueing. Although the data are suggestive, they do not offer conclusive evidence for feedback. Care must be taken when interpreting ERP and fMRI activity, since they may be influenced by such factors as attention, arousal, and imagery (see Driver and Noesselt, 2008).

Secondary association cortex

Several regions of sensory association cortex are involved in multimodal processing. The orbitofrontal cortex in macaque monkeys contains multimodal neurons that respond to smell and taste, or smell and vision, or taste and vision (Rolls & Baylis, 1994), so this area is likely to mediate multimodal flavor perception.

Bimodal neurons in the superior temporal cortex (in the dorsal processing stream) may well play a role in the integration of visual and vestibular signals about body motion (Gu, Angelaki, & DeAngelis, 2008). Functional neuroimaging experiments show that the lateral area of temporal cortex integrates visual and auditory signals (Amedi et al., 2005).

Cross-modal integration of visual and tactile object processing takes place in an area of association cortex lying at the border between temporal and occipital cortex (in the ventral stream; Amedi et al., 2001). So multimodal processing occurs in both major cortical processing streams.

In summary, a range of psychophysical and neuroscientific experiments offers convincing evidence that multimodal processing is much more pervasive than it was once thought, embracing cortical areas that are traditionally viewed as modality specific. This evidence does not invalidate the prevailing modular theory of sensory processing (see Coltheart, 1999). The established view of the sensory cortex being divided up into several modality-specific subsystems still holds, given the overwhelming evidence for functionally specialized brain regions discussed in earlier chapters. However the evidence also indicates that the modality-specific systems are more sophisticated than we once thought, sharing information between them in order to modulate processing in each modality. Theories of information processing in the brain must accommodate such effects. An important theoretical question arises as to how the information from different modalities is combined. The tutorials section at the end of the chapter discusses one contemporary approach to answering this question.

SYNESTHESIA

WHAT IS SYNESTHESIA?

Synesthesia is an intriguing multisensory phenomenon in which stimulation in one sensory modality causes a sensory experience in another modality. For example, some people experience a color sensation not from visual stimulation, but from auditory stimulation. "Colored hearing" is most commonly associated with speech sounds. Different sounds evoke different colors. There are also reports of color sensations evoked by nonspeech sounds, by touch, and by smell. The most common type of synesthesia ("sensory union") involves ordered sequences such as numbers, letters, or days being perceived as sequences of colors. Ward (2008) reports that, for one subject with synesthesia, the letters A and E evoke shades of red, I, M, and N evoke white, D evokes yellow, and so on. Numbers are similarly distinct in terms of color, as are musical sounds and certain words such as days of the week or city names. This phenomenology is typical of synesthesia.

The phenomenon was once on the fringes of scientific research, considered not worthy of serious investigation, but recent research has moved it into the mainstream. The modern study of synesthesia is informed by and can in turn inform theories of multisensory processing.

SYNESTHESIA AS A SENSORY PHENOMENON

Early doubts about the scientific worth of synesthesia were fueled by those who questioned whether it is a genuinely sensory phenomenon rather than learned associative pairings of colors and sounds, perhaps created while learning to read. Several lines of evidence now argue convincingly against the learning account. Baron-Cohen et al. (1993) investigated the consistency of the pairings over time. Nine experimental subjects who reported synesthetic experiences were asked to report the colors evoked by 122 words, letters, and phrases. A matched group of control subjects were asked spontaneously to generate a color to associate with each stimulus, and encouraged to use a mnemonic to aid recall. Both groups were retested on 10% of the words, one year later in the case of the experimental group, and one week later in the case of the control. Ninety-three percent of the experimental group's color responses were identical on retesting after a year, but only 37.6% of the control group's responses were identical after one week. Baron-Cohen et al. (1993) concluded that the phenomenon is a genuinely sensory one. They found that the initial letter of the word tended to determine the color it evoked. Colors reported were generally idiosyncratic to different individuals, though the vowels "i," "o," and "u" were consistently associated with the same colors in different individuals: "I" was grey, "o" was white, and "u" was yellow.

Further evidence for the sensory nature of synesthesia comes from neuroimaging studies. Paulesu et al. (1995) used PET to study brain activation while blindfolded synesthetic subjects and controls were presented with spoken words or pure tones. Both groups showed activation of cortical language areas when words were presented as opposed to tones. In the case of synesthetic subjects, several secondary visual cortical areas were also activated during word stimulation. Nunn et al. (2002) used fMRI to locate the cortical regions activated by speech in subjects with synesthesia and in control subjects. Synesthetic subjects showed activation in area V4/V8 in the left hemisphere, a region normally activated by color. Control subjects showed no activation in V4/V8 when imagining colors in response to spoken words, despite extensive training on the association. Neuroimaging studies show that a wide range of cortical areas is involved in synesthetic experiences, including sensory and motor cortex and areas in the parietal and frontal cortex (Rouw, Scholte, & Colizoli, 2011).

Synesthesia tends to run in families, and population studies indicate a genetic contribution, though the genes involved have not been identified. For example, Baron-Cohen et al. (1996) found a prevalence of one case of synesthesia in every 2000 people. Many more women than men reported synesthesia; the female:male ratio was 6:1. One third of the cases identified by Baron-Cohen et al. (1996) reported familial aggregation. However genes cannot offer a complete account of the phenomenon, because synesthetic responses can be triggered by blindness. Rao et al. (2007) reported auditory ERPs in the visual cortex of subjects who were totally blind as a result of physical trauma.

THE NEURAL BASIS OF SYNESTHESIA

The most plausible account of synesthesia is that it involves direct neural connections between unimodal cortical areas, of the kind that were discussed in the previous section. These connections may be enhanced genetically in people with synesthesia, or strengthened in response to traumatic sensory deprivation. Recall that one explanation of the phantom limb phenomenon (Chapter 3) is cortical reorganization in response to altered sensory input. The prevalence of associations between colors and words in synesthesia is consistent with this account, because the cortical regions responsible for vision and speech are located near to each other. Wernicke's speech area lies sandwiched between auditory cortex and secondary visual cortex (see Figure 13.3). It is also intriguing to note that synesthetic experiences are particularly associated with ordered sequences (numbers, letters, days, months, etc.). Several researchers have proposed that visual processing involves the execution of *visual routines*, which are special self-contained cognitive processes, rather like computer programs, that perform certain tasks such as counting, indexing, and tracking, and are actively invoked by attention (Ullman, 1984a; Cavanagh, 2004; Roelfsema, 2005). Perhaps the neural circuits mediating synesthetic experiences tend to be activated when counting or indexing routines are executed. Brain-imaging studies indicate that visual counting does involve a neural subsystem in secondary visual cortex, adjacent to speech and auditory areas (Piazza et al., 2002).

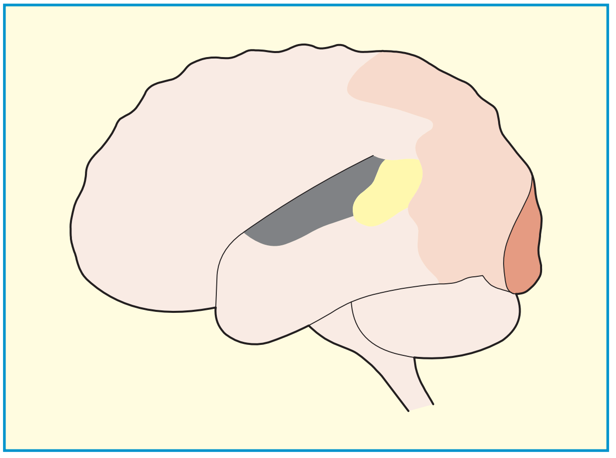

FIGURE 13.3
The dark brown area shows human striate visual cortex, and the light tan area shows extrastriate visual cortex. The dark gray area shows auditory cortex, and the yellow area shows Wernicke's speech area. Estimates of visual cortex were taken from Van Essen et al. (2001). The location of Wernicke's area is discussed in Wise et al. (2001).

THE UTILITY OF SYNESTHESIA

According to the theory of evolution, traits are transmitted to successive generations only when they confer an advantage on the host. What advantages might synesthesia provide? One suggestion is that people with synesthesia are more creative, and therefore more able to generate novel and adaptive ideas which promote survival and reproduction. This idea is contradicted by studies that find no conclusive evidence for differences in creativity between synesthetic and control subjects (Ward et al.,

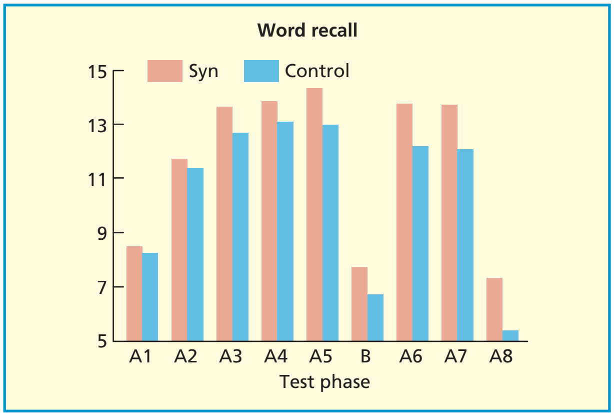

FIGURE 13.4
Word recall in synesthetic and control subjects. Each subject recalled 15 words that were read aloud by the experimenter. List A was read and recalled five times (A1 to A5), before list B was read and recalled (B). Finally list A was recalled three times (A6–A8) without being presented again. Synesthetic subjects consistently recalled more words than control subjects. Data taken from Yaro and Ward (2007), table 1.

2008). A more plausible account of the synesthetic advantage was offered by Yaro and Ward (2007), who compared memory in synesthetic and control subjects. The experiment required subjects to recall two lists of 15 words (A and B) read to them by the experimenter. List A was read and recalled five times, then list B was read and recalled. Finally list A was recalled on three occasions without it being presented again (the final recall test was two weeks later). Figure 13.4 shows the variation in word recall as this procedure progressed. Synesthetic subjects consistently recalled more words than control subjects. A small advantage in memory ability may be sufficient to sustain the synesthesia trait.

CHAPTER SUMMARY

EVIDENCE FOR MULTISENSORY PROCESSING

Multisensory stimulation aids processing in several ways:

- Orienting to novel stimuli is faster and more reliable with multimodal stimulation.
- The different senses combine to alter perception of objects and their properties.
- The apparent location of an event in one sensory modality is influenced by information in another modality.
- Visual and vestibular responses contribute to estimates of body motion.
- Information from several senses is combined to mediate perception of food edibility and flavor.

NEURAL SUBSTRATES OF MULTISENSORY PROCESSING

The responses of multimodal neurons obey three rules:

- Spatial coincidence
- Temporal coincidence
- Inverse effectiveness.

The established view is that multisensory neurons are found only in high-level convergence zones in temporal and parietal cortex, but recent research has found multimodal responses in other brain areas:

- Superior colliculus
- Primary sensory cortex
- Secondary association cortex.

SYNESTHESIA

Evidence for synesthesia as a genuine sensory phenomenon:

- Color-word pairings in synesthetes are too stable and idiosyncratic to reflect learned associations.
- Brain-imaging studies find differences between the brains of synesthetic and control subjects.
- Family and population studies indicate a genetic contribution.

Synesthesia is probably due to direct neural connections between unimodal cortical areas. The prevalence of color-word pairings may be due to the anatomical proximity of cortical areas encoding color and speech.

Synesthesia may have survived during evolution because it is accompanied by enhanced memory for words and colors.

TUTORIALS

BAYESIAN MODELS OF MULTISENSORY PROCESSING

Evidence for pervasive multisensory effects in perception indicates that object properties are often created by combining multiple sources of information. This begs the question as to how the information is combined. For example, if the brain receives both visual and auditory information as to the location of an object, such as a barking dog, it is unlikely that the location estimates in the two modalities will agree exactly because the two subsystems perform completely different computations on different information (for example, stereoscopic disparity in vision and interaural time differences in audition). So how does the brain compute a single estimate of location, which it must do in order to inform a decision on the best course of action (the dog can only have one location at any given moment)? The traditional view is that one sense, usually vision, tends to dominate over all others. More recently researchers have been using Bayesian inference in more sophisticated models of multisensory integration.

The basic principle in Bayesian inference theories (which were introduced in the tutorials section of Chapter 10) is to draw an inference about the world by combining different sources of evidence, while taking their reliability into account. Figure 13.5 shows a simple functional flow diagram of Bayesian inference in the context of multisensory processing. Each sensory modality (shown at the bottom of the diagram) supplies its own estimate of an object property such as location. Each estimate has a likelihood attached, which indicates its reliability. For example, acuity for position is much higher in vision than in audition, so the visual estimate should be

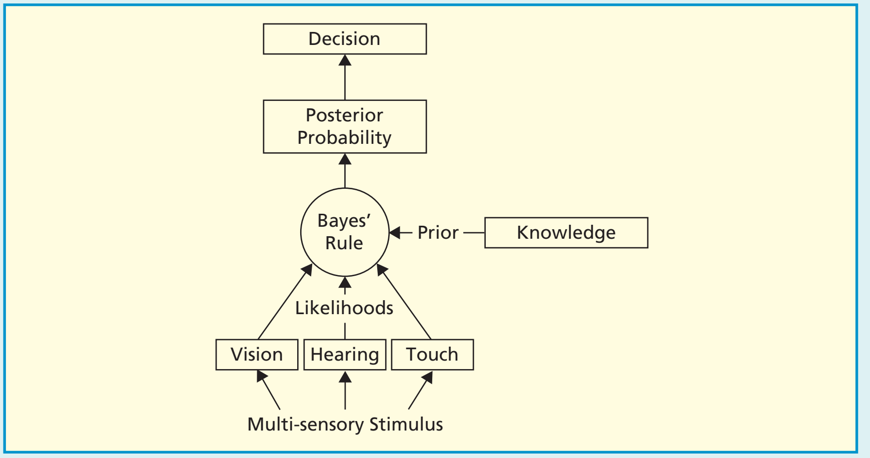

FIGURE 13.5
Flow diagram for Bayesian inference in multisensory processing. Each modality-specific module (bottom) supplies an estimate of a particular object property (likelihoods depend on the reliability of each estimate). The different estimates are combined with prior knowledge to produce a perceptual inference (posterior probability).

accurate and may therefore have a higher likelihood value. The different estimates are combined using Bayes' rule, which takes the average of the estimates but weights the contribution of each by that estimate's likelihood. Prior knowledge about the world also influences the final estimate, as described in Chapter 10. The result of the combination is a measure of the probability that a particular interpretation is correct (posterior probability); in the case of the earlier example, this would correspond to a particular location estimate.

This kind of model can be tested experimentally in studies of multimodal perceptual judgments. If the experimenter manipulates the relative reliability of different cues in the task, then according to Bayesian models the participants should alter the weight they attach to the cues when making their judgments. Alais & Burr (2004) asked participants to judge the location of a stimulus on the basis of combined visual and auditory cues (patches of light and brief audible clicks). They introduced small conflicts between the cues, displacing the position of one relative to the other, and also manipulated the reliability of the visual cue by blurring it so that it was spread over a wide area and consequently less precise. Results showed that unblurred light patches tended to dominate location judgments, but when blurred patches were used the auditory cue tended to dominate.

Ernst and Banks (2002) investigated the integration of visual and haptic (touch) cues in estimates of object size. They used virtual reality goggles to present images of small blocks, and a force feedback device to simulate the felt size of the blocks. Ernst and Banks (2002) manipulated the reliability of the visual cue. Similar to the effect reported later by Alais and Burr (2004), vision tended to dominate size judgments when it was a very reliable cue, but touch become more dominant when the visual cue was unreliable. The pattern of results in both experiments is consistent with a model of integration based on optimal Bayesian cue combination.