A model of auditory attention

Technical Report CS-00-07

June 2000

Stuart N Wrigley
s.wrigley@dcs.shef.ac.uk

Supervisor: Dr Guy J Brown
g.brown@dcs.shef.ac.uk

Speech and Hearing Research Group,
Department of Computer Science,
University of Sheffield

Contents

CHAPTER 1 Introduction

1.1. Auditory Scene Analysis

In typical situations, a mixture of sounds reach the ears. For example, a party with multiple concurrent conversations in the listener's vicinity, a musical recording or simply walking along a busy road. Despite this, the human listener can attend to a particular voice or instrument, implying they can separate the complex mixture.

Bregman (1990) has convincingly argued that the acoustic signal is subject to a similar form of scene analysis as vision. Such *auditory scene analysis* takes place in two stages. Firstly, the signal is decomposed into a number of discrete sensory *elements*. These are then recombined into *streams* on the basis of the likelihood of them having arisen from the same physical source.

The perceptual grouping of sensory elements into streams can occur by two methods: *primitive grouping* and *schema-driven grouping*. Primitive grouping is data-driven whereas schema-driven grouping employs knowledge acquired through experience of varied acoustic environments. Bregman explains primitive grouping in terms of Gestalt principles of perceptual organisation (e.g. Koffka, 1936). For example, the relationship between frequency proximity and temporal proximity has been studied extensively using the two tone streaming phenomenon (see Bregman, 1990 for a review). The closer in frequency two tones are, the more likely it is that they are grouped into the same stream. Similarly, the proximity of two tones in *time*, determines likelihood of streaming. As presentation rate increases, tones of similar frequency group together.

Additional Gestalt grouping factors include *good continuation*: sounds which tend to change smoothly in frequency intensity and spatial location are likely to form a single stream; and *common fate* whereby elements which change in the same way at the same time tend to group together. Common fate properties include common onset/offset, common amplitude modulation (AM) and common frequency modulation (FM).

Attempts to create computer models that mimic auditory scene analysis has led to a new field of study known as computational auditory scene analysis (CASA). There has been work varying from the simple voice separation techniques of Denbigh and Zhao (1992) to the broader CASA research of Cooke (1994), Brown (1992) and Ellis (1996). However, such techniques are functional in approach: some form of time-frequency analysis generally followed by a high-level inference engine to group elements into perceptual streams.

The difficulty involved in producing a computational solution is related to the mismatch between theories of perception, such as Bregman's, and the physiological processing substrate. Consider the two tone streaming stimulus (figure 1). Theories of perception are implied from experimental observations. Applying such mechanisms to figure 1, one can conclude that as δ*f* decreases, it is more likely that the tones will be grouped together. Similarly, as *TRT* decreases, sequential tones will also be more likely to group.

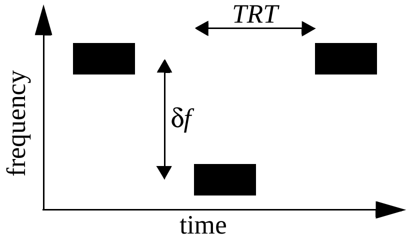


Figure 1. Portion of a two tone streaming stimulus consisting of high-low-high pure tones.

However, the neurophysiological mechanisms underlying auditory stream formation are poorly understood and it is not known how groups of features are coded and communicated within the auditory system. What does it mean to talk of 'frequency proximity' or 'temporal proximity'? The human brain relies solely on time varying electrical impulses with no 'symbolic' input as suggested by Bregman's theory.

The primary objective of this study is to create a physiologically based account of auditory scene analysis. If such a model can be shown to produce data with a high correlation to psychoacoustic experiments, it would provide evidence that the model is indeed processing sound in a similar way to the human auditory system. In essence, the goal of this work is to generate insights into the nature of the auditory system and to improve the effectiveness of current CASA technology.

A long term objective of this field of study is to improve the performance of automatic speech recognition (ASR) systems. Most systems rely on the incoming speech having been pre-segregated or consisting of only one speaker. In a realistic environment, this is not possible and so the process requires automation. A successful computational auditory scene analysis implementation would produce a considerable improvement in current ASR technology.

1.2. Attention

William James (1890) stated ‘*everyone knows what attention is*’. Indeed, the term attention is commonly encountered in ordinary language and people seem to understand what it means. Unlike other fields of research, people have strong convictions on its precise nature which haven't been arrived at by researchers: it is a fundamental part of their daily life and therefore something about which they know a great deal.

In common usage, attention usually refers both selectivity and capacity limitation. It is widely accepted that conscious perception is selective and that perception encompasses only a small fraction of the information impinging upon the senses. The second phenomenon - that of capacity limitation - can be illustrated by the fact that two tasks when performed individually pose no problem; however, when they are attempted simultaneously, they become very difficult. This occurs even when the two tasks are not physically incompatible such as reading a book and listening to the radio. It is this that leads to the common conclusion that attention is a finite resource.

Awareness of stimuli only occurs if they are attended to and the finite nature of attention leads to capacity limitation: when attending to one task there is less attention to devote to other tasks. Devotion of attention to one task is assumed to enhance performance (‘pay attention to your driving’) but can also be detrimental in some highly automatic tasks such as tying one's shoes.

Selectivity of perception, voluntary control of this selection and capacity limits are the core phenomenon addressed by attentional research. Unfortunately, so common is the use of the word ‘attention’, it can dangerously cloud one's thinking of these phenomena and subsequent explanations. In reading the following chapters, special effort should be made to avoid the everyday meaning of attention from obscuring any descriptions or findings.

The next section introduces some of the key terms and findings associated with attention research. Chapter 3 develops these findings and discusses a conceptual model to explain some of these findings. Chapter 4 concludes the report with a timetable for future research.

CHAPTER 2

Literature Review

This chapter aims to draw together a number of perceptual and physiological experiments conducted to gain an insight into the behaviour of human visual and auditory attention.

2.1. Single site allocation of attention

2.1.1. Frequency

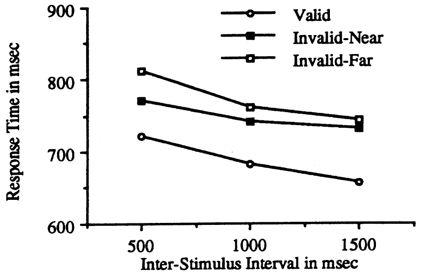
It has been known for some time that listeners are better able to detect expected tones as opposed to unexpected tones. Greenberg and Larkin (1968) developed a probe-signal paradigm to assess the extent of this expectancy effect. Subjects were presented with two intervals, both filled with white noise, one of which contained a pure tone. Listeners were instructed to indicate which interval contained the tone. The subject were led to expect the tone to be of a particular frequency (this signal was termed the *primary*). However, on less than a third of trials, the tone presented was of an unexpected frequency (this signal was termed the *probe*). Greenberg and Larkin found that detection performance was best for primary signals, intermediate for probes within the critical band and worst for probes outside the critical band.

This form of experimental design has been used by a number of researchers since (e.g. Schlauch and Hafter, 1991) to study the effect of expectancy on detection performance. In order to extend and substantiate such experiments, Mondor and Bregman (1994) proposed to investigate frequency selectivity within the context of an identification paradigm. Instead of embedding the tone in noise, the signal was presented in isolation and listeners were requested to indicate whether the target tone was longer or shorter in duration than the cue tone. On valid trials, the target and cue tones were of the same frequency. On invalid trials, the frequency separation of the two tones was manipulated to investigate the role of frequency similarity. In addition to this, Mondor and Bregman adjusted the interval between cue and target in order to determine whether the frequency selectivity effect was dependent on time available to allocate attention to the cued frequency region.

Validly cued targets were equally likely to be one of three different frequencies. Invalid cued targets were also equally likely to be any of these frequencies. This ensured that the experiment would be able to determine whether superior performance on valid trials was due to differential familiarity with the target or allocation of attention to a cued frequency region. In addition, the choice of two frequencies on invalid trials allowed two frequency separations to be used.

Figure 2 shows that identification performance (indicated by the median time from target onset to listener response for each trial) declines as frequency separation increases. It can also be seen that increasing the duration of the cue-target interval improves performance suggesting that a finite amount of time is required before attention is fully allocated to a particular frequency region.

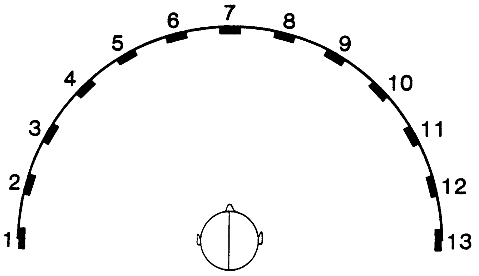
Figure 2. Response time as a function of trial type and interstimulus interval. From Mondor and Bregman (1994), figure 1.



2.1.2. Location

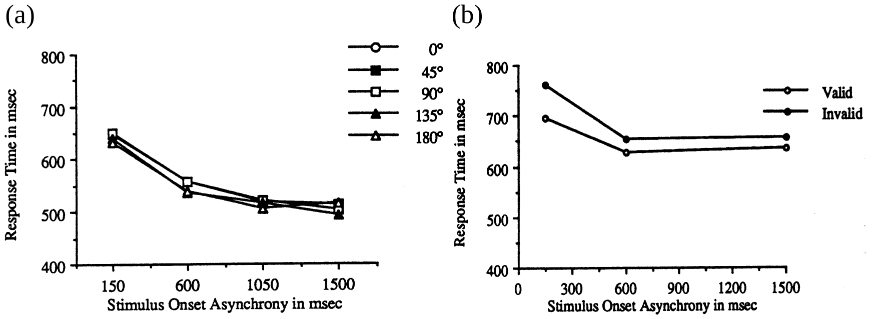
As indicated in section 2.1.1, cues that provide accurate information lead to faster and more accurate target identification than do cues that provide inaccurate information. This is also true of cues providing spatial information with regard to the localisation of an acoustic target. Mondor and Zatorre (1995) examined whether the time required to perform a shift of attentional energy was proportional to the distance of the shift. Shift distance is defined as the spatial separation in degrees between the location of the fixation sequence and the location of the cue and target. Each subject was placed at the centre of a semicircle of speakers (figure 3). A fixation sequence was used to control the focus of attention at the beginning of each trial: listeners were instructed to detect a drop in intensity of a steady tone presented from a particular location. Following this, a brief noise burst was delivered as a spatial cue from the spatial location from which the target tone would sound.

Figure 3. Schematic description of the speaker array used in the experiments of Mondor and Zatorre. From Mondor and Zatorre (1995), figure 1.



Despite all trials being *valid*, the cue-target interval was varied to control the amount of time available to orient attention. Figure 4a shows a similar trend to that of figure 2, in which performance increases as cue-target interval increases.

Figure 4. (a) Response time as a function of attentional shift distance and cue-target onset interval. From Mondor and Zatorre (1995), figure 2.
(b) Response time as a function of cue validity and cue-target onset interval. From Mondor and Zatorre (1995), figure 3.



However, it is important to note that the shift distance has no effect on performance. This is in direct contrast to evidence gathered by Rhodes (1987) in which she found that response times increased linearly with spatial separation up to a certain point, beyond which response times were similar. From this, Rhodes concluded that analogical shifts occurred for relatively short distances and discrete movements occurred for larger shifts. Despite being a controversial topic of debate, Mondor and Zatorre remark that “none of the investigators of visual attention has argued in favour of a model that incorporates both analogical and discrete movements”. In fact, Mondor and Zatorre suggest that Rhodes’ misinterpretation of her data was due to a correction procedure which failed to eliminate the effect of azimuthal position on localisation performance.

The most consistent explanation of this performance is that the cue is causing attention to be oriented to that location. However, one could also argue that the cue was simply acting as a general alert to the listener. In a minor alteration to the first experiment, Mondor and Zatorre introduced a number of inaccurate spatial cues on a portion of the trials. If an auditory cue acts solely to alert the listener, then both valid and invalid cues ought to result in identical performance. However, if auditory attention is oriented on the basis of the spatial cue, increased performance ought to be observed with valid cues. As expected, targets preceded by a valid cue were identified more quickly than those preceded by invalid cues. This provides strong evidence that listeners orient attention to the position in which the cue sounds.

In summary, Mondor and Zatorre found that performance improved as time available to shift attention to a cued spatial position increased. Furthermore, accurate spatial cues facilitated performance more than inaccurate ones: performance declined as the distance of an unexpected target from a cue spatial location increased. Their evidence is also consistent with a discrete attention-allocation model.

2.2. Multiple site allocation of attention

2.2.1. Frequency

In addition to studies concentrating on the role of single site expectancy (or attention allocation), work has also been conducted in which there is uncertainty about the frequency of the signal to be detected. Results from these types of experiments are often compared to *ideal* listeners defined by assumptions regarding the nature of the detecting mechanism (for example, parameters of a bank of bandpass filters). A popular model to arise from this is based on the listener who monitors *M* orthogonal bands (MOB), only one of which contains the target signal (Green and Swets, 1966). Solutions from these models agree qualitatively with the observation from human listeners that the signal level of the target must be increased for detection as the number of monitored bands increases. However, Green (1960, 1961) had found that the predicted loss in sensitivity as *M* increased was in fact larger than that seen in human listeners. He accounted for this by proposing that there was an intrinsically high amount of uncertainty present in all conditions. According to Schlauch and Hafter (1991) Green’s experiments were flawed: they did not control for the possible cognitive influences on listening bands. A listening band is defined as the band on whose output a listener decides whether or not a signal has occurred. Green employed a traditional probe-signal method in which the cue signal was not presented on a trial-by-trial basis forcing subjects to rely on memory to monitor the appropriate frequency bands. The use of only two frequencies to monitor also limited the amount of loss due to uncertainty.

Schlauch and Hafter (1991) controlled for these factors by employing trial-by-trial cuing of subjects (Greenberg and Larkin, 1968) and selecting the frequencies to be monitored at random from a wide range of possibilities to avoid memory effects. Finally, to increase the possible amount of loss due to uncertainty, the number of bands to be monitored was doubled to four. Figure 5 shows the format of their signals.

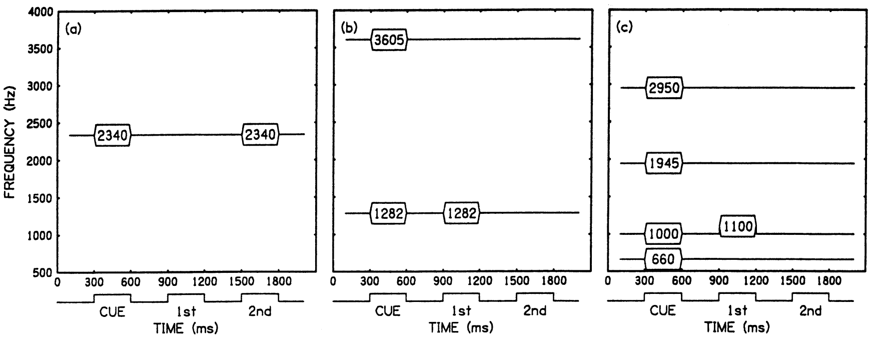


Figure 5. Schematic of three types of trials. (a) M=1 and an expected target. (b) M=2 and an expected target. (c) M=4 and a probe signal. Probe signals are always 1.1 times one of cue frequencies. From Schlauch and Hafter (1991), figure 1.

As observed in single site expectancy experiments, detection performance dropped for frequencies above and below the expected frequencies (figure 6a). In addition to this data, Schlauch and Hefter also calculated the *psychometric function* for each cued condition. It has been classically considered that a threshold is that intensity above which a stimulus can be heard and below which it can not. This is an oversimplification: if the intensity of a stimulus is slowly increased from a low value, there is no well-defined point at which subjects suddenly report the stimulus to be detectable. Instead, there is a range of intensities over which the subject will sometimes declare the stimulus detectable and at other times declare it undetectable. When the responses to a number of trials are plotted with percent ‘detectable’ responses on the ordinate and signal magnitude on the abscissa, a distinctive sigmoidal shape is produced. This plot is termed a psychometric function. This allowed the performance in dB of probes relative to expected targets to be inferred. Additionally, this allowed the hypothesis that the slope of the psychometric function increases with *M* (Green and Swets, 1966) to be tested.

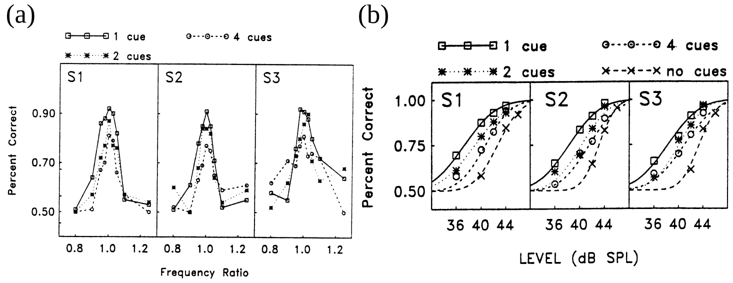


Figure 6. (a) Data from three subjects using one-, two- and four-tone complexes as a cue. Abscissae represent ratio of signal to be detected to target frequency. (b) Psychometric functions for the three subjects. Abscissae represent dB SPL level of a 500Hz tone. From Schlauch and Hafter (1991).

Figure 6b shows that the slope of the psychometric functions does indeed increase as *M* increases which is consistent with evidence from Johnson and Hafter (1980) reporting increased slopes as uncertainty increased.

The performance in dB of probes relative to expected targets also produced interesting results.

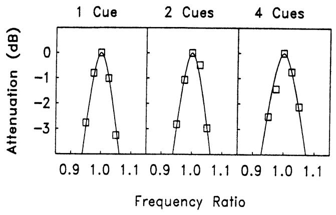


Figure 7. Average listening bands for one, two and four cues. Best fit lines are provided by a ROEX(p) function. From Schlauch and Hafter (1991) figure 4.

The functions shown in figure 7 were provided by adjusting the value of *p* in the ROEX(*p*) filter (Patterson and Moore, 1986). Patterson and Moore, (1986) showed that the equivalent rectangular bandwidth (ERB) of a filter is (4/*p*)*F* where *F* is the centre frequency. From this, it was found that the width of the filters were 12%, 12.4% and 13.7% of the centre frequency for cases *M*=1, *M*=2 and *M*=4 respectively which are essentially the same as those obtained using notch-noise masking (Moore and Glasberg, 1983).

In summary, Schlauch and Hafter (1991) show that there is a significant loss in detection performance due to increasing *M* but that it was indeed possible to monitor a number of harmonically unrelated frequency bands simultaneously. The performance decrease as probes move out of these bands is consistent with the hypothesis that attention is allocated to a number of discrete frequency regions. It is also interesting to note that the width of listening bands increases as the number of frequency regions to be attended rises. However, if the ability to allocate attention to multiple frequency regions does indeed exist, this is in contrast to that apparent in vision, where observers can attend to only one location at a time (e.g. Eriksen and Webb, 1989). Such a difference may suggest that visual and auditory attentional mechanisms operate within different structures in the cortex.

2.2.2. Space-frequency

Previous sections have demonstrated that the performance of target identification can be influenced by prior cues regarding frequency or location of an imminent target. These results are taken to be evidence for the allocation of some form of auditory attention to either single (Mondor and Bregman, 1994; Mondor and Zatorre, 1995) or multiple (Schlauch and Hafter, 1991) sites. However, these studies have not considered what relationship, if any, exists between these two modalities. If a cue contained information from more than one modality, which would take precedence? Deutsch (1974; Deutsch and Roll, 1976) investigated this question by presenting listeners with a succession of pure tones dichotically. In one ear, an 800Hz tone was presented three times followed by two presentations of a 400Hz tone. In the other ear, a 400Hz tone was presented three times followed by two 800Hz tones. The particular frequency heard and the location from which that frequency apparently originated seemed to be governed by separate processes: the frequency heard was that presented to the listeners dominant ear and the location was that of the high tone. However, Bregman and Steiger (1980) argued that this *illusion* was likely to be a conflict of two perceptual organisation principles: grouping by frequency and grouping by location. It was suggested that this conflict was forcing the listeners' auditory system to call upon another, more reliable, form of grouping: the use of a higher harmonic to determine the location of a complex. It was unfortunate that Deutsch used an 800Hz tone which can be viewed as a harmonic of the lower, 400Hz tone. From this, Bregman and Steiger concluded that Deutsch's illusion was in fact the emergent behaviour of a preattentive process in which perceptual features are combined into auditory *objects*. Indeed, Bregman (1990) remarked,

"*The perceptual stream-forming process has the job of grouping those acoustic features that are likely to have arisen from the same physical source. Since it is profitable to attend to real qualities, locations, and so on, rather than to arbitrary sets of features, attention should be strongly biased toward listening to streams*" (p. 138).

Mondor *et al.* (1998) investigated this interdependence of frequency and spatial information to determine if attention is directed at such streams. Using the same form of speaker array as their earlier experiments (Mondor and Zatorre, 1995), as shown in figure 3, listeners were given the task of categorising pure tones on the basis of frequency (low vs. high) or spatial location (central vs. peripheral). In controlled conditions, no variation was made in the ‘irrelevant’ dimension. However, in the selective attention conditions, variations were made in the irrelevant dimension which were uncorrelated to variations in the relevant condition. If auditory attention is allocated separately to the location and frequency dimensions, then no performance degradation should be observed. Alternatively, if auditory attention cannot be allocated separately to the two dimensions, the performance will suffer interference from the variations in the irrelevant dimension.

Consistent with the notion that attention is directed toward streams, listeners found it impossible to ignore variation on an uninformative dimension while making classification judgements on the basis of a second dimension. Mondor *et al.* conclude that “*auditory attention acts to select streams*” (p. 68). When the relative salience of the frequency and location dimension was investigated, listeners were unable to guide selection independently by location or frequency. In other words, neither dimension dominates. Similar modality interdependencies have been observed for pitch, timbre and loudness (Melara and Marks, 1990). This is in conflict with theories of selective attention which have arisen from visual experiments. The feature integration theory (FIT) of Treisman and colleagues (Treisman and Gelade, 1980; Treisman and Gormican, 1988) postulates that selective attention is required to perform a discrimination or detection task only when two or more features are in conjunction and not in variations of one feature. However, evidence collected by Mondor *et al.* (1998) suggests that classification cannot be based on a single feature. Furthermore, no evidence was found to support the dominant role of location present in the FIT.

2.2.3. Physiological evidence for the space-frequency conjunction

Mondor *et al.* (1998) found that listeners were unable to attend to the location of an acoustic stimulus independently of its spectral characteristic, and *vice versa*. If this is true and auditory attention acts on objects rather than individual features, attending to different stimulus features which are integrated ought to result in similar cerebral activity. Zatorre *et al.* (1999) aimed to investigate this prediction by instructing listeners to perform a task which required detection of tones of a specified frequency or at a specified spatial location while undergoing positron emission tomography (PET) scanning. Their findings indicate that an auditory attentional task engages a specialised network of right-hemisphere regions, in particular the joint participation of the right parietal, frontal and temporal cortex. As expected, changes in the CBF were very similar when subjects attended to spatial or to spectral features of the acoustic input supporting the model of Mondor *et al.* (1998) in which an initial stage of feature integration precedes selection on the basis of such streams.

2.3. Attentional 'shape'

In the light of the evidence supporting the allocation of attentional resources to one or more sites simultaneously, it is interesting to consider the 'shape' of the attentional deployment. Two general classes of model have been proposed to describe the focus of attention. *Spotlight* models propose that attention is allocated to a discrete range of frequencies with an even distribution within this range. The edges of this spotlight are characterised by a sharp demarcation between attended and unattended frequencies. Alternatively, the attentional focus may be defined as a *gradient* with the density of the attentional resources being the greatest at the cued frequency and declining gradually with frequency separation from the focal point of attention. In a similar experiment to that described in section, Mondor and Bregman (1994) increased the number of possible frequency separations used on invalid trials to three.

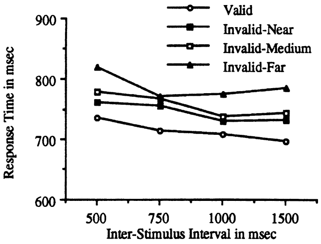


Figure 8. Response time as a function of trial type and cue-target interval. From Mondor and Bregman (1994), figure 2.

Not only is another strong cue validity effect observed (figure 8) but the effect of frequency separation is only consistent with a gradient of attention. As frequency separation increases, so too does the response time. No model incorporating a spotlight of attention with abrupt changes between attended and unattended frequencies could account for this result. This finding is also supported by evidence gathered by Mondor and Zatorre (1995).

2.4. Two forms of attention

It has been suggested that visual attention may be oriented by two different mechanisms which rely on differing amounts of conscious intervention by the listener. The *exogenous* system is considered to take place automatically under pure stimulus control: attention is drawn to the site of the stimulus. *Endogenous* attention is considered to be under control of the listener, whereby attention can be consciously oriented to a particular site (Jonides and Yantis, 1988; Müller and Rabbitt, 1989). In other words, the exogenous system is engaged by peripheral cues such as a cue which signals the probable location of a forthcoming target. In contrast, the endogenous system is engaged by symbolic cues which have to be processed and interpreted before attention can be oriented. Spence and Driver (1994) have argued that these systems are also present in the allocation of auditory spatial attention. If this is true, the studies investigating the frequency sensitivity (e.g. Mondor and Bregman, 1994; Schlauch and Hafter, 1991) and spatial sensitivity effects (e.g. Mondor and Zatorre, 1995) are, in fact, examining the allocation of exogenous attention.

Support for these two mechanisms can be found in the data collected by Hafter *et al.* (1993) in which the effectiveness of two types of cues for reducing frequency uncertainty was studied: *iconic* cues and *relative* cues. Iconic cues are those usually employed in the probe-signal methods described in previous sections. Relative cues were set to be two thirds the frequency of the expected signal - they acted as the symbolic cues which would stimulate the endogenous system. Hafter *et al.* found that both relative and iconic cues were successful in reducing the amount of uncertainty compared to the no-cue situation. However, it also emerged that the listening bands used with relative cues were wider than typically measured by a factor of roughly 1.6. This suggests that the use of iconic and relative cues does indeed engage different mechanisms of attention allocation.

2.5. Summary

This chapter has presented evidence that visual and auditory attention is deployed in a number of interesting ways. Fundamentally, attention can be directed to one (e.g. Greenberg and Larkin, 1968; Mondor and Bregman, 1994; Mondor and Zatorre, 1995) or more (e.g. Green 1960, 1961; Green and Swets, 1966; Schlauch and Hafter, 1991) sites of interest identified by some form of cuing. Furthermore, Mondor *et al.* (1998) have argued that there is an interdependence of frequency and spatial information consistent with the hypothesis that attention is directed at streams. Brain imaging studies (e.g. Zatorre *et al.*, 1999) have provided data consistent with this conclusion. These researchers have also shown that it is highly likely that the focus of attention can be described by a gradient model in which the density of the attentional resources is the greatest at the cued frequency and declines gradually with frequency separation from the focal point of attention. In addition to this, Jonides and Yantis (1988) and Müller and Rabbitt (1989) have argued that attention can be split into two mechanisms: unconscious and conscious allocation (exogenous and endogenous, respectively).

CHAPTER 3

Conceptual Model

In the previous chapter, we looked at a number of basic expectancy and attentional phenomena such as the allocation of attention to particular regions of frequency or space. This chapter will cover a number of more detailed psychophysical and physiological studies which will form the basis of the conceptual model described at the end of the chapter.

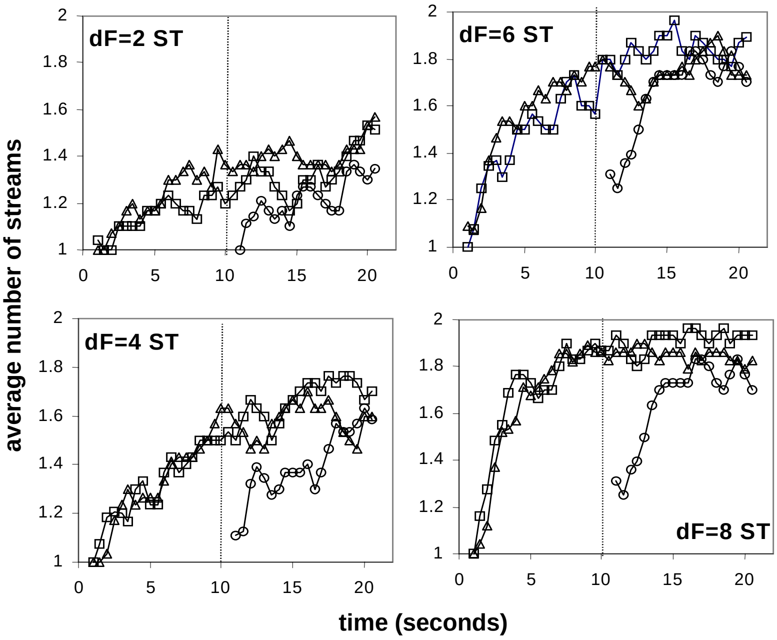
3.1. Stream perception in the presence of competing stimuli

Until recently, it was thought stream formation such as that involved in the two tone streaming phenomenon (Bregman and Campbell, 1971; van Noorden, 1975) was passive in nature: streaming occurred whether the tone sequence was attended to or not. Attention was considered useful only in guiding a particular stream into the attentive ‘foreground’. However, recent work by Carlyon *et al.* (1999) suggests that attention does indeed play an important role in stream formation.

In Carlyon’s experiment, a 21s sequence of A and B pure tones alternating in an ABA-ABA sequence was presented to the left ear. In the ‘baseline’ condition, no stimulus was presented to the right ear. Subjects were instructed to indicate whether they heard a galloping rhythm or two separate streams. In the ‘two-task’ condition, a series of bandpass filtered noise bursts were presented to the right ear for the first 10s of the stimulus. The noise bursts were labelled as either *approaching* (linear increase in amplitude) or *departing* (the approaching burst reversed in time). For the initial 10s, subjects were instructed to ignore the tones in the left ear and simply concentrate on labelling the noise bursts. After 10s the subjects switched to the streaming task. In the ‘one-task-with-distractor’ condition the noise bursts were presented to the right ear, as in the two-task condition, but subjects were told to ignore them and to perform the streaming task on the tones in the left ear throughout the 21s sequence. Consistent with Anstis and Saida (1985), subjects heard a single stream at the beginning of each sequence with an increased tendency to hear two streams as the sequence progressed in time. However, for the two-task condition the amount of streaming after ten seconds is similar to that at the beginning of the baseline sequence - in the absence of attention, streaming had not built up (figure 9).

Furthermore, a second experiment required listeners to assess the nature of the tones which made up the sequence - ‘fast’ or ‘slow’ amplitude modulation - in order to show that the lack of steam segregation in the first experiment was not a result of attending to a different ear. Again, stream segregation did not occur in the presence of the attended auditory task.

Figure 9. Build up of streaming over time for four frequency differences. Scores are averaged across listeners and repetitions for the baseline (triangles), two-task (circles), and one-task-with-distractor (squares) conditions. From Carlyon (1999) figure 3.



To summarise, the findings of Carlyon *et al.* suggest that attention is required for streaming to occur.

3.2. Perception without attention

Many theories of visual perception assume that the extraction of perceptual primitives occurs preattentively: object-based theories maintain that the visual scene is first parsed in accordance with Gestalt principles and then attention is directed to the perceptual objects that result from the parsing process (e.g. Duncan, 1984). In contrast with this view, recent work (e.g. Mack *et al.,* 1992) has suggested that little grouping, if any, occurs preattentively. Mack *et al.* developed an experimental method to investigate what can be perceived under conditions of inattention (the stimuli are within a person's visual field but no attention has been directed toward them). Subjects were presented with a difficult perceptual task, such as identifying the longest arm on a briefly presented cross, superimposed on a background of coloured dots. On the majority of trials, these dots were randomly black or white. On one trial, however, these dots would form a salient pattern if grouping occurred. At the end of the experiment, subjects were unexpectedly asked to make a forced choice decision about the background pattern. If grouping occurs without attention, then despite not attending to the background, subjects would still be able to report the pattern that had occurred.

The results from Mack *et al.*'s study show that accuracy was at chance, with a number of participants even denying that a pattern had even occurred. This is in contrast with traditional object-based theories which would have predicted a high pattern identification accuracy.

Although the subjects were unable to report the pattern, it is not necessarily the case that grouping did not occur. A possibility is that the subject may not be able to remember the pattern: the perception of the background on the critical trial may not have been encoded into memory. To control for this factor, Moore and Egeth (1997) employed a difficult line-length discrimination task on two horizontally oriented line segments superimposed on a background of dots which occasionally formed a pattern.

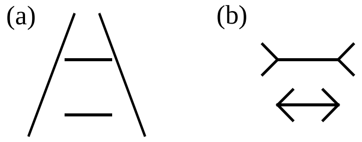


Figure 10. Two perceptual illusions. (a) The Ponzo illusion: the line segment closer to the converging lines appears longer than the identical other. (b) The Müller-Lyer illusion: the line segment with arrowheads that point in appears longer than the identical line segment with arrowheads that point out. Redrawn from Moore and Egeth (1997) figure 2.

If perceived, these patterns could influence the perceived lengths of the lines such as in the Ponzo and Müller-Lyer illusions (figure 10). On trials in which the background pattern formed an illusion, the horizontal line segments were identical: if the background dots were not grouped, the discrimination task performance would be at chance. However, if the dots were grouped, this ought to influence the perceptual task and the line nearest the convergence (for the Ponzo illusion) or the inward pointing arrowheads (Müller-Lyer illusion) would be reported as longest more often than chance. Indeed, this was the case. At the end of the experiment, the subjects were asked if they saw a pattern and were given a forced choice decision over which pattern they saw. A negligible number of participants reported observing a pattern and the identification accuracy was chance. This strongly suggests that grouping did occur during the experiment but that the patterns were either forgotten or never successfully encoded into memory.

In summary, Mack *et al.* (1992) suggested that even salient grouping patterns are not perceived when not directly attended. Moore and Egeth (1997) extend this finding by showing that grouping of the unattended stimuli does occur: the grouped patterns influence the perceptual task. However, such preattentive grouping cannot be later reported. Moore and Egeth suggest that '*attention may be required not for perceptual organisation but for encoding the results of that organisation in memory*'.

Support for preattentive grouping can be found in the mismatch negativity (MMN) studies of Sussman, Ritter and Vaughan Jr (1998, 1999). The MMN is a component of event-related potentials (ERPs) which provides information about preattentive auditory processing. It is believed that the MMN is the outcome of a comparison process when the incoming stimulus differs from the memory of the stimulus in the recent past. It is considered preattentive because attention is not required to elicit a response. In these studies, Sussman *et al.* presented listeners with an AB sequence (similar to that used by Carlyon et al., 1999) to investigate the effect of attention on two tone streaming (Bregman and Campbell, 1971; van Noorden, 1975). In this well-studied phenomenon, listeners are more likely to hear two streams when the tones are presented at a high rate (one stream being A-A-A and the other being B-B-B). At low rates, streaming fails to occur and subjects continue to hear the AB sequence. Sussman *et al.* (1999) reasoned that if grouping occurs preattentively, it may be possible to illicit a MMN response to a deviant which could only be perceived if streaming had occurred. Furthermore, no MMN ought to be observed in the slow presentation situation. Figure 11 shows the standard AB sequence and the deviants predicted to illicit a MMN response. In all cases, the subjects performed another task and were told to ignore the stimuli. As expected, when the tones were presented at the fast pace, MMNs were detected in response to the deviant sequences occurring in both the high and low tones. When the tones were presented at the slow pace, no MMNs occurred for either the high or low tones.

From this, Sussman *et al.* concluded that the streaming effect occurred automatically, (independently of attention) at, or before, the level of the MMN system.

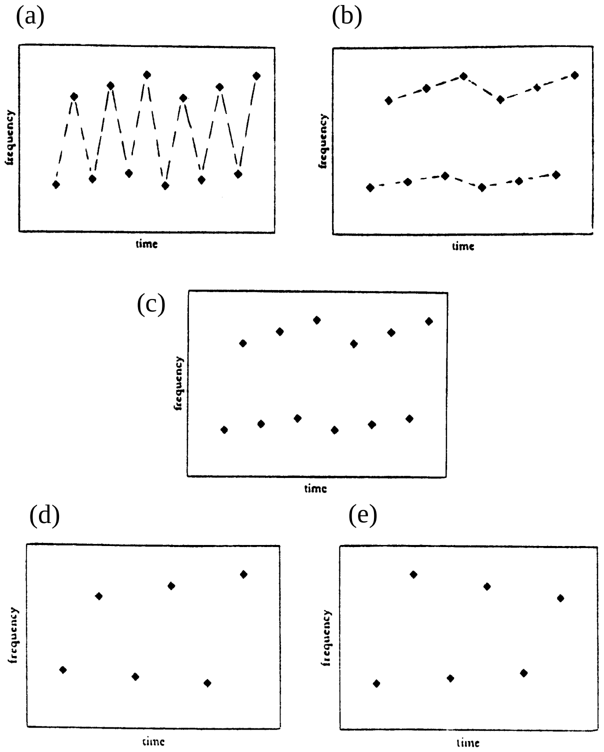


Figure 11. (a) Perception of the AB sequence under slow presentation; (b) perception of the same sequence under rapid presentation. (c) The standard AB cycle of tones. Note both high and low tones exhibit a rising frequency trend. (d) and (e) Deviants to the standard cycle in the low tones and high tones respectively. Adapted from Sussman *et al.* (1999) figure 1.

Further work on these stimuli (Sussman *et al.*, 1998) suggested that attention could even force streaming to occur in certain occurrences of the slow pace situation. Listeners were instructed to attend to only the high tones and signal when they detected a deviant sequence. The results of their study show listeners could keep track of the standard three tone pattern within the high tones by employing highly focused attention. Additionally, MMNs were detected for both the attended (high) tones and the unattended (low) tones suggesting that selective attention can alter the organisation of sensory input. The observance of a MMN in the unattended stream may indicate that the preattentive input to the MMN system has been altered by attention.

Caution should be exercised with respect to the extent to which attention was diverted away from the tone sequences in Sussman *et al.*'s experiments. In the situation when subjects were instructed to ignore the stimuli, the distracting task was to read a book - a passive, visual exercise. It was nevertheless the case that the tones were the only sounds present in the experiment. Duncan, Martens and Ward (1997) have suggested that the visual and auditory attentional systems are largely independent and so one could speculate that some degree of attention was still being directed toward the auditory stimuli. The observance of streaming by Sussman *et al.* (1998, 1999) in a situation of inattention is in contrast to that of Carlyon *et al.* (1999) in which a more rigorous distractor mechanism was employed prevented streaming from occurring. Finally, despite the claim that mismatch negativity responses provides information about preattentive processing, it is interesting to note that Sussman *et al.* (1998) have concluded that attention can influence MMN responses (see also Alain and Woods, 1997; Trejo *et al.*, 1995).

3.3. The conceptual model

The studies presented in the above section and also the previous chapter highlight a number of factors which must be explained and emulated by the conceptual model. Most importantly is the finding by Carlyon *et al.* (1999) that the distraction of auditory attention prevents the percept of streaming from occurring. This led Carlyon *et al.* to conclude that streaming requires attention. However, it is possible that this is not the case. Moore and Egeth (1997) and Mack et al. (1992) convincingly argue that attention is not required for perceptual organisation *per se* but that it is required for the *encoding in memory* of the perceptual organisation. In other words, subjects taking part in the experiments of Carlyon *et al.* may have begun to stream the tone sequences but since attention was diverted to a difficult perceptual task this was not encoded into memory and the build up of streaming (Anstis and Saida, 1985) did not occur. Mismatch negativity research conducted by Sussman *et al*. (1998, 1999) also supports the classical view that grouping and streaming are preattentive activities (Bregman, 1990).

It has been suggested that visual attention can be considered to be engaged by two different mechanisms: the *exogenous* (subconscious) and *endogenous* (conscious) systems (Jonides and Yantis, 1988; Müller and Rabbitt, 1989; see also James, 1890). These two types of attention form the basis of the model: exogenous attention accounts for subconscious (preattentive[1]) processing and the endogenous system controls which organisations are perceived and/or processed further.

3.3.1. The model

The model presented here assumes that it is possible to have a number of simultaneous exogenous processes occurring and that endogenous attention is required to allow the outcome of one of these processes to be perceived. A percept only occurs when a perceptual organisation is encoded into memory (Moore and Egeth, 1997). Figure 12 shows this framework.

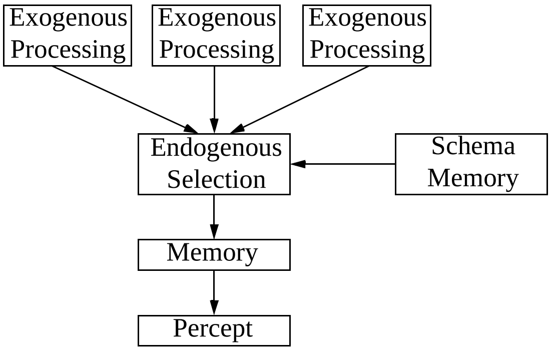


Figure 12. Structure of the attentional model. Note that endogenous interaction is required before exogenous perceptual organisations can be encoded into memory and perceived.

Within this framework, endogenous interaction acts as a kind of filter on the exogenous inputs. However, it should be emphasised that this is different from the traditional attentional spotlight as proposed by Crick (1984) in that a particular organisation is selected and encoded into memory in order to be perceived rather than selective attenuation occurring.

The focus of conscious, endogenous, attention is not controlled entirely by the subject. Schema (Bregman, 1990) driven processing can be exemplified by the cocktail party effect (Cherry, 1953) in which the listener's attention can be unexpectedly redirected by another speaker mentioning their name. A more common form of unconscious redirection of conscious attention can happen in virtually any environment: a loud, usually transient, sound occurring unexpectedly such as a bang or a crack. Therefore, the endogenous selection process requires more than conscious input. It is proposed that endogenous attention is also

[1] For the purposes of this model, it is necessary to clarify the meaning of the term *preattentive*. The classical meaning of this term refers to processing which occurs before conscious intervention by the subject. In the framework of endogenous and exogenous attention, preattentive processing can be considered to be equivalent to exogenous attentive processing.

influenced by other processes such as difference detectors, schema recognisers, etc. Figure 13 shows a possible mechanism by which such a selector may work.

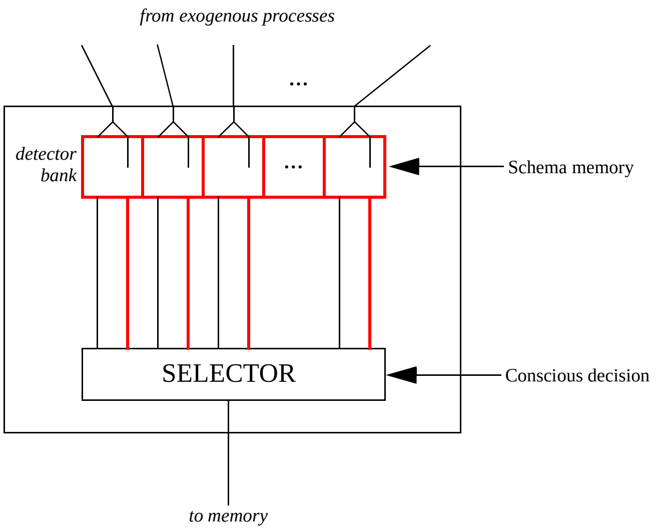


Figure 13. Enlargement of the endogenous selection mechanism from figure 12. The *detector bank* indicated in the diagram corresponds to the proposed difference detectors and schema detectors of the model.

One or more exogenous processes are responsible for grouping and streaming of stimuli reaching the ears. The perceptual organisations of these processes are directed into the endogenous selection module. It is at this stage that both conscious decisions and salient information about the incoming organisations are combined and a selection made. Salient information could be gathered from a number of sources such as difference detectors and schema recognisers (both contained within the *detector bank* abstraction of figure 13): changes in intensity or overall grouping structure may be considered in the difference detectors. Such an assumption finds support in the mismatch negativity studies conducted by Sussman *et al.* (1998, 1999) which suggest that a component of event-related potentials indicates the outcome of a comparison process when the incoming stimulus differs from the memory of the stimulus in the recent past. The fact that conscious attention is not required to illicit such a response is consistent with the assumption made in this model that such difference detectors function continually and may influence the selection process. In a similar fashion, schema driven processing can influence the direction of endogenous attention independently of the particular stream attended (Bregman, 1990; see also Cherry, 1953). This implies that schema recognition is being carried out on all exogenous process inputs and influences the final selection process.

The last stage of the endogenous selection process is the combination of all the decision factors: schema and difference detection outcomes and conscious choice. In the absence of any evidence from the detector banks, the decision will be consistent with the conscious choice. However, should evidence appear that important information is present in a currently unattended stream, the selector overrules the conscious choice and directs attention to that stream.

3.3.2. Thought experiments

Carlyon et al.'s stimuli

Consider a listener who is participating in the experiment used by Carlyon *et al.* (1999) in which noise bursts are presented simultaneously with a galloping ABA-ABA tone sequence. The initial state of perception is that of fusion: a single stream is perceived to which endogenous attention is directed by default. The listener rapidly (almost instantaneously) becomes aware of the two types of sounds: tones and noise bursts. At this stage, the subject has to make a decision over which perceptual object to attend (figure 14).

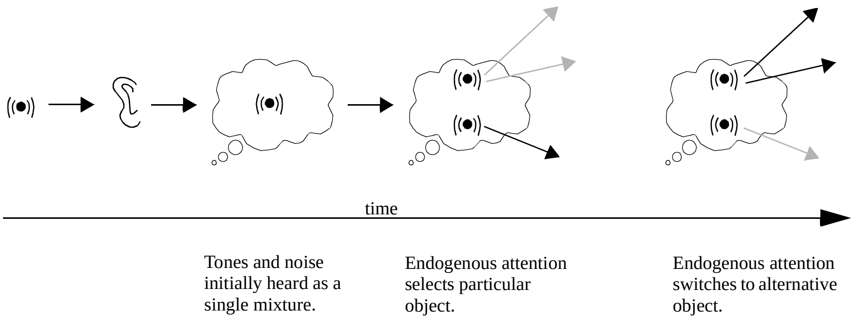

time
Tones and noise initially heard as a single mixture.
Endogenous attention selects particular object.
Endogenous attention switches to alternative object.

Figure 14. Behaviour of the model in response to the Carlyon *et al.* stimulus of tone sequences and noise bursts. Gray arrows indicate exogenous processing which is not encoded into memory due endogenous attention being directed toward other exogenous processing (black arrows).

In this example, the subject is instructed to attend to the noise bursts and perform a perceptual task on them: classifying them as approaching (linear increase in amplitude) or departing (the approaching burst reversed in time). The model predicts that exogenous processing is occurring all the time on all streams. However, only streams subject to endogenous attention are encoded into memory and perceived. Therefore, in this state of diverted attention the build up of streaming, although initiated, does not occur as the tones are not subject to endogenous attention. Later in the experiment, the listener is instructed to switch tasks and concentrate on the alternating tones. In this situation of endogenous attention, the streaming process is encoded into memory and so the build up streaming can occur.

Cocktail Party effect

Imagine you are at a friend’s birthday party. The alcohol has been flowing for some time and level of hubbub is so loud that you are having difficulty following your conversation with the host. In this situation, an exogenous process is responsible for building and maintaining the *host-conversation* stream. In addition to this, there may be other exogenous processes occurring simultaneously - dealing with music, for example. However, the conscious input to the stream selector process continues to encode the conversation stream into memory allowing perception of the speech to continue. Despite the high level of concentration employed to listen to the dialogue, it is still possible that another speaker, on mentioning your name in a different conversation, can unexpectedly redirect your attention to that new conversation. The new conversation is dealt with by a separate exogenous process. The bank of schema recognisers proposed in the model accounts for this unconscious redirection of attention by signalling the endogenous selector that a new source of potentially important information has been detected. Acting upon this, the selector ‘overrides’ the conscious input and selects the new exogenous process.

Allocation of attention to multiple orthogonal bands

Schlauch and Hafter (1991) showed that it is possible to monitor a number of harmonically unrelated frequency bands simultaneously. This led to the conclusion that listeners have the ability to allocate attention to multiple frequency regions. This does not conflict with the model presented here if one makes a simple assumption: an exogenous process can deal with, and form a stream from, a number of frequency bands. It has been proposed that the default condition in the auditory system is fusion (Bregman, 1990), meaning that elements are only segregated when there is evidence to do so. One cue used in grouping is harmonicity: the larger the number of auditory elements consistent with a particular F0, the stronger the evidence to support grouping these elements into their own stream. In the case of Schlauch and Hafter’s (1991) stimuli in which all of the tones were harmonically unrelated, there is no evidence to suggest a particular F0 and therefore no motivation to segregate harmonically *unrelated* tones from harmonically *related* tones. The lack of support for segregation leads to the (default) perception of a single stream. Informal listening suggests that this is the case. It is to this *stream* that endogenous attention is directed and not to the individual frequency bands *per se*. This hypothesis makes the model consistent with evidence from the visual domain in which observers can attend to only one location at a time (e.g. Eriksen and Webb, 1989).

3.4. Summary

The model presented in this chapter provides an explanation of the role of attention in the auditory system. It is based upon the two forms of attention first proposed by William James (1890) and more recently by Jonides and Yantis (1988) and Müller and Rabbitt (1989). In this framework, exogenous processes are responsible for the propagation of individual streams. An endogenous mechanism then selects one of these processes allowing that stream to be encoded into memory and thus perceived. Perception of the other streams does not occur.

The endogenous mechanism consists of a combination of conscious and data-driven decision factors: the decision of the subject; outcome measures from difference detectors operating on each stream; and schema recognition. Data-driven cues can ‘override’ conscious decisions provided they are sufficiently salient.

CHAPTER 4

Issues for Future Research

The next stages of modelling involve producing a number of models of basic auditory grouping. The longer term goal is implement the model described in the previous chapter which will select and 'perceive' the streams produced by exogenous processing.

4.1. Exogenous processing

This processing produces the fundamental data on which the rest of the attention model operates. Although producing models of exogenous grouping is not the primary goal of this research, it forms an essential part. In order to produce representative data, a subset of exogenous tasks will be modelled such as grouping by frequency proximity. Work at this stage will also concentrate on the possible mechanisms for grouping and also means of signalling this grouping to later stages in the model such as oscillations (e.g. Wang, 1996; Brown and Wang, 1999, Wang and Brown, 1999).

4.2. Detection banks

The detection banks are inspired by evidence that a comparison process occurs to identify changes in the stimulus over time (Sussman *et al.*, 1998, 1999). Further research needs to be conducted on identifying the modalities of these comparisons. Initial assumptions are that difference detectors operate on intensity and frequency. Related to this is the use of schema detectors which are constantly searching for information salient to the subject. Given the broad range of possible schemata, it is envisaged that only a small number of schemata will be incorporated to make the model representative.

4.3. Memory

The final stage of the model is the encoding of a selected stream into memory. In order to model a form of short term memory, its structure, longevity and capacity need to be investigated. In addition to this, the method by which information in encoded and retrieved are linked to the method by which streams are encoded by exogenous processes and recalled for perception. It may also be necessary to examine how short term memory interacts with long term memory. It is hoped that part of this research package will be conducted at the EU Advanced Course in Computational Neuroscience in Trieste.

4.4. Selector mechanism

The means of selecting a single stream to be encoded into memory relies on the combination of three types of input: exogenous stream input; detector bank outcomes; and conscious input. It will be necessary to investigate the nature of exogenous overrides of conscious decisions: are all exogenous overrides successful or do they have to be above a salience threshold? Although conscious decision making cannot be modelled, it a means of choosing a particular stream will be modelled. It is likely that this will be achieved by external input to the model.

4.5. Model integration and simulation

The final stage of research will be combining all the above work packages to produce a fully operative model of auditory attention. It is envisaged that model integration and simulation will experience a large amount of overlap since simulation will act as a test of integration success. It is hoped that simulation experiments will consist of complex tasks such as the Carlyon *et al.* (1999) stimuli as well as simpler tasks. In the time plan, a second period of exogenous processing research is shown. Provided there is sufficient time, further models of grouping and streaming phenomena will be added to the model to enable it to handle a broader range of stimuli.

CHAPTER 5

References

Alain, C and Woods, DL (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology* **34** 534-546.

Anstis, S and Saida, S (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception Performance* **11** 257-271.

Bregman, AS (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*, MIT Press.

Bregman, AS and Campbell, J (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology* **89** 244-249.

Bregman, AS and Steiger, H (1980). Auditory streaming and vertical localisation: Interdependence of “what” and “where” decisions in audition. *Perception & Psychophysics* **28** 539-546.

Brown, GJ (1992). *Computational auditory scene analysis: A representational approach*, Ph.D. thesis CS-92-22, CS dept., Univ. of Sheffield.

Brown, GJ and Wang, DL (1999). Timing is of the essence: Neural oscillator models of auditory grouping. In *Listening to Speech*, edited by S.Greenberg and W.Ainsworth, Oxford University Press, in press.

Carlyon, RP, Cusack, R, Foxton, JM and Robertson, IH (1999). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance* (submitted).

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

Cooke, MP (1993). *Modelling auditory processing and organisation*. Cambridge University Press.

Crick, F (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences of USA* 81 4586-4590.

Denbigh, PN and Zhao, J (1992). Pitch extraction and separation of overlapping speech. *Speech Communication* **11**(2-3) 119-125.

Deutsch, D (1974). An auditory illusion. *Journal of the Acoustical Society of America* **55** 518-519.

Deutsch, D and Roll, P (1976). Separate “what” and “where” decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance* **2** 23-29.

Duncan, J (1984). Selective attention and the organisation of visual information. *Journal of Experimental Psychology: General* **113** 501-517.

Duncan, J, Martens, S and Ward, R (1997). Restricted attentional capacity within but not between sensory modalities. *Nature* **387** 808-810.

Ellis DPW (1996). *Prediction-driven computational auditory scene analysis*, Ph.D. thesis, MIT Department of Electrical Engineering and Computer Science.

Eriksen, CW and Webb, JM (1989). Shifting of attentional focus within and about a visual display. *Perception & Psychophysics* **45** 175-183.

Greenberg, GZ and Larkin, WD (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *Journal of the Acoustical Society of America* **44** 1513-1523.

Green, DM (1960). Psychoacoustics and detection theory, *Journal of the Acoustical Society of America* **32** 1189-1203.

Green, DM (1961). Detection of auditory sinusoids of uncertain frequency, *Journal of the Acoustical Society of America* **33** 904-911.

Green, DM and Swets, JA (1966). *Signal Detection Theory and Psychophysics*, Wiley.

Hafter, ER, Schlauch, RS and Tang, J (1993). Attending to auditory filters that were not stimulated directly. *Journal of the Acoustical Society of America* **94** 743-747.

James, W (1890/1950). *The Principles of Psychology*, Volume 1. Dover.

Johnson, DM and Hafter, ER (1980). Uncertain-frequency detection: Cuing and condition of observation. *Perceptual Psychophysics* **28** 143-149.

Jonides, J and Yantis, S (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics* **43** 346-354.

Koffka, K (1936). *Principles of Gestalt psychology.* Harcourt and Brace, New York.

Mack, A, Tang, B, Tuma, Regina and Kahn, S (1992). Perceptual Organisation and Attention. *Cognitive Psychology* **24** 475-501.

Melara, RD and Marks, LE (1990). Perceptual primacy of dimensions: Support for a model of dimension interaction. *Journal of Experimental Psychology: Human Perception and Performance* **16** 398-414.

Mondor, TA and Bregman, AS (1994). Allocating attention to frequency regions. *Perception & Psychophysics* **56**(3) 268-276.

Mondor, TA and Zatorre, RJ (1995). Shifting and Focusing Auditory Spatial Attention. *Journal of Experimental Psychology: Human Perception and Performance* **21**(2) 387-409.

Mondor, TA, Zatorre, RJ and Terrio, NA (1998). Constraints on the Selection of Auditory Information. *Journal of Experimental Psychology: Human Perception and Performance* **24**(1) 66-79.

Moore, BCJ and Glasberg, BR (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* **74** 750-753.

Moore, CM and Egeth, H (1997). Perception Without Attention: Evidence of Grouping Under Conditions of Inattention. *Journal of Experimental Psychology: Human Perception and Performance* **23** 339-352.

Müller, HJ and Rabbitt, PMA (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance* **15** 315-330.

Patterson, RD and Moore, BCJ (1986). Auditory filters and excitation patterns as representations of frequency resolution, in *Frequency Selectivity in Hearing* (ed. BCJ Moore). Academic Press, 123-177.

Rhodes, G (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics* **42** 1-14.

Schlauch, RS and Hafter, ER (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *Journal of the Acoustical Society of America* **90** 1332-1339.

Spence, CJ and Driver, J (1994). Covert Spatial Orienting in Audition: Exogenous and Endogenous Mechanisms. *Journal of Experimental Psychology: Human Perception and Performance* **20**(3) 555-574.

Sussman, E, Ritter, W and Vaughan Jr, HG (1998). Attention affects the organisation of auditory input associated with the mismatch negativity system. *Brain Research* **789** 130-138.

Sussman, E, Ritter, W and Vaughan Jr, HG (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* **36** 22-34.

Trejo, LJ, Ryan-Jones, DL and Kramer, AF (1995). Attentional modulation of the mismatch negativity elicited by frequency differences between binaurally presented tone bursts. *Psychophysiology* **32** 319-328.

Treisman, A and Gelade, G (1980). A feature integration theory of attention. *Cognitive Psychology* **12** 97-136.

Treisman, A and Gormican, S (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review* **95** 15-48.

van Noorden, LPAS (1975). *Temporal coherence in the perception of tone sequences*. Doctoral thesis, Institute for Perceptual Research, Eindhoven, NL.

Wang, DL (1996). Primitive auditory segregation based on oscillatory correlation. *Cognitive Science* **20** 409-456.

Wang, DL and Brown, GJ (1999). Separation of speech from interfering sounds based on oscillatory correlation. *IEEE Transactions on Neural Networks* **10** 684-697.

Zatorre, RJ, Mondor, TA and Evans, AC (1999). Auditory Attention to Space and Frequency Activates Similar Cerebral Systems. *NeuroImage* **10** 544-554.