The Effects of Stimulus Similarity on Auditory Search: An Examination of Target-Distractor, Distractor-Distractor, and Target-Target Similarity

By

Chris E. Tysiaczny

University of Manitoba

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts
in the
Department of Psychology

UNIVERSITY OF MANITOBA

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08982-2

Your file Votre référence
ISBN:
Our file Notre reterence
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

The Effects of Stimulus Similarity on Auditory Search: An Examination of Target-Distractor, Distractor-Distractor, and Target-Target Similarity

BY

Chris E. Tysiaczny

A Thesis/Practicum submitted to the Faculty of Graduate Studies of the University of

Manitoba in partial fulfillment of the requirement of the degree

of

Master of Arts

Abstract

In auditory search, listeners are required to detect the presence of a target sound embedded in a sequence of distractor sounds. Two experiments were conducted to determine whether, (a) the similarity between the target and the distractor sounds, and between the distractor sounds and one another, affects a listener's ability to detect the target, and (b) when listeners are required to detect either of two targets, performance is affected by the degree of similarity between them. The results indicated that target-distractor, distractor-distractor, and target-target similarity all influence target detection. Listeners were most accurate at detecting a target when target-distractor similarity was low, and distractor-distractor and target-target similarity were high. Accuracy decreased as target-distractor similarity was increased, and as distractor-distractor and target-target similarity were decreased. A general theory of auditory search, derived primarily from attentional engagement theory (Duncan & Humphreys, 1989), is proposed and future implications of this research are discussed.

Acknowledgements

First, I would like to thank my advisor, Dr. Todd Mondor, for his support and mentorship throughout this entire process. His wisdom, constructive criticism, and guidance have instilled in me the ability to achieve my potential as a researcher. Second, I would like to thank my committee members, Dr. Jason Leboe and Dr. Pourang Irani, for the helpful suggestions they provided on earlier drafts, all of which contributed to successful revisions of this thesis. Finally, I would like to express my appreciation to the Natural Sciences and Engineering Research Council of Canada and the University of Manitoba Graduate Fellowship program for providing the funding necessary to undertake this project.

Dedication

I would like to dedicate this thesis to my father, Dr. Mark Tysiaczny. Thank you, for your inspiration, encouragement, and support.

Table of Contents

List of Tables

List of Figures

Introduction

When we scan our environment with the goal of locating specific objects, whether they are visual or auditory in nature, we are engaging in a process called *search.* While it is quite clear that our internal mechanisms are highly developed to accomplish this task, it is rather unclear as to precisely how such a deceptively simple, yet highly complex, undertaking is achieved. Our environment is filled with distracting sights and sounds, each with the ability to attract our attention and lead us astray from our target. Yet, we are consistently able to overlook this bombardment of objects resembling our target, and to locate the proverbial 'needle in a haystack' that is our target (our keys on the disorganized kitchen counter, for instance).

A Review of Current Search Theories

Feature Integration Theory

As one might imagine, most of the research that has been accomplished in this area is directly related to our most commonly investigated sense: vision. One of the pioneers in the field of *visual search*, Anne Treisman, developed an intriguing theory to explain the common finding that targets comprised of a combination (or conjunction) of the features that make up surrounding non-targets (or distractors) are often identified more slowly and less accurately than targets defined by a single feature (Treisman & Gelade, 1980). Feature integration theory (FIT), as it has come to be known, addresses the dispute between Associationists, who believe that the perceptual experience of whole objects is preceded by the combination of less complex parts, and the Gestaltists, who argue that we initially perceive whole objects, after which we may decompose these objects into their component parts (Treisman & Gelade, 1980). According to the theory, the first of two successive stages in visual search involves the parallel processing of each feature of an object in our visual field in separate spatiotopically organized *maps*. These maps are representations of the visual field, each sensitive to the presence of a particular elementary feature of the stimulus. For example, one map may code where the colour green occurs whereas another map may code where a circular shape occurs. While this process is occurring in parallel across the visual field, it provides no useful information about the conjunction of these separable features. For example, we may be sensitive to the fact that the visual field contains yellowness, greenness, a circle shape, and a square shape, but we are not aware that there is a yellow circle and a green square. It is only after a second stage, when focal attention is directed to a specific location, that an object's elementary features may be combined to yield what we perceive as a unified visual event (Treisman & Gelade, 1980). That is, when attention is focused on a particular area of the visual field, the information in this position held by the various feature maps is conjoined to yield what we perceive as a whole object (e.g., a yellow circle).

The most compelling evidence for this contention comes from instances in which the features of objects have been incorrectly combined (i.e., if we were to recall a green circle or a yellow square in the previous example). Such 'illusory conjunctions' are said to occur when insufficient time is available for attention to bind individual features together to form an object representation. In this situation, these features are considered to be *free floating*, and when the object is recombined after the stimuli are no longer present, the resulting perception of the object is an illusory conjunction (or incorrect combination) of the object's features (Treisman & Gelade, 1980). These illusory conjunctions are inferred by examining the number of feature errors as compared to the number of conjunction errors. This phenomenon supports feature integration theory in that it is believed to be the application of focused attention that allows us to combine features of objects into a perceptual representation of the actual event. Thus, if focused attention is prevented, that illusory conjunctions occur provides evidence in support of FIT (Treisman & Gelade, 1980; Treisman & Schmidt, 1982; Treisman & Souther, 1985).

Treisman and Gelade (1980) proposed that separable features may be detected without focal attention and thus, a target that is defined by such a feature (e.g., green or horizontal) should be detected quickly, regardless of the number of distractor items that are presented concurrently. According to Treisman and Gelade, separable features should be detectable by parallel search (i.e., an examination of multiple stimuli simultaneously). They claim that separable features are susceptible to illusory conjunction errors when attention is absent and can be identified without necessarily being located (i.e., the presence of critical features may be detected before or even without knowing their precise location). They also predicted that separable features will easily lead to texture segregation and feature-grouping (the detection of spatial discontinuities between groups of stimuli). In addition, Treisman and Gelade predicted that since unattended stimuli should only be registered at the feature level, only separable features should have behavioural effects (either interference or facilitation) because these effects depend only on the features that make up the stimuli, not the conjunctions or combinations of these features. As for targets that are defined by a conjunction of features (e.g., a horizontal green line in a background of vertical green and horizontal red lines), focal attention is required to conjoin the features. This process is said to be serial, whereby in order to have correctly identified a particular conjunction it is first necessary to have located it before focused attention can permit the integration of its features. Thus, attending to the location of a stimulus must precede identification of conjunctions. Treisman and Gelade proposed that in conjunction search, texture segregation should prove quite ineffective, and search time should increase as the number of distractors increases.

In their classic experiment, Treisman, Sykes, and Gelade (1977) compared search for a target defined by a single feature (e.g., *pink* in *brown* and *purple* distractors and the letter *O* in *N* and *T* distractors) with targets defined by a conjunction of features (e.g., a *pink O* in *green O* and *pink N* distractors). They found display size (i.e., the number of distractors in a visual array) to be irrelevant or to have very little effect on search time in the single feature target conditions. However, search times increased linearly with display size when the target was a conjunction of features. These results were taken as support for the claim that visual search is parallel for separable features, yet becomes serial for stimuli involving conjunctions of separable features.

Treisman and Gelade (1980) sought to evaluate FIT further. They tested their predictions using two separable dimensions; shape and colour. Experiment 1 was designed to test whether the detection of a conjunction of shape and colour (e.g., a *green T*) could be changed from serial to parallel with increased practice, thus leading to the formation of a *special detector* for the newly created unitary target, which would no longer require focal attention to be identified. Each session consisted of a feature search condition and a conjunction search condition and display sizes varied from 1 to 30 items in each condition. Treisman and Gelade reported that even after 13 practice sessions (each consisting of 128 trials in both conjunction and feature conditions) with the same stimuli, the increased practice did not lead to the formation of a unitary target. The reason for such a conclusion comes from the examination of the search slopes. The search slope plots the degree of change in reaction time as a function of the display size (reaction time is plotted on the y-axis and display size is plotted on the x-axis). If the reaction time to detect a target increases as the display size (number of distractors) increases (positive slope), then this provides evidence for a serial search. If the reaction time remains fairly constant as the display size increases (a flat or horizontal slope), then this is evidence for parallel search. Slopes remained positive for the conjunction condition throughout the 13 practice sessions. However, due to the time constraints of such a long series of experiments, only two of the six participants continued past seven practice sessions thus adding a degree of uncertainty (in terms of reliability) to their findings. In addition, Treisman and Gelade reported that the possibility exists for the formation of a unitary target if the number of practice sessions is increased beyond 13, if different targets are utilized, or if a different training method is incorporated.

In Experiment 2, Treisman and Gelade reasoned that the speed of detecting a conjunction target in a display might be affected by altering the degree of similarity between the shapes and colours of the two distractors items, which combine to form the conjunction target (perhaps even leading to a parallel search). They tested two conditions, similar distractors (e.g., search for a *green T* target in *green X* and *blue T* distractors) and distinct distractors (e.g., search for a *red O* target in *green O* and *red N* distractors). Despite the fact that the target was found almost three-times as quickly in more distinct distractors than in more similar distractors, the researchers reported that search remained serial (i.e., as display size increased so did search time) for similar and distinct distractor conditions. Thus, it was concluded that search does not simply become serial as the task becomes more difficult because positive slopes were reported in both conditions. The question that this experiment raises, however, is the extent to which the similarity between the target and the distractors plays a role in the search process. That is, target-distractor similarity was greater in the similar distractor condition than in the distinct distractor condition (i.e., whereas the *green T* target is similar in colour to both of the distractors, the *red O* target is similar in colour to only one distractor). In addition, the similarity between the distractors and each other was (intentionally) greater in the similar distractor condition, and this may also have had a significant affect on performance. Both of these factors (i.e., target-distractor and distractor-distractor similarity) will be addressed later in the paper.

Experiment 3 was designed to test the possibility that the conjunction condition is associated with slower detection times because the conjunction target shares one feature with each of the other distractors in the display, while the feature target shares one feature with only half of the distractors. That is, the feature targets might be a *green S* or a *blue T* in *green X* and *brown T* distractors, while the conjunction target might be a *green T* in the same set of distractors. To examine this possibility, Treisman and Gelade used ellipses of five different sizes. When viewed next to each other (in size order), the ellipse on the far left was the smallest (*ellipse one*) and the ellipse on the far right was the largest (*ellipse five*). The distractors in this experiment were *ellipse two* and *ellipse four*. *Ellipse three* (the 'medium' ellipse) was one potential target and was considered to be equivalent to the conjunction target in that this ellipse was 'similar' to each of the distractors. *Ellipse one* and *ellipse five* ('small' and 'large' ellipses respectively) were two other potential targets that were similar to only one distractor (supposedly comparable to the feature targets in previous experiments). The authors reported that search was parallel for the 'medium' ellipse target (the supposed conjunction equivalent ellipse) not serial as was the case for the conjunction target in the previous experiment. According to the researchers, this experiment established that the conjunction effect (i.e. serial processing) did not simply happen because the target was similar to both distractors. However, the conclusion that was drawn from this experiment seems a little premature. First, as display size increased in this experiment so did search time (for all three target ellipses), at a rate that is usually indicative of serial search. Also, the slopes for the 'medium' ellipse target fell between the slopes of the other two targets, which is not a problem for the research if indeed flat slopes were found, but by examining Figure 2 on p. 112 of their paper, they appear to be positive. To accurately demonstrate the point, the 'meduim' ellipse would have had the greatest slope, but in fact, the 'large' ellipse appears to have a greatest slope. The actual slopes, however, were never reported. It is only mentioned that deviations from linearity (which is not defined) are found to be significant. Second, the logic that was used (i.e., that each ellipse was 'similar' to one or two others in terms of their size) is suspect in that in previous experiments, each stimulus shared one attribute with at least one other stimulus (e.g., the colour *green* or the letter *T*). In this experiment, the 'shared' attributes (i.e., ellipse size) are never actually shared; they are only similar (i.e., no two ellipses are the same size). A comparable experiment would be to use a *green T* five-times and vary the shade of green in terms of its darkness. The middle (third) *green T* would supposedly be the conjunction because it 'shares' an attribute (i.e., it closely resembles it in colour) with the second and fourth *Ts*, but the first and fifth *Ts* supposedly only 'share' an attribute with one other stimulus. While these predictions may be plausible, an experiment such as this has not been reported, and parallel search has never been found for the conjunction. Thus, Treisman and Gelade's conclusions for Experiment 3 appear to be unwarranted.

Experiment 4 was designed to test the possibility that letters (because they are very familiar stimuli) are simply integral perceptual stimuli that can be distinguished by *unitary detectors*. Treisman and Gelade however, believed that the features of letters must be combined by focused attention to yield the whole stimulus. The researchers reported that search should be serial and self-terminating if a target letter can be formed by conjoining elements from both of the distractors because of the possibility of conjunction errors. However, if only one of the letters contains critical features of the target, then we should see parallel search. For example, the target letter *R* was used in one of the conditions with either *P* and Q or *P* and *B* distractors (*R/P*Q or *R/PB*). Since the slash that intersects the Q could be combined with the *P* to form an *R*, this was considered to be the conjunction condition. In contrast, no feature of the *B* can be added to the *P* to yield an *R*, so the *R/PB* condition was considered to be the similarity condition (another condition was tested using *T/IZ* and *T/IY* stimuli). Treisman and Gelade reported that since the slopes for the reaction time increased linearly with display size for the conjunction conditions (*R/P*Q and *T/IZ*), search was serial. Search was classified as parallel for the similarity condition (*R/PB* and *T/IY*) due to "…a flat function or a nonlinearly increasing function…" (p. 118). However, the slopes were never flat as display size increased, but were in fact 5.3 and 9.7 ms/item (compared to 12.2 and 27.2 ms/item for the conjunction condition). According to previous logic, this actually indicates serial search in both conditions of the experiment. In addition, in the feature conditions, there was always a feature of the target that subjects could search for which was distinct from the target and likely gave away its presence (i.e., the diagonal slash of the *R* or the horizontal line that forms the top of the *T*). No such feature was present in the conjunction condition, thus feature search could simply have involved search for the distinct target feature, while conjunction search may simply have required a more extensive examination of the stimuli. This point will be addressed in the review of attentional engagement theory (Duncan & Humphreys, 1989), however at this point it can be noted that performance may have been related to task difficulty, with the conjunction conditions perhaps being more difficult than the feature conditions.

In Experiment 5, group and texture segregation were tested using index cards that were designed in such a manner to create an 'imaginary line' or boundary (either horizontal or vertical) that divided a 5 x 5 square matrix of coloured letters so that stimuli on either side of the boundary were defined by either a distinct colour, shape, or a conjunction of both. It was reported that subjects were faster at sorting a deck of cards into two piles (one containing cards with a horizontal boundary and the other containing cards with a vertical boundary) when stimuli on either side of the boundary differed in terms of separable features, but not when they differed in conjunctions of features. The same stimuli were used for both the feature and conjunction conditions (*red O*, *blue O*, *red V*, *blue V*). In the feature condition, the boundary divided *red* stimuli from *blue* stimuli, or *Vs* from *Os*. In the conjunction condition, the *red Os* and *blue Vs* were separated from the *blue Os* and *red Vs*. The mean sorting time for both of the feature decks of cards was significantly faster than for the conjunction deck. Treisman and

Gelade claimed that this advantage for feature decks demonstrated that since texture segregation is dependent on early parallel detection of similarities, easier segregation is found in feature conditions (when groups are separated by one or more separable features) and it is not easily found in conjunction conditions (when groups differ only in the conjunction of features). Thus, they suggested that the boundary is not directly noticed during conjunction search, but instead must be found by scanning several individual stimuli serially.

There are, however, alternative explanations for the findings of Experiment 5. According to Treisman and Gelade (1980), the results could be explained, at least in part, by the fact that the feature deck had only one dimension that was relevant (i.e., shape or colour) and were homogenous on that dimension on each side of the boundary. In contrast, in the conjunction deck, participants had to pay attention to both dimensions. Thus, Experiment 6 was designed to determine if the addition of a disjunctive pack to the task would clarify the findings. Treisman and Gelade created a disjunctive deck of cards consisting of *red Os* and *green Hs* on one side of the boundary, and *blue Os* and *green Vs* on the other. Thus, in the disjunctive deck, no letter from one side of the boundary could be formed by combining the colour and form of two letters on the other side of the boundary (as they could in the conjunction deck), but each side of the boundary was relevant on both dimensions (which was not the case in the feature deck). Participants were significantly faster at sorting the disjunctive deck than the conjunction deck used in Experiment 5, but were significantly slower at sorting the disjunctive deck than the feature deck (but by only a small margin). Treisman and Gelade claimed that this result demonstrated that the greater homogeneity of stimuli and the necessity to rely on two dimensions in the conjunction deck only explained a small proportion of the difference between feature and conjunction search.

In Experiment 7, subjects were shown 5 x 5 matrices of all black, similar letters, in either a single feature condition with letters containing short diagonal lines on one side of the boundary and not on the other (*QR/PO* or *FQ/EO*) or a conjunction condition with no simple features to distinguish letters on one side of the boundary from those on the other (*PQ/OR* or *FK/EX*). Subjects were asked to indicate whether the imaginary boundary created by dividing similar letters (*PR*, *EF*, *OQ*, and *XK*) was horizontal or vertical. In general, significantly faster performance was found on the feature sets than on the conjunction sets. These results led the researchers to conclude that texture segregation is dependent on whether the boundary divided areas that differed in a single feature or only on a conjunction of features (Treisman & Gelade, 1980).

As previously mentioned, feature integration theory predicts that for a subject to correctly identify a conjunction, its features must have been bound through focal attention to a specific location, whereas the identification of a feature does not necessarily depend on the allocation of attention to a specific location. In Experiment 8, the subjects were required to identify one of two possible targets and its location in a 2 x 8 matrix of letters (one of the two targets was present on each trial). The presentation time (i.e., the time that the subjects needed to examine the stimuli to identify the targets accurately) for both the feature and conjunction conditions was determined in advance by a preliminary testing session and varied from subject to subject. Much shorter presentation times were needed when identifying the location of a disjunctive feature target (i.e., a *pink* or *blue H* or an *orange X* or *O* in *pink O* and *blue X* distractors) than when identifying the location of a conjunction target (i.e., a *pink X* or a *blue O* in the same distractors) while maintaining an 80% accuracy rate (65 ms vs. 414 ms respectively). However, correct conjunction targets were identified more accurately than correct feature targets when the location response was also correct, but less accurately when an adjacent location (one horizontal or vertical position from the correct position), or a distant location (all other positions) was reported. In the conjunction condition, the conditional probability of correctly identifying a target in a correct, adjacent, or distant location was .930, .723, and .500, respectively (overall accuracy was .793). In the feature condition, the conditional probability of correctly identifying a target in a correct, adjacent, or distant location was .897, .821, and .678, respectively (overall accuracy was .786). This was taken as support for the notion that conjunction targets must be located before they can be identified, whereas feature targets can be identified without necessarily being located (Treisman & Gelade, 1980).

In Experiment 9, the same conditions as the previous experiment were tested, but equal presentation times for the feature and conjunction search were used. Subjects performed two sessions, each consisting of three blocks of 32 trials on a feature condition and three blocks of trials on a conjunction condition. In the first session, half of the subjects performed three blocks of the feature condition then three blocks of the conjunction condition, while the other half performed three blocks of the conjunction condition then three blocks of the feature condition. Presentation times began at 150 ms for the first block. Presentation time on the second block was contingent on performance on the first block (it was either increased or decreased depending on accuracy), and presentation time for the third block was contingent on performance in the second block.

The presentation times for the next three blocks were matched to the presentation times of the previous three blocks. Presentation times on individual blocks varied from 40 to 200 ms. A second session was run in the exact same manner except the condition order (feature or conjunction search) of the first three blocks was reversed (i.e., if feature search were performed first in the first session then conjunction search was performed first in the second session and vice versa). According to Treisman and Gelade (1980), the reason for this procedure was to ensure that the longer exposure time in the conjunction condition of Experiment 8 did not affect performance "…in some qualitative way." (p. 128). The subjects in Experiment 9 were far more accurate in detecting targets defined by disjunctive features than by conjunctions of features. They were also more likely to report the correct location of the target, whereas in Experiment 8 they were more accurate (overall) in the conjunction condition than in the feature condition (keeping in mind that much more time to respond was allowed in the conjunction condition of Experiment 8). In Experiment 9, subjects were more likely to identify the correct target in the feature condition, regardless of whether or not the location was accurate. Conditional probabilities in the conjunction condition for a correct target in a correct, adjacent, or distant location were .840, .582, and .453 respectively (overall accuracy was .587). In the feature condition, the conditional probabilities of identifying a target in a correct, adjacent, or distant location were .979, .925, and .748 respectively (overall accuracy was .916). Subjects performed more accurately in choosing the correct target in feature conditions when an adjacent or even distant location was reported. In conjunction conditions, however, accuracy was significantly lower if an adjacent location or a distant location for the correct target was reported (usually at chance levels). These results were taken as support for the notion that identification must be preceded by localization for conjunctions, but identification can occur without localization for features. However, in the conjunction condition of Experiment 8, the targets were correctly identified in an adjacent location at a greater than chance level (.723). Treisman and Gelade reported that this result suggests that only an approximate perception of the location appears to be needed before a conjunction can be identified, however, that result was not replicated in Experiment 9. In addition, it appears that the method for equating exposure times for feature and conjunction conditions is suspect. When subjects performed the three blocks of feature search first, they set an unfair standard for conjunction search in that it is simply more difficult to detect conjunctions (because of their similarity to both of the distractors) and thus requires more time, so performance was sure to be poor on the subsequent conjunction conditions. This point is evident in the fact that the correct target was identified in the correct location far less accurately in the conjunction condition than in the disjunctive feature condition (accuracy in these conditions should be roughly equal otherwise it is obvious that conjunction search was made to be more difficult by decreasing presentation times). On the trials where conjunction condition performance set the standard for feature search exposure times, subjects started at 150 ms exposure time (probably too short) then it was likely increased to a maximum of 200 ms (still quite difficult for conjunction search). The subsequent feature trials then likely received the benefit of the difficult conjunction search trials in that exposure time for these trials would have been much more than was necessary for identification (only 65 ms presentation times were used in Experiment 8), causing the results to be inflated in favour of feature search. Though this prediction cannot be proven without the actual data, Treisman and Gelade failed to report whether there were any differences between the presentation times established by conjunction search trials versus feature search trials (only the mean presentation time of 117 ms was reported).

In a later article (Treisman & Gormican, 1988) FIT was modified slightly to explain the finding that feature search is sometimes serial when target-distractor similarity is high. Given a situation in which the targets and distractors are only slightly different in their colour, the distractors might excite the target map, making the judgement of the presence of a target somewhat difficult (i.e., it becomes difficult to determine if the net activity of the target map is sufficient to indicate the presence of the target accurately). They suggested that as the number of distractors (i.e., display size) is increased, the activity produced by the target (in proportion to the total activity produced by all stimuli) is reduced, and thus decision difficulty increases. Treisman and Gormican proposed that clumps of stimuli are searched serially by means of focused attention, and the size of the clumps is directly proportional to the similarity between the target and the distractors (i.e., clumps are chosen such that the net activity produced by the target in the map is sufficient to consistently detect the presence of a target). Thus, the more distinct the target is from the distractors, the larger the clump that can be searched. In the previous version of the theory the entire display was treated as a single, large clump.

Taken together, the findings of Treisman and Gelade (1980) and Treisman and Gormican (1988) provide some evidence for feature integration theory to explain visual search. There are, however, alternative explanations for their results.

Attentional Engagement Theory

Duncan and Humphreys (1989) presented a theory that could account for conjunction effects and even demonstrate how feature and conjunction search could be equally difficult. They were able to create conditions in which search for both feature and conjunction targets could fall into the range suggesting either serial or parallel search, a finding which is, of course, inconsistent with FIT. They began by examining the application of feature integration theory to letter search (i.e., the conjunction of different line segments to form letters). Duncan and Humphreys made a distinction between within-object conjunctions (e.g., two letters, such as L and T, with the same features differing slightly in their spatial arrangement) and across-object conjunctions (e.g., when a target can be formed by conjoining the features from several distractors, such as search for an *R* among *P*s and Qs). They noted several problems feature integration theory fails to explain. While some studies (such as Treisman & Gelade, 1980) do report large effects of display size, other studies that have investigated within-object conjunctions (see Humphreys, Riddoch & Quinlan, 1985), have found that as display size is increased, there appears to be little effect on search time. These results appear to contradict predictions drawn from feature integration theory, despite their resemblance in terms of the task and methodology (i.e., search for an *inverted T* among *upright Ts*) to several other studies that support FIT (see Beck & Ambler, 1973; Bergen & Julesz, 1983). It was, however, demonstrated that the variable of interest in such studies appears to be the ratio of letter size to retinal eccentricity (Humphreys, Quinlan & Riddoch, 1989). That is, the use of relatively small letters produced effects of display size that were not noted with the use of larger letters. This is due to the fact that decreasing the size/eccentricity ratio makes it more difficult to notice differences between the letters. This difficulty in detecting differences evidently becomes a problem for conjunction targets more so than for feature targets because conjunction targets are typically more similar to the distractors. Thus, according to Duncan and Humphreys (1989), the findings of previous experiments that supported feature integration theory may have been (at least in part) due to their use of relatively small letters. As for across-object conjunctions (e.g., Treisman & Gelade, 1980), Duncan and Humphreys note that the smallest effects of display size on search time were reported when distractors were homogenous and the largest effects were noted with heterogeneous targets. Thus their theory, the attentional engagement theory, essentially takes into account the interaction effects of two stimulus variables: the similarity between the target and the distractors and the similarity between the distractors and each other (Duncan & Humphreys, 1992).

In their first two experiments, Duncan and Humphreys (1989) demonstrated that while slopes may vary somewhat between feature and conjunction search, the overall pattern of results is the same. That is, slopes approach zero when the ratio of letter size to retinal eccentricity is large (1/3), whereas slopes increase as the ratio decreases (1/6 or 1/12). They tested subjects using a replication of Beck and Ambler's (1973) investigation in which *Ls* and tilted *Ts* were the targets that were to be detected amongst *upright Ts* or *Ts* that had been rotated 90° clockwise. In Experiment 1, distractors were either homogenous (either all *upright Ts* or all *rotated Ts*) or heterogeneous (approximately half *upright Ts* and half *rotated Ts*). Size/eccentricity ratios were either 1/3 or 1/12. Stimuli were presented for 180 ms. In Experiment 2, distractors were always homogenous (*upright Ts*) and letter size/eccentricity was either 1/3 or 1/6. Stimuli were presented for as long as the subjects took to respond. They found in both experiments, and for both feature (*tilted T*) and conjunction (*L*) targets, that slopes were always in the range to suggest parallel search (as defined by Treisman & Souther, 1985) when letter size/eccentricity was large, whereas the use of small often leads to a search times generally interpreted as serial search. Thus, small letters tend to produce a greater effect of display size (i.e., had greater slopes), whereas search for larger letters was almost independent of display size, despite the fact that targets and distractors consistently differed only in their conjunction of letter strokes. These results demonstrate that the failure to account for letter size/eccentricity in the previous experiments that support FIT is a serious concern for the theory.

Duncan and Humphreys (1989) anticipated the interpretation that proponents of feature integration theory would give to their findings. Specifically, it is possible that, since an L contains attributes not possessed by a T (such as the corner of the L), one could argue that some elementary feature (not necessarily the strokes that make up the letter) was detected in the first, parallel stage of processing, which may have resulted in slopes in the range to suggest parallel search. To address this issue, Experiment 3 used distractors that were identical to the target except for their orientation. The researchers used a target *L* and compared the detection of this target in two conditions. In one condition (homogeneous distractors) all of the distractors were the letter *L* rotated 90° clockwise, and in the other condition (heterogeneous distractors) some of the distractors were the letter *L* rotated 90°, while others were the letter *L* rotated 270° clockwise. Not only was it found that search time for a target among homogenous rotations of the target letter can be in the range that indicates parallel search, but the task became extremely difficult as heterogeneous distractors were included in the display as well. Thus, search was demonstrated to be more difficult as distractor-distractor similarity is decreased.

Experiment 4 built on the findings of the previous experiments. Increasing the display size affected search for a target *L* among rotations of distractor *Ts* when small letters were used, or as target-distractor similarity was increased and as distractor-distractor similarity was decreased, but this was not the case for large letters. In a preliminary experiment, Duncan and Humphreys found that keypress identification of a single letter (*L* or *T*) was significantly faster when the *T* was rotated either 0° or 90° clockwise and was made more difficult when the *T* was rotated either 180° or 270° clockwise. That is, identification was easier when the *T* differed from the *L* by a greater number of strokes (i.e., 180° and 270° rotations are identical to the *L* aside from a half stroke left or down; 0° and 90° rotations differ from *L* on this same stroke plus an additional stroke). Thus, 180° and 270° rotations of *T* are more similar to *L* than are 0° and 90° rotations. For Experiment 4, the more similar (more difficult) pair of *Ts* was used. Search for an *L* among the more similar *Ts* proved to be more difficult (as indicated by significantly greater slopes) as size/eccentricity was decreased (1/3 versus 1/6). In addition, significantly greater slopes were noted (for both size/eccentricity conditions) as distractor-distractor similarity was decreased. However, if the distractors used are more similar to the target as in Experiment 4 (180° and 270° rotations of *T*) rather than less similar as in Experiment 1 (0° and 90° rotations of *T*), the effects of distractor heterogeneity are greater (i.e., slopes in 1/3 size/eccentricity condition of Experiment 4 are greater than the slopes in the same size/eccentricity condition of Experiment 1). Further, as demonstrated in Experiment 3, there is an extremely large effect of distractor heterogeneity when distractors are simply rotations of the target. Together, these results demonstrate not only that target-distractor and distractor-distractor similarity are important in search, but also that these factors interact.

In contrast with feature integration theory, the attentional engagement theory is not concerned with the distinction between serial and parallel search per se. Rather search efficiency is the variable of interest, and this efficiency is measured on a continuum. The experiments in Duncan and Humphreys' (1989) paper demonstrate that search efficiency can be reduced by increasing target-distractor similarity, or by decreasing distractor-distractor similarity, but the most important changes are noted when there is an interaction between both of these factors.

According to Duncan and Humphreys (1989), attentional engagement theory is composed of three parts (or stages). The first stage involves *perceptual description*, whereby the input from the entire visual scene is processed in a "parallel, hierarchically structured representation" (p. 445). A part of this representation can subsequently be accessed to influence current behaviour. Segmentation occurs in that stimuli are naturally grouped by properties or features such as proximity, colour, shape, or orientation, and boundaries are created between stimuli that are dissimilar in such features. It is then by the repetition of segmentation at different levels of scale that a fully hierarchical representation is formed. These segments, called *structural units*, are further subdivided and described by their own properties, as a tree can be divided into a trunk, branches, and leaves, and a hand can be divided into a palm and fingers (Marr & Nishihara, 1978). The top of the hierarchy is a representation of the whole scene, segmented on the basis of a rough set of features and each structural unit is described with its own set of features.

This entire process occurs in parallel, is resource free, and is outside of awareness but it has no direct impact on behaviour (Duncan & Humphreys, 1989).

According to Duncan and Humphreys (1989), for a stimulus (a whole structural unit) to lead to an action, it must first be selected for access to visual short-term memory (VSTM) for which space is quite limited. Thus, it is proposed that some limited-capacity processor is involved, within which there is a finite amount of resources that may be assigned to any structural unit received as an input. A resource is defined as "…any factor competitively distributed among structural units" (p. 446), and may potentially be the degree of activation that is produced in the brain. The more resources that are allocated to one structural unit, the faster and more likely it is that it will gain access to visual short-term memory, and the less likely it is that another structural unit will gain access. This is because each structural unit is assigned a weight (possibly based on excitatory and inhibitory inputs) depending on its similarity to a target template (which could specify only one attribute or a combination of many attributes) and selection is the result of the strength of the matching input to current templates. The nature of this template must be elaborate enough so that a target, when present, is likely to gain access, while non-targets are likely to be excluded. Thus, good matches will increase the weight while poor matches decrease it (Duncan & Humphreys, 1989). Simply put, resources are allocated as a result of input-template matching, and selection weights are assigned as the degree of match is increased.

As the weight of one structural unit increases, the weights of all others change accordingly (i.e., similar items are allocated a greater weight and dissimilar items are allocated a lesser weight). Duncan and Humphreys call this process *weight linkage*.

According to the researchers, a second factor that affects selection weights is *spreading suppression,* or the rejection of similarly grouped distractors. Any change in weight for one input is passed on to others that are of a similar perceptual grouping. So, members of the same group of distractors tend to be rejected together, whereas separate groups of distractors must be rejected independently. As a result, increasing target-distractor similarity is detrimental to search because it fosters competition for access to visual short-term memory by increasing the resources (hence the weight) that is assigned to each distractor, and subsequently decreasing the available resources for the target. In addition, decreasing distractor-distractor similarity is detrimental because it hinders the possibility of spreading suppression through weight linkage (Duncan & Humphreys, 1989). A target, when present, is ultimately identified with a positive response when it enters into visual short-term memory.

In the context of the classic Treisman et al. (1977) experiment, the attentional engagement theory explains the findings as due to the fact that similarity between the distractors and the target template was not matched across conditions. In the feature search condition, each distractor shared no attributes with either target template (*pink* or *O*), but in the conjunction search, the distractors each shared one attribute with the template (*pink O*). Thus, the attentional engagement theory predicts best performance for a condition that includes low target-distractor similarity and high distractor-distractor similarity (e.g. a *red X* in *blue R* and *green P* distractors). Conversely, worst performance is predicted for a condition that includes high target-distractor similarity and low distractor-distractor similarity (e.g. a *red X* among *red O* and *green X* distractors). In essence, feature and conjunction search are quite similar, but it is the inherent complexity in conjunction search (i.e., more resources are allocated to distractors because of their heightened similarity to the target template) that explains the relative difficulty of conjunction search. In addition, the theory explains the relative irrelevance of display size on reaction times when target-distractor similarity is low as resulting because the distractors are attracting no (or very few) resources. Thus, adding additional distractors does not affect the ability of the target to gain access to VSTM. However, when the number of items in the visual array is increased, an effect on search time can be noted if the distractors are added in such positions as to create new perceptual groups (thereby decreasing distractor-distractor similarity), but should not affect search time if it is added to an already existing group (Duncan & Humphreys, 1989). Bundesen and Pedersen (1983) reported such a finding. They found little effect of adding additional distractors to already existing groups (approximately 2 ms/item) but a much larger effect was found when the addition of a new distractor caused an increase in the number of perceptual groups (approximately 12 ms/item).

According to attentional engagement theory (Duncan & Humphreys, 1989), when it comes to negative responses (i.e., a response indicating that a target was not presented) there are a number of different possibilities, but in general, negative responses should result in a slower response time than positive responses because usually they are made by default after comparing each element against a target template. For example, if target and distractors have an equal (or very similar) access to VSTM, a definite, negative response will only occur after all display elements have passed into (or through) VSTM or have been rejected by the template (Duncan & Humphreys, 1989). However, the researchers also note the possibility that the presence of a target produces an effect so as to alter the display as a whole perceptual unit. In this case, the need to compare each element of the array with a target template is unnecessary. Instead, the entire display can be compared to a corresponding template and a response can be based on the degree of match. In such a case, negative responses will not take longer because the task has now become a search for one of two possible properties of the display (i.e., target present or target absent). Thus, distractor-distractor homogeneity should produce much faster negative responses than distractor-distractor heterogeneity (Duncan & Humphreys, 1989).

The first four experiments reported by Duncan and Humphreys (1989) concerned within-object conjunctions. More specifically, they examined the effect of having distractors consisting of various combinations of the strokes that make up the target letter (i.e., L or T). In Experiment 5, Duncan and Humphreys examined the effects of across-object conjunctions in an attempt to challenge feature integration theory's explanation of the difficulty of conjunction search. As previously mentioned, it may be that conjunction search is often more difficult than feature search simply because the target shares more attributes with the distractors than in feature search. Thus, in feature search, often a single attribute of the target can be used for identification, whereas in conjunction search the target template must be more elaborate because each distractor contains one of the two features that are combined to make up the target (this point was made previously in response to the results of Experiment 4 in Treisman & Gelade, 1980). Recall that in Experiment 4, Treisman and Gelade's (1980) subjects found it easier to find an *R* in *Ps* and *Bs* than in *Ps* and *Qs*. In the first case, the target has a unique diagonal stroke, whereas in the second case, the target is only unique in its conjunction of strokes. Thus, in Experiment 5, Duncan and Humphreys modified the letters to reduce target-distractor similarity and had their subjects search for either a *caret* target (an R with the loop omitted) in *I* and Q distractors, or an *R* target in *P* and Q distractors. Search for the *caret* in *I* and Q distractors is equivalent to search for an *R* in *P* and Q distractors with the only difference being the missing loop from the *R* and *Ps*, thus decreasing target-distractor similarity and subsequently making search for the *caret* target easier. They found a much larger effect of display size in the search for an *R* than in the search for the *caret* (slopes of 25 ms/item and 3 ms/item respectively for target-present conditions). Thus, they were able to make conjunction search (i.e., search for *caret* in *I* and Q distractors) a relatively easy or a relatively difficult task by manipulating the degree of target-distractor similarity.

A final addition to the theory, included to explain further the difference between feature and conjunction search, is that the target template need not specify the entire target shape. Rather, in search for an *R* in *Ps* and *Bs* for example, the diagonal line might be a sufficient template. Now, this raises the question as to why the *caret* was not the target template in the search for an *R* among *Ps* and Q*s*. Duncan and Humphreys explain that the *caret* is not a natural divisible part of the letter *R*. They asked 18 subjects to divide an *R* into its most natural elements, and 15/18 identified the diagonal stroke as one of the natural elements of an *R*, whereas 0/18 identified the *caret*. The main point, however, is that decreasing target-distractor similarity can make even conjunction search quite easy.

Guided Search

Another search theory that emerged at approximately the same time as Duncan and Humphreys' (1989) AET is called guided search (Wolfe, Cave & Franzel, 1989). It is based on the same logical foundation as FIT with regard to the processes involved in conjoining features during serial search. However, there is a notable difference. According to Wolfe et al.'s (1989) paradigm, information obtained during the parallel processing stage can be used by the serial processes to facilitate search (recall that according to FIT, parallel and serial processes are autonomous). Thus, during the initial parallel processing stage, before features are conjoined, separable features are registered and can actually guide the subsequent serial search. For example, if one were searching for a target that was a *red X*, amongst *red* and *green* distractors, information from the parallel search would allow a guided serial search to only *red* objects. Thus, while the location of the target cannot be determined by parallel search, the location of all *red* stimuli will be excited, as will be the location of all *Xs*, and the *attentional spotlight* can be directed to these points of excitation. Supposedly a *red X* would produce the most excitation and thus should be attended to relatively quickly (Wolfe et al., 1989), but as AET predicts, this would be dependent on the similarity of the distractors to the target. By turning the display in to a series of distractors and target candidates, the theory of guided search demonstrates how it is not always necessary for a random serial search of all stimuli (as was predicted by FIT), and it allows for the possibility of conjunction search at speeds which FIT claims indicate parallel processing (Wolfe et al., 1989).

Guided search is very similar to AET in that the similarity between the targets and the distractors can lead to varying search speeds. This account, however, does not take into consideration the effect of the similarity between distractors and each other on search times. In addition, FIT predicts that there is an added difficulty presented by the need to conjoin features, whereas AET predicts that there is not. Guided search does not make any specific predictions regarding this aspect of the debate.

While a newer version of the theory, guided search 2.0 (Wolfe, 1994), did provide a more in-depth version of the theory, it still left some points unanswered. The idea of distractor-distractor similarity was mentioned, but largely as it applied to feature, not conjunction search (e.g., p. 232). In addition, Wolfe et al. didn't take either side of the debate regarding the added difficulty of conjunction search (i.e., in the need to form conjunctions). It was, however, mentioned that target-distractor similarity has an impact on the steepness of the search slopes (e.g., p. 225). That is, strong guidance from the parallel process is suggested to result in more efficient searches and shallower slopes, whereas when guidance from the parallel process is weak, search becomes more inefficient and slopes are steeper. But, as previously mentioned, the theory predicts that the degree of guidance comes from the degree of similarity between the target and the distractors. Thus, while the theory has dramatically improved since its conception and is still worth mentioning, it is not necessary to discuss it any further for the purposes of this thesis because questions that the present study intends to answer focus on aspects of search theory on which guided search does not make firm predictions (i.e., the importance of distractor-distractor similarity and the debate surrounding the inherent difficulty of conjunction search).

A Tale of Two Theories

Duncan and Humphreys (1989) were not the first, nor were they the only researchers to find parallel search functions or flat slopes for conjunction search. Nakayama and Silverman (1986) found very close to parallel search when they paired every combination of highly discriminable stimuli including colour, size, spatial frequency, binocular disparity, and direction of contrast. As reported in Treisman and Sato (1990), numerous researchers have found flat (or almost flat) slopes for many other conjunctions including combinations of shape, direction of motion, orientation, lateral separation, colour, etc. (e.g., McLeod, Driver & Crisp, 1988; Steinman, 1987; Wolfe, Cave & Franzel, 1989). Some researchers have also reported that with display sizes of fewer than eight items, search is often not serial and self-terminating, but is instead parallel (Houck & Hoffman, 1985; Pashler, 1987).

The appearance of numerous reports of parallel search for conjunctions were evidently a stunning blow to FIT, and subsequently prompted a swift response from Treisman and Sato (1990) in its defence. They attempted to explain these all too common findings that were inconsistent with FIT by testing three possible hypotheses that could account for the data. The first and simplest explanation proposed by the researchers was the possibility of specialized conjunction detectors that are specifically tuned to respond to certain combinations of features and function at preattentive levels. This account was essentially disregarded due to numerous problems based on the data that the researchers accumulated (see p. 474 of Treisman & Sato, 1990 for further explanation).

The second possible account for the findings, called the segregation hypothesis, builds on an idea first proposed by Egeth, Virzi, and Garbart (1984) that attention can be narrowed by linking one unimportant feature to an entire set of distractors, causing their exclusion, and allowing for the possibility that a set of distractors may be searched in parallel. Treisman and Sato (1990) suggested that we might see inhibition on a feature map that codes a relevant distractor feature. They claim that this could lead to decreased activity in the locations of the master map in which that feature currently exists. Parallel search of the remaining distractors for the unique feature that the target possesses is then possible. This explanation is similar to spreading suppression, however the inhibition of feature maps would occur in parallel, whereas spreading suppression occurs serially (though it can be extremely efficient). However, the researchers suspected that this explanation might not accurately capture the entire picture. They reasoned that since there was an additive component of the feature effect on conjunction search, this might suggest that each of the two features is contributing independently to search time. That is, each feature of the stimuli in the display adds a certain amount of time to the search, thus increasing the slope (see Experiment 2). For instance, in a display consisting of a fixed number of items, adding the dimension of size might increase search time by 6.5 ms, whereas adding the dimension of colour might increase it by 7.5 ms. Given this finding, Treisman and Sato reasoned that the inhibition of only one set of distractors doesn't provide a sufficient explanation for the increasingly common reports of parallel search for conjunctions. While inhibiting only one feature map is likely to reduce search time by a predictable amount, evidence from Experiment 2 suggested that it is unlikely that it would reduce it by enough to find evidence of parallel search for a conjunction target.

The third possibility, the feature inhibition strategy, is more consistent with the finding of additivity of individual features. Rather than simply having reduced activity on one feature map, feature inhibition could occur on two or more feature maps, reducing the activity of all distractors. Treisman and Sato describe the extreme instance whereby all activity of distractors generated in the master map is eliminated, causing the target to 'pop out' no matter how large the display size. If inhibition does not occur, a serial scan is then performed through the master map of locations.

According to the researchers, in order for object representations currently in the attention window to be formed, a pooled response from each feature map is collected which represents the activation produced by attended stimuli. It is subsequently evaluated based on the likelihood that the activation accurately indicates the presence of that particular feature. Thus, the more instances of a particular feature that are present, the more likely it is that its presence will be detected. Inhibition occurs not only for the activation level of the distractor items, but also for the presence of a target that shares their features (i.e., master-map locations). The more distinctive a target's features are, the more accurately it will be detected by this pooled response. That is, the less activation of the target's feature map that occurs by the distractors features, the more likely it is that the target will be detected. To begin with, it is obvious that Treisman and Sato have considered Wolfe et al.'s (1989) theory of guided search, but it is also becoming evident that the fundamental processes being used to explain FIT are very closely approximating those that underlie AET.

The number of adjacent elements that are taken into the attention window is said to be dependent on the summation of the activation that a particular group might cause, keeping the level of activation below a certain criterion amount. If a particular group of distractors contained very strongly inhibited stimuli, then they could be rejected as a group. However, if the distractors were composed of a very slightly inhibited feature, one closely resembling that of the target, only one or two at a time would be taken into the attention window (Treisman & Sato, 1990).

Treisman and Sato also used feature inhibition to explain the difficulty of search for a target which lacks a feature that is present in all of the distractors. If, for example, we are searching for a circle among distractor circles with a slash through them, search is relatively more difficult than if we are searching for a circle with an intersecting slash among circle distractors. The logic used by Treisman and Sato is as follows: If we were to inhibit the slash (a feature only possessed by the distractors) in the first case, both the target and the distractors would be identical, and thus search would become difficult. However, if we were to inhibit the circle (a feature possessed by both the distractors and the target) in the second case, only the slash remains, leaving a clear indication of the presence of the target.

To summarize the most recent version of FIT, when a target is defined by a single feature, one which is not shared by any of the distractors, it can be detected by the presence of activity in a specialized feature map and thus search is parallel and rather quick. If, however, the distractors share one or more features with the target, or the target is defined by a conjunction of features, the distractors will activate the same feature map(s) as the target. In this instance, attention is narrowed to search only a few or even one item at a time within a master map of locations. Only features currently in the attended location may be combined to yield a unified object percept and thus search is normally serial and self-terminating. The size of the attention window and the role of the feature inhibition hypothesis that were described above are the final aspects of the theory.

In addition to responding to the numerous findings of parallel search slopes for conjunctions, Treisman (1991) also responded directly to Duncan and Humphreys' (1989) AET. Treisman sought evidence that conjunction search is not only difficult because the target resembles each distractor and the distractors are different from each other, but also because there is an additional difficulty that is created by the need to conjoin features. In a series of experiments, Treisman attempted to make feature and conjunction search equivalent by manipulating the target-distractor and distractor-distractor similarity (in terms of their colour and orientation) in such a way so that the same target and distractors could be used in both feature and conjunction search and thus all distractors had to be compared against the same target template (see Treisman, 1991 or Duncan & Humphreys, 1992 for a review). While the precise details of these experiments are beyond the scope of this thesis, it was clear that if feature and conjunction search were indeed equated in these experiments, conjunction search did prove to be more difficult, supporting FIT. Duncan and Humphreys (1992) however, demonstrated that Treisman's (1991) feature and conjunction search were not identical. Due to the fact that in feature displays the distractors differed from each other by 0 steps on one dimension (i.e., they were identical on this dimension) and 4 steps on another dimension (i.e., they were very different on this dimension), and conjunction displays distractors differed from each other by 2 steps on each of the two dimensions (i.e., they were moderately different on both dimensions), feature and conjunction search were in fact not equated, and target detection in feature displays proved easier than in conjunction displays. There proved to be an advantage in terms of summed spreading suppression of distractors in feature displays over conjunction displays. Duncan and Humphreys (1992) were able to manipulate the stimuli in such a manner as to control both spreading suppression and input-template matching (the degree of match between the targets and distractors), and found feature and conjunction search to be equally difficult. While it has been demonstrated that on a number of levels, FIT and AET are quite similar, there are still fundamental differences in the two theories. These differences centre on the processes by which the features of a stimulus are bound. FIT requires that a viewer focus attention on a specific location to bind features, while AET assumes that features are bound preattentively. The evidence suggests, however, that AET appears to provide a more compelling description of the visual search process. Duncan and Humphreys (1989) were able to demonstrate that a variety of different within- and across-object conjunctions could be detected preattentively (indicated by parallel search slopes), suggesting that the feature binding process occurs at the parallel processing stage and does not appear to require focal attention.

A Note on Illusory Conjunctions

Illusory conjunctions present a potential problem for AET. Recall that an illusory conjunction is an incorrect combination of the features of two or more objects, and they typically occur when the stimulus presentation is so rapid that focal attention is not possible. Illusory conjunctions suggest then that since focal attention has been prevented, the features of the stimuli are not bound, and thus may be incorrectly combined to yield a perceptual representation of an object that is not actually present. Thus, this phenomenon provides evidence in support of the requirement of focal attention to bind stimulus features. Recently, however, there has been a debate as to the mere existence of illusory conjunctions (for a review, see Donk, 1999, 2001; Prinzmetal, Diedrichsen, & Ivry, 2001). Donk (1999, 2001) explains that it is not possible to directly observe illusory conjunctions. Instead, claims of their existence rely on the examination of error rates (i.e., the number of feature errors vs. the number of conjunction errors). Donk believes that this method of comparison does not necessarily provide an accurate indication of the rate of illusory conjunctions. While a thorough review of the procedures utilized by Donk (1999) is beyond the scope of this thesis, he essentially was able to demonstrate that illusory conjunctions are simply the interpretation that is given to the effects of target-distractor confusion, and guessing errors, rather than a problem with feature binding. For example, when target-distractor similarity is low, the number of feature errors and the number of illusory conjunction errors are roughly equal, but when target-distractor similarity is high, the number of illusory conjunction errors greatly exceeds the number of feature errors (Donk, 1999). Thus, if the target were a *red P* and the distractors were similar to the target, perhaps some were red and some were Ps, a viewer might perceive the colour red and subsequently guess that the *red P* target had also been presented. Donk notes that the forced target present/absent choice, which the viewer must make, is done after the visual array has been presented and is gone (illusory conjunctions are reported more consistently when the stimulus presentation is brief). Wolfe (1994) even discounts the likelihood of *free floating* features in his newer version of guided search (2.0). He reasons that when a stimulus is removed, our mental representation of it begins to decay, and that illusory conjunctions may simply be a product of that decay. That is, similar to the reasoning of Donk (1999), after a few moments have passed since the stimuli were presented, if asked about the presence of a *red P*, attention may recall the colour red and the letter P in a similar location, and subsequently construct a *red P* that was never actually present. This error, however, is different from an illusory conjunction in that it is a memory error, not a perceptual one. So, while this debate may continue, there is strong evidence from numerous sources (also see Jamieson, Thompson, Cuddy, & Mewhort, 2003 for an explanation of illusory conjunction errors in audition) suggesting that illusory conjunctions do not result from imperfect feature binding. Thus, AET appears to be safe from criticism of this nature.

Auditory Search

Now, the challenge for auditory researchers has been to examine what we have learned about the visual search process and determine whether it applies to the area of auditory cognition. Only a handful of studies have tested these theories in audition, presumably because of the difficulty that auditory researchers face in finding stimuli that can be manipulated on as many levels as in vision. One can make subtle yet noticeable changes in the colour, orientation, shape, or size of a visually perceived object, yet these manipulations are somewhat different when working with sounds. The features that auditory researchers tend to examine are the pitch, duration, location, timbre, and intensity.

Auditory Selective Attention

Some of the earliest research on auditory selective attention (i.e., focusing on one of several simultaneously presented stimuli) was performed by Cherry (1953). He had participants perform a dichotic listening task whereby listeners paid attention to input entering one ear and ignored input entering the other ear (voluntary guidance of attention). He found that listeners could shadow (repeat aloud) information presented to the attended ear while ignoring information presented to the unattended ear (i.e., they could choose which message to shadow). Spieth, Curtis, and Webster (1954) found that listeners were quite good at answering a question posed by the speaker of a message in an attended ear while they ignored a message presented to the unattended ear (i.e., they could spatially separate the two messages).

Research in auditory selective attention has since come a long way. It has been demonstrated in numerous studies that auditory attention can be guided both voluntarily and involuntarily, and that identification of an auditory target can be influenced by advance information that provides a cue about either the location or the frequency of a target sound (e.g., Mondor & Bregman, 1994; Mondor & Zatorre, 1995). Performance in these studies was improved if the cue provided accurate information about the probable location or frequency of the target relative to trials when the cues did not provide accurate information regarding the target. The target in these experiments was always defined by a feature other than that which defined the cue (e.g., duration). With these findings in mind, the question may now be raised as to whether or not the selection of auditory attention can be guided independently via location and frequency channels. More specifically, are the dimensions of location and frequency separable or integral? Mondor, Zatorre, and Terrio (1998) sought to investigate this issue. In their article, they noted that Woods, Alho, and Algazi (1994) found evidence that the conjunction of location and frequency information actually occurs before the independent analysis of these features is complete. Thus, it is quite possible that, because these features are conjoined so quickly that attention is focused on a particular feature (e.g., location), then attention may be subsequently directed to the other feature as well (e.g., frequency). To test this theory, Mondor et al. (1998) ran a series of experiments in which performance on a listening task was shown to depend on both frequency and location information, even if one of the features was unimportant to the goal of the task. Experiment 1 was designed to test whether location and frequency are separable or integral dimensions. On each trial, listeners were required to decide whether a pure tone came from either a central or a peripheral location (locations could be either 15° apart on the more difficult condition or 45° apart on the less difficult condition), or whether the tone was either a high or a low frequency (frequencies could differ by either 35 or 50 Hz in the more difficult condition or 447 or 412 Hz on the less difficult condition). Performance was tested with either no variation on the irrelevant dimension (control condition) or uncorrelated variation on the irrelevant dimension (i.e., variation on the uninformative dimension was independent of variation on the informative dimension). That is, in the control conditions listeners were required to categorize tones according to either pitch or location, while the other dimension was held constant. In the experimental conditions listeners were required to make the same judgements, but the irrelevant dimension varied in either frequency or location. The logic was that a listener can allocate attention individually to separable dimensions, but integral dimensions cannot be attended independently (see Garner, 1970, 1974, 1987). In Experiment 1, subjects were unable to ignore the variation on the uninformative dimension while making decisions about the tone based on the important dimension. Thus, the results indicate that, since subjects could not selectively attend to one dimension while ignoring a second dimension, the dimensions of location and frequency appear to be integral. Similar results by Melara and Marks (1990) demonstrated that pitch, timbre, and loudness all appear to be integral as well. As listeners, we appear to attend to sound streams rather than attending to the location, frequency, etc. of sounds independently. However, Mondor, Zatorre, and Terrio (1998) reviewed a theory proposed by Kubovy (1981), which suggested that, just as location and time appears to play a dominant role in guiding visual attention, so might frequency play such a role in audition. Kubovy explained that if, for example, two sounds are presented at the same time and in the same location, thus differing only in frequency, then they will be perceived as separate events. The same perception results if two sounds are of the same frequency and are presented in the same location, but differ only in time. However, if two sounds are of the same frequency and are presented simultaneously, differing only in location, then they will be perceived as a single auditory event, the origin of which is estimated somewhere in the middle.

In a two-part experiment, Mondor et al. (1998) sought to determine whether or not attention could be guided independently by location and frequency. They reasoned that if location and frequency could be attended to separately, then a cue on an uninformative dimension would not have a negative effect on performance on a target identification task. The participants in their experiments were required to judge which of two targets, distinguishable by rise time (the speed at which maximal amplitude is reached), was presented on a given trial. Each trial involved the presentation of a cue followed by a target. In Experiment 2A the rise times of the cues were constant whereas in Experiment 2B the rise times of the cues was varied randomly at 5, 15, 25, 35, or 45 ms to ensure that the cue could not be used as a standard against which to compare the target. Both frequency and location dominance conditions were tested wherein the cue and the target were identical on one dimension and different on the other. The results demonstrated that participants were faster at responding when the cue and the target were similar on the uninformative dimension, thus it would appear that listeners could not guide selection by location and frequency independently, but rather the two dimensions are integral with respect to selective attention (i.e., selection requires attending to both dimensions). Mondor et al. (1998) drew a comparison between their results and AET. They proposed that in a cue-target paradigm, such as the ones just described, an attentional template that includes both location and frequency information (because of the finding that attention cannot be guided independently via either channel) may be determined automatically by the cue in a bottom-up manner.

Woods, Alain and Ogawa (1998) reported evidence that a listener's ability to detect a target defined by a conjunction of frequency and location was affected by frequency information but not location information provided by the preceding tone. Participants responded more slowly when the frequency of the sound preceding the target and the target were identical than when they were different. This result differs from the findings of Experiments 2A and 2B in Mondor et al. (1998) in which a match in frequency between the cue and target facilitated target identification. It also differs from Mondor et al.'s (1998) suggestion that an attentional template for the selection of sounds includes information regarding both location and frequency. Woods et al.'s (1998) findings leave room for speculation that if a target were imbedded in a series of distractor sounds, target-distractor similarity in terms of location is unimportant, or perhaps less important than frequency. Thus in Experiment 3, Mondor et al. examined whether detection of a target embedded within a series of distractors is dependent on the similarity of the target to the distractors with respect to both frequency and location information. Listeners were presented with a target defined by both frequency and location (right location and 1000 Hz) in a series of 24 pure tones. The distractors were varied across four conditions with respect to their similarity to the target sound (either different location-different frequency, different location-similar frequency, similar location-different frequency, or similar location-similar frequency). Distractors were presented from the left in the different location conditions and from the right in the similar location conditions. Distractors were 418, 440, 464, and 488 Hz in the different frequency conditions and were 734, 773, 815, and 857 Hz in the similar frequency conditions. Speakers were positioned at 30° left and right of the individual. The results indicated that target detection was dependent on target-distractor similarity. Participants were fastest when target-distractor similarity was low and slowest when target-distractor similarity was high. In addition, the interaction between frequency similarity and location similarity was significant. Reaction time was substantially affected by both frequency and location similarity, and thus did not depend on only one feature. This finding provided evidence that both frequency and location are important in defining the attentional template in auditory search, but the results of this study do not indicate whether an advantage exists for frequency or location information.

A number of researchers, however, have made claims as to an advantage that frequency or location may have in guiding auditory selective attention. For example, Näätänen, Porkka, Merisalo, and Ahtola (1980) found evidence that if they presented listeners with a target tone that could be distinguished by either the ear of delivery or by frequency, people were faster at identifying targets defined by ear of delivery. This finding would suggest that during auditory selective attention, location has an inherent advantage. However, Woods, Alain, Diaz, Rhodes and Ogawa (2001) claimed that frequency is more important than location in guiding auditory attention. In Experiment 1A and 1B of their study, they had their participants listen for a target sound within lengthy, rapid, irregular sequences (3.5 min.) of more than 500 pure tones. The distractor sounds were one of two potential frequencies (250 or 1500 Hz) in the frequency condition and one of two potential locations (left or right ear in Experiment 1A and 90° azimuth left or right in Experiment 1B) in the location condition. The target sounds, which occurred randomly during the sequence, were one of the potential frequencies of the distractors (i.e., either 250 or 1500 Hz, but never the same frequency as the distractors) during frequency conditions and one of the potential locations of the distractors (i.e., either left or right ear or 90° azimuth left or right, but never the same location as the distractors) during location conditions. A conjunction condition was also tested in which targets were a specific frequency from a specific location. They found that significantly less time was needed to detect targets that were defined by frequency than by location, and concluded that frequency is the primary feature through which information about auditory objects is conveyed. Woods et al. even proposed a frequency-based feature integration theory to account for their findings (FB-FIT). According to FB-FIT, auditory features remain unbound until focal attention is directed to the frequency of the stimulus.

The report by Woods et al. is inconsistent with that of Mondor, Zatorre, and Terrio (1998) who found that frequency and location are integral dimensions. In addition, there were numerous problems with the conclusions reached by Woods et al. (2001; as reported in Mondor, Giannuzzi, & Thorne, 2003). It is possible that performance was determined by the relative ease or difficulty of distinguishing the alternative frequencies and locations used. According to Woods et al., the psychological difference (measured in just noticeable differences or jnds) between the frequencies used was 400 jnds whereas the psychological difference between the locations used was only 40 jnds. Thus, it is obvious that the frequencies may have simply been easier to discriminate than were the locations.

In addition, in Experiment 4, Woods et al. (2001) found a 'conjunction benefit', in which participants had an easier time detecting targets defined by a conjunction of frequency and location than targets defined by location alone. This result is quite unexpected considering that feature integration theory (Treisman & Gelade, 1980), attentional engagement theory (Duncan & Humphreys, 1989), and the guided search model (Wolfe et al., 1989) all predict a more difficult time, in general, when detecting targets defined by a conjunction of features. Mondor, Giannuzzi, and Thorne (2003) sought to address the problematic conclusions proposed by Woods and colleagues and demonstrate that the acoustic feature most useful in guiding auditory attention is the one that permits the easiest discrimination of the target sound from the distractor sounds (as predicted by attentional engagement theory). Each of their experiments consisted of 140 trials of rapid sequences each made up of 24 pure tones. The target was defined by a specific frequency, location, or combination of both, in addition to the presence of either a silent gap or a slow rise time to achieve a discriminable target. In Experiment 1, Mondor et al. equated the psychological difference between frequencies and locations. They reasoned that performance might be equal in frequency and location conditions if they equated the difficulty that listeners had in discriminating target sounds from distractors (measured in jnds). The results actually demonstrated that when jnds were equated for both features, it was location, not frequency that had an advantage. Experiment 2 was based upon information gained from preliminary research whereby location and frequency were equated in terms of the speed and accuracy with which listeners were able to detect these features. It was determined that frequencies of 350 Hz and 1367 Hz are detected as quickly and accurately as locations of 45° left and 45° right. The logic behind this experiment was that if these features were indeed equated as to their difficulty in discriminability, then the accuracy of target detection in both of the feature conditions should be equivalent. The results demonstrated that this was indeed the case. Experiment 3 used frequencies that were much easier to discriminate than the locations used. In this instance the results were similar to the Woods et al. (2001) experiment in that performance was better in the frequency condition due to the discriminability advantage that was created. Experiment 4 tested to see whether a frequency advantage was present when sequences were rapid and irregular (as in Experiment 3 where the stimulus onset asynchrony, or SOA, varied from 150-450 ms) as opposed to a regular sequence (as in Experiments 1 and 2 where the SOA was 300 ms). These were the SOAs used in the Woods et al. (2001) study. No such advantage was evident in Mondor et al.'s results.

Thus, Mondor et al. (2003) determined that neither the frequency nor the location of a target is more important in guiding auditory attention per se, contradicting the predictions of a frequency-based FIT (Woods et al., 2001). Instead, they found that the acoustic feature most useful in guiding auditory attention was the one that allowed for the easiest discrimination between the target sound and the distractor sounds (in line with the predictions of attentional engagement theory). In addition, Mondor et al. never found a significant difference between the more difficult feature condition and the conjunction condition. This finding contradicts the predictions of FIT, for, according to the theory, a target defined by a single feature should be detected better than a target defined by a conjunction of features (Treisman & Gormician, 1988). This may be an indication that focal attention is not required to form conjunctions in audition. The Mondor et al. experiments, however, only examined target-distractor similarity (i.e., the discriminability between the target sound and the distractor sounds). This leaves room for investigation into the effects of manipulating the similarity between the distractor sounds and each other.

Experiment 1

Experiment 1 of the present study was intended to examine attentional search in audition, focusing specifically on the effects of target-distractor similarity and of distractor-distractor similarity. A similar task and methodology as that used by Mondor et al. (1998) was adopted for this experiment.

Though both 'target-absent' and 'target-foil' trials were used to ensure that participants were able to accurately identify the target, the participants' accuracy on 'target-present' trials was of the most interest for this study. It was anticipated that the participants would have the most difficulty (lowest accuracy) when target-distractor similarity was high and distractor-distractor similarity was low, because the distractors were quite similar to the target template and quite dissimilar from each other. According to AET, this factor should make the task of rejecting distractors (and thus detecting the target) quite difficult for two reasons. First, the weight that each distractor is given (due to it's similarity to the target template) is similar to the weight given to the target, thus making rejection of distractors difficult. Second, it is unlikely that spreading suppression (the quick rejection of previously rejected distractors) will occur because most of the distractors that occur sequentially are dissimilar. Thus, a previously rejected distractor is less likely to be similar to a subsequent sound, making a weight linkage (and thus quick rejection) less likely.

It was anticipated that the participants would have the easiest time discriminating the target when target-distractor similarity was low and distractor-distractor similarity was high, because the distractors were not very similar to the target template (and thus shouldn't carry much weight) and spreading suppression should occur between similar distractor sounds. That is, the weight of a previously rejected distractor is more likely to be linked to the weight of a subsequent sound, leading to significantly faster rejection of the subsequent distractor sound. It is proposed that perhaps in audition, this process is carried out by a preattentive grouping mechanism that tags sounds with respect to their probable membership in an existing sound stream (Mondor & Terrio, 1998). A further discussion of this theory will follow in the General Discussion section.

According to AET, if both target-distractor similarity and distractor-distractor similarity are low, it should be slightly more difficult to detect the target than if target-distractor similarity is low and distractor-distractor similarity is high. In the former situation, the distractors are dissimilar to the target and thus will not be assigned much weight. However, spreading suppression between the distractors is unlikely to occur because it is less likely for the distractors to share the same weight, due to their dissimilarities. It follows that when both target-distractor similarity and distractor-distractor similarity are high, target detection should be more accurate than when target-distractor similarity is high and distractor-distractor similarity is low (due to spreading suppression of similar distractors). However, we should see less accurate performance than if both target-distractor similarity and distractor-distractor similarity are low (because the weight assigned to each distractor will be greater, due to their similarity to the target template).

In a visual search task, both AET and FIT predict that target detection will become more difficult as target-distractor similarity is increased and distractor-distractor similarity is decreased. While the predictions of AET appear to be applicable in an auditory search task, the predictions of FIT for this experiment may not, and will be discussed in the Results and Discussion sections of the paper.

Method

Participants

Fifty undergraduate psychology students volunteered to participate in Experiment 1 in exchange for course credit. None of them reported any corrected or uncorrected hearing impairment. The listeners were randomly assigned to participate in one of four possible conditions defined by the similarity of the target to the distractors and by the similarity of the distractors to one another.

Materials

Sounds. Pure tones of 250, 810, 900, 1000, 1100, 1210, and 4000 Hz, each 100 ms in duration, were synthesized at a sampling rate of 44,100 Hz using the Cool Edit Software System (Syntrillium Software Corporation, 1999). All target and distractor sounds had onset and offset amplitude ramps to ensure that there were no onset- or offset-clicks. The target sound had 95 ms onset-ramp and a 3 ms offset-ramp. Half of the distractor sounds had 3 ms onset-ramps and half had 95 ms onset-ramps, and all had 3 ms offset amplitude ramps.

Computer and sound system. The experiment was controlled by a Dell Dimension L800R Pentium III computer running the E-Studio Software System (Psychology Software Tools Inc., 1999). Sounds were presented from Polk Audio R15 speakers. A custom-made device was used to control the speaker from which a sound was presented.

Design and Procedure

Before the experiment was conducted, the participants were required to complete 20 familiarization trials in which they had to differentiate between the 1000 Hz target and the 1000 Hz target-foil to ensure that they were able to discriminate between them. They were presented with a single sound on each of these trials. They were required to make a keypress response of '1' for target and '0' for target-foil on a computer keyboard. A minimum of 75% accuracy on these familiarization trials was required to continue the experiment. Two subjects were not permitted to continue to the actual experiment because they did not meet this minimum accuracy requirement on these familiarization trials.

The actual experiment consisted of 24 practice trials in which accuracy feedback was provided to the listener, and 120 experimental trials with no feedback. All practice and experimental trials consisted of 24 pure tones presented in a rapid, irregular sequence. The time between the onsets of consecutive sounds (or stimulus onset asynchrony, SOA) varied randomly between 150 and 450 ms. This interval is similar to that used by Mondor et al. (2003) and by Woods et al. (2001). Participants were required to listen to the entire sequence and then to respond to a visually presented question asking whether the target had been presented. The target sound, defined by a 95 ms onset amplitude ramp (slow rise time), was presented on 50% of trials, always came from a speaker positioned at 0° azimuth, and was always 1000 Hz. On 25% of the trials, a 1000 Hz sound with a 3 ms onset amplitude ramp was presented in a randomly determined position (target-foil). The target-foil was used to ensure that the listeners were not simply responding to any sound from the 0° speaker. On the remaining 25% of trials no sound was presented from the 0° azimuth speaker. When presented, that target could occur in the 6th, 8th, 10th, 12th, 14th, 16th, or 18th temporal position, the precise position varied randomly from trial to trial. Participants were required to press '1' on a keyboard if they had heard the target and '0' if they had not. The computer recorded the participant's accuracy in making the judgement. Distractor sounds came from different locations than the target. The positions and frequencies of the distractor sounds varied across conditions. However, half of the distractor sounds had 3 ms onset-ramps and half had 95 ms onset-ramps to ensure that rise time was not the defining feature of the target.

Target detection was studied in four separate conditions. In the 'low target-distractor similarity and low distractor-distractor similarity condition' (Low T-D/Low D-D) the target differed substantially from the distractors and the distractors differed substantially from one another. Distractor sounds were presented from 45° azimuth left and right and were pure tones of 250 Hz and 4000Hz (the location and frequency of each distractor sound was randomly determined by the computer program).

In the 'low target-distractor similarity and high distractor-distractor similarity condition' (Low T-D/High D-D), whereas the target differed substantially from the distractors, the distractors were similar to one another. Experiment 1 from Duncan and Humphreys (1989) provided evidence that increasing or decreasing distractor-distractor similarity has little impact on search performance when target-distractor similarity is low.

Duncan and Humphreys argued that this is the case because the distractors attract very few resources, leaving sufficient resources for quick target detection. Thus, to test this possibility in different blocks of trials the distractors were identical in location and frequency. There were four blocks of trials, each defined by the location and frequency of the distractors (45° left & 250 Hz, 45° right & 250 Hz, 45° left & 4000 Hz, and 45° right & 4000 Hz). Different subjects were randomly assigned to each of the four blocks of trials.

In the 'high target-distractor similarity and low distractor-distractor similarity condition' (High T-D/Low D-D), whereas the target was similar to the distractors, the distractors were quite different from one another. In different blocks of trials, distractors were presented from either 15° left and 45° right or 15° right and 45° left, and were pure tones of either 900 Hz and 4000 Hz or 1100 Hz and 250 Hz. Once again, counterbalancing produced four possible sub-conditions comprised of all possible combinations of the potential frequencies and locations. Different subjects were randomly assigned to each of the blocks of trials.

Finally, in the 'high target-distractor similarity and high distractor-distractor similarity condition' (High T-D/High D-D), the target was similar to the distractors and the distractors were similar to one another. Distractors were presented from 15° left and 15° right. In different blocks of trials, the distractors were pure tones of either 810 Hz and 900 Hz or 1100 Hz and 1210 Hz. Once again, different groups of subjects were randomly assigned to each of the blocks of trials.

Results

A detailed summary of performance in Experiment 1 is provided Table 1 and in Figure 1. A two-way between-subjects ANOVA (Target-Distractor Similarity [high, low], Distractor-Distractor Similarity [high, low]) was performed using the proportion of correct responses on target-present trials as the dependent variable. This analysis revealed a significant main effect of target-distractor similarity, $F(1, 44) = 27.44, p < .001$, and of distractor-distractor similarity, $F(1, 44) = 4.07, p < .05$. There was no significant interaction, $F < 1$. As predicted, better performance was found in the two conditions with low target-distractor similarity than in the two conditions with high target-distractor similarity. In addition, in each of the two sets of conditions with the same target-distractor similarity, better performance was found when distractor-distractor similarity was high than when it was low.

False Alarms

An examination of false alarms also revealed some interesting results. A summary of false alarms based on trials on which the target was absent (target-absent trials) or on the trials on which the 1000 Hz non-target sound was presented from the 0° (centre) speaker (target-foil trials) is also provided in Table 1. A two-way between-subjects ANOVA on proportion correct for target-absent trials revealed a significant main effect of target-distractor similarity, $F(1, 44) = 21.52, p < .001$. There was no significant main effect of distractor-distractor similarity, $F < 1$, nor was there a significant interaction, $F < 1$. This pattern of results suggests that listeners perceived a target being present more often in High T-D similarity conditions than in Low T-D conditions, a result that would likely be anticipated by both AET and FIT. However, neither Duncan and Humphreys

(1989), nor Treisman and Gelade (1980) made any specific predictions regarding target-absent trials because they were mainly concerned with reaction times on trials consisting of entire visual arrays, which often could be viewed for as long as necessary to make a response, so errors were rare. This experiment involved serial presentation of sounds, thus failing to detect a target sound, or confusing a distractor sound with the target is more likely. There were no significant differences in false alarm rates as a function of distractor-distractor similarity. This finding is somewhat surprising, for it might be anticipated that when distractors are more dissimilar, it would be easier to confuse one of them with the target, whereas when they are more similar, confusion would be less likely due to weight linkage and spreading suppression occurring between similar distractor sounds, if these factors do indeed exist in auditory search.

A two-way between-subjects ANOVA on target-foil trials revealed that there was a significant main effect of target-distractor similarity, $F(1, 44) = 60.39, p < .001$, but on these trials, there was a significant main effect of distractor-distractor similarity, $F(1, 44) = 4.79, p > .05$. There was no significant interaction, $F < 1$.

Though the finding of no significant main effect of distractor-distractor similarity on target-absent trials is somewhat surprising, the answer may lie in the degree of target-distractor similarity used in this experiment. Listeners may not have needed to rely as much on spreading suppression on target-absent trials to detect the absence of a target because the average target-distractor similarity across all conditions was not incredibly high in this experiment. Instead, listeners could potentially have relied on the input-template matching process to conclude that no target was presented on target-absent trials. Though spreading suppression between similar distractors would still occur on target-absent trials, it would be of less use with a lower target-distractor similarity. That is, the grouping of similar distractors is less important in determining that no target was presented when, in general, target-distractor similarity is relatively low. If, however, the average target-distractor similarity were higher across all conditions on this experiment, then the effects of distractor-distractor similarity on target-absent trials might have become an increasingly important factor. Further explanation of these findings will follow in the Discussion section of Experiment 2.

Signal Detection Analyses

Signal detection analysis was also used to evaluate performance. d' and β were calculated separately using false alarm rates based on both target-absent trials and on target-foil trials (see Table 1). When target-absent false positive error rates were used in the calculation of d', a two-way between-subjects ANOVA revealed that both the main effects of target-distractor similarity, $F(1, 44) = 48.5, p < .001$, and of distractor-distractor similarity, $F(1, 44) = 4.04, p < .05$, were significant. The interaction was not significant, $F < 1$. Similarly, when target-foil false alarm rates were used in the calculation of d', both the main effects of target-distractor similarity, $F(1, 44) = 59.25, p < .001$, and of distractor-distractor similarity, $F(1, 44) = 8.69, p = .005$, were significant. The interaction was not significant, $F < 1$. d' is a measure of sensitivity which is uninfluenced by response bias on the part of the listeners who may have unconsciously used a cost-benefit analysis to respond in a particular way. That is, it is an analysis that controls for the possibility that the respondents may have had a tendency to either indicate that the target was present or absent. Had this analysis revealed non-significant main effects, there would be reason to question the interpretation of the original

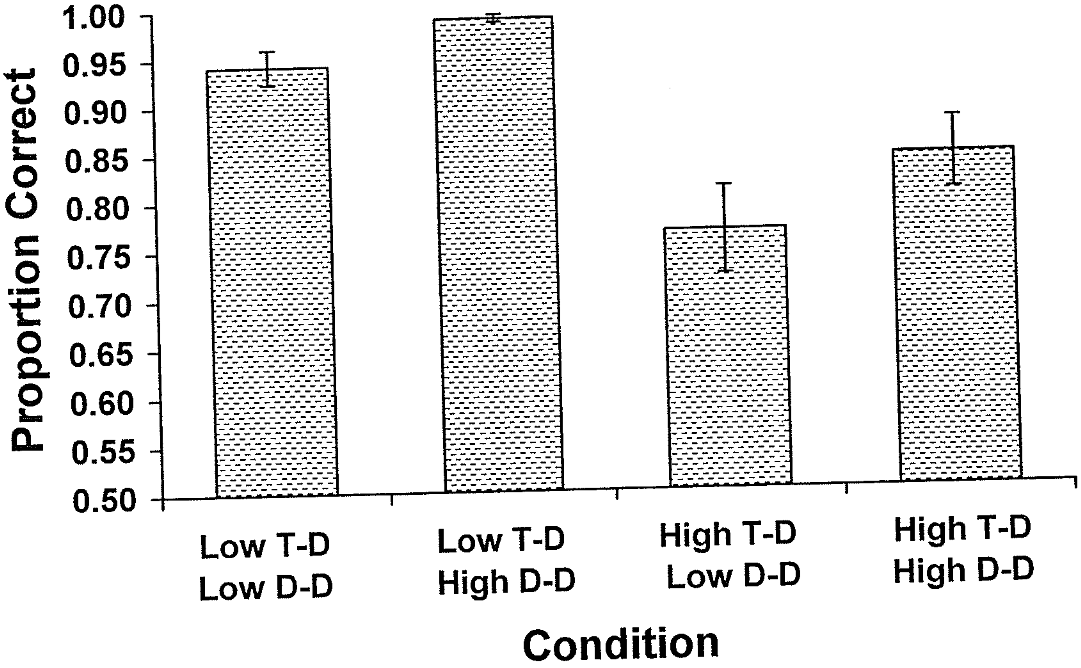
ANOVA, however, this was not the case. Thus, this *d′* analysis reinforces the original interpretation.

Separate ANOVAs revealed that β, calculated using either target-absent or target-foil error rates, did not vary significantly across conditions ($p > .05$ in all cases). This result indicates that there was not a tendency to respond differently (i.e., to indicate that the target was present or absent) across conditions. Using false alarms from target-absent trials, the mean β value was 2.16. This would indicate a tendency for the listeners to respond that the target was not present. When β was calculated using false alarms from target-foil trials, the mean β value was 0.95, which indicates that there was practically no response bias on the part of the listeners.

Table 1: Performance for each condition in Experiment 1 is described in terms of proportion correct on target-present, target-absent, and target-foil trials. *d'* and *β* were calculated using false alarms from both target-absent and from target-foil trials. Standard errors for each measure are given in parentheses.

	Target-Present	Target-Absent	Target-Foil	*d'* (Target-Absent)	*β* (Target-Absent)	*d'* (Target-Foil)	*β* (Target-Foil)
Low T-D Low D-D	.941 (.018)	.993 (.004)	.925 (.015)	4.04 (.170)	3.20 (.813)	3.27 (.266)	0.88 (.156)
Low T-T High D-D	.989 (.004)	1.00 (.000)	.980 (.009)	4.50 (.059)	1.55 (.209)	4.25 (.134)	1.13 (.238)
High T-D Low D-D	.766 (.045)	.780 (.070)	.608 (.047)	2.02 (.380)	2.18 (.486)	1.22 (.300)	0.80 (.109)
High T-D High D-D	.842 (.037)	.798 (.056)	.717 (.055)	2.67 (.359)	1.71 (.558)	1.92 (.382)	0.98 (.220)

Figure 1. Mean proportion correct for target-present trials of Experiment 1 as a function of condition.



It is clear that both target-distractor and distractor-distractor similarity play an important role in a listener's ability to detect a target. Upon examination of both Table 1 and Figure 1, it would appear that the more important factor is target-distractor similarity, as listeners in the two conditions with Low T-D similarity performed significantly better than listeners in the two High T-D similarity conditions (see Figure 1). However, distractor-distractor similarity also appears to play an important role as well, for whether T-D similarity was high or low, listeners were better on conditions with high D-D than on conditions with low D-D similarity, respectively. A calculation of the effect size on target-present trials confirmed this observation. It revealed that 38.4% of the variability in accuracy was accounted for by target-distractor similarity, $\eta^2 = .384$, while 8.4% of the variability in accuracy was accounted for by distractor-distractor similarity, $\eta^2 = .084$. Almost none of the variability in accuracy was accounted for by the interaction, $\eta^2 = .004$.

Discussion

All of the predictions of Experiment 1 were confirmed. It appears that in audition, as in vision, both target-distractor similarity and distractor-distractor similarity play a role in target detection. These results are, of course, consistent with AET (Duncan & Humphreys, 1989; 1992). Both input-template matching and quick rejection due to spreading suppression could potentially affect our ability to perform an auditory search task. When we are listening for the presence of a target embedded in a stream of sounds, it appears that we may be comparing each sound to a template or representation of the target that is stored in memory (see Mondor, Zatorre, & Terrio, 1998, for a discussion). If the sounds encountered are similar to the target, then as suggested by AET, they may be assigned a high weight and be more difficult to reject. In contrast, sounds that are very different from the target would be assigned a low weight and would be easier to reject. This interpretation is consistent with the pattern of results obtained in Experiment 1 in that participants performed significantly better on trials with low T-D similarity than on trials with high T-D similarity.

A version of spreading suppression can be adapted from AET to help explain the finding that, while holding T-D similarity constant, listeners were more accurate on trials with high D-D similarity than they were on trials with low D-D similarity. According to AET, each input is given a weight based on its similarity to the target template. As suggested by Duncan & Humphreys (1998), this weight could perhaps be based on an excitatory or inhibitory input produced by the sound. If two or more sequential distractor sounds carry the same or a similar weight, then the first sound would be analyzed and rejected based on its weight. The subsequent sound, which would carry a similar weight to the previous one, could be quickly rejected due to a process similar to spreading suppression (because of weight linkage). Perhaps in audition, this process is carried out by a preattentive grouping mechanism that operates at this stage to tag sounds with respect to their probable membership in an existing sound stream (Mondor & Terrio, 1998). This extension of AET could explain the effects of both target-distractor and distractor-distractor similarity on auditory search observed in this experiment. Further explanation of this theory as it pertains to the results is provided in the General Discussion section of this paper.

Feature integration theory (Treisman, 1991; Treisman & Gelade, 1980) also appears to be consistent with the effects of target-distractor similarity apparent in

Experiment 1. According to a modified version of FIT, presentation of each sound in a sequence should activate a location and a frequency on a master map that contains the specific features of the target. If the target features are not activated (as in the low T-D conditions), then a quick rejection of that sound is likely to occur. If one of the target features were activated (as in the high T-D conditions), then supposedly, focal attention would be necessary to conjoin all of the remaining features of the sound before a target/distractor distinction could be made. Thus, it could be predicted that listeners would be less accurate in conditions with high target-distractor similarity than in conditions with low target-distractor similarity, because when target-distractor similarity is high, FIT would predict there to be more sounds that activate the target features, which require focal attention to be conjoined. Thus, when target-distractor similarity is high, this process would take longer and errors would be more likely.

It appears, however, that FIT may be inconsistent with the effects of distractor-distractor similarity obtained in this experiment. According to FIT (Treisman & Sato, 1990), high distractor-distractor similarity may assist in limiting the serial search that follows the initial parallel processing. More specifically, spatial locations that do not contain target features can be ignored in the initial parallel stage of processing, leaving only locations that contain potential targets to be searched serially. However, when sounds are presented sequentially (i.e., serially), there can be no such restriction of the search to sounds likely to be the target because it is impossible for a listener to predetermine precise frequency and location of an upcoming sound. Without the entire set of stimuli being presented at the onset of the search process, this explanation falls short. Thus, it appears reasonable to assume that the current version of FIT would predict no effect of distractor-distractor similarity on search. This, of course, is inconsistent with the results of Experiment 1. All in all it appears that AET provides a better account of performance in Experiment 1 than does FIT.

Experiment 2

In Experiment 1 it was determined that both target-distractor similarity and distractor-distractor similarity play an important role in the auditory search process. Another aspect of the listening situation that could potentially affect search performance is the existence of multiple targets. Attentional engagement theory predicts a detrimental effect of having multiple targets, due both to the increase in the memory load, and to the increase in the complexity of the attentional template. Further, it stands to reason that as the heterogeneity of the targets increases, the complexity of the template required to distinguish them from the distractor sounds should increase as well. In addition, if as AET suggests, the selection weight for each distractor is determined by its shared attributes with a target template, then as the number of attributes specified by the template is increased, it follows that the average distractor will likely share more attributes with the target, thus making search less efficient (Duncan & Humphreys, 1989). This hypothesis was tested in the present experiment using a procedure similar to that used in the previous experiment. However, rather than having listeners detect the presence of one potential target, they were required to detect the presence of either of two potential targets. Thus, in Experiment 2 the effect of target-target similarity was manipulated to determine whether a reduction in target-target similarity (i.e., necessitating a more elaborate or complex target template) will have a negative effect on performance of a target detection task, and whether there is an additional effect of decreasing distractor-distractor similarity. Overall, target-distractor similarity was held constant across all conditions of this experiment.

As in the previous experiment, participants' accuracy on 'target-present' trials was of the most interest for this study. It was predicted that performance in the two conditions in which target-target similarity was high (High T-T/High D-D and High T-T/Low D-D) should be fairly accurate because the two potential targets were quite similar. There should, however, be a difference between these two conditions because of the effect of distractor-distractor similarity. In Experiment 1 it was determined that performance is best when distractor-distractor similarity is high. Thus, performance in the High T-T/High D-D condition should be significantly better than in the High T-T/Low D-D condition due to the effects of decreasing distractor-distractor similarity.

It was also anticipated that performance in the two Low T-T similarity conditions should be worse than in the two High T-T similarity conditions because the decrease in target-target similarity should cause the need for a more elaborate or complex target template. This prediction is based on the assumption that more time is required to determine whether a sound matches a more elaborate template, and that this translates into an elevated error rate.

Finally, performance in the Low T-T/High D-D condition was predicted to be significantly better than in the Low T-T/Low D-D condition due to the negative consequences of decreasing distractor-distractor similarity that were apparent in Experiment 1.

The previous explanations of the anticipated findings are all based on an interpretation of AET (Duncan & Humphreys, 1989). Although, to my knowledge, these issues have not been addressed by FIT, an attempt to relate the obtained results of FIT will be undertaken in the Discussion section of this experiment.

Method

Participants

Forty-two undergraduate psychology students participated in this experiment in exchange for course credit. None of them reported any corrected or uncorrected hearing impairment. The participants were randomly assigned to one of four possible conditions defined by the similarity of the two potential targets, and the similarity of the distractors to one another.

Materials

Sounds. Pure tones of 700, 900, 1000, 1200, 1300 Hz, each 100 ms long, were synthesized at a sampling rate of 44,100 Hz using the Adobe Audition Software System (Adobe Systems Incorporated, 2005). All other details of sound synthesis were the same as in Experiment 1.

Computer and sound system. The experiment was controlled by a Dell Dimension L800R Pentium III computer running the E-Studio Software System (Psychology Software Tools Inc., 1999). All of the sounds came from Polk Audio R15 speakers positioned at azimuths of 15° left, 0°, and 15° right.

Design and Procedure

As a result of the findings of Experiment 1, it was anticipated that altering the degree of similarity between the distractors and each other would also have an impact on performance across conditions. Thus, performance was examined in two conditions in which target-target similarity was high and in two conditions in which it was low. The conditions with the same target-target similarity differed in that one consisted of distractors that were similar to one another (high D-D similarity), while the other consisted of distractors that were dissimilar to one another (low D-D similarity). All other methodological details are the same as in Experiment 1. Two subjects were disqualified from continuing on to the actual experiment because they did not meet the minimum accuracy requirement on the familiarization trials.

During the experiment, distractor sounds were presented from 15° left and 15° right of the listener and the target was always presented from 0° azimuth. One of the two possible target sounds was presented on 50% of the trials. A distractor sound of the same frequency as the targets but with a faster rise time (i.e., 3 ms) was presented on 25% of the trials (always from the 0° azimuth speaker), to ensure that listeners were not simply responding to any sound from the 0° azimuth speaker. No sound was presented from the center speaker on the remaining 25% of the trials.

As previously mentioned, target detection was studied in four separate conditions. In the 'high target-target similarity, high distractor-distractor similarity condition' (High T-T/High D-D), the two potential targets were similar to one another and the distractors were similar to one another. The potential target sound on any given trial was either 900 Hz or 1000 Hz (randomly selected by the computer program) and the distractor sounds were 1300 Hz and 1350 Hz. The individual distractor sounds that made up each trial of the experiment, for all four conditions, were selected randomly as was the SOA.

In the 'high target-target similarity, low distractor-distractor similarity condition' (High T-T/Low D-D), whereas the two potential targets were similar to one another, the distractors were quite different from one another. The potential target sound on any given trial was either 900 Hz or 1000 Hz (targets were the same as on the High T-T/High D-D condition), however in this condition, the distractor sounds were 575 Hz and 1325 Hz. In both the High T-T/High D-D and the High T-T/Low D-D conditions, the average frequency separation between the targets and the distractors was 375 Hz. This number was calculated by averaging the frequency separation between each target and each of the distractors, then taking the average of the two numbers.

In the 'low target-target similarity, high distractor-distractor similarity condition' (Low T-T/High D-D), whereas the two potential targets were quite different from one another, the distractors were quite similar to one another. The potential target sound on any given trial was either 700 Hz or 1200 Hz and the distractor sounds were 1300 Hz and 1350 Hz (distractors were same as on the High T-T/High D-D condition).

Finally, in the 'low target-target similarity, low distractor-distractor similarity condition' (Low T-T/Low D-D), the potential targets were quite different from one another and the distractors were quite different from one another. The potential target sound on any given trial was either 700 Hz or 1200 Hz (targets were the same as on the Low T-T/High D-D condition) and distractor sounds were 575 Hz and 1325 Hz (distractors were the same as on the High T-T/Low D-D condition). The average frequency separation between the targets and the distractors once again was 375 Hz (the same as the first two conditions), thus target-distractor similarity was held constant across all conditions of this experiment.

Results

A detailed summary of performance in Experiment 2 is provided in Table 2 and in Figure 2. A two-way between-subjects ANOVA (Target-Target Similarity [high, low], Distractor-Distractor Similarity [high, low]) was performed using the proportion of correct responses on target-present trials as the dependent variable. This analysis revealed a significant main effect of target-target similarity, $F(1, 36) = 32.85, p < .001$, and of distractor-distractor similarity, $F(1, 36) = 7.81, p < .01$. There was no significant interaction, $F < 1$. As predicted, listeners were able to detect the target more accurately in the two conditions with high target-target similarity than in the two conditions with low target-target similarity. In addition, in each of the two sets of conditions with the same target-target similarity, better performance was found in the condition with high distractor-distractor similarity than in the condition with low distractor-distractor similarity.

False Alarms

An examination of the false alarms of Experiment 2 also revealed some interesting findings. A summary of false alarms for both target-absent and target-foil trials is also provided in Table 2. Separate two-way between-subjects ANOVAs of false alarm rates based on both target-absent trials and target-foil trials revealed no significant main effect of target-target similarity in either case, $F(1, 36) = 3.12, p > .05$, and, $F(1, 36) = 3.04, p > .05$, respectively. This finding would be predicted by both AET and FIT because when no target is present, and the degree of match between the targets and the distractors is held constant across all conditions (recall that target-distractor similarity was held constant during Experiment 2), there is no reason to suspect that high T-T similarity would lead to better performance than low T-T similarity. That is, when overall target-distractor similarity is the same in each condition, each distractor will have, on average, a similar degree of match to the target template, so there is no reason to believe that when a target is not present, any given distractor will be more likely to activate the template when T-T similarity is low than when it is high.

There was, however, a significant main effect of distractor-distractor similarity on target-absent trials, $F(1, 36) = 5.29, p < .05$, and on target-foil trials, $F(1, 36) = 4.01, p < .05$. Thus, it appears that when there are no targets present, significantly more target confusion occurs when the distractors are dissimilar from one another than when they are similar. This was the result that would have been expected, but was not found, on target-absent trials in Experiment 1. Further explanation will follow in the Discussion section of this experiment. There was no significant interaction on either target-absent or target-foil trials, $F < 1$ in both cases.

Signal Detection Analyses

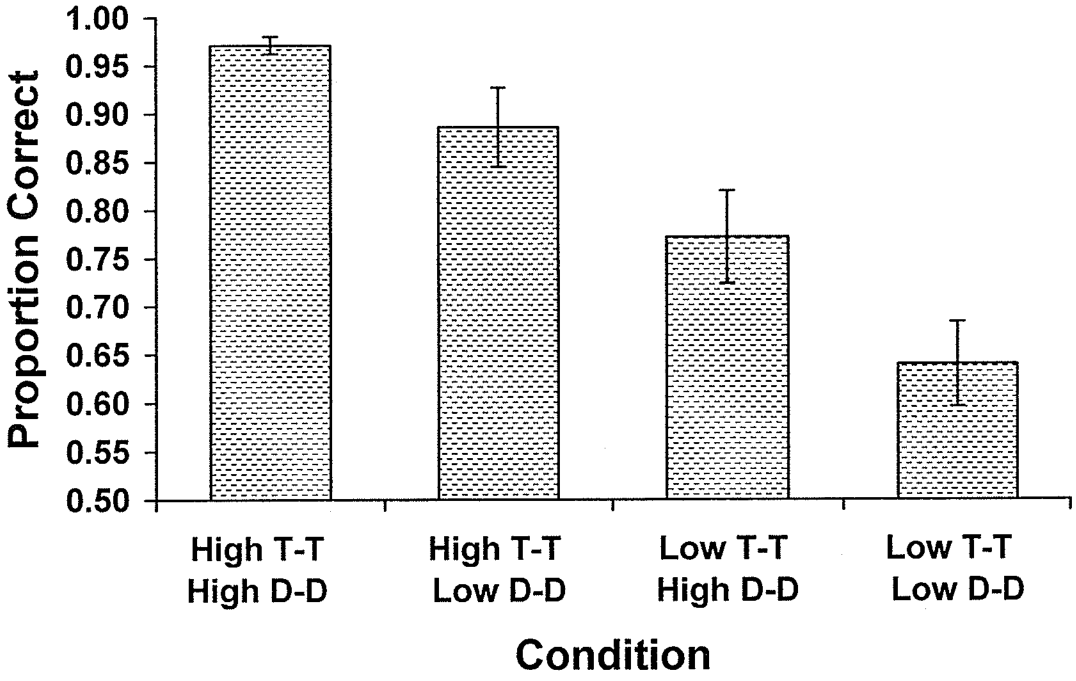
Signal detection analysis was once again used to evaluate performance. Separate two-way between-subjects ANOVAs were used to calculate d' and β using false alarm rates based on target-absent trials and on target-foil trials (see Table 2). When target-absent false alarm rates were used in the calculation of d', both the main effect of target-target similarity, $F(1, 36) = 26.45, p < .001$, and of distractor-distractor similarity, $F(1, 36) = 12.5, p = .001$, were significant. The interaction was not significant, $F < 1$. Similarly, when target-foil false alarm rates were used in the calculation of d', both the main effect of target-distractor similarity, $F(1, 36) = 29.97, p < .001$, and the main effect of distractor-distractor similarity, $F(1, 36) = 9.93, p < .01$, were significant. The interaction was not significant, $F < 1$. Since the results of the d' analysis are consistent with the initial ANOVA for target-present trials, it is evident that response bias on the part of the listeners was not a contributing factor, thus reinforcing the original interpretation.

Separate ANOVAs revealed that β, calculated using either target-absent or target-foil error rates, did not vary significantly across conditions ($p > .05$ in all cases). This result indicates that there was not a tendency for the listeners to respond differently across conditions. Using false alarms from target-absent trials, the mean β value was 3.8. When β was calculated using false alarms from target-foil trials, the mean β value was 2.74. These results indicate that there was a tendency for the listeners to respond that the target was not present.

Table 2: Performance for each condition in Experiment 2 is described in terms of proportion correct on target-present, target-absent, and target-foil trials. d' and β were calculated using false alarms from both target-absent trials and from target-foil trials. Standard errors for each measure are given in parentheses.

	Target-Present	Target-Absent	Target-Foil	d' (Target-Absent)	β (Target-Absent)	d' (Target-Foil)	β *(Target-Foil)*
High T-T	.971	.992	.897	4.27	1.93	3.67	1.39
High D-D	(.009)	(.008)	(.051)	(.194)	(.511)	(.271)	(.589)
High T-T	.886	.955	.895	3.36	2.85	3.00	2.27
Low D-D	(.041)	(.026)	(.042)	(.379)	(.665)	(.366)	(.710)
Low T-T	.772	.974	.873	2.90	5.96	2.30	5.00
High D-D	(.048)	(.010)	(.044)	(.269)	(1.42)	(.258)	(2.08)
Low T-T	.640	.847	.692	1.75	4.46	1.04	2.31
Low D-D	(.044)	(.065)	(.063)	(.290)	(1.29)	(.309)	(1.25)

Figure 2. Mean proportion correct for target-present trials of Experiment 2 as a function of condition.



It is clear from examining both Table 2 and Figure 2 that target-target similarity plays an important role in our ability to detect a target. Listeners were significantly better at detecting the targets in the two high T-T similarity conditions than they were in the two low T-T similarity conditions. However, as in Experiment 1, distractor-distractor similarity also appears to be a factor, with listeners in the two high D-D similarity conditions performing better than listeners in the two low D-D similarity conditions (when T-T similarity is held constant). Target-target similarity appears to have affected listener's abilities to detect the target more so than did distractor-distractor similarity. A calculation of the effect size on target-present trials confirmed this observation. It revealed that 47.7% of the variability in accuracy was accounted for by target-target similarity, $\eta^2 = .477$, while 17.9% of the variability in accuracy was accounted for by distractor-distractor similarity, $\eta^2 = .179$. Only 1.1% of the variability in accuracy was accounted for by the interaction, $\eta^2 = .011$.

Discussion

It was anticipated that search would be more difficult when target-target similarity was low, and this was indeed the case. Listeners more accurately detected the target in conditions with high target-target similarity than in conditions with low target-target similarity. This result is consistent with AET, which states that a more elaborate or complex target template (as would be required when target-target similarity is low) should be associated with more difficult search than would a simpler template (as would be sufficient when target-target similarity is high). According to AET, it becomes more difficult to reject sounds at the template as template complexity is increased. The reason that a more elaborate template leads to more difficult target detection is because of the input-template matching process. If each sound in a sequence is matched against a less elaborate template, then distractor sounds may be rejected more quickly than if a more elaborate template is used. This effect of target-target similarity may be apparent because there are more individual frequencies that exist between the two targets when target-target similarity is low. When target-target similarity is high, relatively few sounds could reach a frequency in between the two target frequencies, thus a less elaborate target template is necessary to ensure that a sound is not one of the targets. On the other hand, when target-target similarity is low, a more elaborate template is necessary to detect the targets. This is because if a sound is presented at a higher pitch than one of our targets, it must still be determined whether or not it is a lower pitch than our other target. This issue is less likely to be a problem when target-target similarity is high.

The design of this experiment also afforded an opportunity to replicate the effect of distractor-distractor similarity apparent in Experiment 1. Recall that in Experiment 1, performance was significantly better in conditions with high distractor-distractor similarity than in conditions with low distractor-distractor similarity (when target-distractor similarity was held constant). This effect was replicated in Experiment 2 demonstrating again the importance of distractor-distractor similarity as a factor influencing auditory search.

Once again, the results of this experiment appear to be beyond any explanation that the most recent version of FIT has provided. It stands to reason that FIT would predict that there would be no significant difference between high T-T and low T-T similarity conditions. According to FIT, as an object is processed, its features activate locations on feature maps where each particular feature is likely to occur. Focal attention is directed to the locations where features of the target are likely to occur, and it is only after this focal attention is directed to a specific location that the features of an input can be combined to yield either a target or a distractor (Triesman & Gelade, 1980). Considering the nature of FIT, it shouldn't make any difference whether or not the targets are similar to each other. So long as T-D similarity is held constant, if any input were to activate a feature map, focused attention would be directed to that input, regardless of its degree of match to another potential target. Thus, according to FIT, there would be no reason to suspect differences between High and Low T-T similarity conditions. In addition, the results of Experiment 2 replicated those of Experiment 1 in terms of D-D similarity, with listeners performing better in conditions with High D-D similarity than on conditions with Low D-D similarity (when T-D and T-T similarity were held constant). Once again, this finding is not likely to have been predicted by FIT.

It is interesting to note that with regard to false alarms, a main effect of distractor-distractor similarity on target-absent trials was found in Experiment 2 but not in Experiment 1. The significant effect of distractor-distractor similarity likely arose because the targets and the distractors were, on average, more similar to one another in Experiment 2 than they were in Experiment 1. Whereas in Experiment 2, the average frequency difference between the targets and the distractors was held constant at 375 Hz, in Experiment 1 this difference varied from 95 Hz to 1875 Hz in different conditions (with an average frequency separation of approximately 1300 Hz, much less than in Experiment 2). In addition the average spatial separation between the target and distractors was held constant at 15° in Experiment 2, but varied from 15° to 45° in Experiment 1 (with an average spatial separation of approximately 40°). The fact that overall target-distractor similarity was greater in Experiment 2 than in Experiment 1 would very likely increase the likelihood of an effect of distractor-distractor similarity emerging because the higher the target-distractor similarity, the more a distractor would activate a target template. Thus, if there is more activation of the target template, then spreading suppression due to weight linkage becomes an increasingly important factor in target detection (and distractor rejection). That is, when similar distractors are occurring in succession, subsequent distractors can be rejected more quickly. Whereas when target-distractor similarity is lower, a target is far more likely than a distractor to activate a target template and will more easily gain access to ASTM, leading to a reduced role for spreading suppression, in this case on target-absent trials. This factor is quite likely the reason why an effect of distractor-distractor similarity was evident on target absent trials of Experiment 2, but not Experiment 1. The fact remains, however, that distractor-distractor similarity played a significant role in the detection of a target when it was present in both of the experiments reported in this paper.

General Discussion

The experiments described in this thesis were designed to improve our understanding of the auditory search process and the factors that influence it, and to provide clues that can be utilized in the development of a general theory of auditory search. The results of these experiments demonstrate that our ability to perform an auditory search task is largely dependent on the similarity between the sounds that we encounter and the targets sound(s) for which we are listening. The most important factor that influences our ability to detect a target appears to be the degree of similarity between the target and the distractors. This was evident in Experiment 1 in which search was accomplished more accurately in the two conditions with low target-distractor similarity than in the two conditions with high target-distractor similarity. If distractor-distractor similarity were the more important factor, then performance in the two high distractor-distractor similarity conditions would have been more accurate than in the two low distractor-distractor similarity conditions. This was, of course, not the result that was obtained (recall that η^2 for the effect of target-distractor similarity was .384, whereas η^2 for the effect of distractor-distractor similarity was .084).

It is clear, however, that auditory search is also influenced by the degree of similarity between the distractors and each other. This was evident in both Experiment 1 and Experiment 2, in that targets were detected more accurately when distractor-distractor similarity was high than when distractor-distractor similarity was low.

Finally, the results of Experiment 2 showed that the degree of similarity between two potential targets also influences the accuracy of auditory search. Thus, in Experiment 2, greater similarity between the two targets was associated with better performance. Target-target similarity also appeared to be a more important factor than distractor-distractor similarity in determining a listener's ability to detect a target in Experiment 2. The effect sizes for target-target similarity was larger than for distractor-distractor similarity, $\eta^2 = .477$ and .179 respectively.

Although more research is necessary before the precise details of a general theory of auditory search can be developed, the results of this study provide support for AET in terms of its applicability to auditory events. The theory appears to interpret many of the findings that we have observed in auditory research accurately and fits quite reasonably with the tasks that we encounter in normal audition.

An initial attempt to utilize modified aspects of AET and apply them to audition should begin with the concept of a target template, and describe how such a mental representation of the target(s) is developed. From anecdotal evidence obtained by observation of participants in these experiments, and pure common sense, it can be assumed that listeners develop some sort of mental representation of the target(s) the first time they encounter it (them) in isolation from other sounds. This representation, however, is likely subject to change once the listener is exposed to the other sounds (i.e., distractors) in the sequence. For instance, if the target were a 1000 Hz sound, an initial template for this target would be created and stored in memory upon hearing the sound independently from the distractors. If, however, all of the other sounds in the sequence were 4000 Hz or greater, the listener's template could be altered to simply specify any sound lower than 4000 Hz as the target. On the other hand, if the distractors that were present with that same 1000 Hz target were 4000 Hz and 200 Hz, a more precise representation of the target would be necessary for accurate identification. That is, the template would have to specify a sound that is lower than 4000 Hz but higher than 200 Hz as the target. Thus, listeners may exert top-down control over selection (i.e., they may have a prespecified mental representation of the target), however it is likely that the parameters of the attentional template are also set up in a bottom-up fashion based on the nature of the distractor sounds in the sequence (see Mondor, Zatorre, & Terrio, 1998).

It is clear that the notion of a target template can be adapted from AET and applied to a theory of auditory search. The complexity of the template can help to explain the effects of target-target similarity that were reported in Experiment 2. The more complex or elaborate the template is (complexity is increased as target-target similarity is decreased), the more difficult it becomes to determine whether or not an input can be rejected at the template. This point leads directly to another aspect of a general theory of auditory search that can be adopted from AET; the idea of input-template matching. This concept can be used as a means to explain the findings of T-D similarity reported in Experiment 1 and throughout other auditory studies (e.g., Mondor, Giannuzzi, & Thorne, 2003; Mondor, Zatorre, & Terrio, 1998; Woods et al., 2001). As each sound in a sequence enters the ear and is processed, it is given a weight based on its similarity (i.e., degree of match) to the target template. Sounds that are very similar to the target are given a greater weight than sounds that are more dissimilar from the target. Access to auditory short-term memory (ASTM; Baddeley, 1986, 1990) is then determined based on the weights of the inputs. If a sound is given a weight sufficient enough to warrant further investigation (i.e., it shares similar features with the target), then it is passed on to ASTM. It is only once an input has entered ASTM that a positive response (i.e., acceptance as the target) can be made (though negative responses can be made at both the template and after gaining access to ASTM). Since only sounds that are given a sufficient weight are passed on to ASTM, it becomes more difficult to reject distractors as target-distractor similarity is increased due to an increased number of inputs gaining access. Alternatively, target-distractor similarity can be so low that only the target would gain access, making for an extremely efficient search in which only the target is passed on to ASTM.

It should be noted that the conjunction of frequency and location information occurs quite rapidly in audition. Woods, Alho, and Algazi (1994) found evidence using evoked potentials generated during a selective listening task that the conjunction of these features occurs even before the independent analysis of the individual features is complete. Also, as previously mentioned, Mondor, Zatorre, and Terrio (1998) reported that a listener cannot ignore one feature of a sound (e.g., the frequency or location), and rely solely on the other feature to guide attention. Duncan and Humphreys (1989, 1992) proposed that search will be accomplished using the simplest attentional template possible (i.e., one dimension such as colour, if possible). However, in an auditory search task in which the dimensions of both frequency and location are present in the target, it appears that, while listeners could potentially simplify their target template, they are constrained by the fact that the attentional template must be defined by both frequency and location information (Mondor et al., 1998).

This concept of input template matching can also help to explain the findings of target-target similarity that were reported in Experiment 2. If the two targets are very dissimilar from one another, then the distractors in the experiment have a greater chance of activating some of the features of the target template. That is, if one target is higher pitched and the other target is lower pitched, then a low-pitched distractor will produce relatively little activation of the high-pitched features of the target template, but it will produce a stronger activation of the low-pitched features of the target template. If, however, the two targets are very similar to one another, a distractor that does not activate the high-pitched features of the template will likely not activate the low-pitched features of the template (because target-target similarity is high). Thus, as was demonstrated in Experiment 2, listeners will tend to have an easier time rejecting distractors when target-target similarity is high than when it is low.

Finally, the concepts of weight linkage and spreading suppression can be altered slightly from their original definitions (Duncan & Humphreys, 1989) to understand why it is easier to identify the presence of a target when distractor-distractor similarity is high as compared to when it is low (as demonstrated in Experiments 1 and 2 of the present study). Recall that the original concepts of weight linkage and spreading suppression were applied to a visual array in which all of the stimuli were presented simultaneously to the viewer. When items in the array shared similar features they were given similar weights, and the process of attaching a weight to each item occurs more quickly when the items are similar to one another (weight linkage). If numerous items occur in close spatial proximity, and share a similar weight, then they can be rejected as a group, once again speeding up rejection of non-target items (spreading suppression). This study demonstrates that a similar process could be occurring in audition, even though the presentation of the stimuli is sequential rather than simultaneous. When similar items (i.e., with respect to their frequency, location, or both) occur in sequential order, it may be the case that they are given similar weights, based on the degree of match to the target template, more quickly than if they are dissimilar from one another, in a process that is parallel to AET's weight linkage. Thus, similar items occurring sequentially in audition would each be given a weight faster than would dissimilar items. If similar items occurring sequentially were given similar weights, then the rejection of the first item could lead to a faster rejection of subsequent similar items. When distractor-distractor similarity is high, there is more likelihood of similar distractors occurring sequentially, which increases the likelihood that spreading suppression due to weight linkage will occur. When distractor-distractor similarity is low, it is less likely that successive sounds will have similar features, thus weight linkage is less likely to occur (individual sounds must be evaluated independently of one another). In this case, spreading suppression will be at a minimum, because sequential items are less likely to share a similar weight, and this should lead to a less efficient search. The question, however, is precisely how the processes of weight linkage and spreading suppression occur in audition?

First, it must be noted that the processes of weight linkage and spreading suppression in audition must occur before access to ASTM is allowed, otherwise, if all sounds were given access to ASTM there would be no (or relatively little) difference in performance between high and low distractor-distractor similarity conditions. The answer to this question may lie in evidence from auditory scene analysis, whereby the effects of distractor-distractor similarity on target detection could be caused by a preattentive grouping mechanism that assigns information to streams based on the likelihood that the information originated from the same source (see Bregman, 1978, 1990, 1993, for a discussion on primitive stream segregation). For instance, Mondor and Terrio (1998) theorized about the possibility of such a preattentive perceptual process as a potential explanation for their findings that a target which deviated most from the pattern structure of a sound sequence was easiest to detect. In a series of experiments, the researchers had listeners determine whether or not a target was present in a series of sounds that ascended or descended in frequency by equal log units (from 500 Hz to 1500 Hz). It was found that listeners were, in general, faster and more accurate at detecting the target the more it deviated from the pattern structure. Listeners performed poorest when the target was consistent with the pattern structure (i.e., when the target was the expected frequency in the ascending or descending pattern). Thus, it would appear that this preattentive perceptual grouping process had a significant effect on the selection of a target amongst distractors. According to Mondor and Terrio (1998), this grouping mechanism operates by evaluating each tone with regard to the likelihood that it is a member of a larger pattern, and target detection is affected by the stream tag that it is assigned to each item. If a target is assigned a tag suggesting that it is part of a pre-existing stream, then the target loses its salience and it becomes more difficult to detect. This proposal is also consistent with attentional engagement theory in that it is assumed that perceptual organization occurs before selection. This account of Mondor and Terrio's (1998) findings could be used to explain the effects of distractor-distractor similarity that were reported in this study and to understand the process by which weight linkage and spreading suppression may operate in audition. For instance, a preattentive grouping mechanism could have tagged each input, on the trials of both experiments of this study, as to its probable membership in a stream. If sounds that occurred sequentially were quite similar to one another (either on the dimensions of location, frequency, or both), then they could be tagged as members of the same stream and essentially grouped together. After a weight is assigned to the first sound in the group, a similar weight could be assigned to each subsequent sound fairly quickly based largely on the pre-existing tag. If the weight of the first sound in the group were such that it did not warrant passage on to ASTM, then it would be suppressed as a non-target, as would be any subsequent sound of a similar weight, at a very efficient speed (provided they were all members of the same group). It follows from the evidence of Mondor and Terrio's (1998) study that a target that is not tagged as part of this group should be detected more quickly and accurately than one that was assigned membership to a group. In fact, if a target were tagged as a member of a group (i.e., if T-D similarity were high), this would have a very detrimental effect on search, as it did in Mondor and Terrio's (1998) study.

Applying this explanation to the present study, it would be predicted that participants in the Low T-D/High D-D similarity condition of Experiment 1 would more accurately detect targets than those who participated in the HighT-D/High D-D condition. This was, of course, the case. If, however, distractor-distractor similarity were low, then there would be less of a chance that this grouping mechanism could tag distractors as part of a stream, thus reducing or eliminating the effect of spreading suppression. This would mean that listeners should be more accurate in the Low T-D/High D-D similarity condition than in the Low T-D/Low D-D similarity condition, which was again a result that was obtained in Experiment 1. Worst performance would still be predicted in High T-D/Low D-D similarity conditions because in this condition, as compared to the High T-D/ High D-D condition, there is much less probability that streaming and thus spreading suppression of similar distractors would occur. Thus, more independent sounds must be assigned a weight in the High T-D/Low D-D condition (of which more will be assigned a similar weight to the target), and more are likely to be allowed access to ASTM. This would result in a very inefficient and error prone search, as was the case for this condition on Experiment 1.

Thus, to recap, in an auditory search task, a listener begins by creating a target template against which inputs are matched. This template can be adjusted depending on the nature of the distractor sounds that a listener is likely to encounter. The complexity of the template can be affected by factors such as the similarity between two potential targets. The greater the similarity between the targets, the more complex the template must be to detect them. A more complex template requires a more thorough evaluation of inputs at the template, thus leading to a less efficient search. Inputs are matched against the template and assigned a weight based on their similarity to the target. Inputs with lower weights are rejected at the template while inputs with a greater weight are given access to ASTM. During the initial search process, however, a preattentive grouping mechanism (Mondor & Terrio, 1989) is responsible for tagging inputs with respect to the likelihood that they originated from the same source (i.e., belong to a sound stream). Inputs that likely belong to the same stream can be rejected as a group in a process akin to spreading suppression via weight linkage, which was originally described in AET (Duncan & Humphreys, 1989). Inputs that do not belong to a stream must be independently matched against the target template and this process takes longer and will typically result in a less efficient, more error prone search. Therefore, search is most efficient and accurate when a target is given a weight sufficient enough (relative to the distractors) that it is the only input gaining access to ASTM (i.e., when T-D similarity is very low) and becomes less efficient as more inputs are allowed access (i.e., as T-D similarity is increased). Once again, access to ASTM is highly competitive and based on the weight that is assigned to each input. However, distractors that share similar weights to the target may be allowed access to ASTM. For this reason, a second selection process occurs in ASTM whereby each input can be either rejected or accepted as the target. In addition, search efficiency is affected by the ability to reject multiple stimuli belonging to the same group. Target detection should be best as the likelihood that this grouping process will occur is increased (i.e., as D-D similarity is increased), and should worsen as the likelihood that this grouping of similar distractors will occur is decreased (i.e., as D-D similarity is decreased).

Thus, this study serves as an initial attempt to accurately describe the auditory search process in a controlled environment. While more research is necessary to apply this paradigm to environments outside of the laboratory, the findings and explanations of the present study appear to be promising.

Future Applications of this Research

As a final note regarding the importance of this and other related studies, it is evident that research of this nature has numerous real-world applications. These applications include the development and implementation of effective warning alarms for use in settings such as airplane cockpits and hospital operating rooms. In these environments, we are surrounded by many acoustic distractions, and the failure to detect an alarm quickly and accurately could result in the loss of life. Alarms can be developed that take advantage of the information gained in this and other related studies to ensure that they are maximally distinct from other environmental sounds (such as noises that are occurring normally in an airplane cockpit) that might otherwise tax necessary resources. For example, in an F-18 cockpit, it would be beneficial to create warning alarms that are of a different pitch than the normally occurring sound in that particular environment (e.g., the frequency of the voices coming through on the radio). One particular strategy that has received some research attention is the use of a female voice as a warning alarm (Tysiaczny, 2005). There are relatively few female F-18 pilots, thus most of the radio communication that a pilot encounters during flight are male voices. A female voice (obviously higher pitched than the typical male voice) should be more distinct and thus capture attention better. As more females enter the cockpit, however, this female voice could potentially become less effective as a warning alarm system. Perhaps manipulating the female pilots voices to sound more masculine on the radio would be a practical solution as the gender gap is reduced for this particular occupation.

The effects of target-target similarity can also be transferred to the fighter jet cockpit. For instance, in a situation where a pilot is flying too low, the warning message might say, "Altitude, altitude, altitude." The pilot's response would be to pull back on the stick to increase his/her altitude. If there were another instance in which the response of the pilot was intended to be the same (i.e., the aircraft was too low to fire on the designated target), then it would be beneficial to use the same warning message (i.e., "Altitude, altitude, altitude.") versus a different message (i.e., "Too low, too low, too low."). The results of this paper have demonstrated that the more similar the targets (when the response to the presence of either of them is the same) the more likely it is that a listener will be able to detect the target and make the correct response. On the other side of the coin, if the desired response to each target were to be different (i.e., if in Experiment 2 the task was to have listeners distinguish between the two targets), it could be anticipated that the more distinct the potential targets were from one another, the more likely it is that they could be distinguished, and the more accurate the responses should be. In a fighter jet cockpit for instance, if one warning message was intended to indicate that the pilot was flying too low, and the other that the pilot was flying too high, the message, "Altitude, altitude, altitude" would not be the best choice for both circumstances. A more specific warning such as, "Pull up, pull up, pull up", in the former situation, and, "Too high, too high, too high", in the latter situation would likely be more effective.

An experiment that could provide useful information regarding the applicability of this research to human speech might involve the use of nonsense words (to minimize the effects of word recognition; e.g., ‘gop’, ‘nop’, ‘tez’, ‘bez’, etc.) in conditions similar to those of Experiment 1 of the present study. The participants could be instructed to detect the presence of a target nonsense word (e.g., ‘tor’), and the target-distractor and distractor-distractor similarity could be manipulated to determine the extent to which these factors apply to human speech sounds. For example, a low target-distractor, high distractor-distractor similarity condition could involve distractors such as ‘gaz’ and ‘daz’, whereas a high target-distractor, low distractor-distractor similarity condition could involve distractors such as ‘tur’ and ‘zor’. If the results were similar to those reported in Experiment 1 of the present study, then it could potentially be useful in developing the most efficient voice operated warning alarm systems. For instance, all essential words used in F-18 pilot communication could be analyzed, and words that are most dissimilar from the commonly used words or phrases could be incorporated and used for emergency purposes only. This system would likely lead to a better recognition of emergency situations and a faster reaction time on the part of the pilots.

Thus, while the development of a general theory of auditory search is still in its early stages, it is clear that the findings of the experiments in this study should be taken into consideration when creating such a theory. It is also clear that research of this nature can be applied to countless situations beyond the experimental setting.

References

Adobe Systems Incorporated. (2005). Adobe Audition 1.5. San Jose, CA.

Baddeley, A. (1986). *Working memory.* Oxford, England: Clarendon Press.

Baddeley, A. (1990). *Human memory: Theory and practice.* Boston: Allyn & Bacon.

Beck, J., & Ambler, B. (1973). The effects of concentrated and distributed attention on peripheral acuity. *Perception & Psychophysics, 1,* 300-302.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature (London), 303,* 696-698.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII* (pp. 63-75). Hillsdale, NJ: Erlbaum.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 10-36). Oxford, England: Clarendon Press.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics, 33,* 487-493.

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

Donk, M. (1999). Illusory conjunctions are an illusion: The effects of target-nontarget similarity on conjunction and feature errors. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1207-1233.

Donk, M. (2001). Illusory conjunctions die hard: A reply to Prinzmetal, Diedrichsen, and Ivry (2001). *Journal of Experimental Psychology: Human Perception and Performance, 27,* 542-546.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 578-588.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Garner, W. R. (1987). Location and color as cuing dimensions in contingent classification. *Perception & Psychophysics, 41,* 202-210.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1,* 225-241.

Houck, M. R., & Hoffman, J. E. (1985). Conjunction of color and form without attention: Evidence from an orientation-contingent color after effect. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 186-199.

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General, 118(3)*, 258-279.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1985). Interactive processes in perceptual organization: Evidence from visual agnosia. In M. I. Posner & O. S.

M. Martin (Eds.), *Attention and Performance XI* (pp. 301-318). Hillsdale, NJ: Erlbaum.

Jamieson, R. K., Thompson, W. F., Cuddy, L. L., & Mewhort, D. J. K. (2003). Do conjunction errors in auditory recognition imply feature migration? *Canadian Journal of Experimental Psychology, 57(2)*, 125-130.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society London B, 200,* 269-294.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature (London), 332,* 154-155.

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 398-414.

Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics, 56,* 268-276.

Mondor, T., Giannuzzi, T., & Thorne, L. (2003). Voluntary guidance of auditory attention by frequency and location: The importance of discriminablility. Unpublished manuscript.

Mondor, T. A., & Terrio, N. A. (1998). Mechanisms of perceptual organization and auditory selective attention: The role of pattern structure. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1628-1641.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 387-409.

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 66-79.

Näätänen, R., Porkka, R., Merisalo, A., & Ahtola S. (1980). Location vs. frequency of pure tones as a basis of fast discrimination. *Acta Psychologica, 44*, 31-40.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264-265.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception & Psychophysics, 41,* 191-201.

Prinzmetal, W., Diedrichsen, J., & Ivry, R. B. (2001). Illusory conjunctions are alive and well: A reply to Donk (1999). *Journal of Experimental Psychology: Human Perception and Performance, 27,* 538-541.

Psychology Software Tools, Inc. (1999). E-Studio Software System. Pittsburgh, PA.

Spieth W., Curtis J. F., & Webster J. C. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America, 26,* 391-396.

Steinman, S. B. (1987). Serial and parallel search in pattern vision. *Perception, 16,* 389-399.

Syntrillium Software Corporation. (1999). Cool Edit Software System. Scottsdale, AZ.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 652-676.

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

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

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459-478.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107-141.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114,* 285-310.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI.* (pp.333-361). Hillsdale, NJ: Lawrence Erlbaum.

Tysiaczny, M. J. (2005, July). Personal communication with the former Deputy Director General of the Defence and Civil Institute of Environmental Medicine, Toronto.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202-238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the modified feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419-433.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 65-74.

Woods, D. L., Alain, C., & Ogawa, K. H., (1998). Auditory and visual feature conjunction during high-rate serial presentation: Processing two features can be faster than processing one. *Perception & Psychophysics, 60,* 239-249.

Woods, D. L., Alho, K., & Algazi, A. (1994). Stages of auditory feature conjunction: An event-related potential study. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 81-94.