

2003

Schema-based segregation in the cocktail party situation and its effects upon two differing types of auditory masking.

Shawn M. Johnson
University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Johnson, Shawn M., "Schema-based segregation in the cocktail party situation and its effects upon two differing types of auditory masking." (2003). *Masters Theses 1911 - February 2014*. 2403.
Retrieved from <https://scholarworks.umass.edu/theses/2403>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.



UMASS/AMHERST

312066 0288 0973 6

SCHEMA-BASED SEGREGATION IN THE COCKTAIL PARTY
SITUATION AND ITS EFFECTS UPON TWO DIFFERING TYPES
OF AUDITORY MASKING

A Thesis Presented

by

SHAWN M. JOHNSON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

February 2003

Graduate Program of Psychology
Division II: Cognitive, Developmental and Educational Psychology

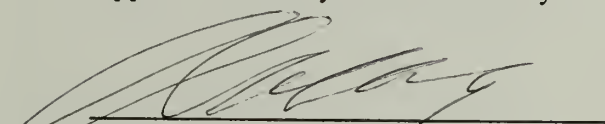
SCHEMA-BASED SEGREGATION IN THE COCKTAIL PARTY
SITUATION AND ITS EFFECTS UPON TWO DIFFERING TYPES
OF AUDITORY MASKING

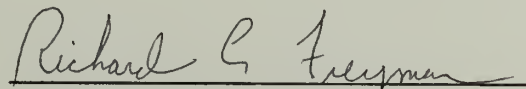
A Thesis Presented

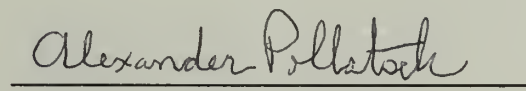
by

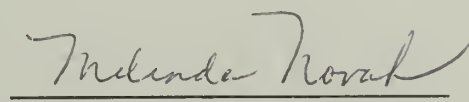
SHAWN M. JOHNSON

Approved as to style and content by:


Charles Clifton Jr., Chair


Richard Freyman, Member


Alexander Pollatsek, Member


Melinda Novak, Chair
Psychology Department

CONTENTS

	Page
LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
CHAPTER	
INTRODUCTION.....	1
1. EXPERIMENT 1.....	8
Design.....	9
Participants.....	9
Apparatus and Procedures.....	10
Experimental Design.....	11
Results and Discussion.....	14
2. EXPERIMENT 2.....	18
Design.....	25
Participants.....	25
Apparatus and Procedures.....	25
Experimental Design.....	26
Results and Discussion.....	29
3. EXPERIMENT 3.....	31
Design.....	31
Participants.....	31
Apparatus and Procedures.....	31
Experimental Design.....	32
Results and Discussion.....	33

4. EXPERIMENT 4.....	39
Design.....	39
Participants.....	39
Apparatus and Procedures.....	39
Experimental Design.....	40
Results.....	43
GENERAL DISCUSSION.....	47
APPENDIX : WORDS USED AS STIMULI (SEPARATED BY TALKER).....	50
BIBLIOGRAPHY.....	53

LIST OF TABLES

Table	Page
1. P(c) Means and Marginal Means for Experiment 1.....	17
2. P(c) Means and Marginal Means for Experiment 2.....	30
3. P(c) Means and Marginal Means for Experiment 3.....	34
4. P(c) Means and Marginal Means for Experiments 2 and 3	35
5. P(c) Means and Marginal Means for Experiment 4.....	44
6. Mean D-prime and Criterion (in parentheses) for Experiment 4.....	45
7. Reaction Time Means and Marginal Means for Experiment 4 (in ms.).....	46

LIST OF FIGURES

Figure	Page
1. Trial assembly process for the various conditions of Experiment 1.....	14
2. Trial assembly process for the various conditions of Experiment 2	28
3. Trial assembly process for the various conditions of Experiment 4	42

INTRODUCTION

One of the most striking properties of the human auditory system is its ability to process the complex acoustic waveform present at each of the ears and obtain the more simplified and distinct streams and objects (after Bregman, 1990) which compose the auditory percept. When walking down the street, one is able to perceive distinct and separable sounds which are identifiable as 'a bird', 'a car', 'children playing' and so forth, despite the fact these signals are relatively confounded upon entering the ear. When listening to a piece of music, we do not necessarily perceive a solidified and impenetrable din of sound created by all the instruments. Each individual instrument can be attended to as a particular and intelligible component of the piece as a whole. Finally, when in a crowded room, one does not hear the unified product of all the voices as a massive and singular babbling. Instead, we are capable of perceiving multiple talkers. We seem to be able to choose or select individual voices, such as a conversational partner or eavesdropping target, to attend to and perceive.

The final example as listed above should be immediately familiar to any individual who has taken an introductory course in cognitive psychology as the 'cocktail-party problem' (or 'phenomena'). Cherry (1953) coined this term and was the first to take note of this phenomena, asking, "(H)ow do we recognize what one person is saying when others are speaking at the same time?" (p. 976). After noting that, among other things, various low-level factors in the speech signals such as spatial separation and pitch certainly contribute to our ability to segregate speech, Cherry observes that we are still able to perform this task in the absence of such cues. Participants listened to a

(presumably monaural) recording of two simultaneous messages being spoken by the same talker and were asked to repeat, or shadow only one of the messages present. Despite the absence of any spatially-derived cues or any gross variance in pitch or talker characteristics, participants were able to perform this task moderately well after prolonged exposure to the stimuli. The errors that participants did make exhibited a general tendency to maintain syntactic and semantic consistency with portions of the message that were correctly shadowed. This led Cherry to conclude that relatively higher-order mental processes, or as Cherry calls them, “Transition-probabilities (subject matter, voice dynamics, syntax....)” (p. 976), were at least partially driving the participants’ ability to segregate simultaneous speech messages.

After establishing the involvement of higher-order mental processes in the resolution of the cocktail party problem, Cherry (1953) dedicated the remaining portion of the article to situations in which one of the low-level factors previously discussed is able to provide a cue for segregation. More specifically, Cherry investigated the effect of delivering two speech messages to a listener, with one message being sent only to the left ear and the other being sent only through the right. While the importance of low-level cues at the auditory periphery in the resolution of the auditory scene should not be underestimated, this direction of Cherry’s research had the unfortunate effect of associating the cocktail party problem almost exclusively with low-level auditory factors. In a review of literature dedicated the investigation of the cocktail party problem, Bronkhorst (2000) presented an overview of the ways in which various factors have been found to affect the speech reception threshold for speech in the presence of competing

speech (p. 125). Of the nine factors presented (spectral differences, fluctuations, voice similarity, spatial separation [single source], spatial separation [multiple sources], best ear v. binaural, reverberation, divided attention, and moderate hearing impairment), only one, divided attention, would seem related to the sort of higher-level processing established by Cherry to play a role in speech-on-speech resolution. Broadbent's (1958) theory of attention, possibly the most influential body of research to arise from Cherry's work, also concerns itself with the low-level physical features of signals which can provide a basis for subsequent "filtering". Despite the connotations generally given to "attention", Broadbent provides a notable lack of experimentation designed to investigate the higher-order mental processes which must account for, at least some of, our ability to segregate auditory scenes and resolve the cocktail party problem.

The engineering/information-processing backgrounds of both Cherry and Broadbent may have been responsible for their tendency to deal with auditory segregation almost exclusively in terms of the physical properties of the signal. However, other researchers have not been so hesitant to theorize about the effects higher mental processes may have upon our ability to segregate an auditory scene. In particular, Bregman (1990) has formulated a wide-reaching theory, largely informed by the work of the Gestalt psychologists, that concerns itself with the effects that higher-order mental processes have in the formation of our auditory perceptions. According to Bregman, many of the same Gestalt principles shown to organize our visual perceptions have

correlates in the auditory world. Just as our visual system sees:

O O O O O O

as two groups of three O's, rather than 6 disparate O's, our auditory system interpret a similarly grouped pattern of 6 simple tones as two triplets rather than six isolated and unrelated auditory events. This can be demonstrated, as Bregman points out, by equating the spatial distance, or proximity of the above O's with auditory separation in either frequency or time. Without any characteristics inherent in the peripheral auditory system or in the signal itself to cause such a grouping, it is reasonable to assume that it is our mind that "constructs" our perception of these distinct auditory objects.

The Gestalt law of continuation is another principle generally applied to visual perception which Bregman feels has parallels in the auditory realm. A group of three distinct pitch glides separated by silence will be perceived as one larger and unified pitch glide when the portions of silence between them are replaced with bursts of noise (Dannenbring, 1976 as cited by Bregman p. 28). Bregman points out the similarity between this phenomenon and certain tendencies of the visual system. In vision, two distinct lines both abutting a shape will generally be perceived as a single line which "continues" under the obstructing shape. Thus, the line formed by a building's roof is perceived as being continuous even when it passes behind an obstructing tree or utility pole. In both the auditory and visual examples, the mind seems to prefer to perceive continuous and unbroken objects, rather than broken and disjoint ones, even when there is no specific sensory evidence to support such a position.

Bregman views this tendency of the mind to segregate the acoustic input present at the ear into distinct objects and continuous streams as the impetus for our ability to make sense of what he calls the “auditory scene” (p. 3). He distinguishes between two separate types of stream segregation which can be performed upon the acoustic input at the peripheral auditory system. The first method is “primitive stream segregation” defined by Bregman as “simpler, probably innate and driven by incoming acoustic data” (p. 397). Though this may seem similar to Cherry and Broadbent’s view of auditory segregation being largely dependent upon low-level factors, primitive stream segregation is, for Bregman, qualitatively different. As Bregman states, “...we must introduce scene analysis as a preliminary process that groups the low-level properties that the auditory system extracts and builds separate mental descriptions of individual voices or non-vocal sounds, each with its own location, timbre, pitch and so on. Only then does it make sense to say that our attention can select a voice on the basis of one these qualities.” (p. 530-531). However, these low-level factors are not the only basis of our ability to segregate the auditory scene, according to Bregman. We are also able use “schema-based segregation” for certain specific types of auditory scenes. Bregman states that schema-based segregation “involve(s) the activation of stored knowledge of familiar patterns or schemas in the acoustic environment and of a search for confirming stimulation in the auditory input” (p. 397). Bregman also states that schema-based segregation processes can be distinguished from primitive segregation processes by its employment of both “voluntary attention” and/or “past learning” (p. 398).

The involvement of all the factors Bregman associated with schema-based segregation can be seen in an experiment performed by Deutsch (1972). Common folk-tunes were recorded with the following systematic distortion: each note was randomly moved up one octave, down one octave or remained the same. Thus, a tune like 'Mary Had A Little Lamb' would retain the note sequence of A-G-F-G-A-A-A, but the absolute relationship between the pitch of the notes would be destroyed due to the random-octave distortion. Deutsch found that participants listening to these distorted tunes were unable to recognize them until they were told what tune to listen for, at which point the tunes became perceptually salient. One can see how each of the factors Bregman associated with schema-based segregation would allow for this phenomenon. The active attention of a listener is certainly required to perceive an auditory stream which was previously unclear and participants' past experience or learning associated with a tune like 'Mary Had A Little Lamb' or the like allows for its search and subsequent activation in the auditory scene. Clearly this is very different from the sort of primitive segregation which might occur when one hears each shot of ongoing machine-gun fire as a stream of machine-gun fire based upon the temporal proximity and similar acoustic makeup of each of the shots.

Upon examination, it becomes clear that one type of auditory stimulus which should be particularly subject to schema-based segregation processes is human speech. The semantically and syntactically consistent errors that Cherry reported in his initial experimentation on the cocktail party effect are proof that our expectations and knowledge of language are particularly crucial in segregating speech from the auditory

scene. Bregman admits as much, citing the phonemic restoration effect (Warren 1970, Samuels 1981) as another example of the involvement of schema-based segregation in speech perception. However, Bregman chooses to devote the remaining space of his chapter on schema-based segregation to non-speech phenomena. Further, his chapter entitled "The Auditory Organization of Speech Perception" deals exclusively with primitive-based methods by which speech streams can be segregated.

The following experiments attempt to remedy this omission by Bregman. More specifically, they attempt to confirm some predictions which fall out of Bregman's discussion of schema-based segregation as applied to a cocktail party-like auditory scene. We will attempt to use priming as a method of making one particular speech stream more attentionally salient by increasing the activation of the stored forms present in the prime. At the same time we will attempt to minimize the number of primitive-based segregation and low-level physical cues available to the listener. In doing so, we hope to explore the characteristics of schema-based segregation, as defined by Bregman, in a speech-on-speech listening situation.

CHAPTER 1

EXPERIMENT 1

In the first series of experiments, participants were asked to listen to a composite of three monaurally-presented simultaneous female talkers, each of whom spoke a series of five or seven fairly common one syllable words. By using only monaural audio files, we hoped to eliminate the influence of spatial separation in speech-on-speech segregation. By using only female voices, we hoped to minimize the role which frequency-based segregation might play in a speech-on-speech listening situation. By using word strings, rather than proper sentences, we hoped to minimize the influence of linguistic knowledge in speech-on-speech segregation, which may tempt listeners to infer (whether consciously or unconsciously) the identity of certain obstructed signals. At the end of each trial, participants were asked to indicate which of two visually-presented similar words (such as “bomb” and “mom”) they had heard. On half the trials, the presentation of the three-speaker composite signal was preceded by an auditory preview of all the words to be spoken by the speaker who says the target word, with the exception of the target word itself. This condition will be known as the “prime” condition. On an orthogonal half of the trials, the composite signal began with two unobstructed words spoken by the target talker before any interfering speech begins. This condition was created to observe the influence of listeners being given a cue as to the target talker (i.e. which voice they should attempt to follow) aside from the influence of the prime (which it could be said provides a target-talker cue). This condition was known as “stagger”. Also, in order to eliminate any effect which might arise from repeated exposures to target

words in one particular region of the word strings, target position was divided in the trials between the 4th, 5th and 6th positions of the word strings. Assuming that the auditory preview (“prime”) increases the salience and/or distinctiveness of the stream that contains the target word, in all target positions accuracy of recognizing the target word should be higher in the prime condition than in a condition for which auditory preview is unavailable. Similarly, if the unobstructed initial words in the stagger condition permits attention to be directed to the target speech stream, accuracy should be higher in the stagger condition than in a condition without stagger. There is no clear basis for predicting whether these two effects should be additive or interactive. However, it should be noted that our condition of primary interest is the “prime” condition, which provides a situation which Bregman’s definition of schema-based segregation predicts increased perceptual salience and subsequent identification accuracy. Any increase in accuracy due to cuing a listener as to which talker to target would be present in both the prime and stagger conditions.

Design

Participants

24 undergraduate students, graduate students and post-doctoral fellows from the University of Massachusetts-Amherst were given course credit or cash in exchange for their participation in this study. Participants were native speakers of English with no known hearing difficulties. Prior to participation, participants read a instruction sheet detailing the particulars of the experiment and signed a form expressing their informed

consent to participate in the experiment. Subsequent to participation, participants received both written and oral feedback detailing the nature of the experimentation and contact information should they have any questions.

Apparatus and Procedures

Three female graduate students from the University of Massachusetts-Amherst each recorded 144 unique single-syllable words (see appendix A) culled from the California Consonant Test. These recordings were made in a sound proof booth using a headset microphone and a portable digital audio tape (DAT) recorder. These words were then transferred to PC and separated into individual 16-bit monaural *.WAV files with a 22,050 Hz sampling rate using Syntrillium's Cool Edit Pro program. Each word was then normalized using Cool Edit Pro to insure an approximation of level between words. Eighteen pairs of words judged to be acoustically similar to one another were then selected from each of the three speakers words. These words were designated to be "target" words and were confirmed to occur in the Francis and Kucera (1982) corpus of speech at a rate of 10 times per million or greater. Words were allowed to vary in length in accordance with the way they were naturally spoken (from 177 to 838 ms).

Participants listened individually to experimental stimuli in a Industrial Acoustics Company sound proof booth. Stimuli was delivered from a PC computer via a Sony STR-DE 135 stereo receiver through 2 Realistic Minimus 7 speakers. Sound level was adjusted to according to the comfort of listeners. Any level fluctuations caused by different volume settings in the receiver or minute participant placement differences within the booth would be affect all portions of the stimuli (both target and interfering

speech) equally due to their previously mentioned normalization. Participants initiated each trial with the pull of a trigger. Participants were then presented with a 7-word string of words spoken by the target talker, and simultaneously presented with either 5 or 7-word strings spoken by two masking talkers (5-word strings were offset by two words in comparison to the target talker to constitute the 'stagger' condition). These masking situations were either preceded by silence or the six non-target words spoken by the target talker constituting the 'prime' condition. Subsequent to the presentation of each masking situation, acoustically similar target pairs (of which one member had been in the trial) were presented as a forced choice at the end of each trial visually on a Amdek monochromatic computer monitor positioned approximately 1.5 feet in front of the participant. Answer choices were recorded by pulling the trigger on the button box corresponding to the position of the answer on the computer screen (i.e. right-hand target word selection was indicated by pulling the right-hand trigger and vice versa).

Experimental Design

A 2 X 2 X 3 randomized and counterbalanced within-subjects design was employed. As previously mentioned the three experimental variables were "prime" vs "no-prime", "stagger" vs "no stagger", and "position 4" vs "position 5" vs "position 6" . In the "prime" condition, the participant listened to all words, save the target word, that the target speaker would speak in the presence of competing speech before the competing speech was initiated. The "no-prime" condition provided no sort of preview prior to exposure to the test string. In the "stagger" condition the listener was allowed to hear the first two words of the target speaker's word string without interference before

the competing speech was introduced, while in the “no stagger” condition the target speaker and distractor speakers began synchronously. The “position” variable varied whether the target was presented in the 4th, 5th, or 6th position of the target speaker’s word string. On each trial a computer program determined the combination of conditions that would be presented. The program then randomly selected a target word from one of the talkers and six random non-target words (target words were never used as non-target words and vice versa). These words were used to create the target talkers word string. Five or seven non-target words (depending upon whether the “stagger” condition was present) were randomly selected from each of the two non-target speakers and used to construct the interfering word strings. All randomly selected words for both target and non-target strings were selected without replacement. Words in both target and non-target word strings were systematically separated by 100 ms of digitally created silence. Protocol was added to insure that the same word would not be selected twice for any given word string. Non-target words were only resampled after each word in the set had been used once. This procedure insured that both the combination of non-target words chosen and the way in which the three talkers word strings were synchronized (or unsynchronized) with one another was randomized from trial to trial. Figure 1 details this trial assembly process and the four conditions which might result if target position ‘4’ were selected. Just as the spaces between these written words are occasionally both synchronous and asynchronous with the spaces of other talkers, so were the spaces between the talker’s spoken words.

The experiment was administered in two separate blocks with an optional approximate 5 minute break for participants in between. In each block exactly half of each talker's target word pairs were presented. The remaining half of each talker's target pair were presented in the second block, resulting in 54 trials per block and 108 trials total.

Figure 1: Trial assembly process for the various conditions of Experiment 1.

TARGET WORD = "BOMB" IN POSITION 4; TARGET TALKER = T2.

<p><u>PRIME/STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: ————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-BOMB-See-Take-Thing</p> <p>T3: ————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>PRIME/NO STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-BOMB-See-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>
<p><u>NO PRIME/STAGGER</u></p> <p>no prime...</p> <p>T1: ————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-BOMB-See-Take-Thing</p> <p>T3: ————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>NO PRIME/NO STAGGER</u></p> <p>no prime...</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-BOMB-See-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>

Afterwards, the participant would be asked to choose whether **BOMB**, or an acoustically similar word like **MOM** was presented.

Results and Discussion

Mean proportion correct ($p(c)$) and marginal mean $p(c)$ for the twelve experimental conditions are presented in Table 1. A repeated measures ANOVA was performed upon the $p(c)$ using subjects as the random variable.

Participants generally performed better as the target approached the end of the word string ($F = 6.138, (p < .005)$). Overall $p(c)$ for performance where the target occurred in the fourth position was .644. For the fifth and six target positions, participants' $p(c)$ was .686 and .741, respectively. The most feasible explanation for this pattern seems to

be that an increase in memory load associated with any speech subsequent to target presentation lowered performance levels. . Regardless, the same general pattern of results for the “stagger” and “prime” conditions was observed across all target positions.

Allowing participants to listen to a few words of the target talker’s word string before the interfering speech was initiated (the “stagger” condition) improved accuracy ($F = 17.076, (p < .001)$). Participants’ overall $p(c)$ for stagger and non-stagger trials was .719 and .655, respectively. This suggests that participants were certainly able to use the information about which talker to target to better their performance in the task.

Our principal question of interest was whether priming a participant with a particular talker’s word string minus the target word could enhance the subsequent perception of a target word when inserted into that string. In other words, we are interested in whether a new and previously unheard word can “piggy-back” its way into a speech stream whose members have had their activation raised through the use of a prime. Performance was increased in our “prime” condition ($F = 27.485, (p < .001)$), strongly suggesting that this is, indeed, the case. Participant’s $p(c)$ for primed and non-primed trials were .741 and .639, respectively. All interactions were statistically insignificant.

Thus, we seem to have confirmed one of the characteristics of schema-based segregation, as defined by Bregman, in a speech-on-speech listening situation. An auditory stream subject to schema-based segregation (in this case speech) can be made more attentionally salient through the activation of its stored forms. In the case of the present experiment, even an unactivated (non-primed) form can be made more

attentionally salient by inserting it in a speech stream whose forms have been activated. In the following experiment we will attempt to show that increasing the attentional salience of a speech stream in such as fashion can only result in increased perceptual salience when the interfering sound forms an informational rather than energetic masker. Energetic and informational masking are terms whose meaning should become apparent in the subsequent discussion.

Table 1: P(c) Means and Marginal Means for Experiment 1.

TARGET	STAGGER	NO	
POSITION 4		STAGGER	
PRIME	0.764	0.617	0.691
NO	0.622	0.571	0.597
PRIME			
	0.693	0.594	0.644
TARGET	STAGGER	NO	
POSITION 5		STAGGER	
PRIME	0.744	0.767	0.756
NO	0.659	0.572	0.616
PRIME			
	0.702	0.67	0.686
TARGET	STAGGER	NO	
POSITION 6		STAGGER	
PRIME	0.797	0.754	0.776
NO	0.762	0.647	0.705
PRIME			
	0.78	0.701	0.741

CHAPTER 2

EXPERIMENT 2

It is certainly clear that masking (defined by Pickles (1988) as when "...one stimulus obscures or reduces the response to another" (p.103)) is involved in the cocktail party problem. However, while the masking present in the cocktail party problem is certainly consistent with the above definition, it isn't clear that the masking present in the cocktail party problem is consistent with the connotation most commonly associated with masking. Rather, the vast majority of the psychoacoustic literature dedicated to masking seems to give treatment to masking of an "energetic" type. In energetic masking, the energy level of one stimulus is so large it effectually negates or swamps the effect another stimulus might have upon the auditory system at the level of neural encoding. This would be the type of masking you might experience if you tried to listen to the speaking voice of a friend who just happened to be on the other side of an operating jet engine. The energy level of the sound produced by the jet engine would be so large that the auditory system would be unable to code any further information. A second, less discussed type of masking has been termed "informational" masking (Pollack, 1975 as cited by Watson, 1976). In informational masking, competing stimuli interfere at higher levels of processing rather than at the peripheral level of the auditory system. Thus, in the traditional cocktail party situation (unless it is a very loud party) the peripheral auditory system should have access to the sound being produced by a

conversational partner. Any subsequent difficulty in the segregation and retrieval of the voice of a conversational partner must then be due to informational rather than energetic masking.

Over the years, a handful of studies have investigated those variables which provide relief from informational masking. Much of the initial experimentation designed to explore informational masking was performed by Watson and colleagues (1975 and 1976) in a series of articles entitled *Factors in the Discrimination of Tonal Patterns*. While we will generally concern ourselves with the second and third articles of this series, it is worth noting that Watson, Kelly and Wroton (1976) concluded that the pattern of effects present in the first article (1975) were consistent with an informational type of masking. We question this conclusion. Watson et al (1975) asked participants to listen to two tonal sequences each consisting of ten 40 ms components which ranged in frequency from ten preselected static values between 256 to 892 Hz. The two sequences were separated by 500 ms of silence. The probability that the two tonal sequences of a trial were the same was $p = .5$, with both positive and negative random values of Δf being generated for the target tones of the remaining trials (increments were chosen to maintain accuracy at 75% to 80% correct). Listeners were asked to determine whether the pair of tone sequences were the “same” or “different”. Watson et al noted that the $\Delta f/f$ necessary to maintain performance at $d' = 1$ was dependent upon the serial position of the tone being altered. More specifically, tones later in temporal sequence required a much lower value of $\Delta f/f$ to maintain a $d' = 1$ level of performance.

Watson et al equate this finding with the “recency effect” found in memory literature and state that this effect is “similar to other instances of backward-acting interference..., which are variously referred to as “recognition masking”, “blanking”, “informational masking”, or “temporal interference”.” (Watson et al, 1976, p. 1176). While a recency effect was no doubt observed (which was, in fact, very similar to the effect which serial position of target words had in experiment 1 of the current study), we question whether this sort of effect should be grouped under the heading “informational masking”. It seems prudent to segregate the retroactive, masking effects that memory might have on our recollected perceptions from those sorts of masking which actually affect our perceptions in an “online” fashion. Otherwise, even the fallibility of long-term memory could be seen as a sort of perceptual masking (i.e. my memory of my 6th year birthday cake is long gone, but it shouldn’t be said that it has been “masked” in any perceptual sense). Our definition of informational masking will instead focus upon those instances where the resolution abilities of higher levels of processing cause difficulty in an immediate perceptual sense for the listener.

Further articles by Watson et al in the series are more in accordance with this view of informational masking. Watson et al (1976) performed a similar experiment showing that high certainty stimuli are an effective method of blocking informational masking. Participants were again asked to listen to sequences of ten tones from both the same frequency range and of the same duration as those presented in Watson et al (1975). However, in this instance the same tonal sequence was used throughout the entirety of the experiment and the only tonal component subject to change was the

second component of the sequence. *Delta-f* values were again selected to maintain accuracy within a given range, in this case 60%-95% correct. Participants were asked to listen to seven 100 trial blocks per day for 14 consecutive days. On eight of these days (initial 6 and final 2) participants engaged in a same-different task similar to that of Watson et al (1975). On the remaining six days, participants engaged in a method of adjustment paradigm. A pair of tone sequences was repeated and participants were asked to manually adjust the second component of the second sequence until it matched the second component of the first sequence. Participants were allowed to listen to the pair of sequences as many times as necessary, until they were satisfied that the second components of both tonal sequences were equal in frequency. Watson et al reported that, aside from data trends that seemed to arise from the same/different to method-of-adjustment paradigm switch, the *delta-f/f* necessary to produce a $d' = 1$ level of performance steadily decreased as the participants gained subsequent days of experience in the task.

Watson et al claim the ability of participants to resolve frequencies at higher and higher resolutions while maintaining an equivalent level performance is due to an increasing level of certainty on the part of the listener. A post-hoc analysis of previous similar research is also presented to illustrate other ways by which increased certainty on the part of the listener results in increased resolution abilities. This analysis suggests that as the number of tonal-patterns, the number of components subject to change, and the number of signal frequencies used to create tonal patterns are reduced, a psychophysical minimum in uncertainty is achieved and informational masking is reduced. Watson et al

hypothesize this effect of certainty upon listeners' performance is due to the ability to "...discriminate between stimuli we've met before, or which we expect, or which we're directed to look for, much more accurately than between those that are unexpected or unfamiliar" (p. 1185). Thus, we have our first hints that attentional mechanisms are to play a significant role in the resolution of listening situations where informational masking is present. In the case of Watson et al (1976), a listener's certainty about where in a tonal sequence attentional focus is required seems to be capable of resolving listening situations where informational masking is present.

While Watson relied exclusively on non-speech stimuli for his investigations into informational masking, he did choose stimuli that correlated with the "duration of the shortest phonemes in the most rapid intelligible speech" (P. 375, Watson and Foyle, 1985), presumably with the hope of having data that would be applicable to situations involving informational masking in speech. However, recent experimentation performed by Freyman and colleagues has shown that the effects of informational masking can be observed in speech on speech listening reminiscent of original "cocktail-party" situation. Freyman, Helfer, McCall and Clifton (1999) extended a series of experiments performed by Kidd and colleagues (Kidd et al, 1994; Kidd, Mason and Rohtla, 1995) which showed that spatial separation was a useful cue for segregation in listening situations assumed to produce informational masking but not in listening situations assumed to produce energetic masking. Whereas Kidd et al's studies used tonal stimuli, Freyman et al's experimentation utilized speech presented in nonsense sentences (e.g. "The thorn can wake the kettle") as stimuli. These sorts of stimuli were presented as targets in the

presence of either speech spectrum noise or more nonsense sentences produced by a second female talker. In addition to using actual spatial separation, Freyman et al created conditions of simulated spatial separation in an anechoic chamber through the use of “the precedence effect”, where a “...signal and its (simulated) reflections are gathered into a single image perceived near the location of the original source..” (1999, p. 3579). Noise and masking signals were delivered in one of four different ways. A “front/front” condition delivered both signal and masker from a single speaker directly in front of the listener (0 degrees). A “front/right” condition delivered the signal from the speaker directly in front of the listener and noise from a position 60 degrees to the right of the listener and was designed to determine the effects of true spatial separation. Two additional conditions (“front/front-right” and “front/right-front”) delivered the signal from the front speaker, while noise was presented in a lead-lag pair designed to call upon the precedence effect, where the lead signal was presented from either the front speaker or the right speaker and then subsequently presented in the alternate speaker. In conditions designed to call upon the precedence effect, the “reflection” signal lagged the “original” lead signal by 4 ms. When the masking stimulus was the speech spectrum masker, Freyman et al report a 8.2 dB advantage for identification of key words in the presence of the speech spectrum masker for the F/R presentation (true spatial separation) when compared to the remaining presentation conditions, which were approximately equal (p. 3582). However, when the signal was presented along with the female talker masker a 13.7 dB advantage for the F/R condition was observed when compared to the F/F

condition (p.3583). Thus Kidd et al's finding that spatial separation is a more useful cue for overcoming listening situations involving informational masking than for situations involving energetic masking was replicated.

More interestingly, the conditions assumed to produce the precedence effect (F/F-R and F/R-F) produced quite different data when the masker was another female talker rather than speech spectrum noise. While performance for the conditions designed to invoke the precedence effect did not equal performance in the F/R condition (in which true spatial separation was present), performance was strikingly better than that of the F/F condition (in which signal and masker were presented from the same source) and did approach performance in the F/R condition at lower S/N ratios . Participants were approximately 20% more accurate in the F/F-R and F/R-F presentation conditions compared to performance in the F/F presentation condition (excepting the -12 dB S-N ratio condition in which performance in the F/F-R, F/R-F, and F/F conditions were approximately equal). This is especially impressive given the fact that the F/R-F and F/R-F conditions are identical to the F/F condition excepting the addition of extra acoustic information (the channel delayed by 4 ms) which could only have increased the overall dB level at the present at the ears, and thus the amount of overall energetic masking present. Freyman et al interpret this result to suggest that even perceived spatial separation is effective in overcoming listening situations where informational masking is present (e.g. female talker interference) but not in those situations where energetic masking is present (e.g. speech spectrum noise).

The following experiment will attempt to show that such an advantage for speech listening situations involving informational masking versus energetic masking can be produced by manipulating attentional factors rather than the perceived spatial separation factors utilized by Freyman et al. Experiment 2 is nearly identical to Experiment 1, with the exception that conditions were added to compare speech (informational) and non-speech (energetic) maskers.

Design

Participants

40 undergraduate students, graduate students and post-doctoral fellows from the University of Massachusetts-Amherst were given course credit or cash in exchange for their participation in this study. Participants were native speakers of English with no known hearing difficulties. Prior to participation, participants read a instruction sheet detailing the particulars of the experiment and signed a form expressing their informed consent to participate in the experiment. Subsequent to participation, participants received both written and oral feedback detailing the nature of the experimentation and contact information should they have any questions.

Apparatus and Procedures

The same recorded words spoken by three female talkers detailed in the first experiment were used as stimuli. Target pairs also remained the same. A computer program inspired by Schroeder (1968), which randomly switched the sign of each sample or allowed it to remain the same, was devised to create an energetic masking situation. The energetic maskers created through this noise transform were also increased in

amplitude by a magnitude of three in order to yield approximately equal levels of performance by participants in both energetic and informational masking situations. Participants were presented with stimuli with the same apparatus described in the first experiment. Trial initiation and responses were made in the same fashion as well.

Experimental Design

A 2 X 2 X 2 randomized and counterbalanced within-subjects design was employed. The three experimental variables were “prime” vs “no-prime”, “stagger” vs “no stagger”, and “speech masker” vs “non-speech (noise) masker”. In the “prime” condition, the participant listened to all words, save the target word, that the target speaker would speak in the presence of competing speech before the competing speech was initiated. In the “no-prime” condition the speech to noise transform detailed earlier was performed upon the words listeners would have had access to in the “prime” condition. In the “stagger” condition the listener was allowed to hear the first two words of the target speaker’s word string without interference before the competing speech was introduced, while in the “no stagger” condition the target speaker and distractor speakers began synchronously. In the “speech masker” condition, the signal was presented along with the two female non-target talkers, as in all conditions of the first experiment. In the “noise masker” condition each of the non-target female talkers stored waveforms had the speech to noise transform performed upon it. On each trial a computer program determined the combination of conditions that would be presented. The program then randomly selected a target word from one of the talkers and six random non-target words (target words were never used as non-target words and vice versa). These words were

used to create the target talkers word string. Five or seven non-target words (depending upon whether the “stagger” condition was present) were randomly selected from each of the two non-target speakers and used to construct the interfering word or noise strings. All randomly selected words for both target and non-target strings were selected without replacement. Words in both target and non-target word strings were systematically separated by 100 ms of digitally created silence. Protocol was added to insure that the same word would not be selected twice for any given word string. Non-target words were only resampled after each word in the set had been used once. This procedure insured that both the combination of non-target words chosen and they way in which the three talkers word strings were synchronized (or unsynchronized) with one another was randomized from trial to trial. Figure 2 details this trial assembly process and six possible conditions which might result given the selection of a set of words. Once again, just as the spaces between these written words are occasionally both synchronous and asynchronous with the spaces of other talkers, so were the spaces between the talker’s spoken words. Participants were exposed to all stimuli in a single block of approximately 45 minutes during experiment 2, though an optional break could be taken at any time.

Figure 2: Trial assembly process for the various conditions of Experiment 2.

TARGET WORD = "BOMB" ; TARGET TALKER = T2.

Speech Interference

<p><u>PRIME/STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: ————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See- BOMB-Take-Thing</p> <p>T3: ————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>PRIME/NO STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>
<p><u>NO PRIME/STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1: ————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: ————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>NO PRIME/NO STAGGER</u></p> <p>Speech-transformed noise of prime....</p> <p>then all speakers commence:</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>

Noise Interference

<p><u>PRIME/STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: ————Speech transformed noise of T1</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: ————Speech transformed noise of T3</p>	<p><u>PRIME/NO STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: Speech transformed noise of T1*****</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Speech transformed noise of T3*****</p>
<p><u>NO PRIME/STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1:———Speech transformed noise of T1</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: ———Speech transformed noise of T3</p>	<p><u>NO PRIME/NO STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1: Speech transformed noise of T1*****</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Speech transformed noise of T3*****</p>

Afterwards, the participant would be asked to choose whether **BOMB**, or an acoustically similar word like **MOM** was presented.

Results and Discussion

Mean $p(c)$ for the eight experimental conditions are presented in Table 2. A repeated measures ANOVA was performed upon the $p(c)$ using subjects as the random variable. Once again, participants were able to correctly identify the target word with significantly greater accuracy in trials where the prime was present compared to trials in which the prime was not ($F = 14.51, (p < .001)$). Trials in which the prime was present resulted in an approximate 5 % advantage in accuracy. Trials in which the stagger was present resulted in an approximate 2 % advantage in participants' accuracy. The advantage produced by the stagger condition was not statistically significant ($F = 1.33, (p = .26)$). Nor was participants' performance in energetically and informationally masked trials significantly different ($F = 1.44, (p = .24)$). The general pattern of data were consistent with the experimental predictions (namely that priming would produce greater accuracy in the presence of a speech (informational) masker than in the presence of non-speech (energetic) masker). The presence of the prime resulted in an approximate advantage of 7% in accuracy for the informationally masked listening situations, while producing a 3% advantage for accuracy in the informationally masked situations. However, these factors failed to produce an interaction of significance ($F = 2.06, (p = .16)$).

A concern arose that holding the position of the target constant (recall that the target words in experiment 2 were always presented in position 5) could produce the sort of reduction of informational masking Watson et al (1976) found in the face of reduced uncertainty on the part of the listener. A reduction in informational masking due to

increased certainty could have hypothetically reduced any effects priming may have had in an alleviating informational masking. It was decided that, at the expense of an increased number of conditions, an additional experiment identical in most respects to experiment 2 should be performed in which the target position was varied in a fashion similar to experiment 1. It was also decided that in the analysis of experiment 3, the data would be collapsed across the position factor, both since memory effects aren't a question of interest in the present study and since the same general pattern of results were observed across all position factors in experiment 1.

Table 2: P(c) Means and Marginal Means for Experiment 2.

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.762	0.761	0.762
NO	0.732	0.647	0.69
PRIME	0.747	0.704	0.726
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.715	0.729	0.722
NO	0.688	0.695	0.692
PRIME	0.702	0.712	0.707

CHAPTER 3

EXPERIMENT 3

Design

Participants

24 undergraduate students, graduate students and post-doctoral fellows from the University of Massachusetts-Amherst were given course credit or cash in exchange for their participation in this study. Participants were native speakers of English with no known hearing difficulties. Prior to participation, participants read a instruction sheet detailing the particulars of the experiment and signed a form expressing their informed consent to participate in the experiment. Subsequent to participation, participants received both written and oral feedback detailing the nature of the experimentation and contact information should they have any questions.

Apparatus and Procedures

The same recorded words spoken by three female talkers detailed in the first two experiments were used as stimuli. Target pairs also remained the same. The same computer program which randomly switched the sign of each sample or allowed it to remain the same was utilized to create energetic masking situations. These energetic maskers were once again increased in amplitude by a factor of three in order to approximate participants' performance in energetic and informational masking situations. Participants were presented with stimuli with the same apparatus described in the first two experiments. Trial initiation and responses were made in the same fashion as well.

Experimental Design

A 2 X 2 X 2 X 3 randomized and counterbalanced within-subjects design was employed. The twenty four experimental variables were “prime” vs “no-prime”, “stagger” vs “no stagger”, “speech masker” vs “non-speech (noise) masker” and target positions “4”, “5”, and “6”. All experimental variables were identical to those mentioned in experiments 1 and 2. Once again, on each trial a computer program determined the combination of conditions that would be presented. The program then randomly selected a target word from one of the talkers and six random non-target words (target words were never used as non-target words and vice versa). These words were used to create the target talkers word string. Five or seven non-target words (depending upon whether the “stagger” condition was present) were randomly selected from each of the two non-target speakers and used to construct the interfering word or noise strings. All randomly selected words for both target and non-target strings were selected without replacement. Words in both target and non-target word strings were systematically separated by 100 ms of digitally created silence. Protocol was added to insure that the same word would not be selected twice for any given word string. Non-target words were only resampled after each word in the set had been used once. This procedure insured that both the combination of non-target words chosen and they way in which the three talkers word strings were synchronized (or unsynchronized) with one another was randomized from trial to trial. Participants were exposed to stimuli in two approximately ½ hour blocks, with a mandatory break in between blocks.

Results and Discussion

Mean $p(c)$ for the various conditions (collapsed across position) are presented in Table 3. Trials in which participants were presented with the prime prior to the masked listening situation produced an approximate 7 % advantage in accuracy. The advantage produced by the prime was statistically significant ($F = 12.85, (p < .001)$). There was no statistically significant difference between trials containing stagger and no stagger ($F = .64, (p = .43)$) or between informationally and energetically masked trials ($F = 3.40, (p = .08)$), though masker-type as a factor did approach significance. Once again, the general pattern of results were consistent with predictions regarding attentional salience in informational masking versus energetic masking (i.e. priming showed a greater effect upon participant's $p(c)$ in listening situations involving speech maskers than in those listening situations involving speech-transformed noise). The presence of the prime in informationally masked listening situations gave participants an approximate advantage of 9 %, while the presence of the prime only gave listeners an approximate advantage of 6% in the energetically masked situations. Once again, this interaction failed to obtain levels of statistical significance ($F = .79, (p = .38)$).

Table 3: P(c) Means and Marginal Means for Experiment 3.

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.760	0.759	0.760
NO PRIME	0.693	0.646	0.670
	0.727	0.703	0.715
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.780	0.775	0.778
NO PRIME	0.722	0.719	0.721
	0.751	0.747	0.749

A post-hoc analysis was performed upon the combined data (60 subjects) from experiments 1 and 2, in which the experiment number was manipulated as a between-subjects factor. Mean p(c) for the various combined conditions are presented in Table 4. The presentation of a prime prior to exposure to the masked listening situation produced an approximate 6 % advantage across trials. The advantage produced by the prime was statistically significant ($F = 28.08, (p < .001)$). The 'stagger/no stagger' factor failed to attain statistical significance ($F = 1.77, (p = .19)$) as did the 'informationally masked/energetically masked' factor ($F = .44, (p = .51)$). Once again the general

pattern of results were consistent with experimental predictions. Exposure to the prime prior to an informationally masked listening situation produced an approximate advantage of 9 %, while exposure to the prime prior to an energetically masked listening situation produced an approximate advantage of 5 %. While “experiment number”, as a factor, failed to reach significance ($F = 1.55$, ($p = .22$)), our interaction of interest (info/energetic masking X no prime/prime) did approach statistical significance ($F = 2.5$, ($p = .12$)) when data was combined across experiments. In addition, this effect’s interaction with the “experiment number” factor failed to achieve any sort of statistical significance ($F = .04$, ($p = .85$)).

Table 4: P(c) Means and Marginal Means for Experiment 2 and 3.

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.761	0.760	0.761
NO PRIME	0.713	0.647	0.680
	0.737	0.704	0.721
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.748	0.752	0.750
NO PRIME	0.705	0.707	0.706
	0.727	0.730	0.729

While the general pattern of results for experiments 2, 3 and the results for their combined data were consistent with our expectations regarding the role of attention in listening situations involving informational masking when compared to listening situations involving energetic masking, the interaction consistent with this prediction was not statistically significant for experiments 2 and 3 and only approached statistical significance for their combined data. It is possible that the task utilized in these experiments limited the magnitude of the predicted interaction. Recall that in both experiment 1 and Watson et al (1975) memory effects were observed in the data. Watson et al (1975) found that items later in the temporal sequence of tones required a much lower *delta-f/f* to maintain a level of performance in which $d'=1$. In experiment 1, the “position” factor had a statistically significant effect upon participant’s ability to correctly choose the target word from the word pair presented at the end of each trial. Performance was systematically worse for those target words occurring earlier in the target word string. These data patterns lead us to believe that the task used in experiments 2 and 3 may have tempted participants to utilize a memory intensive strategy, which may have confounded any effects which could have been attributed to differences between informational and energetic masking. In trials containing the prime, it is feasible to imagine a participant subvocally rehearsing the words of the prime and listening for any deviation from these words. This strategy is somewhat different than the selective attention strategy we had hoped participants would engage in.

In an attempt to alleviate this concern, Experiment 4 relied on a task assumed not to be susceptible to the same memory based strategies which may have flawed experiments 2 and 3 (recall that we are interested in the immediate perceptual salience of a target word in the presence of the two types of maskers). Namely, a single target word was presented visually to the participant before each and every trial and the participant was instructed to pull a trigger upon hearing the word, if and when the word should be spoken. On half of the trials this target word was present in the target word string. On the remaining half of the trials the target word was not present in the target word string as a probe, and was replaced by an acoustically similar word (to be determined by the minimal word pairs used in experiments 1,2 and 3). In this way, participants were assumed to only hold the target word in memory, and not engage in any elaborate rehearsal processes which may have confounded any effects associated with our hypothesis concerning attention's differential role in informational and energetic masking situations.

The trial structure of experiments 2 and 3 may have been another possible experimental artifact which may have affected any differences which might have been observed for priming in listening situations involving informational masking versus energetic masking. Recall that the masker present on each trial (either energetic or informational) was randomly selected (though counter-balanced). It is conceivable that listeners' exposure to energetically-masked trials (in which increased attention presumably wouldn't have been of much assistance) may have suppressed their attentional engagement in informationally-masked trials (where attention is hypothesized

to play a large role in unmasking). For this reason, in Experiment 4, listeners were exposed to the different types of masking in two approximately ½ hour blocks. Participants were alternately exposed to either the informationally-masked trials or the energetically-masked trials followed by a break and then an equal block of the opposite type of masking trials.

CHAPTER 4

EXPERIMENT 4

Design

Participants

32 undergraduate students, graduate students and post-doctoral fellows from the University of Massachusetts-Amherst were given course credit or cash in exchange for their participation in this study. Participants were native speakers of English with no known hearing difficulties. Prior to participation, participants read a instruction sheet detailing the particulars of the experiment and signed a form expressing their informed consent to participate in the experiment. Subsequent to participation, participants received both written and oral feedback detailing the nature of the experimentation and contact information should they have any questions.

Apparatus and Procedures

The same recorded words spoken by three female talkers detailed in the first three experiments were used as stimuli. Target pairs also remained the same. The same computer program which randomly switched the sign of each sample or allowed it to remain the same was utilized to create energetic masking situations. These energetic maskers were once again increased in amplitude by a factor of three in order to approximate participants' performance in energetic and informational masking situations. Each trial was initiated by the pull of a trigger. Participants were then presented visually with a target word (a member of the target pairs) via the Amdek monochrome monitor and instructed to pull a response trigger as soon as possible if they

heard that target word in the subsequent masked listening situation. They were instructed not to pull the trigger if the target word was not heard. The target word remained upon the screen for the duration of the trial. After each trial, participants received feedback via the monitor as to whether their performance was 'very fast', 'fast', 'slow', or 'too slow'. The particular feedback delivered was determined based on the participants trigger response relative to the onset of the target word. All other details pertaining to apparatus and stimuli are identical to those presented for the first three experiments.

Experimental Design

A 2 X 2 X 2 X 3 randomized and counterbalanced within-subjects design was employed. The twenty four experimental variables were "prime"vs "no-prime", "stagger"vs "no stagger", "speech masker" vs "non-speech (noise) masker" and target positions "4", "5", and "6". All experimental variables were identical to those mentioned in Experiment 3. Once again, on each trial a computer program determined the combination of conditions that would be presented. The program then randomly selected a target word from one of the talkers and six random non-target words (target words were never used as non-target words and vice versa). These words were used to create the target talker's word string. Five or seven non-target words (depending upon whether the "stagger" condition was present) were randomly selected from each of the two non-target speakers and used to construct the interfering word or noise strings. All randomly selected words for both target and non-target strings were selected without replacement. Words in both target and non-target word strings were systematically separated by 100 ms of digitally created silence. Protocol was added to insure that the

same word would not be selected twice for any given word string. Non-target words were only resampled after each word in the set had been used once. This procedure insured that both the combination of non-target words chosen and the way in which the three talkers word strings were synchronized (or unsynchronized) with one another was randomized from trial to trial. The assembly process for each of the conditions (excepting position) is detailed in Figure 3. Participants were exposed to stimuli in two approximately ½ hour blocks, with a mandatory break in between blocks.

Figure 3: Trial assembly process for the various conditions of Experiment 4.

TARGET WORD = "BOMB" Presented to participant visually for duration of trial.

Speech Interference

<p><u>PRIME/STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: —————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See- BOMB-Take-Thing</p> <p>T3: —————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>PRIME/NO STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>
<p><u>NO PRIME/STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1: —————Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: —————Slam-Walk-Wheel-Cheese-Plane</p>	<p><u>NO PRIME/NO STAGGER</u></p> <p>Speech-transformed noise of prime.....</p> <p>then all speakers commence:</p> <p>T1: Ball-Swim-Dot-Jump-Grass-Want-Chase</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Cake-Free-Slam-Walk-Wheel-Cheese-Plane</p>

Noise Interference

<p><u>PRIME/STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: —————Speech transformed noise of T1</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: —————Speech transformed noise of T3</p>	<p><u>PRIME/NO STAGGER</u></p> <p>first the prime(T2): Pen-Chair-Love-See-Take-Thing</p> <p>then all speakers commence:</p> <p>T1: Speech transformed noise of T1*****</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Speech transformed noise of T3*****</p>
<p><u>NO PRIME/STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1: —————Speech transformed noise of T1</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: —————Speech transformed noise of T3</p>	<p><u>NO PRIME/NO STAGGER</u></p> <p>Speech-transformed noise of prime...</p> <p>then all speakers commence:</p> <p>T1: Speech transformed noise of T1*****</p> <p>T2: Pen-Chair-Love-See-BOMB-Take-Thing</p> <p>T3: Speech transformed noise of T3*****</p>

The participant would attempt to pull the trigger as quickly as possible upon encountering the target word 'BOMB'.

Results

Mean $p(c)$ for the various conditions (collapsed across position) are presented in Table 5. Participants' performance was significantly different ($F = 33.102, (p < .001)$) for energetically masked (overall $p(c) = .882$ for positive probes) and the informationally masked trials (overall $p(c) = .803$ for positive probes). Participants also performed significantly better in trials containing the prime, than in those trials with no prime ($F = 10.027, (p < .005)$). 'Stagger/No Stagger' failed to attain significance as a factor ($F = .004, (p = .952)$). Once again, the interaction between 'prime/no prime' and 'energetic/informational masker' failed to attain statistical significance ($F = 1.77, (p = .193)$). Still, the general trend of the data is consistent with our predictions concerning schema-driven processing's differing role in informationally and energetically masked listening situations. In the informationally masked listening situation, the presence of a prime increased listener's accuracy in detecting the target word by approximately 7%, whereas the prime only improved listener accuracy by about 3.5% in conditions designed to approximate an energetically masked listening situation.

This familiar pattern of data (recall Experiments 2,3 and their combined data) makes it difficult to reject our hypothesis that schema-driven processes of auditory segregation should be more effective in informationally masked listening situations than in energetically masked situations, despite the data's refusal to yield any more concrete affirmations. However, a signal-detection based analysis of the data suggests this pattern may be illusory. D -prime values (obtained by averaging correct positive probe trials across serial position to obtain hit rate and using incorrect negative probe trials to obtain

false alarm rate) suggest that participants' may have used the prime to a slightly greater advantage in energetically masked trials, and that the prime advantage seen in $p(c)$ for informationally-masked positive probe trials may have been an artifact of reduced bias (d' -prime and criterion values for the various conditions are presented in Table 6).

Average reaction time data (measured in milliseconds from the offset of the target word) for the various conditions are presented in Table 7. However, reaction time data failed to achieve statistical significance in any of the experimental conditions or their respective interactions.

Table 5: $P(c)$ Means and Marginal Means for Experiment 4

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.854	0.820	0.837
NO PRIME	0.767	0.771	0.769
	0.811	0.796	0.803
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	0.899	0.899	0.899
NO PRIME	0.851	0.879	0.865
	0.875	0.889	0.882

Table 6: Mean D-prime and Criterion (in parentheses) for Experiment 4.

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	2.07	1.84	1.96
	(1.05)	(0.95)	(1.00)
NO PRIME	1.90	1.99	1.95
	(1.20)	(1.30)	(1.25)
	1.99	1.92	1.96
	(1.13)	(1.13)	(1.13)
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	2.14	2.12	2.13
	(0.88)	(0.90)	(0.89)
NO PRIME	2.26	1.79	2.03
	(1.25)	(0.68)	(0.97)
	2.20	1.96	2.08
	(1.07)	(0.79)	(0.93)

Table 7: Reaction Time Means and Marginal Means for Experiment 4 (in ms.).

SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	120	123	122
NO PRIME	110	111	111
	115	117	116
NON SPEECH MASKER	STAGGER	NO STAGGER	
PRIME	64	118	91
NO PRIME	106	123	115
	85	121	103

GENERAL DISCUSSION

These frustratingly non-significant, yet recurrent patterns of data occurred with variations in both experimental stimuli (Experiment 2's constant target word position vs. Experiment 3 and 4's varying target word position), experimental design (varying the types of masker presented to listeners within blocks in Experiments 2 and 3 vs. varying the types of masker between blocks in Experiment 4) and even experimental task (the memory task in Experiments 2 and 3 vs. the probe task in Experiment 4). In each experiment, the priming manipulation hypothesized to affect attentional salience of the speech stream, and therefore enhance schema-driven segregation, improved listeners' performance in the informationally masked listening situations to a greater extent than it did for the energetically masked situation.

These consistencies across experiments make it tempting to suggest a few unexplored possibilities which could have acted as artifacts across Experiments 2, 3 and 4. While it is possible that schema-driven segregation is non-preferential in its assessment of informationally and energetically masked listening situations, there are reasons to believe this might not be the case. Aside from obvious inconsistencies with the aforementioned recurrent pattern of data, explanations accepting the null hypothesis also remain unattractive due to an incompatibility with the theories and data of the current literature. Two alternative explanations seem far more likely, and suggest possible areas for subsequent research.

Firstly, it remains a possibility that the experimental tasks chosen were too laden with artifacts to sufficiently disentangle any experimental effects from noise. Perhaps an alternative task would be far more powerful in establishing the role of schema driven processing in informationally masked listening situations. Both the memory and probe based tasks used in Experiments 1 through 4 allowed participants far too much leeway in terms of the experimental strategies they could adopt. Despite encouraging participants to utilize the prime to their advantage and to use selective attention strategies whenever possible, it was never possible to be certain they were actually using these strategies. Experimental scenarios where participants gained multiple exposures to a prime prior to each trial or where the number of trials containing primes were greater than trials without might encourage participants to utilize the prime in a fashion more in accordance with our expectations. The consistent use of the prime by all experimental participants might alleviate some of the noise associated with the general pattern of results collected across Experiments 2, 3 and 4. Experimental scenarios with slightly more engaging and realistic stimuli, such as proper sentences, might also have the effect of engaging participants' selective attention to a greater degree, and negating the use of divided attention as a strategy.

Secondly, it is possible that the informational and energetic maskers created for this experiment were not "pure" enough to exhibit clear cut differences between informational and energetic masking listening situations. The informational (speech) maskers used may have contained some degree of energetic masking, particularly at moments when both masking talkers produced high-energy speech, while the target

talker produced low-energy speech. In the same fashion, the energetic maskers did not completely energetically mask speech, in an effort to maintain comparable levels of accuracy across masking conditions. It also remains possible that any residual similarity to speech the energetic maskers may have retained subsequent to being transformed could have resulted in some degree of informational masking. While it was assumed at the initiation of these experiments that the use of “impure” maskers would take the form of random noise introduced into the experiment, perhaps more stringently produced stimuli might be more effective in establishing that informationally masked listening situations can be overcome through the use of schema driven processing. Speech-strings carefully controlled for any incidental energetic masking could reduce the “impurity” of the informational maskers in the experiment. The use of an energetic masker correlated with some other component of human speech, such as steady-state pink-noise (which shares the general frequency spectrum of human speech but not the gross amplitude and envelope characteristics of the energetic maskers of these experiments) might be more effective in creating a more “pure” energetic masker.

APPENDIX

WORDS USED AS STIMULI (SEPARATED BY TALKER)

<u>Talker 1:</u>					<u>TALKER 2:</u>
	clown	half	park	soup	
ace	coat	high	paste	splash	
ache	cool	him	path	stick	babe
aisle	could	his	pearl	stove	bail
all	crab	hit	peg	suit	ban
as	dad	hunt	phone	talk	bank
axe	date	it	pie	tea	bare
bar	day	jam	please	team	bat
barn	deaf	jug	plow	them	batch
base	dodge	knees	race	there	beach
bathe	door	knife	ran	thin	beak
bear	dusk	late	rat	thing	beat
bee	earn	laugh	red	thumb	big
beg	east	law	ring	toe	bill
bells	fair	lid	road	tray	boat
bet	falls	life	rose	true	bomb
bless	fan	loud	rug	turn	boot
bowl	fat	low	rush	twins	box
broom	feed	luck	sack	void	buck
cab	felt	mop	seal	wait	budge
cage	flag	mouse	search	waste	buff
cake	fresh	nail	see	weed	but
camp	give	name	she	wet	buzz
carve	good	neat	sheep	what	can
cause	got	none	ship	when	care
chat	grab	not	sieze	wire	cart
cheek	grew	note	sing	yes	cash
chew	grey	owl	skin	you	catch
church	gun	page	socks	youth	chair

cheap	have	paw	shin	beef	fill
chief	hick	pays	shoe	black	find
chill	hiss	peach	shore	blind	five
chin	hitch	peak	shun	bounce	floor
chop	jail	peep	sick	boy	fold
core	key	pick	sip	bread	fox
cuff	kid	pill	slice	break	freeze
cup	kin	pin	sore	bud	frog
cuss	kiss	pool	sun	bug	fun
died	lap	pop	tail	bus	gate
dies	lash	rack	tame	bush	germ
dine	latch	rage	tan	case	goes
dive	leach	raid	tick	cat	grade
face	leaf	rail	tie	chick	great
fail	league	raise	till	choose	higher
faith	leak	rake	tin	chore	hot
fake	lean	rap	top	class	knee
fame	lease	rash	tore	cost	know
fate	leash	rave	two	crib	lay
feet	lice	reach	van	cut	leave
fell	man	reef	vine	desk	lip
fin	map	reek		dice	may
fled	mass	rib	<u>TALKER 3:</u>	did	me
for	math	rid	air	dish	mess
gale	mom	robe	arm	ditch	most
game	much	rove	back	dog	mouth
gave	mush	scene	bad	drop	neck
goat	mutt	seat	bag	else	need
hack	paid	shame	bait	end	nest
hail	pat	share	ball	eye	next
has	patch	sheath	bath	fed	nice
hatch	pave	sheet	bead	few	nuts

oar	set	west
on	shell	wheel
own	shirt	white
pain	shop	wide
pale	sin	wing
pan	sis	wreck
pants	sit	yard
pass	sled	
pinch	slip	
pink	smile	
plane	smoke	
pole	stare	
pond	straw	
poor	street	
press	string	
pun	such	
purse	take	
put	teach	
quick	tell	
rag	than	
rain	thank	
raw	thick	
rice	tire	
rich	train	
ride	tree	
rig	up	
room	us	
rush	use	
scab	vase	
school	walk	
seed	wash	
sell	ways	

BIBLIOGRAPHY

- Bregman, A. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, Mass. : MIT Press.
- Broadbent, D. (1958). *Perception and communication*. London: Pergamon.
- Bronkhurst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acoustica*, 86, 117-128.
- Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, 25, 975-979.
- Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling transitions. *Canadian Journal of Psychology*, 30, 99-114.
- Deutsch, D. (1972). Octave generalization and tune recognition. *Perception and Psychophysics*, 11, 411-412.
- Francis, W.N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.
- Freyman, R.L., Helfer, K.S., McCall, D.D., & Clifton, R.K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustic Society of America*, 106, 3578-3588.
- Kidd, G., Mason, C.R., Deliwal, P.S., Woods, W.S. & Colburn, H.S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustic Society of America*, 95, 3475-3480.
- Kidd, G., Mason, C.R. & Rohtla, T.L. (1995). Binaural advantage for sound pattern identification. *Journal of the Acoustic Society of America*, 98, 1977-1986.
- Pickles, J.O. (1982). *An introduction to the physiology of hearing*. London; New York: Academic Press.
- Pollack, I. (1965). Auditory information masking. *Journal of the Acoustic Society of America*, 57, S5.
- Samuel, A.G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1124-1131.

- Schroeder, M.R. (1968). Reference signal for signal quality studies. *Journal of the Acoustic Society of America*, 44, 1735-1736.
- Warren, R.M. (1968). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393.
- Watson, C.S., & Foyle, D.C. (1985). Central factors in the discrimination and identification of complex sounds. *Journal of the Acoustic Society of America*, 78, 375-380.
- Watson, C.S., Kelly, W.J., & Wroton, H.W. (1976). Factors in the discrimination of tonal patterns: II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustic Society of America*, 60, 1176-1185.
- Watson, C.S., Wroton, H.W., Kelly, W.J. & Benbassat, C.A. (1975). Factors in the discrimination of tonal patterns. I. Component frequency, temporal position and silent intervals. *Journal of the Acoustic Society of America*, 57, 1175-1185.

