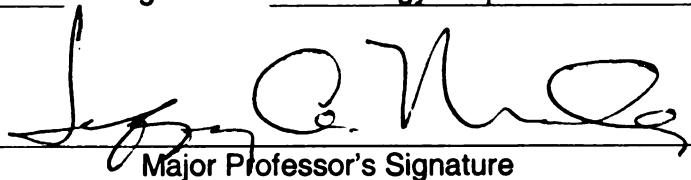This is to certify that the
thesis entitled

Statistical Speech Segmentation: A Neuropsychological
Investigation of Auditory Object Formation

presented by

Daniel Fogerty

has been accepted towards fulfillment
of the requirements for the

_____M.A._____ degree in ___Audiology & Speech Sciences___

_____
Major Professor's Signature

_____09/28/04_____
Date

MSU is an Affirmative Action/Equal Opportunity Institution

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

6/01 c:/CIRC/DateDue.p65-p.15

STATISTICAL SPEECH SEGMENTATION: A NEUROPSYCHOLOGICAL
INVESTIGATION OF AUDITORY OBJECT FORMATION

By

Daniel Fogerty

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Audiology & Speech Sciences

2004

ABSTRACT

STATISTICAL SPEECH SEGMENTATION: A NEUROPSYCHOLOGICAL INVESTIGATION OF AUDITORY OBJECT FORMATION

By

Daniel Fogerty

During language acquisition, the development of a speech segmentation strategy has classically been investigated by how language learners discover word boundaries. In addition to certain acoustic cues (e.g., stress, phonotactics, etc.), statistical regularities have been identified as an early perceptual cue to segmentation. Sensitivity to statistical regularities appears to be a domain-general process that is founded in neurophysiological mechanisms. This study examined the ability to use statistical regularities to identify word boundaries in an artificial language. It also expanded upon previous research by investigating whether statistical regularities are capable of forming auditory objects. In speech, these objects are the early, prelexical word-forms used to segment speech online. The creation of word-forms was examined through a pretest-posttest design and compared to optimal performance on isomorphic English tasks. The manipulability of these word-forms was also investigated through test protocols that measured a hierarchy of processing levels and through a higher-order cognitive task of phonological synthesis. Results support the ability of statistical regularities to identify word boundaries and create perceptual word-forms for speech segmentation. The results also suggest that these word-forms may be represented at a higher level of processing that is cognitively accessible. Domain-general theory of statistical learning, second language acquisition, and therapeutic intervention techniques are discussed.

To Him who created our ears to hear and our brains to understand. This is my attempt to learn of your miracle, that we might help those who struggle with this wonderful gift – that of communication.

and…

A mon amie le plus proches et au compagnon le plus cher…Jonni.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Conceptualizing the recognition of speech is a daunting task. It may be examined either through breaking down of a continuous speech stream into meaningful segments, or joining multiple acoustic signals into auditory objects. It is through an adequate system of speech perception that meaning emerges and language develops. This paper investigates a specific process to acquiring a system of speech perception that is capable of identifying words. How is a continuous stream of speech, containing endless acoustic features, initially organized into meaningful perceptual objects? Or conversely, how is that same speech stream, containing no acoustic breaks to mark the end of words, broken into meaningful segments? Through a historic line of research, processes of speech perception have been identified, as well as rejected.

This paper posits that individual speech stimuli are first processed as isolated acoustic events. Later, the statistical regularities of speech stimuli allow for the integration of them into larger perceptual objects. An early sensory memory store is implicated in the ability to compare perceptual features for these objects. It facilitates online parsing of the continuous speech stream into individual objects that can later be discriminated in isolation or reformed from temporally segmented object features. These perceptual objects may, through later processes, eventually become lexicalized as words. Improvement in identifying perceptual units within the continuous stream would suggest

that perceptual abilities change due to exposure to the statistical probabilities inherent in the speech stream, perhaps due to the formation of auditory objects.

Kubovy and Van Valkenburg (2001) argued for the concept of an auditory object. According to them, a perceptual object "is that which is susceptible to figure-ground segregation" (p. 102). Historically studied by Gestalt psychologists within the visual modality, Kubovy and Van Valkenburg describe figure-ground segregation in terms of domain-general objects. Early pre-attentive processes of grouping and feature integration occur following the principles of Gestalt psychology. Stimulus elements are formed into perceptual organizations. Later, attention selects one organization (or a set of them) to become the figure, while all other information is relegated to the ground. Thus, those perceptual organizations that form the figure are perceptual objects, segmented from all other undifferentiated information of the ground. However, while visual information is segmented from the spatial scene, auditory information is segmented within a temporal envelope. Therefore, this paper defines an auditory object as a set of stimulus elements that are pre-attentively organized, grouped, and differentially selected (or segmented) from all other acoustic information. It is the position of this paper that individual speech stimuli may be organized, grouped, and selected through statistical means to form an auditory object, which can then be differentiated from the rest of the speech stream. These auditory objects serve as the initial parsing of the continuous speech stream into identifiable units. However, even if these auditory objects coincide with the boundaries of syllables or words, such information is prelexical; the meaning of such objects has not necessarily been accessed (or even acquired) at this time in processing. Therefore, this is a study of speech object formation and perception, not lexical access.

The concept of a domain in this paper goes beyond the different sensory processes (or modalities) of vision and audition to include such cognitive processes as morphology or reasoning. Domain-generality states that the same perceptual/cognitive processes that underlie one specific function also underlie other similar functions. This concept is typically expanded to support distributed network models of neural functioning, rather than specific modules within the brain that account for only one function. Therefore, when domain-generality is discussed in this paper, I am referring to cognitive/perceptual processes which may apply equally well to visual or auditory domains (and possibly others).

Infants, as well as other animals (monkeys and birds) have been shown to poses an innate ability to discriminate differences between speech sounds (Kluender, Diehl, & Killeen, 1987; Kuhl & Miller, 1975; Kuhl & Padden, 1983; Steinschneider, Schroeder, Arezzo, & Vaughan, 1995). However, what appears necessary for speech perception is the acquisition of a system for recognizing and organizing speech patterns. Simos and Molfese (1997) discuss an innate bases and early reorganization of the neurophysiological mechanisms involved in categorical speech perception. Infants soon acquire the ability to ignore certain phonetic contrasts that are not meaningful to their language experience and to acquire sharper contrasts in other areas (Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho, & Naatanen, 1998). This ability to organize sound variations into phonetic categories is essential for speech perception; thereby, allowing for allophonic and speaker variations to be perceived as belonging to the same phonetic category.

Phonetic categories allow for the formation of meaningful perceptual units that can indicate a change in word meaning; these units are recognized as phonemes. However, listening to a continuous speech stream composed of endless phonemes is no more worthwhile than when it was composed of endless acoustic features. These phonemes must again be organized into larger perceptual units. Statistical learning is a process which creates perceptual units based on the probabilities that one feature (a phoneme in this case, or a syllable as presented by Saffran, Newport & Aslin, 1996) occurs adjacent to another one. Weak probabilities will likely signal the end of words. Stress and other prosodic features likely offer additional cues (see Jusczyk, 1999 for an overview).

The exact process of the perceptual identification of words in a novel language is not fully understood. The historic problem which has obfuscated this identification process is the absence of consistent pauses marking word boundaries (Cole & Jakimik, 1980). To address this problem, acoustic speech cues have been proposed to assist in the segmentation process. Markers such as metrical stress (Cutler & Norris, 1988; Jusczyk, Houston, & Newsome, 1999), phonotactics (Mattys & Jusczyk, 2001a; Vitevitch & Luce, 1999), and allophonic variations (Jusczyk, Hohne, & Bauman, 1999) have all been proposed. Additionally, the method of statistical learning, which calculates the conditional probability between speech syllables, termed *transitional probabilities*, has also been proposed (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Goodsitt, Morgan, & Kuhl, 1993; Saffran, Aslin, et al., 1996). Transitional probabilities are calculated from the co-occurring frequency of two stimuli. In speech segmentation research, these stimuli are typically syllables. The use of transitional probability as a tool to speech

segmentation is predicated on the premise that intra-word syllables have a higher probability of co-occurring than do syllables that span word boundaries. For example, Saffran (2003) discussed the phrase *pretty baby*. The syllables *pre* and *ty* occur together much more frequently than the syllables which span the word boundary, *ty-ba*. In fact, in the speech directed to young infants, *pre* is followed by *ty* about 80% of the time; whereas, *ty* is followed by *ba* only about 0.03% of the time (Saffran, 2003). Thus, weak transitional probabilities may be used to indicate the ends of words. Metrical stress and statistical cues to word boundaries appear to be the earliest segmentation strategies used by infants (Jusczyk, 1999), emerging around 7.5 months (Jusczyk, et al., 1999) and 8 months (Saffran, Aslin, at al., 1996) respectively. Several computational models of segmentation learning have attempted to examine these segmentation processes (Brent, 1999) though not one clear method has emerged. In addition, the relative strength of contribution between different proposed methods of segmentation has received a cursory examination with varying results (for an investigation of statistics versus stress, see Johnson & Jusczyk, 2001; Saffran, Newport, & Aslin, 1996).

The use of statistical cues appear to be a domain-general technique for learning, as its properties have been demonstrated for visual stimuli (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002), nonlinguistic acoustic stimuli (Saffran, Johnson, Aslin, & Newport, 1999), artificial grammar acquisition (Gomez & Gerkin, 1999, 2001; Saffran & Wilson, 2003), and is present in animal learning (Hauser, Newport, & Aslin, 2001). These studies suggest that sensitivity to statistical structure is neither found solely in a linguistic domain, nor is it unique to humans. The use of statistical cues may be a natural technique used to identify perceptual objects within a set context. As such,

statistical cues identify and segment auditory objects within the speech stream. A further investigation of the perceptual and cognitive processes undergone during this segmentation strategy is warranted, which this paper proposes to address.

Speech Segmentation Strategies

Infants as young as one month can use the acoustic properties of speech to differentiate phonetic categories (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). By 6 months of age, this ability is altered through linguistic experience to reflect phonetic perception patterns of the native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). The beginning of successful speech segmentation starts around 7.5 months of age (Jusczyk & Aslin, 1995) and continues to develop as additional strategies of segmentation are applied (Jusczyk, 1999). Infants even apply these segmentation strategies to other languages that have similar structures (Houston, Jusczyk, Kuijupers, Coolen, & Cutler, 2000). By 8 months of age, infants exhibit some ability to remember frequently occurring words (Jusczyk & Hohne, 1997).

A technique that is frequently used by researchers to examine infant perceptual abilities is the headturn preference procedure (HPP; see Jusczyk, 1998). The HPP, an effective technique in studying infants between 4.5 and 21 months of age (Jusczyk, 1998), is particularly useful to establish which of two stimuli the infant prefers (Houston, et al., 2000). Infants are typically familiarized with isolated words and then tested on passages which either do or do not have the familiarized words. Longer listening times to the passages containing the familiarized words are interpreted as demonstrating successful segmentation processing by the infant (Mattys & Jusczyk, 2001b).

Jusczyk and Aslin (1995) were one of the first to adapt this technique to study speech segmentation processes. They familiarized 7.5 month-old infants to repetitions of two different words, and then tested whether the infants would listen longer to speech passages containing the familiarized words. Significant evidence for this was found. Moreover, when infants were first exposed to the passages and subsequently tested on repetitions of words which occurred frequently in the passages, infants again listened longer to the repetitions of familiarized words as compared to repetitions of words that did not occur in the passages. In contrast to the performance of these 7.5 month-old infants, repetition of the same procedure with 6 month-olds yielded no significant results. These results suggest acquisition of speech segmentation abilities between 6 and 7.5 months of age. However, this study did not provide evidence for how 7.5 month-old infants segment the speech stream, only that they can. Further studies are continuing to examine the exact mechanism(s) for how speech segmentation is performed.

Metrical Stress

One such mechanism that has been presented and supported as a cue to word boundaries is metrical stress (Cutler & Norris, 1988; Jusczyk, et al., 1999; Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996; Norris, McQueen, & Cutler, 1995). Cutler and Carter (1987) reported that English is a language that is overwhelmingly characterized by initial syllable stress, with 90% of lexical items beginning with strong syllables. This allows for the possibility of using initial syllable stress to mark word boundaries and divide up the speech stream into manageable units that can be analyzed. Indeed, studies have shown that infants and adults are able to segment speech based upon

stress patterns (Cutler & Norris, 1988; Jusczyk, et al., 1999; McQueen, Norris, & Cutler, 1994), even when other potential cues to segmentation contradict stress (Johnson & Jusczyk, 2001). Based upon evidence such as this, Cutler and Norris (1988) proposed a stress-based Metrical Segmentation Strategy (MSS). They proposed this model as an alternative to the left-to-right sequential processing strategy, such as that suggested by Marslen-Wilson (1987), which does not allow following context to impact previous segmentation. Other studies have cast doubt upon sequential models by showing that lexical units within a speech context are frequently recognized only after their acoustic occurrence (Grosjean, 1985). The MSS allows for following context to affect preceding segmentation through identifying likely word boundaries based upon stress patterns. Thus, words such as *run* and *runner* might be accurately segmented using MSS, while a left-to-right processing strategy would incorrectly segment *runner* into *run* and *ner* (Cutler & Norris, 1988). Considering this evidence, sequential processing strategies are no longer actively discussed in the speech segmentation literature.

In a study examining infants' ability to segment speech based upon a metrical stress strategy, Jusczyk, Houston, and Newsome (1999) exposed 7.5 month-old infants to strong/weak and weak/strong bisyllables in fluent speech. They discovered that infants appeared to correctly segment strong/weak bisyllables, but incorrectly segment weak/strong bisyllables. Furthermore, when the distribution probabilities of words were manipulated so that these weak/strong syllables were consistently followed by a specific weak syllable (e.g., "guitar is"), the infants would misperceive these strong/weak words (e.g., "taris"). These results suggest the influence of transitional probabilities on speech segmentation.

Allophonic variation

A second proposed method of segmentation is the ability to use the information provided by different phonetic variants of the same phoneme that are restricted to certain word positions. Thus, the phoneme /p/ in *pat* is an aspirated allophone, but it occurs as an unaspirated allophone in the word-final position in *lap*. Therefore, the speech stream can be segmented by using different allophonic variations as a cue to whether the phonetic variant is in an initial or final position. This information would then signal a word boundary or a word-medial position. Such a strategy relies on the infant's ability to recognize which variants occur frequently or infrequently in certain word positions; thus, the infant must keep track of certain phonetic probabilities, a concept which will be addressed later.

Previous discussions of allophonic variations have proposed that such a strategy is not available immediately, but is only employed after the speech stream has been initially parsed by other segmentation methods (Jusczyk, 1999; Mattys & Jusczyk, 2001b). Therefore, segmentation by allophonic variations may be a reflection of a refining ability to segment the speech stream, and not a primary means (Jusczyk, 1999). Indeed, while metrical stress cues appear to be available by 7.5 months (Jusczyk, et al., 1999), allophonic variation appears to first become available sometime between 9 to 10.5 months (Hohne & Jusczyk, 1994).

Phonotactics

A third means of segmentation examines what sounds are permitted to occur adjacent to each other within a syllable. In English, the cluster /st/ occurs relatively frequently within syllables, while /mg/ is not permitted within a syllable (Jusczyk, 1999); therefore, /mg/ has a high probability of signaling a syllable boundary. Mattys and Jusczyk (2001b) used the headturn preference procedure to assess 9 month old infants' ability to segment the speech stream based upon phonotactic cues. Good phonotactic cues were estimated by computing the frequency of C·C clusters in child-directed speech (where the dot signals a potential word boundary). Two consonants that had a high probability of occurring between words were associated with good phonotactic cues. Results showed that in sentential contexts, infants listened longer to a *CVC* stimulus with good phonotactic cues (meaning that the probability that C·C clusters occurred between words in <u>C·C*VC*·C</u> was high). Similar results occurred when good phonotactic cues only occurred at the onset <u>C·</u>*CVC* or offset *CV*<u>C·C</u> of target words in the utterance, suggesting that infants calculate the probabilities at both onset and offset boundaries.

Mattys and colleagues (1999) also tested phonotactic probabilities between C·C clusters. This time, infants were tested on bisyllabic nonwords (i.e., CVC·CVC). Either the internal C·C cluster had a high probability of occurring at a word juncture or a high probability of occurring within a word. Infants were sensitive to how the phonotactic cues aligned with word boundaries.

Similarly, McQueen (1998) also used aligned and misaligned phonotactic cues to word boundaries in Dutch. His results showed that native Dutch college students are able to detect word boundaries easier when they are aligned with the word boundary, but also suggested that these phonotactic cues are secondary to lexical activation. He presented

the possibility that competing lexical candidates are eliminated by phonotactic violations. Vitevitch and Luce (1999) also concluded that increased phonotactic probabilities facilitate sublexical processing during word competition in adults. However, this method of lexical activation during segmentation is not available to infants or second language learners who have a limited lexicon (unlike the fluent adult speakers in these studies). Therefore, phonotactic cues may also be an advanced, secondary cue to parsing the speech stream.

As with allophonic cues, for infants to successfully use phonotactic information to segment speech, they must calculate the transitional probabilities that occur between different phonemes. The infant can then use these probabilities to determine if the phonemes occur within the same word, or if they are likely to mark a word boundary.

Statistical Regularities

Allophonic and phonotactic segmentation strategies require the infant to monitor the probabilities of allophones occurring in specific word positions and phonemes occurring in certain clusters within syllables. Therefore, for infants to use these strategies, it is likely that an established system is required to monitor these statistical probabilities. Segmentation by allophonic variants and phonotactics are therefore only specific applications of a more general segmentation method based on computing the statistical regularity of information within the speech stream. Also, the segmentation errors of *taris* in *guitar is* (Jusczyk, et al., 1999) reflect that metrical stress cues to segmentation may be mediated by transitional probabilities, as *tar* was always followed

by *is*. Johnson and Jusczyk (2001) also discuss that many of the speech cues that have been proposed are probabilistic in nature and can be framed as statistical cues.

The use of statistical regularities appears to be a domain-general ability for identifying perceptual objects (Kirkham, et al., 2002). Several studies have provided evidence that statistically defined patterns can be used in learning auditory and visual sequences. In the classic study conducted by Saffran, Aslin, et al. (1996), infants were exposed to a continuous artificial speech stream composed of four trisyllabic words. No auditory cues to word boundaries were provided; therefore, the only cues to segmentation were provided by the transitional probabilities between syllables. After only two minutes of exposure to this continuous stream, infants were able to discriminate between words in the artificial language and nonwords (i.e., three syllable patterns that never occurred adjacent to each other during exposure), as well as between words and partwords (i.e., three syllable patterns that spanned word boundaries). This study inspired a line of research supporting sensitivity to transitional probabilities across domains and species (Aslin, Saffran, & Newport, 1998; Fiser & Aslin, 2002; Hauser, et al., 2001; Johnson & Jusczyk, 2001; Kirkham, et al., 2002; Saffran, et al., 1999; Saffran, Newport, et al., 1996). The primitive and ubiquitous nature of statistical sensitivities suggest that a neurophysiological mechanism may underly them.

An explanation of how statistical patterns across domains are implicitly learned might be provided through literature in neural networks and neural plasticity. Research on cross-modal cortical plasticity in ferrets has shown that re-wiring the visual pathway into the auditory cortex, instead of the visual cortex, causes the auditory cortex to not only take over the function of the visual cortex, but to also become arranged into

orientation maps similar to the normal visual cortex (Sharma, Angelucci, & Sur, 2000; Sur & Leamey, 2001). This evidence suggests that experience (the input stimulus) to some degree determines the function of the neural cells which receive the input. In addition, evidence suggests that the more often a stimulus is presented in the environment, the stronger the neural connection becomes which receives that stimulus input (Sur & Leamey, 2001). Thus, higher frequency events are maintained by stronger pathways. The Hebbian rule suggests that if a certain cell repeatedly or persistently stimulates another cell, then a growth process or metabolic change will occur in both or either cell to facilitate future cell firings (Sur & Leamey, 2001).

This evidence relates to statistical learning in the following ways. First, it is the frequency of events which creates strong neural connections within the brain. Therefore, more probable events are more likely to create and travel along well established pathways, facilitating faster and more efficient processing. Second, neural representations in the brain are modified to more closely resemble the incoming stimulus. Thus, it is likely that sensitivity to statistical properties is not detected by a specific area of the brain, but is a generalized property of forming neural connections. In terms of learning segmentation, this may apply in the following way. Continued exposure to certain high probability events (certain syllables occurring together) in the environment cause stronger neural representations to be formed (which over time may result in specific connections). These neural representations are constantly being refined by metabolic or growth processes (i.e. via the Hebbian Rule) to facilitate more efficient firing when the same stimulus is re-encountered. As neural input to some extent guides the development of neural processing (Sharma, Angelucci, & Sur, 2000), these highly

probably stimulus features (the co-occurring syllables) become neurally represented. Thus, statistical learning may be a general neural process that allows for the continued refinement of the brain to interpret environmental stimuli. Kersten and Yuille (2003) discussed a statistical model that suggests feedforward and feedback neural connections (found in primates, see Felleman & Van Essen, 1991). These statistically derived connections allow for higher level processing to refine lower level activity (Kersten & Yuille, 2003), and play a crucial role in perception (Pascal-Leone & Walsh, 2001) and image segmentation (Tu & Zhu, 2002).

The very nature of neurally represented co-occurring stimulus features supports the possibility that these features may become integrated into one neural trace, that of an auditory object. Murray, Kersten, Olshausen, Schrater, and Woods (2002) discovered that grouping features into an object reduces brain activity at lower processing levels; thereby, suggesting easier processing for perceptual objects (Kersten & Yuille, 2003). Reduced activity to grouped objects may facilitate segmentation from ungrouped, novel elements (Murray, et al., 2002), which provides a neural basis for the figure-ground process that Kubvoy and Van Valkenburg (2001) suggested occurs with auditory objects. Clearly, it is expected that auditory objects are processed at a level higher than the processing of their individual features. Indeed, Liebenthal, Binder, Piorkowski, and Remez (2003) discovered behavioral and physiological differences using fMRI between processing spectral information as either individual acoustic features or as a grouped speech object.

Support for the neurophysiology of statistical sensitivity is also derived from animal research; thereby, demonstrating its primitive nature. The use of statistical

information for visual sequences is also an identified ability of pigeons. In 1991, Terrance conducted a study which showed that pigeons are able to combine visual information into one perceptual object when learning serial lists. The ability to combine adjacent items into one object; thereby, reducing the number of discrete items in the list, increased recall ability (Terrance, 1991). As items in the list were consistently presented in the same order, the transitional probabilities between two internal items was $p=1.0$. This statistical regularity provided a mechanism for adjacent items to be perceptually grouped together. This ability to learn sequential information is important for higher organisms to adapt and survive within a temporally bounded environment (Conway & Christainsen, 2001).

Animals not only have the ability to identify sequential visual sequences, but can also use statistical probabilities to identify speech patterns. Hauser, Newport, and Aslin (2001) studied cotton-top tamarins using the same artificial language stimuli Saffran, Aslin, et al. (1996) used in their infant study. Like humans, tamarins are able to discriminate syllable sequences based upon the statistical probability for which they occurred within the speech stream. Tamarins are also able to extract enough information from continuous speech to distinguish between Dutch and Japanese above and beyond speaker variability (Ramus, Hauser, Miller, Morris, & Mehler, 2000).

Statistical sensitivity is also present very early in human development. Kirkham and collegues (2002) habituated the looking times of two-, five-, and eight-month old infants to statistically predictable patterns of visual stimuli. After infant looking times habituated to a continuous stream of randomly ordered visual pairs, the infants were tested on displays of familiar sequences (sequences that maintained the statistical pairs)

and novel sequences (sequence orders to which the infants had never been exposed). They found that when tested later, infants at all ages showed greater interest (longer looking times) to novel visual sequences than familiar sequences. This suggests that infants as young as 2-months are sensitive to statistical properties, as they recognized the violation of a learned statistical pattern. Kirkham and colleagues (2002) suggest that this mechanism may play a role in later cognitive development.

One area where the abilities of humans and animals may begin to diverge is in analyzing the complexity of statistical information. Hunt and Aslin (2001) showed in a visuomotor task that adults can utilize simultaneous sources of statistical information. Infants (Gomez & Gerken, 1999), as well as children and adults (Saffran, 2002), are able to acquire the legal ordering of words in sentences from probabilistic structures, independent of the actual vocabulary used. In addition, infants are able to generalize this statistical structure to novel auditory strings with less than 2 minutes of exposure (Gomez & Gerken, 1999) and adults can learn the statistical regularities between non-adjacent sound sequences (Newport & Aslin, 2004). These studies suggest that statistical information can organize the speech stream beyond simple transitional probabilities and object formation. Statistics not only defines the internal structure of words, but also identifies the grammatical structure of sentences when lexical information is not even available. Therefore, the suggestion that segmentation cues are secondary to lexical activation (see McQueen, 1999) is not necessarily true, but may be an independent and collaborative process. Lexical processing depends upon segmentation to identify auditory objects for lexical access, and later, to statistical properties in identifying the overall grammatical structure of the speech stream.

It is likely that other processes, in addition to the sensitivity to transitional probabilities, are involved in speech segmentation (Johnson & Jusczyk, 2001; Jusczyk, 1997; Saffran, Newport, et al., 1996). Several of these processes may become more available after early structure and auditory objects have been identified using statistical properties. Such processes may be related to lexical activation of neighboring words, or further refinements of statistical sensitivity that go beyond simple transitional probabilities of syllables (such as allophonic variations or phonotactics). Also, this paper does not propose that initially parsed auditory objects are necessarily finalized as real lexical words (though they may be treated as such, see Saffran 2001), but are only an initial parsing of the speech stream. The purpose of this paper is not to discount other viable methods of language acquisition, but to describe the significance that statistical structure plays in early speech perception and how it facilitates language acquisition. Later processes of assigning semantic meaning to auditory objects are required to complete the process of word learning (see Hollich, Hirsh-Pasek, & Golinkoff, 2000, for a further overview of later word-learning processes), and the infant language learner is also engaged in acquiring the structure of objects and events, as well as conceptual structure (Jusczyk, 1997). It may be possible that cross-modal statistical regularities also facilitate the acquisition of word-object associations (Roy & Pentland, 2002). Statistical learning is possibly a very early and primitive pattern detection process which allows for the building of memory traces of frequent events (auditory objects) and later enables the recognition and identification of the same patterns. New research continues to unfold an increasing complexity of statistical learning as an elaborate mechanism for perceiving structure in the environment (Newport & Aslin, 2004).

Adults naturally perceive patterns when presented with continuous stimuli, even when these stimuli are presented at random with no intended pattern (Huettel, Mack, & McCarthy, 2002), and infants are able to implicitly detect these patterns (Saffran, Johnson, Aslin, Newport, 1999). Implicit knowledge results from instilling an abstract representation of the stimulus structure (Reber, 1989), a structure which forms the perceptual object. They are implicitly learned through an automatic and natural process of statistical learning, which may arise from brain neurophysiology.

The review of these studies indicates that the use of statistical information is an ability that spans across species and cognitive/sensory domains, allowing for the discrimination and segmentation of stimulus patterns, even those as abstract as grammar. Infants as young as two-months of age are able to make use of statistical patterns to order their world (Kirkham, et al., 2002). Studies of neural plasticity suggest stronger neural connections are formed to process more frequent events (Sur & Leamey, 2001). Neurophysiological changes may facilitate object segmentation (Murray, et al., 2002) and reduce the processing required (Kersten & Yuille, 2003). As statistical properties are a generalized mode of learning (and perhaps a general property of neural organization), it is no surprise that it is also used to identify syllabic patterns in segmenting speech appropriately into words. Even other suggested methods of speech segmentation, namely allophonic and phonotactic cues, are founded on the properties of statistical patterns. Moreover, an early method of segmentation based upon stress is influenced by probabilistic patterns.

Electrophysiology of Auditory Sequences

This paper suggests that the formation of auditory objects within a continuous speech stream is necessary to identify individual words. The electrophysiological literature offers great insight into the processes of how these auditory objects may be formed, and is therefore worthwhile for our review.

In an effort to identify the perceptual processes necessary in the formation of auditory objects, the structure of auditory patterns has also been investigated using event related potentials (ERP). These auditory objects are analogous to those that the pigeons formed in Terrance's tasks (1991) or the auditory sequences that are identified as words due to high transitional probabilities. The following is a discussion supported in the adult electrophysiological literature.

The ability to segment speech based upon transitional probabilities is the discrimination between two units that occur together frequently (in this case, syllables that occur within words) from two units that occur together relatively infrequently (syllables spanning word boundaries). A parallel for this process is present in research examining an ERP component called the mismatch negativity (MMN), an index of auditory discrimination independent of attention (Naatanen, 1992, 1995). The oddball paradigm is typically used to elicit the MMN by presenting a highly probable stimulus (called the standard) and a highly improbable stimulus (the deviant). The response elicited by the standards is subtracted from the response to the deviant, creating a difference wave that peaks at 100 to 200 ms after stimulus onset.

In this paradigm, the duration of time between the offset of one stimulus and the onset of the next stimuli is referred to as the Stimulus Onset Asynchrony (SOA), or sometimes as an Inter-Stimulus Interval (ISI). The logic behind using SOAs is that by

delaying presentation, decay occurs in the neural representation of the standard (Naatanen, 1992). Also, when SOA durations are reduced, stimuli elicit larger MMN amplitudes at the faster presentation rate (Naatanen, 1992). It is also possible that longer SOA times assist in providing a temporal gap for dissociating stimuli. During continuous speech, there are relatively few SOAs (acoustic breaks/silences) between syllables.

Saffran, et al. (1999) reported that adults can learn auditory sequences, even when uninformed that the stimulus stream contained units and were instructed to avoid analyzing it. Therefore, infrequent probabilities within the speech stream may provide the cues to segmentation independent of attention. The pertinence of the MMN in the study of speech segmentation through statistical means is made stronger by recent evidence that suggests the MMN may be more closely related to auditory regularities than to deviance detection (Winkler & Czigler, 1998). The MMN is also sensitive to the degree of probability, with higher probable standards eliciting larger amplitudes (Naatanen, 1992; Naatanen, Sams, Jarvilehto, & Soininen; 1983; Sato, Yabe, Hiruma, et al., 2000).

Several recent studies using the MMN have begun to examine detection of deviancy within a repeating auditory pattern or patterns. The investigation of pattern violation detection is important to the understanding of auditory object formation and subsequent segmentation from the acoustic environment. It is these auditory patterns that are formed into gestalts, or auditory objects (see Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995 for a discussion of auditory stimuli stored as gestalts in a distinct physiologic form). The detection of a deviant event within an acoustically varying environment (such as that presented by many of these auditory patterns) requires that

these patterns be identified; an auditory object, a figure, needs to emerge from the ground. Violation to this figure can then be readily perceived. The elicitation of the MMN requires detection of the deviant event; and therefore, perception of auditory objects within the varying acoustic environment.

Vaz Pato, Jones, Perez, & Sprague (2002) presented adults with synthesized musical instrument tones in a continuous sequence of four and of five tones in a repeated rising or pseudorandom pattern at a rate of 16 tones/second. Deviant tones occurred after every 20 tones and were of higher pitch. A MMN was elicited to the deviant immediately following the standard pattern. These results suggest the segmentation of sequential information from repetitive and pseudorandom patterns. In this study, the standard was not individual tones, as was the case in many early studies examining the MMN, but was a repeated sequence of tones. The elicitation of the MMN occurred with a violation of this pattern (due to presentation of the deviant, a higher pitched stimulus), suggesting two conclusions. First, the repetitive standard pattern was recognized as one auditory object due to the high probability of those tones occurring together. Second, it suggests that these auditory objects can be formed in the environment of continuous auditory streams, independent of other acoustic cues such as pauses indicating the end of the sequence.

The concept of auditory object formation has further support in the literature. Atienza, Cantero, Grau, Gomez, Dominguez-Marin, and Escera (2003) showed that sequential grouping of auditory stimuli does occur. Also, memory traces of the MMN are not only formed for individual tones, but are also present for larger acoustic events that contain several sounds and span larger temporal scales. This supports the results of Vaz

Pato, et al. (2002) which suggested that patterns of auditory stimuli presented as standards can be grouped perceptually as auditory objects, and Winkler, Schroger, and Cowan (2001) also support the concept of preperceptual organization for auditory events.

Winkler, Cowan, Csepe, Czigler, and Naatanen (1996) demonstrated that these auditory objects are preserved over the course of different presentations of the auditory object, and do not need to be recalculated each time the auditory sequence is encountered. Trains of 6 tones (five standard tones and one deviant) were presented to subjects with an intertrain interval of 9.5 seconds. If the pattern needed to be reestablished each time, then the initial presentation of at least two or three standards would have been required to detect a deviant. However, Winkler and colleagues found that a single reminder of the standard tone reactivates the representation of the standard stimulus; a deviant in the second position elicited the MMN. These results indicate that a sensory memory trace, indexed by the MMN, is able to store complex auditory events for extended periods of time; thereby, allowing for the establishment and maintenance of auditory objects. This paper proposes that these auditory objects are not transient sensory traces subject to almost immediate decay, but remain established in a memory store for later comparison and organizational processes mediated through statistical means.

Auditory objects can also include other information, in addition to the spectral patterns discussed thus far. Alain, Cortese, and Picton (1999) presented a continuously repeating sequence of four tones where deviants varied in either frequency or timing. Both types of deviants generated the MMN with identical scalp topography, suggesting that spectral and temporal relations of auditory patterns are encoded in a unified memory trace (Alain, et al., 1999).

Presented thus far is electrophysiological evidence that individual acoustic events can be grouped into larger auditory objects on the basis of the statistical regularity of their pattern, that this object formation occurs within continuous presentation environments requiring the identification and segmentation of individual auditory patterns, and that auditory objects can simultaneously include spectral and temporal information. Furthermore, the MMN, a preattentional index of auditory memory, is sensitive to probabilities, detects these auditory patterns, and is preserved over time for use on later pattern detection and comparison processes.

Also, multiple acoustic patterns can be maintained at the same time (Brattico, Winkler, Naatanen, Paavilainen, & Tervaniemi, 2002). Brattico, et al. presented two different standard sequences of four tones. Unlike many previous studies that used a different auditory stimulus for the deviant, they constructed their deviant from the first two tones of one standard and the last two tones from the second standard. Therefore, the deviant constructed was not a violation of the auditory event, but was a violation in pattern (similar to the partwords used to detect segmentation by Saffran Aslin & Newport, 1996). Only a system capable of recognizing and organizing these patterns into perceptual objects would be capable of detecting this deviance. The MMN was elicited by this pattern deviance in the absence of a change in the acoustic environment. These results also indicate that multiple sound patterns can be maintained during an automatic stage of auditory processing, as two different standard patterns were maintained. This clearly demonstrates the formation and maintenance of multiple auditory objects that were segmented from the presentation of a continuous auditory stream. This study

warrants further investigation of pattern deviancy detection and multiple sequence maintenance with more complex auditory signals, such as speech.

While studies using pure-tone stimuli have provided an excellent discussion of auditory objects, Rauschecker (1998) discussed their limitation. These studies are unable to examine how complex sounds are re-assembled, lead to perceptual or cognitive performance, or form memory traces for the recognition of complex sounds (Rauschecker, 1998). The current study investigates these issues within a neuropsychological framework.

Electrophysiology and Neuropsychology of Speech Segmentation

Speech segmentation studies have been limited to an investigation of segmentation ability after the speech stream has already been parsed using the proposed speech cues. These studies are therefore unable to establish the time course of the perceptual processes performed (Sanders, Newport, & Neville, 2002). Few studies have attempted to investigate the online perception of these cues, or the process of identifying the auditory objects (word-forms) within the stream. (This paper uses the term word-form to refer to auditory objects that are the prelexical auditory structure of words.) These issues have not been addressed because relatively few behavioral methods have been used to investigate speech segmentation, and these have all been limited to studying offline detection of isolated words. Offline tasks use stimuli isolated from the natural context of a continuous speech stream to assess speech segmentation. This project will investigate segmentation through online protocols that embed the stimuli within continuous presentations.

The concept of the word as an auditory object is not new. As early as 1970, Hayes and Clark discussed the word as a perceptual object of language. From their experiments, they concluded that a primitive mechanism for clustering continuous speech into words exists for segmentation. They state that "this mechanism appears to operate only on the information available in the intercorrelations between successive sounds in the speech stream; it identifies as words the clusters of sounds that consistently recur in an unbroken sequence" (Hayes & Clark, 1970, p. 233). Therefore, the concept of using statistical information to unite individual acoustic events into perceptual word-forms is long standing. Even as early as 1899, Bryan and Harter discussed how repetitious exposure could train people to perceive the individual tones in morse code as whole words and phrases.

Electrophysiological methods have the advantage of directly indexing the neural correlates of perceptual abilities. The results of these studies provide great insight into how the perceptual abilities of the brain work. Therefore, evidence from these studies offer a neurophysiological framework with which to integrate behavioral evidence into a working theory of the mind. This study attempts to take electrophysiological evidence of auditory object formation and explain how it applies to a behavioral investigation of speech segmentation. It is from an accurate understanding of the neurophysiological correlates involved in perception that accurate neuropsychological test protocols may be developed to behaviorally examine these processes in higher-order tasks such as speech segmentation, and from these results infer neuropsychological mechanisms.

Electrophysiology

Perhaps the electrophysiological study most relevant to speech segmentation was conducted by Sanders, et al. (2002) in an adaptation of a behavioral investigation done by Saffran, Aslin, et al. in 1996. The purpose of this study was to confirm that larger N100s are evoked by word onsets. They recorded ERPs before and after participants had learned to segment an artificial speech stream identical to that created by Saffran, Aslin, et al. (1996). Sanders, et al. (2002) discovered a larger N100 for high learners (participants with the greatest behavioral change in segmentation performance) after as compared to before training participants to segment. This study by Sanders, et al. (2002) suggests that early ERPs can index perceptual changes developed through learning and provides an electrophysiological correlate to measure segmentation ability. However, this study did not address the ability of incidental language learning shown in Saffran, Aslin, et al.'s (1996) study, as Sanders, et al. (2002) actively trained participants on the correct segmentation. Also, a larger N100 in this study may be more related to perceptual features at word onsets rather than an actual formulation and detection of auditory objects necessary for word identification. In order to examine this, the actual formation of words as entire auditory objects needs to be assessed; an investigation that can be achieved using the mismatch negativity. This claim is based upon the precedent of using MMN in the study of other similar auditory patterns.

An overlooked area in behavioral speech segmentation research is how perception might change after successful parsing of the speech stream is learned. As the study by Sanders, et al. (2002) indicates through an electrophysiological paradigm, the perception of the speech stream can be recorded before word-forms are identified and then compared to perception after segmentation properties are learned. Such a study would be able to

examine how learning word boundaries might influence perceptual processing.

Experience has already been shown to influence the MMN. In a study by Naatanen, Schroger, Karakas, Tervaniem and Paavilainen (1993) a standard sequence of 8 consecutive segments of different frequencies was presented. In the deviant condition, the sixth segment had a different frequency. During the beginning of the study, the MMN was not elicited to the deviant. However, after continued exposure, participants who learned to discriminate the sequences revealed a MMN to the deviant at the end of exposure. This demonstrates that the MMN can index changes in the ability to discriminate auditory sequences. Evidence such as this can be built into neuropsychological paradigms to measure changes in performance ability.

Neuropsychology

Electrophysiological investigations add to our understanding of the perceptual processes involved in speech segmentation by providing a framework for the time-course of segmentation properties and the examination of these properties online during auditory object acquisition. Evidence from electrophysiological studies provides a larger picture of the perceptual processes that underlie speech segmentation. The unique abilities of electrophysiological methods can be adapted to fit a neuropsychological paradigm. Speech segmentation abilities have been measured using high amplitude sucking, headturn preference procedures, and two-interval forced choice paradigms. However, these methods are only able to show that acoustic or statistical cues were identified during exposure and that these properties are detected later when presented in isolation. They lack the ability to examine the perceptual processes that support the acquisition of

novel words within the speech stream. Though they have shown speech segmentation through various mechanisms, they also lack an adequate theory of how the segmented syllables are clustered into single, unitary objects that may later become lexicalized. This is despite the evidence that these perceptual objects, segmented and formed from the speech stream, are treated as English words (Saffran, 2001).

The distinction between processes of discrimination, recognition, and identification must also be discussed briefly. Griffiths (2003) organized complex sound perception into a processing hierarchy. The levels are: 1) the simple processing of acoustic features, 2) complex processing of temporal, spectral, and spatial patterns (which is the level of the auditory object), and 3) semantic processing (or the symbolic use of auditory objects). It is at this last level that meaning is associated with prelexical objects. Craik and Lockhart (1972) proposed that memory is a bi-product of the level of processing. Thus, different processing may recruit different memory stores. Discrimination, recognition, and identification are interactive and related, but separate levels of processing that are involved in speech segmentation and auditory object formation. Burnham, Earnshaw, and Quinn (1987) state that "...different neural and cognitive processes may be involved in discrimination and identification" (p. 256).

Discrimination is the detection of any difference between two tokens (Burnham, Earnshaw, & Quinn, 1987). Thus, discrimination is a very early process that enables later perception, but does not necessarily lead to it. The method of a two-interval forced-choice task is a discrimination paradigm in which the participant detects the difference between two stimuli. Other methods, such as the head-turn preference procedure or high

28

amplitude sucking, also indicate this preperceptual process of detection. Discrimination of stimuli uses an early sensory memory store (Winkler, et al., 1996).

Recognition on the other hand, is the implicit knowledge that a stimulus has been encountered before (Galotti, 1999). It is the early implicit detection of a recurring structure of stimulus features. Recognition is at higher level than discrimination as it is no longer preperceptual detection, but requires the integration of features into an early structure. According to Craik and Lockhart (1972), this higher processing level would recruit memory resources different than the early sensory memory store implicated in discrimination.

Identification "…involves first the abstraction of common bases for grouping sounds together, and then the comparison of each token with this abstraction (Burnham, Earnshaw, & Quinn, 1987, p. 256). This is no longer a lower, preperceptual process like discrimination, or an implicit detection of pattern structure like recognition. Identification requires an abstraction representation of the object. It examines the holistic characteristics of the stimulus and requires an explicit knowledge structure; or in other words, identification labels the stimulus as an object. This is a higher-order process that recruits memory resources for maintaining an abstract representation of the object, which is penetrable to cognitive tasks.

This study proposes to measure detection, by replicating previous studies, and build upon them by also measuring recognition and identification in two different online tasks where the participant must sort through distraction stimuli to perceive the learned words online. These later two conditions implicate different requirements of memory that are associated with their level of processing.

The neuropsychological methods designed for this project measure recognition and identification of word-forms before and after learning segmentation, examine segmentation abilities online, and investigate the formation of words as auditory objects in a recognition task. This paper makes the distinction between online and offline tasks as such: offline tasks extract the stimulus from the continuous speech stream and present it in isolation; whereas, online tasks present the stimulus within the speech stream. Thus, online tasks more closely model perception during natural speech contexts and actively require speech segmentation processes to occur during testing. Offline tasks already have the stimulus segmented during the test presentation. This study also utilizes a cognitively more complex identification task that requires phonological synthesis (the conceptual reformation of a word-form from its individual elements) of these stored auditory objects. Future studies may wish to examine the electrophysiological substrates that are involved in these processes.

This paper proposes that the online recognition and identification of words suggests successful speech segmentation. This furthers earlier reports using isolated stimuli in discrimination tasks; such tasks may index the detection of familiar patterns or the presence of a less probable feature (a word-boundary), rather than the identification of a full auditory object – the word (Fiser & Aslin, 2002). This is the difference between feature detection and gestalt processing abilities. The current project examined this difference by using tasks structured on a hierarchy of processing difficulty.

The first level of this processing hierarchy examined is discrimination. A two-interval forced-choice (2IFC) paradigm was used as an offline discrimination task like that used by Saffran, Newport, et al. (1996) to confirm the results of this study in

30

comparison to previous studies. Successful performance on this test would confirm the occurrence of segmentation. As stated earlier, this paradigm provides evidence for a feature detection of statistical segmentation cues that were learned during exposure. Unlike the next two testing paradigms, this task does not require gestalt auditory processing (as in the Auditory Probabilistic Word Count) or higher cognitive tasks such as phonological synthesis (as in the Slow Online Segmentation Test), but is a relatively early process of discrimination and can only infer word boundary detection, not word-form learning. Both of these tasks involve higher-order, online processing of learned auditory objects. The combinations of these testing methods thus allow for the dissociation of segmentation, object perception, and the hierarchy of perceptual/cognitive processes.

The Auditory Probabilistic Word Count (APWC) is an online measure of recognition, the second level of the hierarchy. This task was designed from the probabilistic paradigms of ERP studies to examine detection of low occurrence auditory objects (from the artificial language) within highly frequent randomized syllables. Thus, this task is very similar to odd-ball paradigms used for auditory deviance detection within continuous complex auditory sequences. The participant's task is to count the number of artificial words (learned from the exposure period) that they hear within a continuous speech stream presented at a normal rate and composed mostly of randomized isolated syllables. This method conjectures that if memory traces for the exposed words were formed during exposure using the statistical distribution of syllables (which required word segmentation), then this same probabilistic information should be accessible to successfully segment the learned words in a novel language exposure with distracters

(which are randomized isolated syllables from the exposed language). To accurately detect the learned words, the participant must have already integrated the statistical distributions of the syllables into a implicit, early structural formation of a perceptual object (word-form). As such, this is a measure of auditory gestalt processing. An accurate count across subjects is dependent upon successful segmentation performance.

The third level of processing was assessed using the Slow Online Segmentation Test (SOST). This protocol (described in further detail in the methods section) requires the participant to identify word boundaries online during the continuous presentation of isolated syllables that follow the same probabilistic structure of the exposure speech stream. In order to perform successfully at this task, participants are required to combine individual speech syllables into one of the auditory objects learned during exposure. They then must identify the object boundary within a several sentence-length presentations of other syllables belonging to other learned objects or used as distracters. Support for this method comes from the phonological processing literature that is typically used to investigate early reading abilities. However, phonological processing abilities (specifically phonological awareness and memory) are arguably essential for the timely acquisition of a novel language. Phonological awareness is the explicit knowledge of the language's phonological features (Bradley & Bryant, 1985). It is the awareness of individual phonemes that comprise words and requires the same level of mental representation that is required for identification processing. The auditory objects, in contrast to discrimination and recognition processing, need to have an explicit symbolic structure for identification, and may imply the recruitment of prelexical semantic memory resources.

Phonological synthesis, typically studied using Sound Blending and Nonword Blending subtests in neuropsychological test protocols, is the ability to combine an isolated set of phonemes together into a recognizable word (Torgesen, Morgan, & Davis, 1992; Wagner, Torgesen, Rashotte, Hecht, Barker, Burgess, et al., 1997). For an example of a phonological synthesis task, the participant might be asked, "What word is this: /k/, /æ/, /t/?" The correct response would be "cat". Evidence suggests that the syllable is the unit that children first become phonologically aware of (McClure, Ferreira, & Bisanz, 1996; Wagner, et al., 1997), which complements the current study's use of syllables as the individual features of the novel words.

Tests of phonological awareness and auditory processing have traditionally used phonological synthesis as an index of how we integrate and say whole words (CTOPP; Wagner, Torgesen, & Rashotte, 1999; WJ-R; Woodcock & Johnson, 1989, 1990). Wagner, et al. (1997) also used Sound Blending and Nonword Blending to measure phonological synthesis. Though typically measured as correlating with reading ability, this synthetic process enables learning and is not reciprocally facilitated by later acquisition of reading. It is described as an aspect of phonological sensitivity which is a very basic recognition of the phonological elements in oral language (Wagner, Torgesen, & Rashotte, 1994). Thus, it is appropriate to use in an oral language task to study integrating isolated syllables into an auditory object.

The development of the SOST was based upon this concept. In this task, the syllables, presented at a slow rate, must be phonologically synthesized into words (as in the Sound Blending and Nonword Blending tasks). Torgesen and Morgan (1990) discussed a model of phonological synthesis outlined in a presentation by Perfetti, Beck,

and Hughes in 1981. The model suggests several processes that are involved in phonological synthesis. The most basic component is awareness that individual sounds can be combined in various ways to form words. But, these sounds must also be represented and stored in memory and then combined to form a word-like representation. However, during the SOST, participants must be able to do these processes online during the continuous presentation of sentence-length speech streams. Thus, in addition to phonologically synthesizing these syllables into words corresponding to the auditory object memory traces, the participants must also be able to segment these words from the speech stream. Success at this task would support statistical speech segmentation, as well as, the formation of auditory objects. (A standardized screening of phonological memory and phonological awareness were conducted with participants to compare each participant's phonological abilities to a normative sample.)

Auditory perception is influenced by presentation rates, as reflected in electrophysiological studies which vary SOA times (Naatanen, 1992). By testing in a temporal window that differs from exposure rates, the stimuli are no longer perceived in the same relation to each other. Therefore, the successful identification and segmentation of the learned words is dependent upon first synthesizing the independent syllables. This gestalt processing of auditory sequences has been shown to influence the perception of auditory spaces by perceptively shrinking the amount of time between stimuli (Sasaki, Suetomi, Nakajima, & Hoopen, 2002). If only the statistical features are learned during exposure, and no memory traces for perceptual objects are formed, then the isolated syllables will not be able to be phonologically synthesized appropriately within the continuous speech stream.

While the 2IFC was an offline discrimination task, the APWC and SOST assessed participants' ability to segment speech online, during continuous presentations. These online tasks were measured through a pretest-posttest design. Pre- and post-exposure tests are relatively straightforward. As discussed, electrophysiological studies have provided evidence for how perceptual processes change as a result of exposure to stimulus patterns (Naatanen, et al., 1993). In the current study, any perceptual changes that occur are a result of the exposure to the probabilistic structure of the artificial language. Improved performance is therefore related to changes in perceptual ability. This paper predicts that a wider distribution of scores and poorer accuracy will be measured during the pretest as opposed to measurements taken after exposure. It also predicts that processing for the artificial language will become more like native language processing. This will be assessed through comparisons with performance on isomorphic tests using English words.

While these English tests measure native processing abilities, their primary purpose is to serve as a measure of external validity for the APWC and SOST neuropsychological measures. The participants performed these two tasks with low frequency English words. Low frequency words reduce the confound of lexical access facilitating performance (which is not available in the experimental tasks). If the participants were unable to successfully perform these tasks with English words, then they would almost certainly perform poorly with the experimental stimuli. However, if they were able to perform accurately in English, then they should also have the potential to perform accurately on the experimental stimuli, given that they have been able to

segment the artificial language using the statistical distributions. These English tests were considered a measure of optimal segmentation ability.

Research on phonological processing has also revealed differences in the processing and even localization of function between males and females. Data from functional imaging studies has revealed more bilateral activation during phonetic processing tasks in females; whereas, males primarily activated the left hemisphere (Pugh, Shaywitz, Shaywitz, Constable, Skudlarski, Fulbright, et al, 1996; Shaywitz, Shaywitz, Pugh, Constable, & Skudlarski, 1995). Structural MRI studies have also revealed that the left planum temporale is significantly larger in males than females (Kulynych, Vladar, Jones, & Weinberger, 1994), which is believed to be involved with language components that require rapid temporal processing (Lambe, 1999).

Neuropsychological studies have also discovered differences in processing. In dichotic listening tasks, a significantly stronger right ear advantage occurs for right-handed males than for right-handed females (Lake & Bryden, 1976). The right ear advantage is also produced for stop consonants which have rapidly changing acoustic cues, rather than for steady state vowels (Fitch, Miller, & Tallal, 1997). Coney (2002) also observed that females are faster to respond in a nonword rhyming task in which successive rhyming words were presented to different visual fields, which is consistent with the view that females may possess "a greater facility for dual hemispheric processing in phonological operations" (p. 363). These results of anatomical, functional, and behavioral differences in the ways males and females processes phonological information led Lambe (1999) to recommend the separation of males and females in research studies to minimize variance; thus, enabling subtle differences that are

functionally relevant to be detected. In following this suggestion, this study only used females as participants, with the hope that future studies will be able to examine a more diversified population.

Proposal for the Current Study

This study is designed to address several questions within the speech segmentation literature. It examines online auditory object perception related to the statistical distribution of syllables within and between artificial words. The neuropsychological protocols developed measured changes in perceptual ability during the acquisition of speech segmentation abilities through pre- and post-learning measurements. Such evidence would suggest that perceptual processes change due to exposure to the statistical probabilities inherent in the speech stream.

Current trends in the literature support the rationale for investigating these issues using more advanced behavioral methods. While continued support for statistical learning from the behavioral literature is established, an expansion of the neuropsychological methods is required to more fully investigate this phenomenon.

The current proposal investigates these issues through the use of a neuropsychological paradigm. Participants will listen to a continuous auditory stream of an artificial language analogous to the speech stream created by Saffran, Aslin, et al. (1996).

The current study was designed to examine the following areas:

1) Segmentation of complex patterns via statistical regularities

This study will test the use of statistical regularities in analysis of complex stimulus patterns, namely speech syllable distributions, and to segment those patterns within a continuous auditory stream that offers no acoustic breaks.

Participants will be exposed to 3-syllable artificial words that are concatenated together in a pseudo-random order. The only cue to word boundaries will be the statistical regularities between syllables. Syllables within words will always occur together, $p=1.0$. Syllables between words will only occasionally occur, $p=0.33$. Participants must therefore learn the statistical regularity to segment the artificial words.

2) *Replication of discrimination studies*

This study will confirm the ability to discriminate segmented words from foils by expanding upon previous studies.

The Two-Interval Forced Choice paradigm has been used in the literature to indicate that speech segmentation did occur during exposure and that participants are able to identify exposed words in an offline discrimination task (Saffran, Johnson, et al., 1999; Saffran, Newport, et al., 1996). This paradigm will be replicated to confirm results with previous research and identify the particular words that have been learned. It also tests the ability of statistical probabilities to influence a lower-level feature-detection process of discrimination.

3) *Object recognition and identification*

This study will investigate the online ability to recognize and identify auditory objects.

38

Participants' ability to identify the artificial words online during continuous presentation of the speech stream will be measured using the Auditory Probabilistic Word Count and Slow Online Segmentation Test. These tests examine the ability of statistical probabilities to influence higher-level gestalt processes of recognition and identification.

4) *Changes in performance due to enhanced perception*

This study will identify changes in performance related to an enhanced ability to perceive perceptual units.

Participants' ability to identify artificial words will be tested pre- and post-exposure to the continuous stream using the Auditory Probabilistic Word Count and the Slow Online Segmentation Test. Low accuracy scores pretest and significantly improved scores posttest would indicate learning the perceptual structure of the artificial words.

5) *Manipulability of auditory object memory traces*

This study will examine the manipulability of memory traces for the learned auditory objects in higher-order processing.

Temporally separated syllables will be presented to the participants at a continuous rate. As temporal information affects our perception of auditory events, for the participants to identify word boundaries they must first phonologically synthesize the individual syllables into the learned auditory objects maintained as memory traces from exposure. Also, the perceptual hierarchy of discrimination, recognition, and identification will examine what representation levels auditory objects can influence.

The present study investigates these perceptual changes within an artificial speech stream. Neuropsychological protocols will measure perceptual processes pre- and post-segmentation learning. This study hypothesizes that the statistical properties of speech will provide for the formation of auditory objects; thereby, influencing perceptual states and enabling successful behavioral performance in identifying artificial words. Therefore, a quantifiable difference in behavioral performance is predicted to occur before and after speech segmentation is learned.

CHAPTER II

Methods

Participants

Twenty-two college students participated in this study; however, two subjects were eliminated from analyses due to demonstrated and reported fatigue by one subject, and failure to complete the testing protocols by the second. Implications for these exclusions are addressed later in the discussion of this paper. Prior to selection, participants completed a short survey via phone or in person on previous language/auditory training and a medical history. Criteria for exclusion included: fluency in a language other than English; previous speech, language, hearing, or reading difficulties; or presence of a developmental disorder. All participants were monolingual English females by self-report and were either paid for their participation or received course credit. Females were selected due to the demonstrated anatomical, functional, and behavioral differences between males and females that have been demonstrated in the literature (see the introduction for a discussion). All participants passed a bilateral pure-tone hearing screening at octave frequencies 500 to 8000 Hz presented at 20 dB HL (ANSI, 1996) and had otoscopy performed to verify the ear canal was free from occlusions.

To be included in the study, all participants were required to achieve standard scores of an 8 or above (representing average or above phonological processing abilities) on Memory for Digits and Blending Nonwords subtests of the Comprehensive Test of

Phonological Processing (CTOPP, Wagner, Torgesen, & Roshotte, 1999). Participants also completed the Nonword Repetition subtest to achieve a Phonological Memory Composite score. Blending Nonwords is an acknowledged test of phonological synthesis, while Memory for Digits and Nonword Repetition are recognized measures of Phonological Memory (Wagner, Torgesen, & Roshotte, 1999). They also completed the Matrices subtest of the Kaufman Brief Intelligence Test which is a nonverbal measure of fluid thinking (K-BIT, Kaufman & Kaufman, 1990).

Stimuli

Speech Synthesis

Speech syllables were created using the built-in speech synthesizer on a Macintosh OSX system. The speech output was synthesized using a MacInTalk 3 female voice (Kathy) and converted to a digital audio file using Voice Box 1.4. Audio files were then digitally edited to control for amplitude, fundamental frequency, and duration of each syllable. Edited syllables were strung together to form the three-syllable words using a waveform editor and saved as .wav files. Distracter syllables used in the neuropsychological protocols were individual syllables from the medial position of the words. All syllables were 333 ms in duration, had a fundamental frequency of 186 Hz, and an amplitude with an average root mean squared (RMS)=.2200 during the steady state of the vowel and a RMS=.1774 across a concatenation of all 12 syllables. English stimuli consisted of low frequency 3-syllable English words (as provided in the *Word Frequency Book,* Carroll, Davies, & Richman, 1971) to more closely model the artificial language and were also synthesized using this method. Low frequency words were

determined to be both a better control and a measure of optimal performance as they reduced the confound of lexical activation assisting in segmentation.

Pilot data on these syllables was collected using an identification task where participants transcribed each syllable that they heard. The experimental syllables were repeated five times with a 0.167 second ISI before the participant was required to identify the syllable. Each experimental syllable was identified three times for a total of 36 items. Results of this pilot study are discussed in the following chapter.

Language Construction

Speech syllables were assembled to form four tri-syllabic words, as in Saffran, Aslin, et al. (1996). Two counterbalanced conditions were created in order to control for any perceptual biases unrelated to the experimental investigation (Language A: *daropi, pabiku, golatu, tibudo*; Language B: *bikuti, tudaro, pigola, budopa*). Both languages were constructed identically according to the methods described here. An equal number of participants heard each language. Participants listened to the same language during each exposure period and during testing.

Tri-syllabic words were concatenated together in pseudo-random order to form three different blocks of 7-minutes each with the stipulation that the same word never occurred twice in a row. This created a total of 21 minutes of exposure to the probabilistic structure of the artificial language. No pauses or other acoustic cues to word boundaries occurred between syllables. The only cues to word boundaries were the statistical probabilities between words. An orthographic representation of the resulting speech stream is as follows: *pabikutibudogolatudaropi....*

Transitional Probabilities

In Saffran, Aslin, et al. (1996) the internal probability between syllables within a word (e.g., *pabi* and *biku* in *pabiku*) was *p*=1.0 because *pa* was always followed by *bi* (and *bi* always by *ku*) in the speech stream. The external transitional probabilities used by Saffran, Aslin, et al. (1996) for syllables spanning a word boundary (e.g., *ku#ti* in *pabiku#tibudo*) were *p*=0.33. This was because the final syllable (in this case *ku*) could be followed by any of three different initial-position syllables (*ti, go,* or *da*). This external probability of *p*=0.33 was maintained in the present study, as the same four final syllables were each followed by one of three initial syllables, and as a word could not follow itself.

Neuropsychological Test Materials

The *Auditory Probabilistic Word Count* (APWC) consisted of a speech stream of approximately 3 minutes in duration and was presented at the same presentation rate as the exposure speech stream (i.e., 3 syllables per second). Syllables that occurred in the medial position of the exposed words were randomly presented with the 3-syllable words from exposure (either from Language A or Language B). The four exposed, 3-syllable words had a 20% chance of being presented, as lower probabilities yield a stronger deviant, and therefore allowed for easier detection of the attended stimulus (the word). Five different presentations were generated: one English version, and a pre- and posttest each for Language A and B.

The *Slow Online Segmentation Test* (SOST) consisted of 25 isolated speech

streams that were approximately sentence length. Each stream consisted of 15 syllables

which formed 4 words and 3 distracter syllables (to prevent pattern responses) for a total

of 100 words to identify. Syllables were presented at a rate of 2 syllables per second (one

every 500 ms), which is the rate that has been used in phonological synthesis tasks

(Wagner, et al., 1997), and were in a pseudo-random order (with the stipulation that no

stimuli occurred twice in a row). The distracter syllables were from the medial position

of the exposed words; thus, *bi* became a distracter from the word *pabiku*. Each speech

stream started with a tone (a sine wave of 1 second in duration) followed by one second

of silence, and ended with a 500 ms pause immediately followed by a series of 5 beeps

(five triangle waves at 150 Hz with a duration of 200 ms each). An interval of 4 seconds

passed between streams. These prevented participants from losing their place during the

continuous presentation and to control the amount of time available for stream analysis;

as this test was designed to measure online processing, not offline reanalysis of the

stream. Again, five test streams were created: one English version, and pre- and posttests

for the two artificial languages. A response sheet containing 25 rows of boxes (15 boxes

per row) was created for participants, each box represented one syllable and each row

corresponded to each speech stream.

Design & Procedure

Participants attended one session that lasted approximately 120 minutes, including

scheduled breaks. The hearing screening and all experimental tasks were completed in a

sound attenuated room. All protocol directions were audio recorded and presented prior

to each testing section, in addition to the written transcript being provided at the beginning of each protocol response form. All stimulus presentations and test protocols were transferred onto a single CD-Rom, played on a KLH 5-disc CD changer player (DA1602), and presented via ER-3A 50Ω insert earphones with E-A-R LINK foam eartips positioned by the experimenter. The system was calibrated weekly to present the syllables within a 2 dB difference of 60 dB SPL using a Larson Davis Laboratories precision integrating sound level meter (Model 800B). The calibration was done using a sine wave generated at the syllable fundamental frequency of 186 Hz and RMS of 0.22. This calibration setting was then checked during a continuous presentation of the syllables. A biologic check was also performed weekly.

Participants first completed the standardized test screenings, followed by three training tasks designed to familiarize the participants with the synthesized English words and teach them how to complete the SOST and APWC tests. These training tasks were performed before any experimental testing began.

All participants completed pretests of the APWC and SOST. Then, they listened to 3 blocks of a 7-minute artificial language exposure (described above) which was interspersed with English versions of the APWC and SOST. Next, participants received posttests of the APWC and SOST and completed testing with the two-interval forced-choice task which replicated Saffran, Newport, et al.'s (1996) method. Pretest and posttest versions of the APWC and SOST were counterbalanced with each other. The APWC always preceded presentation of the SOST as this was viewed as a possible initial exposure period. (The APWC was selected for first presentation as it did not require the explicit identification of word-forms that the SOST requires). Two 5-minute breaks were

suggested and made available to participants, though they were not mandatory: one occurred following the pretests and the other occurred right before the third exposure to the 7-minute speech stream. The entire testing session lasted approximately 2 hours. Further descriptions of these training tasks and experimental tests follow.

Neuropsychological Training Tasks

The *English Word Familiarization* was a task that familiarized participants with the synthesized nature of the speech they would hear. Subjects were read off a word list in natural speech and its synthetic counterpart of eight words (four English control words and four additional example words). The four English control words were included, as pilot tasks indicated that some participants had difficulty deciphering the synthetic nature of the speech. As the English control tasks were a measure of optimal ability, familiarization of the synthetic quality (which subjects had never heard) was necessary to measure optimal task ability without the confound of perceptual difficulty. However, participants were not explicitly told which of the eight words would be heard during testing. After presentation of the word list, participants completed an eight-item identification task where they numbered each of the eight words in the order that they heard them. This task was repeated if the participant was not successful the first time. All participants reached the 100% accuracy criterion on this task, with the exception of four participants who required two presentations. Participants also heard a reading of the word list prior to each English protocol with natural and synthesized speech.

The *APWC Training Task* consisted of three examples using non-experimental English words. Before each example, the participant was told the words which she would

hear in a continuous speech stream and given the synthesized stimulus. The participant

counted the number of times she heard those words within the sample speech stream.

Following the stream, the correct number of words was announced, and the participant

listened to the same speech stream a second time. This was performed for all three

examples, which built in complexity from listening from just two words in the first

example to four words in the third.

The *SOST Training Task* consisted of four trials and three practice examples. The

training task followed the procedure of the SOST, except that the word(s) to segment

were provided aurally and in writing before each presentation. Trial tasks increased in

complexity and practice examples were directly identical to the SOST testing items with

the exception that the English training stimuli were used.

Behavioral Replication Testing

To assess behavioral performance of learning segmentation for an artificial

language, a 32-item two-alternative-forced-choice (2IFC) test was constructed, analogous

to the test used by Saffran, Newport, and Aslin (1996). For each test item, participants

heard two tri-syllabic strings, separated by an inter stimulus interval (ISI) of 500 ms.

One of the two strings was a word from the exposed language, while the other was a

partword. Partwords were formed from the syllables which spanned word boundaries.

The probability of these words occurring in the exposed speech stream was equal to the

transitional probability between words, p=0.33. The partwords that were tested were the

words from the unexposed language (as used in Aslin, Saffran, & Newport, 1998). For

example, the word *tudaro* from Language B occurred as a partword in Language A from

adjoining the words *golatu#daropi*. Thus, in testing a participant exposed to Language A, *golatu* (a word) was compared against *tudaro* (a partword). Therefore, all participants, regardless of language exposure, received the same behavioral test, although correct responses were exactly the opposite for participants who received different language exposures. The nature of the language allowed for the testing of two different partword types. Initial partwords contained the initial two syllables of a word (e.g., the initial partword *tudaro* contained the initial syllables from *daropi*) and final partwords contained the final two syllables from a word (e.g., the final partword *bikuti* contained the final syllables from the word *pabiku*).

During behavioral testing, participants were provided with a response sheet containing three English training examples and 32 numbered experimental items. All stimuli were presented to the participants via headphones. Participants were asked to indicate which of the two strings was more familiar by circling a "1" or a "2" on the response sheet corresponding to the indicated string. Three practice items were given to the participants on English versus nonword stimuli to familiarize them with the task. As with the SOST, a tone followed by one second of silence was played prior to the stimuli presentation and a series of five beeps occurred at the end of item columns. Four seconds elapsed between item presentations. For half of the items, "1" was the correct answer, and for the other half, "2" was correct. Each word and partword was paired exhaustively with each other to yield a total of 16 items. Participants were tested on each item twice, rendering 32 test items. The entire test lasted approximately six minutes.

Neuropsychological Testing

Auditory Probabilistic Word Count (APWC)

The APWC was given pre- and post-exposure and validity for the APWC was measured using English words (E-APWC). Participants listened to the pre- and posttest continuous syllable streams that were constructed respective of their exposed language, and the English version. A total of 36 words were presented.

The participants' task was to count the number of artificial words (learned from the exposure period) that they heard within the continuous speech stream composed mostly of randomized syllables from the medial position of the words. Participants drew tick marks each time they identified a familiar word. If a memory trace for the exposed words was formed (which required segmenting words) using the statistical distribution of syllables, then this same probabilistic information should have been accessible to successfully segment the learned words in a novel language exposure with distracters (which were randomized isolated medial syllables from the exposed language). To accurately detect the learned words, participants must have integrated the statistical distributions of the syllables into a perceptual object (word). As such, this was a measure of auditory gestalt processing. Participants were provided with a brief trial speech stream of 10 seconds, which contained non-experimental English words embedded in it to familiarize them with the task. The three minute speech stream was divided into six segments of 30 seconds each. Participants recorded the number counted after each 30 second segment. The entire APWC test lasted approximately 5 minutes.

Slow Online Segmentation Test (SOST)

The SOST was given pre- and post-exposure to measure the amount of improvement that could be attributed to exposure sessions. Validity for the SOST was measured using English stimuli (E-SOST), instead of the artificial words to determine optimal performance. This test allowed for testing segmentation ability online, rather than discriminating isolated words, and investigated auditory object formation for artificial words.

Participants, using a colored marker, followed along with the auditory speech stream by placing a dot inside each box on the response sheet for every syllable heard (syllables presented at a slowed rate of two per second), and a slash after a box upon perceiving the end of a word. Thus, this test examined online segmentation of the speech stream by requiring participants to synthesize temporally segmented syllables into the learned auditory objects, and then identify the boundaries of these objects online. Four practice items using non-experimental English training stimuli were provided before beginning each test presentation (pre-, post-, and English), the first two of which had the correct answer provided to the participants so that they had a clear model of how they were expected to complete the task. Responses to the practice items were monitored to ensure appropriate responses (i.e. drawing a slash after word-final syllables). The entire test lasted approximately 9 minutes.

Data Analysis

The alpha level for all test comparisons was set at $p=.05$. In order to measure the percent nonoverlap of the two samples, effect sizes were calculated for significant results. Cohen (1988) defined levels of effect sizes using the d statistic: for small effect sizes

$d=.2$, for medium effect sizes $d=.5$, and for large effect sizes $d=.8$. The pooled sample variance was used in calculating effect sizes. Bonferonni corrections were made for all post-hoc comparisons using the English testing. Inter-rater reliability was also measured by having 20% of the data rescored by an independent rater.

Scores obtained from the 2IFC were the total correct (which was converted to percent correct) and the number of false positives to final and initial partwords. Means and standard deviations were calculated on each of these scores. Analysis consisted of a series of t-tests. The first test compared percent correct between Language A and Language B to determine if any factor other than the statistical nature of the language (e.g., phonetic structure, acoustic characteristics, etc.) differentially impacted performance between groups. Equal variances were tested between the languages and then analyzed using the pooled sample variance. All other analyses were conducted by pooling the two languages together as one group. A single-sample t-test was used to determine if percent correct was significantly above chance (50% correct) and a matched-pairs t-test compared selection errors of initial and final partwords.

The APWC had raw scores for the number of words counted at each of the six speech streams. Absolute differences were calculated between each of the six raw scores and the correct number of words in each stream. These six difference scores were then summed to create the test score used in analyses. An absolute difference score of 0 indicated an approximation of perfect performance (as the exact words counted could not be determined); thus, lower scores indicated better performance. A series of matched-pairs t-tests were conducted on this data: pretest versus posttest, pretest versus English,

and posttest versus English. Means and standard deviations were also calculated for each test presentation.

Raw scores for the SOST included the total number of words identified and the total number of words correct. Percent correct was calculated by dividing the number of words identified by the number correct. Matched-pairs t-tests were performed using the number correct and the percent correct for the following comparisons: pretest versus posttest, pretest versus English, and posttest versus English. Analyses using the English condition as a control for performance were also conducted by dividing the test score by the English score. This data for the pretest versus the posttest were analyzed using t-tests. Means and standard deviations were also calculated for each test presentation on both number correct and percent correct.

CHAPTER III

Results

Syllable Identification of Synthetic Speech

Results of the syllable identification pilot study indicated that participants identified syllables correctly 58% of the time across all syllables. Identification accuracy scores ranged from 0% for /ro/ to 100% for /do/ and /la/. Accuracy scores across all participants and syllables increased to 81% when the most frequent response was considered correct and response accuracy ranged from 47% (/go/) to 100% (/do, la, ro/). Within subject consistency scores (the percent of syllables an individual subject identified the same across all three trials) was an average of 92% across all syllables; ranging from 67% (/pa/) to 100% (/bi, bu, do, la, ro, ti/). Only two syllables were consistently identified as another syllable within the artificial language (/go/ and /pi/ were consistently identified as /do/ and /ti/ respectively). Three other syllables were identified as non-language syllables (/bi/, /ku/, and /ro/ were identified as /di/, /pu/, and /lo/ respectively). Identification results with a smaller sample who heard the syllables within the artificial words had similar results. Implications of these results are addressed in the discussion of this paper.

Inter-rater Reliability

Inter-rater reliability was calculated for the individual test protocols. An independent rater with no knowledge of the theoretical foundations of the tests scored a

sample of 20% of the data as a measure of inter-rater reliability. Raters were in 100% agreement on the APWC and 2IFC protocols, which was likely the result of very easy scoring on the APWC and score rechecking procedures with the 2IFC. Total agreement on the SOST was also high at 99%, with 99% agreement for the total number of words counted and 98% agreement for the number of correct words counted. Due to the high inter-rater agreement, in the event of a disagreement the first score calculated was chosen for use in the analysis.

Standardized Screening Measures

All twenty participants scored above 90 on the Matrices subtest of the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990), indicating at least average nonverbal abilities (M=104, SD=8.0). All participants also scored above an 8 on the Memory for Digits (M=12, SD=1.8) and Blending Nonwords (M=12, SD=2.3) subtests of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Roshotte, 1999), again indicating at least average abilities. The Phonological Memory Composite score had a mean that fell within the average range (M=108, SD=9.9), where a score between 90 and 110 indicates normal performance. The Nonword Repetition subtest (M=11, SD=1.8) was also administered as a non-criterion measure to obtain the composite score.

Table 1. Standardized test scores

	Matrices	Memory for Digits	Nonword Repetition	Memory Composite	Blending Nonwords
Mean	104	12	11	108	12
Standard Deviation	8.0	1.8	1.8	9.9	2.3

Experimental Questions

An alpha level of $p=.05$ was used as the significance level for all statistical tests. Significance probabilities are reported here.

Question 1: Segmentation of Complex Patterns via Statistical Regularities

This is a global question as to the amount of learning which occurred on the three experimental measures; as such, specific results will be discussed in the sections that follow. Language A and Language B were first examined to determine if any factor other than the statistical nature of the language (e.g., phonetic structure, acoustic characteristics, etc.) differentially affected performance between groups. This was examined using a t-test conducted between the Language A and Language B 2IFC scores, as used in previous research (Saffran, Newport, & Aslin, 1996). Results indicated no significant difference between the languages as detected by the 2IFC test, $t(18)=.56$, $p=.58$. These results suggest that there was no perceptual difference between the two languages that impacted participants' performances. As these languages were perceived in the same way as predicted, the scores for the two languages were collapsed in the analyses that follow.

Question 2: Replication of Discrimination Studies

For the *2IFC Discrimination Task*, the mean score was 20.3 of a possible 32 items (63.4%), where chance performance equals 16. A single-sample t-test (two-tailed) revealed an overall performance significantly above chance with a large effect size: $t(19)=5.15, p<.0001, d=1.08$.

Figure 1. Mean scores for each participant on the 2IFC
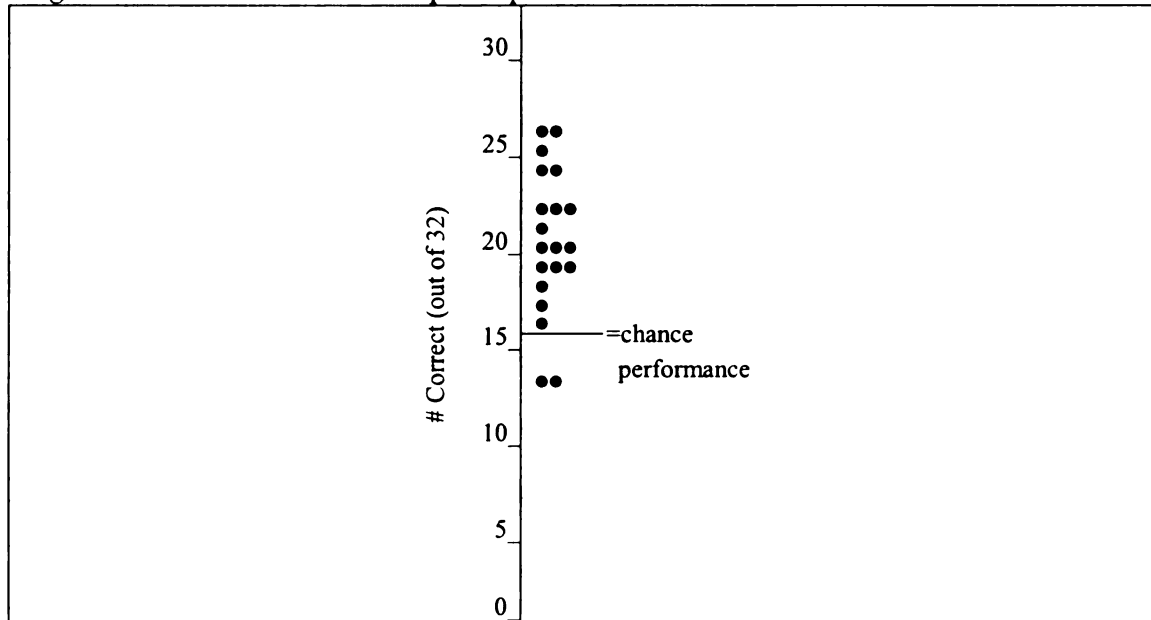
Previous research has demonstrated differential performance between

discrimination pairs when a partword contained the final two syllables of a word versus

when a partword contained the initial two syllables of a word (Saffran, Newport, & Aslin,

1996). A *t*-test comparing the errors of these two types of partwords was not significant:

$t(19)=1.86, p=.07$. The mean number of errors for initial partwords was 6.7, versus 5.1

for final partwords.

Question 3: Object Recognition and Identification

APWC Recognition Task

Results of the optimal English testing on the APWC revealed a mean total word

count of $M=28$, $SD=10.8$. To assess for participants' accuracy in counting the correct

words, the absolute difference (where a difference of 0 indicates a perfect score) was

calculated between the participant's counted score and the correct number of words presented during each of the six different speech streams. These scores were then summed. The mean absolute difference from the correct count was $M=12$, $SD=7.6$. The English test was also compared against both pre- and posttests. These comparisons used a Bonferonni-adjusted significance level of $p=.025$. Results demonstrated significantly better performance on the English test as compared to both pretest: $t(19)=5.03$, $p<.0001$, $d=1.89$, and posttest: $t(19)=3.64$, $p=.002$, $d=1.10$. These results indicate that while participants performed significantly better on the English testing, the greatest effect size was between the English and pretest with a percent nonoverlap of 79.4%. The percent nonoverlap of the posttest and English test was 58.9%. The percent nonoverlap is the area that is not covered by both distributions combined.

Experimental results from the APWC Recognition Task demonstrated improved accuracy and decreased variability between subjects when comparing the pretest to the posttest. For the pretest, participants' mean number of total counted words was 51 (total number of words actually presented during the test was 36) with a standard deviation of 53.8. This is compared to their more accurate and less variable posttest score ($M=39$, $SD=30.6$). The mean absolute difference on the pretest was higher ($M=44$, $SD=37.6$) than the posttest ($M=24$, $SD=19.7$), indicating increased accuracy after exposure for the posttest. Ranges of the absolute difference also decreased with exposure (pretest=137; posttest=83; English=26).

The group data for the APWC, particularly on the pretest scores, had several outliers. While the number of words presented during the test was 36, six participants counted more than 36 words, three of whom counted over 100 words. This lack of

normality did not allow pre- and posttest absolute difference scores to be compared using a t-test. Therefore, the data was transformed through logarithms (Iman, 1994). This allowed the data to fulfill the normality assumption. The log of the pretest mean absolute difference (from the correct number of words) was compared to the log of the posttest mean absolute difference, demonstrating a significant difference between tests with a large effect size: $t(19)=-3.00, p=.007, d=.88$. This indicates more accurate performance on the posttest than on the pretest.

Table 2. APWC means and standard deviations (in parentheses)

APWC	Pretest	Posttest	English
# Counted	51 (53.8)	39 (30.6)	28 (10.8)
Absolute Difference from # Correct	44 (37.6)	24 (19.7)	12 (7.6)
Log of Absolute Difference	1.5 (0.3)	1.3 (0.3)	1.0 (0.3)

SOST Identification Task

The optimal English scores were higher than both experimental measures with a total number counted of $M=81$, $SD=16.6$; correct counted of $M=74$, $SD=23.2$; and percentage correct of $M=88.3\%$, $SD=16.3\%$. The differences for the pre- and posttests with the English test on the total number of correct words identified score were also compared. These comparisons used a Bonferonni-adjusted significance level of $p=.025$. For the pretest, a significant difference was found when compared to participants' English performance: $t(19)=-10.18, p<.0001, d=2.96$. This was also true for the posttest comparison, $t(19)=-9.42, p<.0001, d=2.56$. Comparison of the percent correct was also significantly different for the pretest, $t(19)=-9.60, p<.0001, d=2.85$, and posttest, $t(19)=-7.45, p<.0001, d=2.36$.

Analysis of the experimental SOST Identification Task data demonstrated increased performance on the posttest as compared to the pretest. The mean total number of words identified increased from 46 in the pretest to 59 in the posttest. Of those words identified, a mean of only 19 words were correct in the pretest versus 26 in the posttest. This measure was viewed as an indicator of participant accuracy in completing the task. A matched-pairs t-test between these pre and post measures of the total number of correctly identified words revealed a significant difference with a medium effect size: $t(19)=-2.39, p=.028, d=.52$. The percent correct was also calculated for pre- and posttest scores by taking the number correct over the number counted. This significant difference was still upheld when English test scores were used as a control factor to account for individual subject differences in general SOST test performance, $t(19)=-2.18, p=.04, d=.16$.

The mean percent correct for the pretest was 37.3%; whereas, it was 45.6% in the posttest. This measure was viewed as a possible indicator of participant precision in completing the task. A greater dissociation (and lower percentage) between the number identified and the correct number identified could be a measure of intra-subject reliability. A matched-pairs t-test of these scores was also conducted, but demonstrated no significant difference between the scores: $t(19)=-1.43, p=.17$. However, there was a trend toward increased scores for the posttest. No significance on this measure was found when using the English tests scores as a control factor, $t(19)=-1.59, p=.13$.

Table 3. SOST means and standard deviations (in parentheses)

SOST	Pretest	Posttest	English
Total Identified	46 (23.4)	59 (20.5)	81 (16.6)
Correct Identified	19 (12.8)	26 (13.6)	74 (23.2)

% Correct of Total (%)	37 (20.9)	46 (22.0)	88 (16.3)

Question 4: Changes in Performance due to Enhanced Perception

Perceptual changes were indicated in the above sections by the improved

performance on posttests. Also as outlined above, performance scores on the English

tests indicated higher task accuracy, as was expected from an optimal performance

measure. The percent of nonoverlap between the experimental and English samples

decreased between pre- and posttests: 79.4% pre- to 58.9% posttest on the APWC; 92.8%

pre- to 89.3% posttest on the SOST. The difference was most reduced for the APWC.

As predicted, variability indexed by standard deviations and ranges, also decreased for

the APWC from pretest (SD=37.6, range=137) to posttest (SD=19.7, range=83). These

results are examined in more detail in the discussion of this paper.

Table 4. Significance probabilities and effect sizes for test comparisons

Test Measure	Pretest vs. Posttest	Pretest vs. English	Posttest vs. English
APWC log of absolute difference	p=.007, d=.88	p<.0001, d=1.89	p=.002, d=1.10
SOST # correct	p=.028, d=.52	p<.0001, d=2.96	p<.0001, d=2.56

Question 5: Manipulability of Auditory Object Memory Traces

The examination of task hierarchical difficulty from the predicted easier 2IFC

discrimination task, to more difficult the APWC recognition task, to the most difficult

SOST identification task was not directly testable through a statistical model. However,

examination of effect sizes provides some indication for this hierarchy of processing. For

the discrimination task, the large effect size was d=1.08 with a percent nonoverlap of

58.9% for the difference between participant performance and chance. This size was

reduced for the recognition task, which had an effect size of $d=.88$ and percent

nonoverlap of 51.6% for the difference between pre- and posttests. The medium effect

size for the SOST was the least at $d=.52$ and percent nonoverlap of 33.0% for the pre-

and posttest difference.

Table 5. Effect sizes changing according to task difficulty

	2IFC	APWC	SOST
d value	1.08	.88	.52

CHAPTER IV

Discussion

The segmentation problem, within the context of novel language acquisition, arises from the quasi-continuous acoustic structure of speech. All listeners must learn to solve this acoustic problem in order to learn and comprehend spoken language. The acoustic problem to speech segmentation is twofold. First, how are word boundaries identified from an often unbroken acoustic stream? Second, how are the endless acoustic features of the speech stream combined into word-forms? The work of researchers such as Saffran, Jusczyk, Aslin, Newport, Johnson, among others, has been focused on the first question of how infant and adult listeners learn to identify word boundaries. This project replicated their methodology of using an offline discrimination task to confirm the learning of word boundaries, and expanded upon it by instituting more rigorous acoustic stimulus controls and using online tasks requiring higher-level processing. Research cited in the electrophysiological literature has focused on this second question: integrating multiple features into one object. These studies have looked at primarily early stages of auditory processing with tone patterns. This project attempted to expand upon these findings by using this theory of auditory objects (Hayes & Clark, 1970; Kubvoy & Van Valkenburg, 2001) to explain a mechanism by which statistical structure could integrate individual aspects of speech (syllables) into a preliminary auditory object (prelexical word-form). It is the position of this paper that an early sensory memory store built up by statistical regularities allows for the detection of word boundaries; thus,

63

segmenting speech. Novel individual acoustic features are later integrated into perceptual

objects (at least in part) through these statistical regularities, and allow for the recognition

of unified objects within the continuous speech steam. These objects may later be

entered into higher-order memory to establish interaction with cognitive-linguistic

processing and explicit object identification. Unarguably, cues other than simple

transitional probabilities also facilitate object formation and segmentation. However, it is

the early nature, saliency, and ability of these cues to work in isolation from other cues,

which make transitional probabilities particularly interesting and significant. Through a

series of neuropsychological protocols investigating the recognition and identification of

word-forms, this study investigated statistical auditory object formation.

It was these two questions of how word boundaries are identified and how

acoustic features are combined into word-forms that guided this project. The task

complexity hierarchy of discrimination (2IFC), recognition (APWC), and identification

(SOST), allowed for the unique investigation of these two questions in both online and

offline tasks. The sections that follow will discuss the results and implications this study

has for answering these two questions; as well as, address other relevant issues such as

domain-generality and implications for therapeutic intervention.

Identification of Word Boundaries

Several techniques for identifying word boundaries were reviewed in the

introduction to this paper and have consistently been supported in the literature. Most

segmentation techniques that have been proposed are acoustic cues inherent in the speech

stream. Such characteristics are metrical stress, allophonic variation, and phonotactics.

However, most of these cues are either mediated by (in the case of metrical stress), or are more sophisticated levels of statistical structure (such as phonotactics and allophonic variation). It is statistical regularities that seem to be an essential aspect of perceiving complex scenes (Kersten & Yuille, 2003) and may be an innate process (Canfield & Haith, 1991; Kirkham, et al., 2001). Saffran and colleagues have presented transitional probabilities as one early statistical cue to identifying word boundaries (Aslin, et al., 1998; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). To investigate this question, this paper replicated the discrimination task of Saffran, Newport, et al. (1996) with more highly controlled stimuli.

Question 1: Segmentation of Complex Patterns via Statistical Regularities

The overarching purpose of this study was to test the use of statistical regularities in the analysis of complex stimulus patterns, namely speech syllable distributions, and to segment those patterns within a continuous auditory stream that offered no acoustic breaks.

Humans seem to be able to identify word boundaries on the bases of several different statistical parameters. Sensitivity to these conditions allows individuals to detect what stimulus occurrences are probable, and to discriminate those events from ones that occur rather infrequently. The detection of a word boundary reflects sensitivity to an infrequent event (two syllables which do not occur frequently together or predict each other).

By reviewing the present study comprehensively, the results indicate successful learning of the two artificial languages, with no difference in performance between the

languages (as tested by the 2IFC). This study's results indicate that adults are sensitive to statistical information. They can use this information to identify word boundaries (2IFC condition) and auditory objects (APWC and SOST conditions).

The investigation of Question 1 draws two primary conclusions: 1) Online speech segmentation can be measured by the neuropsychological protocols developed in this study, and 2) statistical regularities are sufficient for initial parsing of the speech stream, identification of word boundaries, and formation of auditory object word-forms.

Question 2: Replication of Discrimination Studies

This second question was designed to confirm participant ability to discriminate segmented words from foils so that conclusions from the neuropsychological protocols could be drawn. Confirmation of participant abilities was addressed through the replication of previous stimulus discrimination methodology. These studies are based upon the underlying assumption that segmented words from the continuous speech stream will be identified at above-chance levels as compared to nonsegmented words. This was tested in an offline presentation of words versus nonwords/partwords. This test examines whether words from the artificial language were discriminatively selected from syllable strings that spanned word boundaries.

Discrimination is the detection of a difference between stimuli (Burnham, et al., 1987). This was the process that was tested during the 2IFC task. An ability to select the appropriate stimuli above chance reflects the participant's learning of the artificial language on a very early, detection level. This task does not require the participant to be able to explicitly identify the words that were a part of the language's lexicon, or

recognize these words within a continuous speech sample. It is instead, the detection of an infrequent, deviant event. Therefore, discrimination success demonstrates sensitivity to the statistical structure of the language and detection of word boundaries. Analyses of the 2IFC indicated a significant selection of the word over the partword stimuli. These propitious results suggest that participants were able to identify word boundaries above chance following statistical exposure.

However, detection of a foil's word boundary does not guarantee recognition of the word as one perceptual object. Fiser and Aslin (2002) suggested that the nature of the 2IFC is such that "it is unclear whether participants implicitly extracted triplets [objects] during the familiarization phase, or whether they simply became sensitive to the pairwise statistical relations present in the stream..." (p. 464). They continued by stating that "...pairwise shape information present during the familiarization phase is completely sufficient to discriminate between familiar and novel triplets" (p. 465). Thus, successful completion of the 2IFC indicates sensitivity to statistical structure and to the identification of word boundaries. However, it is not able to determine if the three syllables forming a word are perceived as one unified object, or if participants are merely recognizing the presence of an infrequent pairwise relationship (a word boundary) in the partword. It is unclear whether participants were selecting the word or selecting against the presence of a word boundary in the partword. To investigate this issue, an analysis of initial and final partwords was conducted.

Previous research in visual (Fiser & Aslin, 2002) and auditory (Saffran, Newport, & Aslin, 1996) domains has indicated a discrimination asymmetry between initial and final partwords. These partwords had either an X-1-2 or a 2-3-X structure (where X is an

"out-of-order" syllable or visual symbol). The partword X-1-2 has a joint probability of the first shape pair which differs from the exposed word; whereas, 2-3-X maintains this same joint probability of the initial pair. Subjects consistently and incorrectly chose partwords that contained the final syllables of words more often than partwords that contained initial syllables. Thus, both studies found above-chance performance in discriminating X-1-2 sequences, but not 2-3-X sequences. Fiser and Aslin (2002) and Saffran, Newport, and Aslin (1996) interpreted these results to suggest that subjects focus on the ends of sequences.

The results of the present study did not support previous results. The analysis of initial versus final partword selection errors indicated no significant difference. This suggests the possibility that participants completed the 2IFC task in this study differently than in previous studies. To perform successfully on the 2IFC, a participant only needs to be able to recognize pairwise relations to detect the word boundary (Fiser & Aslin, 2002). The previous studies showed that participants failed to detect the word boundary in the 2-3-X condition, possibly because of their reliance on pairwise relations. However, subjects in the present study accurately discriminated words from both 2-3-X and X-1-2 partwords in the 2IFC condition (as analyzed by the initial and final partword t-test). This suggests that subjects were not simply using pairwise relations to discriminate words from foils, but were possibly recognizing the words as unified objects. This conclusion was tested using the APWC and SOST protocols which examined object recognition and identification (see Question 3). What made participants in this study perform differently from earlier studies? It is possible that consciously recognizing and identifying words in previous conditions (namely, the APWC and SOST) trained participants to also complete

the 2IFC in an equivalent manner. Thus, a change in actual perception may not have occurred, but rather a change in discrimination strategy occurred from looking for word boundaries to word-forms. Performance ability measured in percent correct does not seem to vary much from previous studies: 65% in Saffran, Aslin, et al. (1996) and 63% in the present study. This hypothesis is supported in this study by no significant difference between 2-3-X and X-1-2 partwords. A simple discrimination strategy of selecting against weak probabilities would support differential performance on these two types of part words (as found by Fiser & Aslin, 2002; Saffran, Newport et al., 1996). A strategy of selecting word-forms predicts no difference between partword types, and was supported in the current study. This segmentation strategy was investigated in this study by testing higher-order abilities of recognition and identification (see next section).

In conclusion, the examination of Question 2 suggests that word boundaries can be detected from continuous statistical presentations and discriminated offline. Also, investigation of partword errors suggests that discrimination success may occur due to word boundary detection (suggested in previous studies) or auditory object detection (as indicated in this study).

Integrating Acoustic Features into Auditory Objects: Creating Word-Forms

The second aspect of the acoustic problem to speech segmentation presents a slightly different question. It is one that has not received much attention with speech stimuli in the segmentation literature. How are the seemingly endless acoustic features of the speech stream combined into word-forms? It is from an adequate understanding of this second question, that the first question can also be more fully understood. Once

word-forms are created and lexicalized, additional information is available to aid in speech segmentation, such as lexical activation (McQueen, 1999).

The grouping of auditory features is important in forming coherent speech percepts (Barker & Cooke, 1999). Hayes and Clark (1970) proposed that the word was one type of auditory object. It is a short cut for reducing the amount of stimuli that need to be processed and considered by the brain; thereby, improving perceptual performance. The perception and formation of auditory objects is essential for the development of adequate speech perception and language acquisition. This study examined how the early formation of these auditory objects within a novel speech environment are recognized and identified online during perception.

Question 3: Object Recognition and Identification

This third question investigated participants' online ability to recognize and identify patterns. The last question addressed used a discrimination task that only examined word boundary detection. This question expanded upon that task by examining exactly how individual syllables might be combined into one object. The protocols used to investigate question three measured recognition and identification of these objects.

The Recognition Task

Recognition is the implicit knowledge that some present stimulus has been encountered before (Galotti, 1999). It requires discriminating the stimulus from other information and implicitly knowing that it is a reoccurring event. Access to the actual symbolic content of the stimulus, or any higher-order structure that might be present is

not required. For example, recognition of the word "dog" requires the perceiver to segment "dog" from any other information and to access the prelinguistic knowledge that the sound structure of "dog" has been heard before. However, recognition does not require explicit access to higher-order structure of the word. Recognition is the implicit knowledge of a recurring sound structure.

Results from the APWC recognition task indicated increased accuracy in counting words following the statistical exposure. The APWC examines the beginning stage of the structural building of perceptual objects. During this task, the participant did not definitively state what syllables formed the word they were counting. However, in order to obtain an accurate count, they needed an implicit recognition of the auditory object structure. This is a prelexical mental representation of the object form. Clearly, participants had no such representation in the pretest (as indicated by posttest and English comparisons). However, the improved accuracy of the posttest APWC, which more closely approximated optimal performance, indicates that early formation of object structures in memory. In addition, the predicted trend for decreased variability between subjects was noted from the pretest to the posttest. This indicates a greater precision for between-subject performance.

The Identification Task

Identification entails recognizing the abstract representation of a perceptual object (Burnham, et al., 1987). This level of processing goes beyond recognition. It requires the perceiver to access the mental representation of the stimulus, either by accessing semantic meaning, or by explicitly accessing the higher-order stimulus structure. The

word "dog" needs to be identified as a word-form, not just as a recurrent sound structure. Identification also implies an ability to explicitly label the stimulus event.

Results of the SOST identification task again indicate improved posttest performance. The SOST expands upon the APWC by requiring the participant to explicitly identify the word-form. While this word-form has no lexical meaning, it must have a clearly formed, holistic, mental representation of the auditory object such that it can be consciously retrieved from memory. Therefore, participants likely had an early cognitive representation. Such a representation was even more important considering that it needed to be phonologically synthesized in memory.

The smaller difference between pre- and posttests on the SOST may also have been due to the APWC serving as an initial exposure period (as the pretest APWC was presented before the SOST pretest). This seems possible as two subjects achieved an average of 82% correct on the SOST pretest; which was not likely due to chance alone. This initial exposure may have reduced the difference between the pre- and posttests. However, despite this possible reduction, participants on average still performed significantly better on the posttest than on the pretest when measuring the total number of correct words counted. Change in perceptual performance for infants has been demonstrated to occur in periods of time that were shorter than the 3-minute APWC test stream used in this experiment (Saffran, Aslin, et al., 1996). As statistical learning is proposed to be a very early, initial parser, it is possible that some learning and perceptual improvement occurred before the presentation of the SOST pretest.

Also worthy of discussion is the difference in significant findings between the analysis of the total number of correct and the percent correct. The total number correct

was significantly higher posttest, while the percent correct showed a clear trend toward improvement. However, it also appears that these two calculations may be measuring two different, though related, phenomena.

The significant increase in number correct suggests that participants entered mental representations of the auditory objects into a memory form that is cognitively accessible. The findings of percent correct indicate that while participants increased in number correct, they also increased more in their total number of responses. This is related to the increase in between-subject variability in the posttest, which is a measure of how precise the participants were as a whole. Thus, while accuracy improved for the posttest, precision did not. This was not a factor on the English test, which had significantly better accuracy and precision. This finding may indicate the very early, tenuous position the mental representations of these auditory objects held in memory. Perhaps information other than a unified acoustic structure needs to be present to further improve performance and implant auditory objects in memory to improve both accuracy and precision. This information could be lexical meaning, as this was present (although to a controlled minimum degree) in the testing of low-frequency English words.

The results of Question 3 suggest that recognition is the early structure building of auditory objects and occurs implicitly during online presentations. Also, identification implicates forming a higher-order mental representation of the auditory object that is accessible to cognitive processing. Statistical probabilities are sufficient in forming these representations, though more exposure or information may be required for native-like processing. Also, performance improved in accuracy on the SOST, while both accuracy and precision increased for the APWC.

Question 4: Changes in Performance due to Enhanced Perception

An enhanced ability to perceive perceptual units was believed to be implicated in performance changes that occurred over the time course of the experimental session. This change in abilities was assessed through the use of a pretest-posttest experimental design. While the English conditions were designed as a measure of validity to determine optimal performance on the novel neuropsychological protocols used in this study, they were also available to serve as a comparison to native language performance. Therefore, this study also compared perceptual changes to the optimal (English) performance level, which would be expected of an adult native language speaker. The results showed that participants performed significantly better on both recognition and identification tasks on the posttest measure; as well as, significantly better on the optimal performance tests. On the recognition task (APWC), participants performed four times less accurately on the pretest measure as compared to their optimal performance. They were only two times less accurate on the posttest. Thus, the posttest had a smaller percent nonoverlap with the optimal measure than the pretest. (This would be expected from more native-like processing.) While posttest identification (SOST) did significantly improve, participants maintained a similar percent nonoverlap with the optimal test from pretest to posttest. This was perhaps due to the increased difficulty of the test and lack of lexical association.

Overall, these results suggest greater proficiency for recognition and identification tasks following statistical exposure to the language. However, participants did not reach a level of native language processing. Greater acquisition of recognition than

identification abilities was demonstrated by a closer approximation of optimal performance during the APWC.

In summary, the results of Question 4 demonstrate perceptual changes can occur in the way humans perceive auditory information. Initial perception of auditory features may, during refinement by statistical regularities, become holistically perceived as auditory objects. The results also suggest that with statistical exposure, recognition ability may more closely approximate optimal performance, while identification ability may require additional information.

Question 5: Manipulability of Auditory Object Memory Traces

Statistical learning of auditory objects may facilitate later cognitive processing. Fritch, Large, and Pisoni (2000) presented adults with nonwords that were created according to the probabilities of the onset and rime constituents of syllables contained in an English dictionary. Subjects rated the wordlikeness of the nonwords containing high probability constituents higher than low probability nonwords. Subjects had better recognition memory performance for the high probability constituent nonwords. These results suggest that adults maintain the probability of linguistic segments within memory and are able to recognize novel combinations of these segments based upon their probabilistic nature. Therefore, statistical information is not accessed only during early sensory processing, but is stored to facilitate later processing of novel sequences.

In the present study, a memory trace to the probabilistic nature of the artificial language was created during exposure. An attempt was made to determine whether these memory traces could be used to facilitate further processing, as Fritch, Large, and Pisoni

(2000) demonstrated could be done in a recognition memory task. This study examined a hierarchy of processing difficulty designed to determine the flexibility of these auditory object memory traces. It also investigated whether these memory traces could integrate temporally separated acoustic information into the learned auditory object through a phonological synthesis task.

Building Higher-Level Processing: From Discrimination to Identification

The manipulability of memory traces for the learned auditory objects was examined through the use of higher-order processing tasks rather than the widely used discrimination task. The hierarchy of perceptual difficulty between tasks can be inferred from participant accuracy. (Direct comparison of the conditions was not possible due to the highly different measures and different theoretical assumptions.) Comparison of significance probabilities between tasks indicates less likelihood that participants' behavior was the result of chance for the discrimination task. The possibility of chance behavior increases for the recognition task and most greatly for the identification task (although participants still increased performance significantly). Additionally, effect sizes were largest for discrimination and smallest for identification, supporting this hierarchy.

Levels of processing were proposed by Craik and Lockhart (1972) and applied to complex auditory processing by Griffiths (2003). The hierarchy of processing examined through the protocols of the current study occurs in the following way. In the discrimination task, participants did not need to have any preliminary form of a perceptual object. They must have only been able to detect statistically infrequent

76

syllable pairs at word boundaries. However, this simple sensitivity to weak transitional probabilities was not enough to accurately complete the recognition task. Participants must have formed a higher level of representation based upon the statistical structure of the language. This representation is at the preliminary structure building of the auditory object where the explicit word-form does not have to be stated, but where an implicit detection of its holistic structure occurs. It is not until identification occurs that an explicit word-form (that is stored as a cognitively penetrable memory form) needs to be accessed. An implicit detection of the preliminary structure of the auditory object would have been insufficient to complete the identification task. Participants were required to explicitly label the exact structure of the word. This was different from implicitly detecting its occurrence (as in the recognition task). A final level of representation, not examined by this study, would be the retrieval of the word-form stored in the lexicon.

Evidence of Phonological Synthesis

The ability to phonologically synthesize auditory objects provides another argument for a higher level of representation than that offered by early sensory discrimination processes. To regenerate a word-form from temporally separated segments, an explicit mental representation of the word-form is required. As this is a phonological task, it might be argued that this word-form is now stored as a prelexical form in memory. Participants demonstrated their ability to complete the phonological synthesis task of the SOST after exposure to the artificial language. This suggests that statistical regularities are sufficient for forming and entering auditory objects into memory as a higher-order representation.

In conclusion, Question 5 suggests that processing of the speech stream may occur on a hierarchy of difficulty building from discrimination to recognition to identification. Each of these levels may entail its own level of representation, may require different involvements of memory, and may to various degrees be accessible to explicit cognitive processing. Auditory objects are susceptible to higher levels of processing, such as phonological synthesis, suggesting the possibility of maintaining a prelexical representation in memory, accessible by cognitive processing.

Summary of the Research Questions

The results derived from the research questions of this project confirmed the presence of a statistical learning mechanism. This project replicated previous studies by verifying the ability to discriminate word-forms from partwords based solely upon statistical exposure. It also expanded upon these previous studies by examining the strength of statistical learning in higher-level processing during recognition and identification tasks. These higher levels of processing required more than a simple detection of low probability word boundaries. They required the formation of auditory objects. The results of the recognition (APWC) and identification (SOST) tasks suggest that statistical learning mechanisms are able to combine multiple features (syllables) into one perceptual object, a word-form, which is available to cognitive processing. These results support a processing hierarchy and provide a mechanism which is capable of creating abstract representations of word-form structure to which meaning can later be applied. The sections that follow will further discuss this statistical learning mechanism,

apply these results to language learning and therapeutic settings, address possible

limitations of this study, and suggest further avenues of research.

Domain-Generality of Statistical Learning and Perceptual Objects

An investigation of the domain-generality of statistical learning can further help

to explain the process by which perceptual objects are formed. Domain-general

processes may work to form and/or support domain-specific modules. Elman and Bates

(1997) discuss statistical learning as an innate default assumption to evaluating all

stimulus input (regardless of domain), which evolved for other reasons, but is now

recruited for language learning. While Cosmides (1989) supports a more specialized

process approach, she states that an organism's behavior will be random unless it has a

reliable and efficient means of extracting information from the environment and a well-

defined system of rules to use the information. She continues by saying that specialized

learning mechanisms organize experience into meaningful units which focus attention,

organize perception and initiate procedural knowledge that lead to domain-specific

processing (Cosmides, 1989). The current paper suggests that statistical learning may be

such a learning mechanism that works to shape and implement future domain-specific

processing modules. Probabilities may be particularly important for initial parsing and

ordering of stimulus events; thereby, enabling the best mode of processing the

information to be determined and assigned.

Statistical sensitivities may work like a person filing a stack of assorted forms in a

law office. The person recognizes the pattern of the letterhead, the shape of the text

body, or the form title, but does not process the actual content of the form. Processing

occurs when the form reaches the office of the specific professional who was trained to work with that form. That specialized person is doing the processing work of a specific domain module. The person received specific information that was identified and assigned on the bases of statistical pattern sorting. Therefore, this paper does not claim that it statistical learning is the only information available to segment speech for the acquisition of language. Other domain-specific processes are also likely to facilitate and carry on this work, especially when a higher-order system is established to specifically recognize and identify perceptual objects after statistical learning has shaped those modules.

This study is unable to directly test domain-generality, as it only studied one domain of speech perception. However, it is relevant to this discussion within the context of previous research, as statistical learning appears to be a ubiquitous process. Investigating the domain-generality of statistical learning requires the examination of processes isomorphic to those examined in the speech domain by this paper. Thus, investigating musical and visual domains can help explain the process by which the brain creates representations of word-forms via statistical learning.

Music is a different form of the same ability that underlies speech perception; it is the ability to organize complex stimuli into temporally ordered sequences (Jourdain, 1997). Electrophysiological results suggest that music and language appear to have similar electrophysiological correlates for auditory and temporal structure (the level of the auditory object), while processing semantic meaning in context appears to be language-specific (Besson & Schön, 2003). Tervaniemi (2003) discussed ERP and MEG evidence that may implicate Broca's Area in processing chord cadences (an auditory

object), but differential processing for individual musical sounds. This evidence suggests that there may be a common way the brain processes auditory objects. Saffran (2003b) continued this discussion by suggesting that the learning and memory for music and language occur without instruction or reinforcement. Humans are able to implicitly learn and remember structured environmental information.

These studies implicate common processing of auditory objects in music and speech domains and in certain cases, specific brain locations. Auditory objects are not specific to speech, but are a domain-generalized process of perception that can be derived implicitly through the natural probabilistic structure the environment. Indeed, tonal patterns have been perceived as whole units based upon the statistical presentation of their structure (Brattico, et al., 2002; Saffran, Johnson, Aslin, & Newport, 1999; Vaz Pato, et al., 2002).

Vision is also processed according to statistics. Kersten and Yuille (2003) wrote that "statistical regularities in natural images and scene properties are essential for taming the complexity and ambiguity of image interpretation" (p. 150). They cite studies of homogeneous textures and object boundaries, as well as, scene and object recognition studies where statistical patterns help to define, constrain, and/or explain the perception of visual objects. Fiser and Aslin (2002) furthered the support for the domain-generality of statistical learning by adapting Saffran, Aslin, et al.'s (1996) study to a visual shape paradigm. Sequences of three visual shapes were presented during a 6-minute movie. Their results replicate past findings in that the three shape sequences were successfully discriminated from novel sequences and part sequences in an offline two-interval forced-choice task.

Evidence from visual studies has provided neurophysiological explanations for how perceptual objects facilitate processing. It appears that the formation of a unified object creates a better perceptual fit (Kersten & Yuille, 2003) and reduces brain activity; thereby, making detection of novel stimuli easier (Murray, et al., 2002). This may be similar to the neural mechanism behind the MMN, which indexes increased activity following novel stimuli that break an auditory pattern.

Applications of the Current Study

In addition to investigating the theoretical aims of this research, this study also has applications to other areas. The theoretical foundations proposed in this study support the importance of statistical and probabilistic structure in learning theory. This theoretical base adds to the understanding of very early, basic language acquisition mechanisms. It provides a rationale for methods of teaching second languages and for various therapeutic interventions for language-delay and neurogenic clinical populations.

At the most basic level, this theory suggests that in order to learn language, we need to hear language—repetitively. We must be exposed to enough recurrences of a particular token to form a mental representation of it; thus, perceiving the probabilistic structure and deriving a perceptual object. Therefore, experience guides perception, which is supported in neural plasticity research (Sharma, Angelucci, and Sur, 2000).

This paper stresses that statistical learning is only one mechanism that the child employs during language development. As discussed by Hollich, Hirsh-Pasek, and Golinkoff (2000) and Jusczyk (1997), many other cues certainly play a role in language development, including higher-level learning structures such as that employed during

social interaction. However, the theoretical foundation of this study suggests that statistics is a highly important cue that infants and adults can tap into. It is available across domains (Kirkham, et al, 2002) and may even be important for integrating cross-modal information, such as pairing word-forms to real world visual objects for lexical development (see Roy & Pentland, 2002).

Second Language Instruction

The results of this study, along with the previous research on statistical regularities suggest that humans are naturally sensitive to patterns; we perceive patterns even when none exist (Huettel, et al., 2002). Deriving those patterns from the environment allows us to understand the structure of language and how those components interact. Our sensitivities also allow us to identify patterns in context, which is essential to fluent spoken language comprehension. While learning a second language through a memorization of isolated vocabulary words is common, an exposure to words in context forces us to derive the patterns of the language (syllabic and grammatical). This concept may be similar to a theory proposed in the problem solving literature (the problem here being the attempt to learn a foreign language). Vollmeyer, Burns, and Holyoak (1996) reported that when participants are presented with specific goals, they will learn how to achieve the goal, but will be unable to generalize to other related tasks. However, when participants were provided a nonspecific goal, they worked to discover the operators by which the problem could be solved. This strategy enabled the participants to generalize their strategies to solve other similar problems. Translating this literature to learning language: a specific goal for the language learner is to memorize the correct translation of

vocabulary words; whereas, a nonspecific goal would be to listen to a speech sample to derive the gist, or overall meaning. It may be that when learning language, a more efficacious approach would be to use a nonspecific goal. This approach is supported by the present study. Our human capacities enable us to identify statistical patterns online within a continuous speech stream. Therefore, we have the ability to derive the patterns of a language, including grammar (Gomez & Gerken, 2001), without explicit instruction on individual words. This is how infants must learn language, and this study suggests that the young adult still has the capacity to do this. Statistical learning suggests why people typically do not fluently learn languages until they are immersed. Goal-specific vocabulary teaching may encourage "translation thinking," while a nonspecific goal immersion approach forces the learner to identify language patterns; thereby, facilitating understanding of the language structure and allowing for "language thinking."

In applying this research to the student learning a foreign language, it is apparent that enough auditory information is required to form an object for word learning. This is evident from numerous antidotal reports that people learn a new language better when completely immersed in the foreign language. Many mechanisms are at play in such a setting, just as they are at play for the infant. Social and situational cues are likely important. But also important is the rich and constant exposure to the structural probabilities of the language. Constant exposure to the statistical structure of the language allows for the perceptual formation of many different auditory objects, which can later be identified and entered into the lexicon when the objects are assigned semantic meaning through social-pragmatic learning processes.

However, the conclusion that adults learn language as efficiently as a child cannot be derived from the results of this study where adults learned to identify prelexical structures. Research suggests that as language learners mature, their ability to learn language broadly deteriorates, thus supporting a critical period (Johnson & Newport, 1991). A critical period for musical ability has also been discussed (Pantev, Oostenveld, Engelien, Ross, Roberts, & Hoke, 1998), possibly suggesting the development of an early neural organization for complex sound perception based upon early environmental exposure to specific sound patterns.

Theraputic Intervention

In addition to second language acquisition, statistical learning has implications within the therapeutic environment. Breitenstein & Knecht (2003) suggest the use of massive, repeated interactive exposures; high concurrence of language and corresponding sensory processes; and intense training frequencies. Creating a rich, stimulating environment is the essence of most treatment techniques. The Power-Law of Practice states that the time required to complete a task decreases over the number of trials; thus, practice is an important component for acquiring a new skill (Lovett, 2002) or rehabilitating an old skill.

The use of highly frequent, statistical presentations is found in a number of different treatment techniques. One general category is behavior modification, based upon operant conditioning. Lovaas (1977) proposed using operant conditioning to facilitate language development in nonverbal children. Behavior modification has also been used in aphasia treatment (Davis, 2000). Another aphasia treatment technique is

85

programmed stimulation, which relies on stimulation to elicit repeated responses (Davis, 2000). Articulation therapy has a treatment technique called Bombardment, which is designed to increase the frequency of a target stimulus in the child's environment (Nemoy & Davis, 1954). Computer-assisted therapy, such as Fast ForWord (R) (Scientific Learning Corporation, 1998), also provides statistical environments for learning.

While many of these (and other) therapy techniques may be effective in changing or modifying behavior, possibly in part due to statistical presentations, it is important to consider other factors when selecting a treatment technique. Behavior modification techniques have been criticized as teaching specific behaviors that do not generalize outside of the therapy environment, and ignoring the actual underlying source of the impairment (Davis, 2000). Treatments other than these cited may be more effective at producing functional behavior that transfers and generalizes to other situations outside of the treatment setting. The Cognitive Stimulation approach to aphasia treatment takes some of this into account by suggesting that therapy tasks should be structured to ensure 60-80% success so that successful cognitive processing is practiced (Davis, 2000). Constant failure only practices the inappropriate processes that lead to the failure.

Statistical learning may also have a role to play in assigning meaning to word-forms during the development of a lexicon. Cross-channel Early Lexical Learning (CELL) is a computer model that was designed to find and model consistent structure from multi-sensory information (Roy & Pentland, 2002). CELL creates a lexicon through segmenting speech and assigning semantic meaning through statistical parings of speech and visual objects. Therefore, word learning may be a function of identifying

statistical regularities cross-modally. This suggests that to facilitate development of a lexicon, words should be consistently paired with real world visual/tactile objects to which they refer. Intensive statistically cross-modal associations may be essential. This strategy of word-object pairings is integrated into naturalistic early intervention methods (Bailey & Wolery, 1992).

Many possibilities exist for the delayed development of language. However, this study proposes one such possibility: a decreased or impaired ability to process the statistical structure of language. Therefore, language therapy for such a child should focus on increasing the salience of other cues to word learning, or to increasing statistical input. By exaggerating other cues, the child may be able to tap into unimpaired learning mechanisms which may thereby assist in the acquisition of language. By increasing statistical cues, the child may be able to more easily detect the probabilities inherent in the language.

It is possible to envision several scenarios of language-delay where statistical regularities may play a role. The child may be in an impoverished language environment where he or she is not receiving enough language exposure to identify statistical structure. This intuitively would suggest that increased language exposure is needed, which is supported now by the theory of statistical learning. Another situation might be that the complexity of the language exposure has decreased probabilities to such an extent that the child is not able to identify structure (which could result from a highly complex and variable semantic and syntactic structure in the environment). Reducing the complexity of the language exposure and increasing perceptual cues, such as provided by melodic intonations characteristic of motherese, may be appropriate.

While these types of therapies are rather intuitive and are already being employed by many speech-language pathologists, this study provides a theory to justify such therapy techniques and explains a possible contributor to the underlying causes of the language delay. While the child's language delay might be the result of many impaired learning processes, such as a lack of appropriate social language models, an impoverished statistical environment may also confound such problems.

This study also examined the transfer of speech sounds stored as a sensor memory trace to encoded auditory objects. Early on, the phonological aspect of word learning may be more important than semantics (Hu, 2003). A breakdown in the temporary store of word-forms may impair construction of phonological representations (Hu, 2003). Underspecification of phonological representations may underlie specific language impairment (Maillart, Schelstraete, & Hupet, 2004).

This study suggests, along with Breitenstein & Knecht (2003), that massive, repeated exposure is important in therapy environments, particularly when the treatment must be performed at a very low, sensory level of processing. However, this statistical, therapeutic environment needs to be structured in such a way as to facilitate behavioral generalization and transfer. It is essential that trained professionals use their clinical expertise to structure appropriate learning environments. Statistical learning is a general tool that is likely to always be available for use in therapy (as it may represent a general brain process). As such, it is a powerful instrument for structuring the environment for early acquisition of abilities and for cognitive retraining after brain damage. This study develops an avenue for future research which can apply the theory of statistical learning to clinical populations and therapies.

Study Limitations and Considerations

Syllable Identification Pilot Data

A preliminary study of the speech syllables used in this study was conducted to investigate how participants would perceive and identify the synthesized speech used in this study. Fifteen participants were presented all 12 syllables across three trials, to which they were required to transcribe the syllable as they perceived it.

Perceptual Learning of Synthetic Speech

The synthesized speech syllables had reduced intelligibility due to the elimination of acoustic cues by controlling for frequency, amplitude, and duration across all syllables. Previous research has indicated that listeners have an increased difficulty for understanding non-native synthesized speech such as the artificial language used in this study (Reynolds, Bond, & Fucci, 1996). Studies suggest that training in perceiving the synthesized speech is required for decreasing error rates (Conroy, 2003; Francis, 1998). Also, listeners will change their judgments of synthesized consonants after training, which may indicate a change in the features where attention is focused (Francis, 1998). Thus, it may be that participants required a training exposure first in order to adequately perceive the synthesized, syllabic stimuli used in this study. This may be why participants first required the initial English familiarization task to perform adequately on the English testing portions of this study. However, it may be that the statistical exposure periods were enough to improve perceptual judgments of the experimental stimuli. Future studies may wish to train participants first on the individual synthesized syllables

that will be used by presenting the syllable in natural speech followed by its synthetic counterpart (as this study did with the English words). Such a procedure should help to improve identification scores of the experimental stimuli and decrease variability between subjects.

Altering the Probability of the Artificial Language

This identification data suggests the possibility that participants perceived a language that was different from the one intended. Based upon the most frequent identification results for each syllable, the resulting languages were the following: Language A: *padipu, tibudo, dolatu, daloti*; Language B: *tudalo, tidola, diputi, budopa*. As there were differences between subjects in the actual transcription they used to identify certain syllables, subject perception of the languages could have varied from these modified languages. As the syllables /do/ and /ti/ now occur twice within the languages in two different words, participants would have had to make use of an altered statistical environment. This language was modeled after Saffran, Aslin, and Newport's (1996) stimuli and synthesized on a newer version of the text-to-speech synthesizer that they used. They reported that transitional probabilities within words were $p=1.0$ and between words was $p=0.33$. In light of this identification data, these transitional probabilities change depending upon the word examined. For example, *padipu* (originally *pabiku*) still maintains the $p=1.0$ / $p=0.33$ probabilities as no syllables within the word were repeated in another word. However, *daloti* (originally *daropi*) repeats the syllable /ti/. This weakens the internal transitional probability of the final syllable pair to $p=0.50$ as it occurs 50% of the time in a different word. However, the external

transitional probability between words following *daloti* is also weakened to *p*=0.167 as /ti/ is followed by /bu/ (in *tibudo*) 50% of the time and the initial syllables /pa/, /ti/, and /do/ the other 50%.

Although the identification data altered the perceived languages and the transitional probabilities, participants still had clear probabilities to learn the languages based on statistics. Saffran, Newport, and Aslin (1996) demonstrated that participants can learn a language with weakened transitional probabilities when they used the same syllables within multiple words, creating internal word probabilities ranging from 0.31 to 1.0. Therefore, the weakened probabilities of the languages used in the present study were not a likely cause for not learning the language to a higher degree of proficiency. As an older computer synthesizer created Saffran, Aslin, et al.'s (1996) stimuli, they also likely had similar identification characteristics that were not identified in a separate identification paradigm as in this pilot study, yet infants were still able to learn the language during a 2-minute exposure presentation.

Although this pilot data does raise some interesting considerations for the use of synthetic speech in perceptual speech research, the impact on this study's outcome and interpretation seems minimal, if not non-existent. Any effect of these identified factors (e.g., variability between subjects on syllabic identification, altered transitional probabilities, etc.) could only have increased variability between subjects and made the language's statistical structure harder to perceive. The significant differences obtained in this study suggest that these characteristics of the synthetic stimuli had little effect. This strengthens support for the incredible ability humans have to recognize statistical structure, even with possibly ambiguous stimuli such as those used in this study.

English Validity Testing

The results on the English control tasks within this study demonstrated that participants can successfully complete the neuropsychological protocols that were developed specifically for this investigation. This suggests that these measures can be used in future studies to further investigate online segmentation in speech or possibly other domains. The discrepancy between scores on the posttest and English measures suggest that a strong competency in perceiving the language is needed. Perhaps additional exposure, statistical exposure in additional domains, or other information (statistical, lexical, or otherwise) is required to achieve the same level of proficiency as with English stimuli.

Additional Considerations

The tasks of this study required motivation and attention to complete. Participant abilities in these areas may have decreased by the end of the session influencing performance on the SOST and 2IFC posttests.

Possible effects of dropping out several subjects from the analyses of this study must also be considered. Nittrouer (2001) stated that "children may become uncooperative precisely because they cannot discriminate the stimuli presented. ...on a different day they usually become uncooperative with the same or similar stimuli" (p.1603). Nittrouer continues by saying that "it would be inappropriate to assume that the dismissed infants would have performed as the infants who were not dismissed" (p.1603). It is possible that a similar phenomenon may have occurred with the adult participants

who were not able to complete the experimental tasks. Indeed, 3 participants counted no words from the artificial language in the pretest of the APWC; as Nittrouer's infants, they offered no response during this impossible task. It was impossible because participants had never been exposed to the words in this language before. How could they be expected to count them from within a continuous speech stream? One participant was eliminated from the analysis because she failed to identify any artificial words in either pretest or posttest of the SOST and APWC. She also did not attempt to discriminate between the words and foils during the 2IFC. Was this lack of response due to a behavioral problem? As she did successfully complete the English testing (identifying 95 out of 100 words correctly on the English SOST, which would have tied for the fourth highest score had she been included), it is not likely that the task structures themselves required too difficult of a response. Therefore, it seems possible that she was unable to discriminate (and therefore also recognize and identify) the words in the artificial language; or, simply confused by the task requirements, even after task training.

While the possibility of this study's results being affected by eliminating subjects exists, it is unlikely that the results of this study are compromised. Of the 22 subjects tested who met inclusion criteria, only two were eliminated from analyses. The first subject was excluded due to demonstrated and reported fatigue during testing, the second subject already discussed failed to complete the testing.

Further Questions and Areas of Research

The neuropsychological protocols developed for this study were an attempt to behaviorally measure segmentation performance online before and after learning a novel

language. The development of an electrophysiological paradigm to test the neurophysiological correlates of statistical learning and the identification of auditory objects online may now be warranted. The MMN may be able to index the change from perceiving the novel continuous speech stream as random syllables to perceiving it as a series of auditory objects that are prelinguistic word-forms. A pretest-posttest design to measure how learning the language influences the electrophysiology of the brain, as well as subsequent discrimination paradigms, could be developed to identify electrophysiological changes which occur when learning a new language.

Further isomorphic studies of the domain-generality of statistical learning and the similarity between auditory object formation and perception with visual object formation and perception may also be warranted. Similarities between speech and music perception may also help to delineate how statistical regularities help to identify perceptual objects as a general process that the brain uses in a variety of domains, as well as investigate where speech, music, and other domains begin to differ in their representation and processing. Do different domains tap into statistical information differently? Or does it provide the same information and means to an end for several different brain processes? Also, neuroimaging studies may wish to examine whether statistical regularities tap into the same anatomical regions, or if they are a more generalized mechanism of brain neurophysiology.

Evidence from artificial grammar studies suggests that the same process may be involved in higher order domains of grammar and vocabulary (Gomez & Gerken, 1999, 2001). Studying the time course of acquisition of these grammatical processes may help

to understand the order of how we learn grammar and provide intervention techniques for individuals who are not developing grammatical skills within a normal timeframe.

Examinations of the degree of complexity that can be extracted from statistical structures is also warranted. Newport and Aslin (2004) have recently discovered that adults are sensitive to non-adjacent patterns. This study has suggested that statistical regularities can form objects and enter them into a level of representation where they are susceptible to cognitive-linguistic processing. What exactly is the upper bound of statistical learning?

Along with this are investigations of higher order grouping. This study examined how perceptual objects are formed through combining individual syllabic features into one unit. Past research on morse code has demonstrated an ability to go from perceiving only individual letters to perceiving whole words and eventually higher language habits (Bryan & Harter, 1899). Studies in perceptual grouping for learning artificial grammars are now ongoing (Gomez & Gerken, 1999, 2001), and grouping structures have also been identified for highly skilled activities, such as chess (Chase & Simon, 1973; Gobet, Lane, Crocker, Cheng, Jones, Oliver & Pine, 2001; Simon & Gobet, 2000). It is possible that clauses, sentences, and even discourse scripts become recognized as units based upon the frequency of the co-occurrence of their individual elements. Therefore, individuals who do not perceive these scripts, or have difficulty forming higher-order objects may benefit from intensive treatment based upon developing these high probabilistic memory traces. It may be warranted to investigate possible intervention techniques once the formation of these higher-order groupings is more clearly defined.

This paper presented that statistical learning is a domain-general process that we do innately and implicitly. It is possible that intense and repetitive exposure to an interactive therapeutic environment would increase client success. To that end, studies may be designed to investigate how to tap into these statistical processes within the intervention environment. As sensitivity to statistical regularities occurs very early on in life, it is likely that this structure could be used successfully with developmentally disabled and neurologically impaired individuals. Studies should examine more diverse populations of race, age, and gender, and their sensitivity to statistical structure.

This study also examined perceptual changes resulting from creating higher-order representations of auditory objects. Some studies have connected this process to language learning (Hu, 2003) and implicated resulting structures to language impairment (Maillart, et al., 2004). Future studies may wish to continue defining how this perceptual process impacts clinical populations and whether statistical learning can be used to facilitate the development of more appropriate representations.

Conclusions

The changes in performance and above-chance word discrimination measured in this study clearly demonstrate the ability to use statistical regularities to segment speech. Adults are able to use this information to detect word boundaries and form prelexical auditory word-forms. Participants in this study were able to solve three different perceptual/cognitive tasks which required three different levels of processing. The successful completion of these tasks suggests that statistical regularities can be used to: 1) perform preperceptual discrimination tasks of infrequent stimulus detection (weak

transitional probabilities), 2) structure early perceptual forms for the implicit recognition of familiar pattern structures, and 3) enter into memory prelexical, mental representations of auditory object word-forms which are explicitly available for language processing in phonological synthesis and identification tasks. In short, statistical structure forms auditory objects, which are the acoustic structure to which meaning is symbolically assigned. Speech segmentation is not just a simple detection of word boundaries. Instead, it is an active process where prelexical word-forms are created. In a naturalistic environment, it is possible to envision how statistical pairings of auditory objects to visual (or tactile) objects associates these word-forms to meaning and enters them into the lexicon. I go so far as to predict that statistical learning is a neurophysiological phenomenon of neural plasticity: it is how our brains form structural representations of the random environment in which we find ourselves. As such, statistical regularities influence domain-general processes, and are an important, innate, and environmentally adaptive process of perception. They form structural representations for later cognitive processing. Statistical learning opens an avenue of research to examine the connection and interaction between perception and cognition, and offers potential insight into new ways of therapeutically adapting the environment to facilitate language learning and cognitive retraining in multiple domains.

This study supports conclusions in the following areas:

1) *Segmentation of complex patterns via statistical regularities*

   a. Online speech segmentation can be measured by the neuropsychological protocols developed in this study.

97

b. Statistical regularities are sufficient for initial parsing of the speech stream, identification of word boundaries, and formation of auditory object word-forms.

2) *Replication of discrimination studies*

a. Word boundaries can be detected from continuous statistical presentations and discriminated offline.

b. Discrimination success may occur due to word boundary detection or auditory object detection, as indicated by different results on partword selection errors.

3) *Object recognition and identification*

a. Recognition is the early structure building of auditory objects and occurs implicitly during online presentations.

b. Identification implicates forming a higher-order mental representation of the auditory object that is accessible to cognitive processing. Statistical probabilities are capable of forming these representations.

4) *Changes in performance due to enhanced perception*

a. Perceptual changes occur in the way humans perceive auditory information. Initial perception of auditory features may, during refinement by statistical regularities, become holistically perceived as auditory objects.

b. With statistical exposure, recognition ability may more closely approximate optimal performance, while identification ability may require additional information.

5) *Manipulability of auditory object memory traces*

a. Processing of the speech stream may occur on a hierarchy of difficulty building from discrimination to recognition to identification. Each of these levels may

entail its own level of representation, may require different involvements of memory, and may to various degrees be accessible to explicit cognitive processing.

b. Auditory objects are susceptible to higher levels of processing, such as phonological synthesis, suggesting the possibility of maintaining a prelexical representation in memory, which may be accessed by linguistic processing.

APPENDICES

APPENDIX A

Participant Forms and Recruitment

PARTICIPANT CONSENT FORM
Statistical Speech Segmentation:
A Neuropsychological Investigation of Auditory Object Formation

You are invited to participate in a study that will examine how second language learners identify words within long speech passages. My name is Dan Fogerty and I am using this study as the foundation of my master's thesis in Speech-Language Pathology at Michigan State University. This study is being advised by Jeff Marler, Ph.D., who is an Assistant Professor in Audiology & Speech Sciences. I hope to learn more about how people identify words while listening to someone speak, particularly if this person is speaking a foreign language. You are eligible to participate in this study if you have normal language and hearing, and if you are fluent in only one language. You will be one of 20 college students to participate in this study.

If you decide to participate in this study, you will:

1) You will be asked questions concerning your language and musical experiences, as well as your medical history. These will require yes/no or open-ended responses.

2) Screen your hearing. This will include three measures: a) you will respond every time you hear a beep-like tone, b) your ear canal will be viewed through otoscopy, and c) your middle ear functioning will be assessed with tympanometry.

3) You will be asked to serve in the current study for one 120-minute session (with short breaks when necessary). During this session, you will receive a hearing evaluation and be asked to complete a standardized screening of your ability to manipulate speech sounds. You will also be asked to sit in a sound-treated room while you wear earphones and listen to continuous artificial syllables. None of the sounds you will hear poses any risk of hearing damage. I will ask you to listen and respond to different speech-like words. I will ask you to identify which artificial word sounds the most familiar or to identify those word sounds while listening to another artificial speech passage. You will also be asked to complete these tasks with English speech passages.

4) This study does not involve risks or harm any greater than those ordinarily encountered in daily life. You may choose to withdraw from further participation at any time during the session (and will be paid for your time up to that point). You should inform the investigator immediately if there is discomfort of any kind during the course of the experiment. The investigator may periodically request such information from you during the session.

5) Potential benefits from participating in this study are to provide a stronger theoretical understanding of speech perception and language acquisition. If requested, you may receive a personal written report at the end of the study.

6) Your privacy will be protected to the maximum extent allowable by law. Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission. All forms containing your information will be stored in a locked cabinet in the Cognitive Auditory Research Lab. Only Dan Fogerty and Jeff Marler, Ph.D, will have access to this cabinet. Any publication resulting from this study will identify you only according to code.

You are being asked to participate in a study investigating the learning of speech segmentation. Your decision to participate will not influence your future relations with Michigan State University. You will receive either $15.00 or course credit for the completion of the experimental session. You may refuse to answer any question during the experimental session without penalty.

Please initial here if you wish to receive course credit: _____

If you have any questions about this study at any time, please contact the investigator Dan Fogerty by phone: 394-7412, email: fogertyd@msu.edu, or regular mail: 221 E. Edgewood Blvd., Apt. F, Lansing, MI 48911; or his advisor Jeff Marler, Ph.D. by phone: 355-7628, fax: (517) 432-2370, email: marler@msu.edu, or regular mail: 378 Comm. Arts & Sci. Bldg., East Lansing, MI 48824. If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously if you wish – Peter Vasilenko, Ph.D., Chair, University Committee on Research Involving Human Subjects (UCRIHS) by phone: 355-2180, fax: (517) 432-4503, email: uchrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

Should you choose to discontinue participation in this study, you may withdraw at any time after signing this form.

Thank you for your time in visiting about the project and your consideration to participate. You will receive a copy of this form to keep for your files.

I voluntarily agree to participate in this study.

_____   _____   _____
Participant Signature   Date   Phone #

_____   _____
Signature of Investigator   Date

Statistical Speech Segmentation:

A Neuropsychological Investigation of Auditory Object Formation

RESEARCH PARTICIPANTS NEEDED

For a study in
Audiology & Speech Sciences

Research will be conducted in a session of about 120- minutes.

Listen to artificial speech...

And get $15.00!

Project Description

Have you ever listened to someone speak a foreign language and have difficulty telling where the words are? We might say that the person is speaking too fast – all the words blend together. This study attempts to learn more about how people identify words while listening to someone speak a foreign language. Participation requires some perceptual testing, a hearing screening, and listening to artificial speech. You are eligible to participate in this study if you have no history of a language, reading, or hearing impairment and if you are fluent in only English. Also, due to demonstrated differences in speech processing, you need to be female. Sorry guys!

PHONE SCREEN

SUBJECT ID:_____

Called Date:_____

Screen Results: IN   OUT

Interviewer's Initials:_____

Hello, this is Dan Fogerty from the Michigan State University Speech Segmentation study. Is (person's name) available? I am calling in response to your expression of interest in the study. I have a few questions that help us to determine if you are eligible for the study. Would you mind spending 5-10 minutes on the phone with me at this time to answer these or is there a better time I could call back? Your confidentiality concerning this interview will be protected to the maximum extent allowable by law. If you have any questions along the way, please let me know, and I will do my best to answer them. You indicate your voluntary agreement to let me ask you these questions by beginning this phone interview. If you qualify for the study and you are still interested in being a participant, you will be asked to read and sign a consent form when you come for the study.

Contact Information

Name:_____
          First   Last

Phone: (Home)_____(Work)_____

Email:_____

Major:_____

Demographics

DOB:_____

Age: _____

Gender: ☐ M   ☐ F

Educational History

What is the highest educational level you have attained?

☐ High school

☐ Some college

☐ Graduated from college

☐ Graduate school

Language Experience

Are you a native English speaker?   ☐ Yes ☐ No

Have you ever received formal instruction in a foreign language?   ☐ Yes ☐ No

Do you speak any language other than English?   ☐ Yes ☐ No

      If yes, please continue with the following questions:

      What language(s)?_____

      Please rate your proficiency (1=limited experience, 3=fluent)   1   2   3

Medical History

Do you have any major health, medical, neurological, or psychiatric conditions? ☐ Yes ☐ No

       If yes, please specify: _____

Have you ever received treatment for speech, language, hearing, or reading? ☐ Yes ☐ No

       If yes, please specify: _____

Any physical handicaps?	☐ Yes	☐ No	Describe:
Normal vision and hearing?	☐ Yes	☐ No	Describe:
Any speech or language difficulties?	☐ Yes	☐ No	Describe:
Is your whole family native English speakers?	☐ Yes	☐ No	Describe:
Any learning disabilities?	☐ Yes	☐ No	Describe:
Previous or existing reading difficulty?	☐ Yes	☐ No	Describe:
Training in either phonetics or phonics?	☐ Yes	☐ No	Describe:

Musical Experience

Have you ever received any musical training?   ☐ Yes ☐ No

       If yes, please continue with the following questions:

       Please describe your experience (e.g., choir, private lessons, etc.): _____

       _____

       How many years did you receive this training?

           ☐ 1-2

           ☐ 3-4

           ☐ 5-6

           ☐ 6+

Thank you for your participation in this initial screening. (If they are eligible for the study, continue): *Your responses show that you are eligible to participate in this study. If you decide to continue with the study you will be asked to make one visit to campus that will last about 2 hours. You will receive either $15.00 or course credit for the completion of the experimental session.*

Would you like to schedule your visit with us now?

                      Date:_____   Time:_____

PARTICIPANT WORKSHEET

ID: _____   LANGUAGE: A   B

DATE: _____   COUNTERBALANCED: A   B

EXPERIMENTAL *Scores*	PRETEST *Scores*	ENGLISH VERSION *Scores*

APWC
  Form: A / B
  ▢ ▢ ▢ ▢ ▢ ▢ 36
SOST
  Form: A / B
 	
  100
2IFC
  Initial:_____
  Final: _____ 32
	APWC
	  Form: A / B
	  ▢ ▢ ▢ ▢ ▢ ▢ 36
	SOST
	  Form: A / B
	
	  100
		E- APWC
		
		  ▢ ▢ ▢ ▢ ▢ ▢ 36
		E- SOST
		
		  100

SUBTEST	Raw Score	%tile	Standard Score
Matrices			
Memory for Digits			
Nonword Repetition			
Blending Nonwords			

COMPOSITES	Sums of SS	%tile	Composite Score
Phonological Memory			

COMMENTS:

_____

_____

_____

_____

_____

_____

APPENDIX B

Pilot Syllable Identification Data

SYLLABLE IDENTIFICATION IN ISOLATION

Syllables Presented for Identification

		bi	bu	da	do	go	ku	la	pa	pi	ro	ti	tu
Syllables within the Artificial Language	bi	3											
	bu		36										
	da			43									
	do				45	21							
	go					14							
	ku						16						2
	la							45					
	pa								22				
	pi									9			
	ro												
	ti						3			36		42	3
	tu						3						40
Additional Syllables Identified in Participant Responses	bo					2							
	chi											3	
	di	42											
	dol					1							
	duh		2										
	ha								7				
	hu						1						
	huh								1				
	i		3										
	ka								5				
	lo										45		
	no					3							
	o		3										
	pu						22						
	ta								10				
	tho					3							
	uh		3										
	NR					1							
Accuracy* (%)		7	80	9	100	31	36	100	49	20	0	93	89
Between Consistency† (%)		93	80	93	100	47	53	100	53	87	100	93	87
Within Consistency‡ (%)		100	100	93	100	80	87	100	67	80	100	100	93

Bold highlights the most frequent response $N=15$

* **Accuracy** reflects percentage of correctly identified syllables over all responses between subjects.

† **Between Consistency** is the percent of subjects who identified the syllable the same over at least 2 trials. e.g., 14 of 15 subjects identified *bi* and *di* at least twice; therefore, *bi* has a 93% consistency value.

‡ **Within Consistency** is the percent of subjects who identified the syllable as the same over all individual trials. e.g., 15 of 15 subjects identified *bi* the same over all trials (14 as *di* and 1 as *bi*); therefore, *bi* has a 100% within subject consistency value. (All subjects were 100% consistent for all syllables over 2 trials.)

SYLLABLE IDENTIFICATION IN WORDS

Syllables Presented for Identification

		bi	bu	da	do	go	ku	la	pa	pi	ro	ti	tu
Syllables within the Artificial Language	bi		1										
	bu		7										
	da			12									
	do				12	8							
	go					4							
	ku												
	la							12					
	pa								4				
	pi						2						
	ro												
	ti	2					2			11		12	4
	tu						1						8
Additional Syllables Identified in Participant Responses	di	9	1										
	ha								4				
	hi	1											
	i		3										
	ki									1			
	lo						1				12		
	pu						6						
	ta								4				
Accuracy* (%)		0	58	100	100	33	0	100	33	0	0	100	67
Between Consistency† (%)		75	58	100	100	67	50	100	33	92	100	100	67
Within Consistency‡ (%)		67	67	100	100	100	67	100	100	83	100	100	100

Bold highlights the most frequent response $N=3$

Word List: (subjects heard each syllable four times, twice in each of two words)

pigola	tibudo
tudaro	pabiku
bikuti	golatu
budopa	daropi

* **Accuracy** reflects percentage of correctly identified syllables over all responses between subjects.

† **Between Consistency** reflects the percent of which the syllable was identified most frequently across subjects. e.g., 9 of 12 times *bi* was identified as *di*; therefore, *bi* has a 75% between word consistency value.

‡ **Within Consistency** reflects the percent in which a syllable was identified as the same over all trials of an individual word. e.g., The three subjects identified *bi* as the same for a total of 4 out of a possible 6 times (as each of 3 subjects heard *bi* in two different words); therefore, *bi* has a 67% within word consistency value.

APPENDIX C

Experimental Procedure and Response Forms

Testing Procedure Sequence

Screening Tests
- Language and medical history interview
- K-BIT
  - Matrices
- CTOPP
  - Memory for Digits
  - Nonword Repetition
  - Nonword Blending
- Otoscopy and Hearing Screening

Test Training
- English Word Familiarization Task
- APWC Test Training
- SOST Test Training

Experimental Testing
- APWC Pretest
- SOST Pretest
<5-minute Break>
- **Exposure 1**
- English APWC
- **Exposure 2**
- English SOST
<5-minute Break>
- **Exposure 3**
- APWC Posttest
- SOST Posttest
- 2IFC

Note: As two different tests were created for the APWC and SOST, these tests were counterbalanced between test presentations.

ENGLISH WORD FAMILIARIZATION

Possible English Word List:
 Tuxedo
 Textural
 Juxtapose
 Republic
 Fixative
 Explosion
 Deception
 Sketchily

Please number the words in the order that you hear them.

 _____Deception

 _____Juxtapose

 _____Republic

 _____Textural

 _____Fixative

 _____Tuxedo

 _____Explosion

 _____Sketchily

A P W C *T r a i n i n g   F o r m*

Instructions: *You will be listening to a continuous speech stream. Your job is to count the number of words that you hear. When you hear a familiar word, press the counter. You will hear the example once, and then it will be repeated. Please write the number counted after each presentation.*

EXAMPLE 1:

Please write the number of words counted (displayed on your counter):

A) _____

B) _____

EXAMPLE 2:

Clear your counter

Please write the number of words counted (displayed on your counter):

A) _____

B) _____

EXAMPLE 3:

Clear your counter

Please write the number of words counted (displayed on your counter):

A) _____

B) _____

S O S T   *T r a i n i n g   F o r m*

Instructions: You will hear a series of syllables through your headphones. Each box on this sheet represents one syllable. Some syllables will be unrelated to each other, while other syllables will combine to form a word in English. Your task is to place a dot in each box as you hear a syllable and to draw a slash *after* the box which represents the last syllable in the word. Each line begins with a tone and is followed by 5 beeps to signal you to move to the next line. Follow along with the answer provided for you in the first example of each trial. Then, complete the second example on your own.

TRIAL 1:   deception

A) ☐ ☐ ☐ ☐/☐ **BEEP**

B) ☐ ☐ ☐ ☐ ☐ **BEEP**

TRIAL 2:   deception, republic

C) ☐ ☐ ☐/☐ ☐ ☐ ☐/☐ **BEEP**

D) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

TRIAL 3:   deception, republic, tuxedo

E) ☐ ☐ ☐/☐ ☐ ☐ ☐/☐ ☐ ☐/ **BEEP**

F) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

TRIAL 4:   deception, republic, tuxedo, explosion

G) ☐ ☐ ☐ ☐/☐ ☐ ☐/☐ ☐ ☐ ☐ ☐/☐ ☐ ☐/ **BEEP**

H) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

PRACTICE EXAMPLES:

1) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

2) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

3) ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ **BEEP**

A P W C   R e s p o n s e   F o r m

Instructions: *You will be listening to a continuous speech stream. Your job is to count the number of words that you hear. When you hear a familiar word, press the counter. When you hear a series of 5 beeps, write the number displayed on your counter in the space provided on your APWC response form. Please practice with the following example.*

EXAMPLE:

After you hear 5 beeps, please write the number of words counted (displayed on your counter):

C) _____

D) _____

EXPERIMENT:

PLEASE CLEAR YOUR COUNTER

You will now begin. Remember; press the counter when you hear a familiar word

Please write the number of words counted (displayed on your counter):

1) _____

2) _____

3) _____

4) _____

5) _____

6) _____

S O S T   R e s p o n s e   F o r m

Instructions: You will hear a series of syllables through your headphones. Each box on this sheet represents one syllable. Some syllables will be unrelated to each other, while other syllables will combine to form a word from this study's artificial language. Your task is to place a dot in each box as you hear a syllable and to draw a slash *after* the box which represents the last syllable in the word. Each line begins with a tone and is followed by 5 beeps to signal you to move to the next line. Please listen and follow along with the first example, then complete the remaining three examples.

EXAMPLES:

A) □ □ □/□ □ □ □ □/□ □ □ □/□ □ □/ BEEP

B) □ □ □ □/□ □ □ □/□ □ □/□ □ □ □/ BEEP

C) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

D) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

Now you will begin the test. Remember: Draw a slash *after* the last syllable in the word.

1) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

2) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

3) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

4) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

5) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

6) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

7) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

8) □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ BEEP

GO ON...
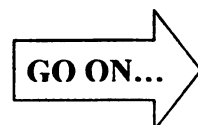
9) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

10) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

11) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

12) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

13) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

14) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

15) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

16) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

17) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

18) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

19) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

20) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

21) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

22) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

23) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

24) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**

25) □ □ □ □ □ □ □ □ □ □ □ □ □ □ **BEEP**   **STOP**

2 I F C   R e s p o n s e   F o r m

Instructions: *You will hear two words for each item. Circle the number corresponding to the "familiar" artificial word. Circle 1 if the word was the first one presented, or 2 if it was the second. Each item will begin with a tone and each column will end with 5 beeps. Please complete the first 3 examples of English words.*

Example 1:   1   2

Example 2:   1   2

Example 3:   1   2

Please continue with the remaining items. Circle the number that corresponds to the familiar word.

1)	1	2	12)	1	2	23)	1	2
2)	1	2	13)	1	2	24)	1	2
3)	1	2	14)	1	2	25)	1	2
4)	1	2	15)	1	2	26)	1	2
5)	1	2	16)	1	2	27)	1	2
6)	1	2	17)	1	2	28)	1	2
7)	1	2	18)	1	2	29)	1	2
8)	1	2	19)	1	2	30)	1	2
9)	1	2	20)	1	2	31)	1	2
10)	1	2	21)	1	2	32)	1	2
11)	1	2	22)	1	2			

BEEPS   **BEEPS**   **BEEPS**

ID: _____   DATE: _____   FORM: _____   SCORE: _____

APPENDIX D

Individual Participant Data

Standardized Criterion Testing

Subject	Matrices	Memory for Digits	Nonword Repetition*	Blending Nonwords	Memory Composite
		Standard Scores			
03A1	120	14	13	13	121
05A1	109	12	13	17	115
06A1	101	13	12	10	115
07A1	98	13	12	11	115
08A2	99	9	8	14	91
09A2	103	12	12	10	112
10A2	96	10	11	11	103
12A2	109	13	11	15	112
15B1	115	12	10	13	106
16B1	101	10	9	14	97
17B2	103	8	7	9	85
18B2	94	11	10	13	103
19B2	116	13	12	16	115
20B1	99	13	12	9	115
21B1	101	11	9	11	100
23A1	94	15	10	11	115
25B2	115	12	9	10	103
26B2	92	13	13	12	118
27A2	104	10	9	15	97
28B1	105	13	12	14	115
MEAN	**104**	**12**	**11**	**12**	**108**
SD	**8.0**	**1.8**	**1.8**	**2.3**	**9.9**

*Nonword Repetition was a non-criterion subtest administered to achieve the composite score.

Two-Interval Forced-Choice (2IFC)

Subject	# Correct	% Correct	Initial Errors	Final Errors
03A1	22	69	8	2
05A1	26	81	5	1
06A1	26	81	1	5
07A1	25	78	6	1
08A2	13	41	11	8
09A2	19	59	7	6
10A2	16	50	9	7
12A2	20	63	5	7
15B1	13	41	13	6
16B1	21	66	8	3
17B2	19	59	8	5
18B2	18	56	5	9
19B2	24	75	4	3
20B1	17	53	6	9
21B1	20	63	5	7
23A1	22	69	9	1
25B2	24	75	3	5
26B2	20	63	6	6
27A2	19	59	10	4
28B1	22	69	4	6
MEAN	**20.3**	**63.4**	**6.7**	**5.1**
SD	**4.0**	**12.0**	**2.9**	**2.5**

Auditory Probabilistic Word Count (APWC): Total Scores

Subject	Pretest Raw	Abs[†]	Log[‡]	Posttest Raw	Abs	Log	English Raw	Abs	Log
03A1	0	36	1.6	5	31	1.5	23	13	1.1
05A1	51	17	1.2	25	15	1.2	17	19	1.3
06A1	54	30	1.5	15	21	1.3	28	8	0.9
07A1	43	15	1.2	60	24	1.4	31	5	0.7
08A2	90	58	1.8	20	16	1.2	31	5	0.7
09A2	13	23	1.4	23	13	1.1	27	9	1.0
10A2	179	151	2.2	126	90	2.0	32	4	0.6
12A2	25	15	1.2	58	22	1.3	65	29	1.5
15B1	0	36	1.6	20	16	1.2	24	12	1.1
16B1	20	22	1.3	23	15	1.2	21	15	1.2
17B2	36	14	1.1	39	15	1.2	28	8	0.9
18B2	83	47	1.7	82	46	1.7	19	17	1.2
19B2	1	35	1.5	34	14	1.1	11	25	1.4
20B1	11	25	1.4	46	16	1.2	15	21	1.3
21B1	148	112	2.0	30	8	0.9	33	3	0.5
23A1	15	21	1.3	0	36	1.6	33	3	0.5
25B2	0	36	1.6	35	7	0.8	33	3	0.5
26B2	67	45	1.7	87	53	1.7	28	14	1.1
27A2	43	35	1.5	24	12	1.1	27	11	1.0
28B1	150	114	2.1	24	12	1.1	30	6	0.8
MEAN	**51**	**44**	**1.6**	**39**	**24**	**1.4**	**27.8**	**12**	**1.1**
SD	**54.0**	**38.0**	**1.6**	**31.0**	**20.0**	**1.3**	**10.8**	**7.6**	**0.9**

†Abs is the Absolute Difference of the Raw Count from the Correct Count
‡Log is the logarithmic transformation values of the Absolute Difference scores. This was calculated to fulfill the normality assumption which was required for the analyses.

Auditory Probabilistic Word Count: Raw Number Counted

| | Pretest | | | | | | | Posttest | | | | | | | English | | | | | | |
| | # Counted for each stream | | | | | | | # Counted for each stream | | | | | | | # Counted for each stream | | | | | | |
Subject	6	4	8	6	8	4	Sum	6	4	8	6	8	4	Sum	6	4	8	6	8	4	Sum
03A1	0	0	0	0	0	0	0	2	0	3	0	0	0	5	4	3	5	3	4	4	36
05A1	10	10	11	7	7	6	51	4	6	2	4	2	7	25	2	2	5	5	4	3	23
06A1	0	14	12	11	9	8	54	3	2	2	3	4	1	15	5	3	5	3	6	3	17
07A1	2	9	9	10	8	5	43	10	7	10	7	15	11	60	5	3	6	6	6	3	28
08A2	6	32	6	16	22	8	90	2	5	4	2	5	2	20	2	6	7	6	8	4	31
09A2	0	3	4	3	3	0	13	4	6	6	4	2	1	23	4	7	5	6	7	4	31
10A2	45	38	21	32	43	0	179	26	17	19	17	31	16	126	4	4	5	6	4	4	27
12A2	2	6	7	5	5	0	25	12	8	9	10	10	9	58	9	11	5	5	10	10	32
15B1	0	0	0	0	0	0	0	2	5	4	3	6	0	20	14	11	11	11	10	10	65
16B1	3	7	3	2	3	2	20	2	3	3	6	5	4	23	1	5	5	5	7	3	24
17B2	6	8	4	9	5	4	36	4	8	10	5	5	7	39	3	2	4	5	3	4	21
18B2	19	16	11	11	15	11	83	18	15	17	7	14	11	82	3	4	6	5	6	4	28
19B2	0	0	0	0	1	0	1	6	7	9	6	4	2	34	1	1	2	4	5	3	19
20B1	3	1	3	2	1	1	11	8	4	11	4	11	8	46	2	2	3	4	3	1	15
21B1	15	25	27	26	25	30	148	7	4	4	6	5	4	30	4	8	6	8	8	4	33
23A1	2	3	3	2	2	3	15	0	6	7	7	9	6	35	4	3	6	6	8	4	33
25B2	0	0	0	0	0	0	0	5	9	12	21	21	19	87	4	4	7	7	8	4	33
26B2	12	14	15	21	5	0	67	6	4	5	3	6	0	24	5	4	6	7	6	3	28
27A2	0	13	15	11	4	0	43	6	5	5	4	4	0	24	5	5	6	7	3	3	27
28B1	19	16	25	28	29	33	150	5	3	5	3	5	3	24	3	3	7	6	7	4	30
MEAN	7.2	11	8.8	9.8	9.3	5.6	51.4	6.4	6	6.9	6.3	7.8	5.5	38.8	3.5	3.7	5.6	5.3	6.2	3.7	27.8
SD	11	11	8.3	10	12	9.5	53.8	6.3	4.5	5.1	7.4	5.1	5.1	30.6	1.6	2.7	2.3	2	1.9	1.7	10.8

Italics indicate the correct number of words in each presentation stream.

**Auditory Probabilistic Word Count: Absolute Difference from Correct Count**

Subject	Pretest # Counted for each stream						Sum	Posttest # Counted for each stream						Sum	English # Counted for each stream						Sum
03A1	6	4	8	6	8	4	36	6	4	5	4	8	4	31	2	1	3	3	4	0	13
05A1	4	6	3	1	1	2	17	2	2	1	2	6	2	15	4	2	5	3	4	1	19
06A1	6	10	4	5	1	4	30	3	2	6	3	4	3	21	1	1	2	1	2	1	8
07A1	4	5	1	4	0	1	15	4	5	2	1	7	5	24	1	1	2	0	1	0	5
08A2	0	28	2	10	14	4	58	4	2	3	2	4	1	16	4	0	1	0	0	0	5
09A2	6	1	4	3	5	4	23	3	2	3	0	4	1	13	4	1	3	0	1	0	9
10A2	39	34	13	26	35	4	151	20	13	11	11	23	12	90	2	0	0	1	1	0	4
12A2	4	2	1	1	3	4	15	6	4	1	4	2	5	22	3	10	3	5	2	6	29
15B1	6	4	8	6	8	4	36	4	1	4	3	4	0	16	3	3	3	1	1	1	12
16B1	3	3	5	4	5	2	22	4	1	4	3	2	1	15	3	2	4	1	3	2	15
17B2	0	4	4	3	3	0	14	2	4	2	1	3	3	15	3	0	2	1	2	0	8
18B2	13	12	3	5	7	7	47	12	11	9	3	4	7	46	4	2	5	2	3	1	17
19B2	6	4	8	6	8	3	35	0	3	8	0	0	3	14	3	3	6	5	6	2	25
20B1	3	3	5	4	7	3	25	2	5	3	1	2	3	16	4	2	5	3	5	2	21
21B1	9	21	19	20	17	26	112	1	0	4	0	3	0	8	2	1	0	0	0	0	3
23A1	4	1	5	4	6	1	21	6	4	8	6	8	4	36	3	0	0	0	0	0	3
25B2	6	4	8	6	8	4	36	1	2	1	1	2	0	7	2	0	1	0	0	0	3
26B2	6	10	7	15	3	4	45	1	5	4	15	13	15	53	3	2	4	1	2	2	14
27A2	6	9	7	5	4	4	35	0	0	5	3	4	0	12	2	1	5	0	2	1	11
28B1	13	12	17	22	21	29	114	1	1	3	3	3	1	12	3	1	1	0	0	0	6
MEAN	**7.2**	**8.9**	**6.6**	**7.8**	**8.2**	**5.7**	**44.3**	**4.1**	**3.6**	**4.4**	**3.3**	**5.3**	**3.5**	**24**	**2.8**	**1.6**	**2.8**	**1.4**	**2**	**1**	**12**
SD	**8.2**	**9**	**4.9**	**7.1**	**8.2**	**7.6**	**37.5**	**4.6**	**3.3**	**2.8**	**3.7**	**5.1**	**4**	**20**	**0.9**	**2.1**	**1.9**	**1.6**	**1.7**	**1.4**	**7.6**

Slow Online Segmentation Test (SOST)

Subject	Pretest			Posttest			English		
	Total	Correct	% Correct	Total	Correct	% Correct	Total	Correct	% Correct
03A1	27	13	48.1	55	38	69.1	71	67	94.4
05A1	59	44	74.6	71	61	85.9	91	91	100.0
06A1	12	5	41.7	58	13	22.4	94	83	88.3
07A1	38	7	18.4	56	5	8.9	92	90	97.8
08A2	71	23	32.4	64	13	20.3	97	89	91.8
09A2	27	10	37.0	35	16	45.7	59	33	55.9
10A2	52	9	17.3	82	23	28.0	96	95	99.0
12A2	39	11	28.2	31	20	64.5	64	40	62.5
15B1	35	7	20.0	80	28	35.0	90	86	95.6
16B1	48	43	89.6	71	42	59.2	76	75	98.7
17B2	56	25	44.	55	24	43.6	68	42	61.8
18B2	28	7	25.0	69	24	34.8	58	33	56.9
19B2	39	17	43.6	38	32	84.2	75	74	98.7
20B1	67	26	38.8	33	10	30.3	45	44	97.8
21B1	92	21	22.8	76	37	48.7	96	92	95.8
23A1	90	37	41.1	96	35	36.5	99	99	100.0
25B2	0	0	0.0	94	33	35.1	98	98	100.0
26B2	43	21	48.8	34	23	67.6	91	86	94.5
27A2	40	15	37.5	36	6	16.7	70	68	97.1
28B1	64	33	51.6	54	28	51.9	98	98	100.0
MEAN	**46.4**	**18.7**	**37.3**	**59.4**	**25.6**	**45.6**	**81.4**	**74.2**	**88.3**
SD	**23.4**	**12.8**	**20.9**	**20.5**	**13.6**	**22.0**	**16.6**	**23.2**	**16.3**

APPENDIX E

Data Figures
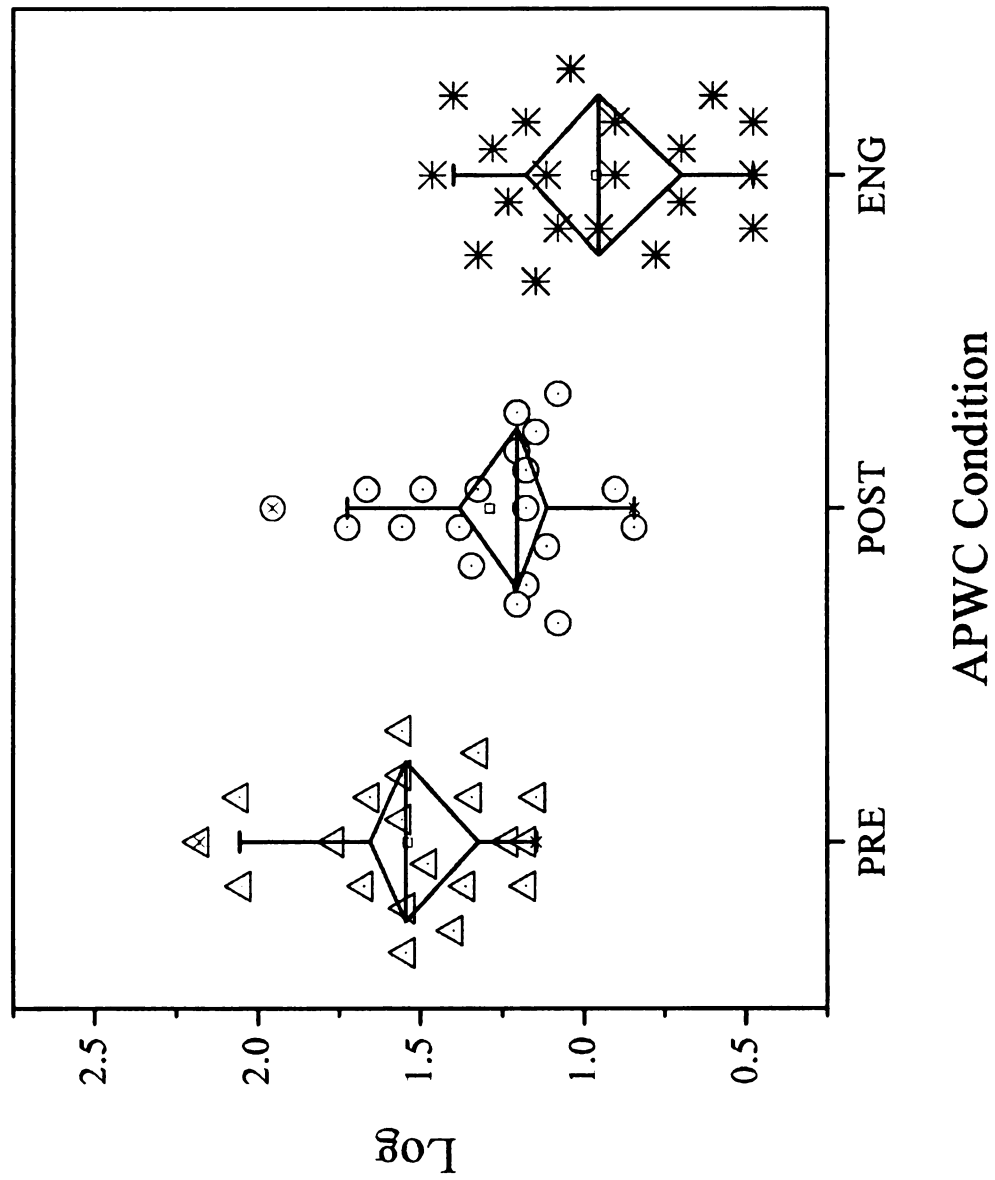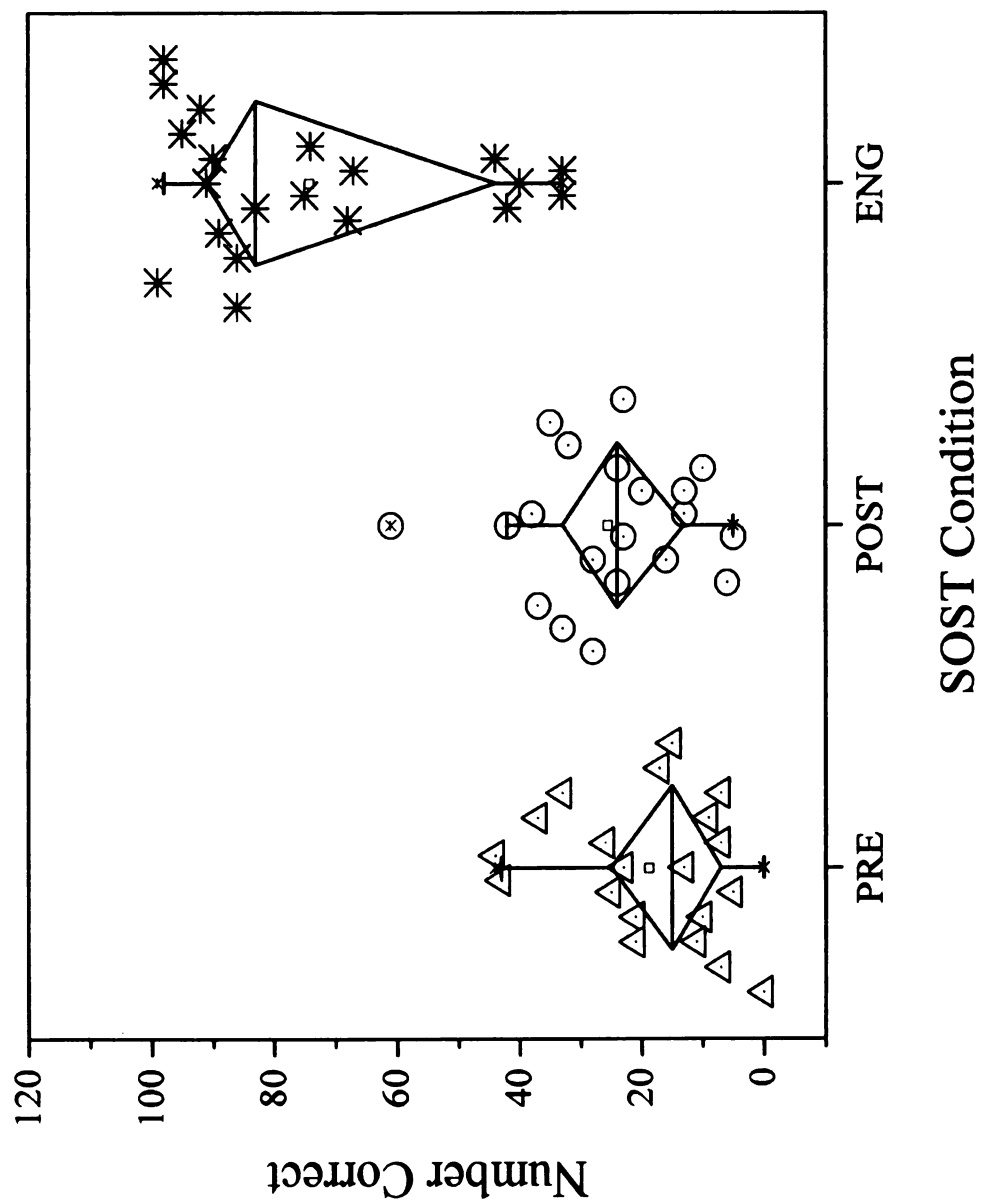
REFERENCES

REFERENCES

Alain, C., Cortese, F., & Picton, T. (1999). Event-related brain activity associated with auditory pattern processing. *Neuroreport, 10*, 2429-2434.

American National Standards Institute (ANSI). (1996). *Specifications for audiometers* (ANSI S3.6-1996). New York: Author.

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8 month-old infants. *Psychological Science, 9*, 321-324.

Atienza, M., Cantero, J., Grau, C., Gomae, C., Dominguex-Marin, E., & Escera, C. (2003). Effects of temporal encoding on auditory object formation: a mismatch negativity study. *Cognitive Brain Research, 16*, 359-371.

Bailey, D. & Wolery, M. (1992). *Teaching Infants and Preschoolers with Disabilities*. New York: Merrill.

Barker, J. & Cooke, M. (1999). Is the sine wave speech cocktail party worth attending? *Speech Communication, 27*, 159-174.

Besson, M. & Schön, D. (2003). Comparison between language and music. In I. Peretz and R. Zatorre (Eds), *The cognitive neuroscience of music* (pp.269-293). London: Oxford University Press.

Bradley, L. & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor: University of Michigan Press.

Brattico, E., Winkler, I., Naatanen, R., Paavilainen, P. & Tervaniemi, M. (2002). Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. *Neuroreport, 13*, 1747-1751.

Breitenstein, C. & Knecht, S. (2003). Spracherwerb und statistisches Lernen [Language acquisition and statistical learning] [Abstract]. *Nervenarzt, 74*, 133-143.

Brent, M. (1999). Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences, 3*, 294-301.

Bryan, W. & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*, 345-375.

Burnham, D., Earnshaw, L., & Quinn, M. (1987). The development of the categorical identification of speech. In B. McKenzie and R. Day (Eds.), *Perceptual development in early infancy: Problems and Issues* (pp. 237-275). Hillsdale, NJ: Lawrence Erlbaum Associates.

131

Canfield, R. & Haith, M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology, 27*, 198-208.

Carroll, J. Davies, P. & Richman, B. (1971). *Word frequency book*. New York: American Heritage Publishing Co, Inc.

Chase, W. & Simon, H. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing* (pp. 215-281). Oxford, England: Academic Press.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K. & Naatanen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*, 351-353.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*, 1570-1580.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Cole, R., & Jakimik, J. (1980). *A model of speech perception*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Coney, J. (2002). Lateral asymmetry in phonological processing: Relating behavioral measures to neuroimaged structures. *Brain and Language, 80*, 355-365.

Conroy, C. (2003). The effects of four variables on the intelligibility of synthesized sentences [Abstract]. *Journal of the Acoustical Society of America, 114*, 2366.

Conway, C., & Christiansen, M. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*, 539-546.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.

Craik, F. & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

Davis, G. (2000). *Aphasiology: Disorders and clinical practice*. Boston: Allyn & Bacon.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in

infants. *Science, 171*, 303-306.

Elman, J. & Bates, E. (1997). Acquiring language [Letter to the editor]. *Science, 276*, 1180.

Felleman, D. & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*, 1-47.

Fiser, J., & Aslin, R. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458-467.

Fitch, R., Miller, S. & Tallal, P. (1997). Neurobiology of speech perception. *Annual Review of Neuroscience, 20*, 331-353.

Francis, A. (1998). Perceptual learning of synthetic speech [Abstract]. *Journal of the Acoustical Society of America, 104*, 1835.

Fritch, S., Large, N., Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*, 481-496.

Galotti, K. (1999). *Cognitive Psychology In and Out of the Laboratory*. Boston: Wadsworth Publishing Company.

Gobet, F., Lane, P., Crocker, S., Cheng, P., Jones, G., Oliver, I. & Pine, J. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Science, 5*, 236-243.

Gomez, R. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gomez, R. & Gerken, L. (2001). Infant artificial language learning and language acquisition. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings* (pp. 42-48). Malden, MA: Blackwell Publishers Ltd.

Goodsitt, J., Morgan, J., & Kuhl, P. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20*, 229-252.

Griffiths, T. (2003). The neural processing of complex sounds. In I. Peretz and R. Zatorre (Eds), *The cognitive neuroscience of music* (pp. 167-177). London: Oxford University Press.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics, 38*, 299-310.

Hauser, M., Newport, E., & Aslin, R. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition, 78*,

133

B53-B64.

Hayes, J. & Clark, H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. Hayes (Ed.), *Cognition and the Development of Language* (pp. 221-223). New York: John Wiley & Sons.

Hohne, E., & Jusczyk, P. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics, 56* 613- 623.

Hollich, G.J., Hirsh-Pasek, K., & Golinkoff, R.M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65* (3).

Houston, D., Jusczyk, P., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds, *Psychonomic Bulletin & Review, 7*, 504-509.

Hu, C.-F. (2003). Phonological Memory, Phonological Awareness, and Foreign Language Word Learning. *Journal of Learning Language, 53*, 429-462.

Huettel, S., Mack, P., & McCarthy, G. (2002). Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience, 5*, 485-490.

Hunt, R. & Aslin, R. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*, 658-680.

Iman, R. (1994). *A data-based approach to statistics*. Belmont, CA: Duxbury Press.

Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548-567.

Johnson, J. & Newport, E. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39*, 215-258.

Jourdain, R. (1997). *Music, the brain and ecstasy*. New York: William Morrow & Co.

Jusczyk, P. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. (1998). Using the headturn preference procedure to study language. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 188-204). Stanford, CT: Ablex.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science, 3*, 323-328.

Jusczyk, P., & Aslin, R. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1-23.

Jusczyk, P., & Hohne, E. (1997). Infants' memory for spoken words. *Science, 277,* 1984-1986.

Jusczyk, P., Hohne, E., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61,* 1465-1476.

Jusczyk, P., Houston, D., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159-207.

Kaufman, A., & Kaufman, N. (1990). *Kaufman brief intelligence test.* Circle Pine, MN: American Guidance Services.

Kersten, D. & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13,* 150-158.

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83,* B35-B42.

Kluender, K., Diehl, R., & Killeen, P. (1987). Japanese quail can learn phonetic categories. *Science, 237,* 1195-1197.

Kubovy, M. & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97-126.

Kuhl, K., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606-608.

Kuhl, P. & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69-72.

Kuhl, P. & Padden, D. (1982). Enhanced determinability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics, 32,* 542-550.

Kulynych, J., Vladar, K., Jones, D., & Weinberger, D. (1994). Gender differences in the normal lateralization of the supratemporal cortex: MRI surface-rendering morphometry of Heschl's gyrus and planum temporale. *Cerebral Cortex, 4,* 107-118.

Lake, D. & Bryden, M. (1976). Handedness and sex differences in hemispheric asymmetry. *Brain and Language, 3,* 266-282.

Lambe, E. (1999). Dyslexia, gender, and brain imaging. *Neuropsychologia, 37,* 521-536.

Liebenthal, E., Binder, J., Piorkowski, R., & Remez, R. (2003). Short-term

reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience, 15,* 549-558.

Lovaas, I. (1977). *The autistic child: language development through behavior modification.* New York: Irvington Publishers.

Lovett, M. (2002). Problem solving. In D. Medin & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology, Third Edition, Volume Two: Memory and Cognitive Processes* (pp. 317-362). New York: John Wiley & Sons.

Maillart, C., Schelstraete, M.-A., & Hupet, M. (2004). Phonological representations in children with SLI: A study of French. *Journal of Speech, Language, and Hearing Research, 47,* 187-198.

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition, *Cognition, 25,* 71-102.

Mattys, S., & Jusczyk, P. (2001a). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78,* 91-121.

Mattys, S., & Jusczyk, P. (2001b). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27,* 644-655.

Mattys, S., Jusczyk, P., Luce, P., & Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465-494.

McClure, K., Ferreira, F., & Bisanz, G. (1996). Effects of grade, syllable segmentation, and speech of presentation on children's word-blending ability. *Journal of Educational Psychology, 88,* 670-681.

McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39,* 21-46.

McQueen, J., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 621-638.

Morgan, J. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35,* 666-688.

Murray, S., Kersten, D., Olshausen, B., Schrater, P., & Woods, D. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Science, 99,* 15164-15169.

Naatanen, R. (1992). *Attention and brain function.* Hillsdale, NJ: Lawrence Erlbaum Associates.

136

Naatanen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience, *Ear & Hearing, 16*, 6-18.

Naatanen, R., Schroger, E., Karakas, S., Tervaniem, M., & Paavilainen, P. (1993). Development of a memory trace for complex sound in the human brain. *Neuroreport, 4*, 503-506.

Nemoy, E. & Davis, S. (1954). *The correction of defective consonant sounds*. Magnolia, MA: Expression.

Newport, E. & Aslin, R. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *Journal of the Acoustical Society of America, 110*, 1598-1605.

Norris, D., McQueen, J., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1209-1228.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811-814.

Pascual-Leone, A. & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*, 510-512.

Pugh, K., Shaywitz, B., Shaywitz, S., Constable, R., Skudlarski, P., Fulbright, R., Bronen, R., Shankweiler, D., Katz, L., Fletcher, J., & Gore, J. (1996). Cerebral organization of component processes in reading. *Brain, 119*, 1221-1238.

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 349-351.

Rauschecker, J. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8*, 516-521.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Reynolds, M., Bond, Z., Fucci, D. (1996). Synthetic speech intelligibility: Comparison of native and non-native speakers of English. *Augmentative and Alternative Communication, 12*, 32-36.

Ritter, W., Deacon, D., Gomes, H., Javitt, D., & Vaughan, H. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review. *Ear & Hearing, 16*, 52-67.

Roy, D. & Pentland, A. (2002). Learning words from sights and sounds: a computational model. *Cognitive Science, 26*, 113-146.

Saffran, J. & Wilson, D. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4,* 273-284.

Saffran, J. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition, 81*, 149-169.

Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Saffran, J. (2003a). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, J. (2003b). Mechanisms of musical memory in infancy. In I. Peretz and R. Zatorre (Eds), *The cognitive neuroscience of music* (pp. 32-41). London: Oxford University Press.

Saffran, J., Aslin, R., & Newport, E (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by infants and adults. *Cognition, 70*, 27-52.

Saffran, J., Newport, E., & Aslin, R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Sanders, L., Newport, E., & Neville, H. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5*, 700-703.

Sasaki, T., Suetomi, D., Nakajima, Y., & ten Hoopen, G. (2002). Time-shrinking, its propagation, and gestalt principles. *Perception & Psychophysics, 64*, 919-931.

Sato, Y., Yabe, H., Hiruma, T., Sutoh, T., Shinozaki, N., Nashida, T., & Kaneko, S. (2000). The effect of deviant stimulus probability on the human mismatch process. *Neuroreport, 11*, 3703-3708.

Scientific Learning Corporation. (1998) *Fast ForWord*™ [Computer Software]. Berkeley, CA: Author.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature, 404*, 841-847.

Shaywitz, B., Shaywitz, S., Pugh, K., Constable, R., & Skudlarski, P. (1995). Sex differences in the functional organization of the brain fro language. *Nature, 373*,

607-609.

Simon, H. & Gobet, F. (2000). Expertise effects in memory recall: Comment of Vicente and Wang (1998). *Psychological Review, 107*, 593-600.

Simos, P. & Molfese, D. (1997). Electrophysiological responses from a temporal order continuum in the newborn infant. *Neurophsychologia, 35*, 89-98.

Steinschneider, M., Schroeder, C., Arezzo, J., & Vaughan, H. (1995). Speech-evoked activity in primary auditory cortex: Effects of voice onset time. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials, 92*, 30-43.

Sur, M. & Leamey, C. (2001). Development and plasticity of cortical areas and networks. *Nature Reviews Neuroscience, 2*, 251-262.

Terrance, H. (1991). Chunking during serial learning by a pigeon: I. basic evidence. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 81-93.

Tervaniemi, M. (2003). Musical sound processing: EEG and MEG evidence. In I. Peretz and R. Zatorre (Eds), *The cognitive neuroscience of music* (pp. 294-309). London: Oxford University Press.

Torgesen, J., & Morgan, S. (1990). Phonological synthesis tasks: A developmental, functional, and componential analysis. In H. Swanson & B. Keogh (Eds.), *Learning disabilities: Theoretical and research issues* (pp. 263-276). Hillsdale, N.J.: L. Erlbaum Associates.

Torgesen, J., Morgan, S., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84*, 364

Tu, Z. & Zhu, S-C. (2002, April). Parsing images into region and curve processes. In A. Heyden, G. Sparr, M. Nielsen, P. Johnasen (Eds.), *Proceedings of the 7th European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag. Abstract retrieved January 21, 2004, from http://link.springer.de/link/service/series/0558/tocs/t2352.htm#toc2352.

Vaz Pato, M., Jones, S., & Sprague, P. (2002). Mismatch negativity to single and multiple pitch-deviant tones in regular and pseudo-random complex tone sequences. *Clinical Neurophysiology, 113*, 519-527.

Vitevitch, M., & Luce, P. (1999). Probabilistic and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374-408.

Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity and systematic of strategies on the acquisition of problem structure. *Cognitive*

Science, 20, 75-100.

Wagner, R., Torgesen, J., & Rashotte, C. (1994). Development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73-87.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). *Comprehensive test of phonological processing.* Austin, TX: PRO-ED.

Wagner, R., Torgesen, J., Rashotte, C., Hecht, S., Barker, T., Burgess, S., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Winkler, I. & Czigler, I. (1998). Mismatch negativity: Deviance detection or the maintenance of the "standard." *Neuroreport, 9,* 3809-3813.

Winkler, I., Cowan, N., Csepe, V., Czigler, I., & Naatanen, R. (1996). Interactions between transient and long-term auditory memory as reflected by mismatch negativity. *Journal of Cognitive Neuroscience, 8,* 405-415.

Winkler, I., Schroger, E., & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience, 13,* 59-71.

Woodcock, R. & Johnson, M. (1989, 1990). *Woodcock-Johnson psycho-educational battery-revised.* Allen, TX: DLM Teaching Resources.