Contextual Influences in the Sequential Integration of Speech

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the
Graduate School of The Ohio State University

By

Adam J. Thomas, B.S.

Graduate Program in Psychology

The Ohio State University

2013

Thesis Committee:

Mark A. Pitt, Advisor

James Todd

Antoine Shahin

Copyright by

Adam J. Thomas

2013

Abstract

The influence of memory on the perceptual integration of speech sounds over time is poorly understood. The current study examined the influence of different linguistic properties on the sequential integration of speech using a novel methodology. Cues to lateralization were varied in order to produce perceptual segregation of an /s/ onset from the following speech sounds within words and nonwords (e.g. “slaz”). In Experiment 1, different sources of contextual (lexical, phonological) information elicited different amounts of perceptual segregation, indicated by a subjective measure of streaming. Experiment 2 used gap detection, a task showing impairment in segregated relative to integrated percepts (Woods, Sorkin, & Boggs, 1979). This experiment strengthened interpretations of the results of Experiment 1 as being due to changes in sequential integration. Gap detection worsened for items reported as streamed, validating the results of Experiment 1. The results of the current study provide evidence for a schematic grouping process in speech perception and provide the groundwork for further investigation, both for the contribution of schemas to the perceptual grouping of speech and their relevance to higher-level processes of spoken word recognition.

Vita

2007 Manzano High School

2011 Rachael Dowler Outstanding Honors Thesis,
UNM Psychology Honors Program

2011 B.S. Psychology, University of New Mexico

2011-2012 University Fellow, The Ohio State University

2012-Present..... Graduate Teaching Associate, Department of
Psychology, The Ohio State University

Fields of Study

Major Field: Psychology

Table of Contents

Abstract.....	ii
Vita	iii
List of Tables	vi
List of Figures	vii
1 Introduction	1
2 Experiment 1.....	15
2.1 Method	21
2.1.1 Participants	21
2.1.2 Stimuli	21
2.1.3 Procedure.....	23
2.2 Results.....	24
2.3 Discussion.....	29
3 Experiment 2.....	32
3.1 Method	33

3.1.1	Participants	33
3.1.2	Stimuli	34
3.1.3	Procedure.....	35
3.2	Results.....	37
3.3	Discussion.....	38
4	General Discussion.....	41
	References	50

List of Tables

Table 1. Conditions of contextual information in Experiment 1.....	16
Table 2. Specific comparisons made across contextual conditions in Experiment 1	17
Table 3. Fixed effects tests for the logistic mixed effects model in Experiment 1	26

List of Figures

Figure 1. A spectrogram of the word “spike”	3
Figure 2. Results of Experiment 1, showing mean proportions of responding “Not /s/-initial” by spatial continuum step and condition of contextual information.....	25

Chapter 1: Introduction

In a typical listening situation, a listener must discriminate between different sounds arriving at the ear simultaneously. Auditory scene analysis (Bregman, 1990) refers to the process a listener carries out in order to parse the complex pattern of auditory information into various perceptual streams. There are two primitive processes for stream formation in auditory scene analysis: simultaneous integration and sequential integration. Simultaneous integration occurs when two sounds share enough acoustic properties- common onset, harmonicity, comodulation- in order to fuse into a single perceptual stream at one moment in time. Sequential integration occurs when sounds are grouped together over time into a single stream. The classic demonstration of sequential integration is playing alternating high and low-frequency pure tones for a listener (i.e. Bregman & Campbell, 1971). At a lower rate of presentation, the listener will hear the high and low tones as occurring in a single stream. As the presentation rate increases, the listener will eventually hear high tones perceptually segregate from low tones, forming two auditory streams. If one begins at a high rate of presentation and gradually slows down the repetition of high and low frequency tones, participants will perceive the tones change from being in two streams to one, meaning they change from segregated to integrated into the same stream. Changes from perceiving one stream segregate into two streams, or perceiving two streams integrate into one stream, are due to changes in the sequential integration of the tones. The fundamental processes of

auditory scene analysis are thought to be used in order to group all acoustic signals into streams, including speech (Bregman, 1990; Darwin, 1997).

The speech signal presents a unique challenge to the basic processes of scene analysis described above. A talker can produce a variety of spectrally dissimilar sounds in quick succession, and the listener must be able to group these sounds together over time in order for the linguistic message they convey to be understood. For example, Figure 1 shows a spectrogram of the word “spike”. Within it is a high-frequency hiss /s/, bursts of noise accompanying /p/ and /k/, as well as a periodic sound /ai/, all of which must be integrated across time and over periods of relative silence in order to be understood as the word “spike”. Failure to sequentially integrate speech phonemes into a coherent stream can result in misperceptions of the word, as demonstrated in studies of the verbal transformation effect (Shoaf and Pitt, 2002).

Speech phonemes are produced at a rate of about 10 per second (Warren, 1982). At this rate of presentation, alternating high and low pure tones can easily segregate into two streams, perhaps implying that the same should be true for the complex sounds of speech. Indeed, there are a few experiments that use repeating cycles of vowels that illustrate the problem of grouping speech sounds across time. Warren & Warren (1970) presented listeners with cycles of four steady-state vowels, such as /i/, /u/, /e/, and /o/, with each vowel presented for 200 milliseconds. Listeners performed no better than chance at reporting the order in which the vowels occurred, an effect that is indicative of stream segregation (Bregman, 1990).

Figure 1. A spectrogram of the word “spike”

Listeners’ performance in judging the order of the vowels only improved to near-perfect accuracy once the researchers used vowels with normal vocal quality and each vowel had a 50 millisecond silence following it. Dorman, Cutting, & Raphael (1975), in an experiment that also used repeating sequences of vowels, corroborated Warren & Warren’s (1970) findings that natural-sounding speech is more resilient to impaired judgments of temporal order. Sequences that had formant transitions between vowels were judged accurately 65% of the time, as opposed to sequences of steady-state vowels with no transitions, which were judged accurately 30% of the time. Formant transitions served to facilitate the sequential integration of these vowel sounds, making judgments of temporal order much easier. These studies demonstrate the importance of perceptual grouping in speech. The rate at which speech sounds are produced, often much faster than the rate used in Warren & Warren (1970) and Dorman et al

(1975), should cause speech sounds to segregate into multiple sub-streams of hisses, clicks, and periodic sounds. The robustness of speech perception is further demonstrated in studies that instead use repeating sequences of nonspeech sounds, in which listeners are unable to accurately judge their order (Warren, Obusek, Farmer, & Warren, 1969). From these studies, particularly in the inclusion of formant transitions in Dorman et al (1975), the role of acoustics in determining how well sounds will sequentially integrate is clearly demonstrated.

Speech is unique relative to other auditory stimuli, however, in that listeners have detailed knowledge that they may be able to use in order to organize incoming auditory information. Different types of stored knowledge, such as phonological and lexical information, could all be used in order to influence the auditory grouping of speech sounds. Indeed, there is reason to suggest that listeners would take advantage of their knowledge of speech in the perceptual grouping process, as there are many cases where entirely signal-driven processing is unlikely to provide sufficient explanations for successful integration in speech (Remez, 2005). For example, studies using speech with highly impoverished cues to perception (Remez et al, 1981; Remez et al, 1994) provide cases where bottom-up processing would be insufficient to achieve successful integration, yet participants are able to perceive speech. Another line of evidence showing the importance of high-level contextual influences on integration is the study of cross-linguistic differences in the perception of speech. Speakers with different native languages have different sets of sounds, or phoneme inventories, that they will need to be able to assimilate into a speech stream. For example, a native speaker of Zulu has to be able to

integrate “click consonants” into the speech stream. Click consonants are produced with a sucking action of the tongue then releasing suction abruptly at the tongue or lips. This articulation results in a brief, spectrally complex sound with rapid acoustic variation, not unlike the acoustic realization of stop-consonants such as /k/ or /t/. A speaker of Zulu will treat these sounds as phonetic whereas a speaker of English will treat these clicks as nonspeech sounds (Best & Avery, 1999). It is interesting to note that integrating such sounds presents a perceptual challenge like the one previously referred to in Figure 1- integrating the rapidly varying, high frequency /s/ noise with the following phonemes. There is no basis for cross-linguistic differences in click consonant perception with bottom-up cues alone. The difference in the perception of sounds such as click consonants as speech or nonspeech is based on the listener’s knowledge of the typical patterns of the speech signal. Compelling examples such as these suggest that understanding the sequential integration of speech as occurring solely due to bottom-up cues is unlikely to give a full understanding of the process.

If listeners are using higher-level representations of speech, such as lexical representations, in order to organize speech sounds, it may be useful to think of their contribution in terms of Bregman’s (1990) schematic grouping process. The schematic grouping process that Bregman (1990) describes is a way of conceptualizing when and how top-down processing in auditory grouping occurs. Schemas extract familiar patterns from a sound mixture and are a powerful tool for organizing sounds into meaningful auditory objects. The primitive scene analysis described above partitions the auditory signal while the attention-driven

schematic process groups together patterns from the sound mixture. Schemas enable a listener to organize acoustic information in terms of one's prior knowledge of a signal. A simple demonstration of schema-driven processing can be found in Deutsch (1972), who designed an experiment wherein she played the tones of familiar melodies, such as "Yankee Doodle", for participants. Melodies were distorted by having each sequential tone go up or down by one octave, or staying in the same octave. Participants were unable to determine what melodies they were hearing from these distorted patterns. However, if they were told in advance which melodies they were hearing, they could follow the melody. In other words, they could integrate tones into a single stream that without having the schema of the melody would have segregated into a meaningless pattern of tones. Their schema of the familiar melody enabled them to overcome the frequency separation of the sequential tones.

Although schema-driven processing would benefit listeners while organizing speech, experiments demonstrating the effects of higher-level (i.e. lexical) schematic grouping on the auditory organization of speech are somewhat uncommon. However, schematic grouping effects on speech have been observed in both simultaneous and sequential integration.

Effects that demonstrate schematic grouping in the simultaneous integration of speech have been demonstrated both in studies of duplex perception and sinewave speech. In duplex perception, a sound is constructed in which three sine tones are used to convey the first three formants of a syllable, typically using /da/ and /ga/. These syllables are distinguished by the third-formant transition onto the syllable. These simple sounds are presented dichotically, with

the third formant transition presented to one ear and the rest of the acoustic information, or the base, presented to the other ear. Duplex perception occurs when listeners are simultaneously able to hear the transition as contributing to the base and hear the formant transition as a chirp in the other ear. One study in duplex perception that is especially relevant is Nygaard (1993). In experiment 3, she used duplex stimuli in which the formant transition presented to one ear led the presentation of the base in the other ear by 0, 40, or 80 milliseconds. This asynchrony between the transition and the base made it less likely that they would cohere, and therefore less likely that a participant would be able to accurately report whether the onset was a /g/ or a /d/. Syllables were either words or nonwords (“dut”, “dusk”, “gut”, “gusk”). Stimuli whose transitions formed a word when grouped with the base in the opposite ear were more accurately categorized than those that formed a nonword. “Dusk” and “gut” were more accurately categorized in terms of their onset phoneme than “gusk” and “dut” because the lexical information helped to integrate the formant transition in one ear with the base in the other ear. The results of Nygaard (1993) lend credence to the idea that higher-level speech representations can contribute to the perceptual organization of speech sounds.

Studies of sinewave speech (Remez et al, 1981; Remez et al, 1994) have also been interpreted as demonstrating schematic grouping effects (Darwin, 2008). Sinewave speech is produced by creating three (or more) sinusoidal tones that reproduce the frequency and amplitude variation of the first three formants of speech, thus preserving spectrotemporal variation of speech while eliminating most other traditional cues to speech perception. When

participants first listen to sinewave speech, they hear it as a strange collection of whistles and tones (Remez et al, 1994). Once told that they are listening to speech, participants are immediately able to organize the sounds as a single stream of phonetic information. Their knowledge of the speech signal enables them to integrate the three sine tones in order to hear a coherent speech stream, again demonstrating the influence of linguistic knowledge on perceptual organization.

Schematic grouping effects have also been observed in the sequential integration of speech. The verbal transformation effect (VTE) is a perceptual phenomenon induced in laboratory conditions that allows the study of sequential integration speech. In a typical experiment, a word is presented to a listener on a loop. Eventually, after several repetitions, the listener may report that their perception of the temporal order of the speech phonemes has changed. For instance, a listener hearing “spike” may eventually report hearing “bikes” instead. A listener may also hear a foreground and background emerge as a result of auditory streaming. In this case, a listener presented with “spike” may eventually hear “bike” emerge as the foreground, while the /s/ is heard as a background noise that is no longer sequentially integrated with the rest of the word. Shoaf and Pitt (2002) compared the VTE for words, pseudowords, and phonotactically-illegal pseudowords. Pseudowords yielded more kinds of transformations than words. Furthermore, participants often regrouped words into their veridical percept during a trial, whereas veridical regrouping did not occur as frequently in pseudowords. These results suggest that lexical information influenced the sequential

integration of speech phonemes during this task. Specifically, the lexical information appears to have contributed to what foreground would be selected from the speech stream. Much like the ability of participants to follow a familiar melody in Deutsch's (1972) experiment, the lexical schemas enabled the listener to sequentially integrate the repeating loop of speech sounds.

Another paradigm in which evidence of a lexical effect on sequential integration has been found is the migration paradigm. In this paradigm, speech materials are presented simultaneously to both ears. Migrations occur when linguistic information presented at one ear is perceived as contributing to the perception of information presented at the other ear. For instance, listeners presented with "pay" and "lay" at opposite ears will often hear them as the fused percept "play", and this effect can be influenced by semantic information in a sentence leading up to the dichotic target (Cutting, 1975). In such a case, the lexical schema for "play" is integrating information across the dichotically presented "pay" and "lay". Mattys and Samuel (1997) found migrations occur more frequently for pseudowords than for words, suggesting that the organization of pseudowords is less stable than for words. Also, migrations that create illusory words occur more frequently than those that create illusory pseudowords (Kolinsky and Morais, 1992, reviewed in Kolinsky & Morais, 1996). These effects of lexical status again suggest that lexical information influences the sequential integration of speech sounds.

Further evidence of contextual influences in sequential integration of speech is found in studies of phoneme restoration, in which lexical effects are observed (Samuel, 1981; Samuel, 1996). Phoneme restoration occurs when phonemes of a word are replaced with noise, and

despite that it is missing, listeners often report that the phoneme is intact. They have perceptually restored the phoneme from the noise. Bregman (1990) argues that the phenomenon of phoneme restoration depends upon a process of sequential integration. A compelling piece of evidence he cites is Warren & Obusek (1971). In this experiment, a participant would have heard a sentence such as, "The state governors met with their respective legislatures convening in the capital city." The /s/ in "legislature" would be replaced by (or combined with) a noise. Upon listening to this sentence, participants would report hearing the replaced phoneme, having restored it from the noise. However, they could not accurately report when the noise occurred within the sentence. Such an impairment to temporal judgments is indicative of the noise being segregated, since temporal judgments across streams are more difficult than within-stream judgments (Bregman, 1990). This decrement in judgments for when the noise occurred demonstrates the involvement of a process of sequential integration. Samuel (1981) found that such restorations are more common for words than pseudowords. If the phoneme replaced by noise can be predicted from contextual information, such as it would be in a word, it will be much easier for the auditory system to restore the phoneme from the noise. Thus, lexical schemas may act to help sequentially integrate components of the noise necessary for a participant to experience phoneme restoration.

As seen in the work reviewed above, most studies that speak to the issue of contextual influences in sequential integration of speech focus solely on a change in lexical status- whether an item is a word or a nonword. Schematic processing's effect on sequential integration could

be far more intricate, with influences from multiple types of linguistic knowledge. Indeed, there are many perceptual effects in spoken word recognition that demonstrate reliance on multiple levels of processing. For example, regressive place assimilation (e.g. “greem boat” being perceived as “green boat”) entails the use of phonological information, giving the phonemic contexts in which the effect is able to occur, as well as lexical information, which contributes to the disambiguation of unclear sounds (Dilley & Pitt, 2007). Part of what allows speech perception to be robust is that listeners have a highly detailed knowledge of speech at multiple levels of description, such as the phonological and lexical levels.

The lack of studies investigating more than the effects of lexical status on sequential integration illustrates how little work has been done in order to understand schematic grouping’s contribution to sequential integration. Few studies explicitly examine the relationship of schematic processes to sequential integration, and few studies distinguish between different types of high-level contextual information (e.g. lexical vs. phonological). The previous literature leaves many questions about the role of schematic processes in the sequential integration of speech unanswered. The current study can help to address two questions of primary importance when discussing schematic grouping’s contribution to sequential integration.

First, how does contextual information contribute to processes of perceptual segregation and integration? It is important to distinguish segregation and integration, and contextual influences on them, because they entail differing contributions from speech schema.

Because the default of the perceptual system is presumably to integrate sounds (Bregman, 1990; Darwin, 2008), influences that increase integration may suggest that contextual information acts as confirmatory evidence to integrate speech sounds. The perceptual system is “biased” toward integrating sounds, so if contextual information contributes to increased integration, the associated schemas could be seen as “confirmatory integrators”. Increased segregation, by contrast, could be seen as due to positive evidence to segregate sounds. Therefore, if contextual information contributes to increased segregation, the associated schemas may be seen as being a source for perceptual evidence to segregate, changing the perceptual grouping of speech sounds. Another important distinction that may be made has to do with the time-scale over which contextual influences may occur in integration or segregation. Whereas low-level cues to integration and segregation operate on a relatively short time-scale, schematic grouping may operate on a larger time-scale, such as in studies of phoneme restoration in which the restored phoneme is influenced by context presented later in the speech (Warren & Warren, 1970). For example, in a sentence such as “It was found that the *eel was on the axle”, a participant is likely to hear the missing phoneme (*) as /w/ due to the context occurring much later in the sentence. It may be the case that schematic influences on integration and segregation are differentially impacted by the time-scale over which these cues are present. There may, for example, be a larger time window in which lexical information contributes to integration, since this would allow for successful word recognition. Segregation, on the other hand, may be more influenced by perceptual evidence of grouping occurring over a

shorter time scale, such as phonological information. Whereas perceptual integration necessarily extends over long durations, segregation has no such constraint. By examining contextual influences on integration as well as segregation, the current study explores the mechanisms by which high-level speech representations operate within a schematic grouping framework.

Second, how are different types of linguistic information (lexical, phonological) able to influence sequential integration? This question is itself intricately tied with the previous question. However, the relationship of contextual influences with segregation and integration may not be straightforward. As noted above, there is reason to expect different types of linguistic information to contribute to sequential integration in differing ways. By virtue of the time-scale over which lexical and phonological information occur alone, it may be reasonable to expect that phonological and lexical information contribute differently to integration and segregation. This study, therefore, also examines how different types of linguistic information contribute to the sequential integration of speech.

In summary, exploring the role of contextual information in the sequential integration of speech may give insights that are crucial to understanding how humans connect sounds in order to form meaningful perceptual objects. More broadly, exploring this topic could potentially provide a meaningful way of connecting the high-level processes of spoken word recognition, like word segmentation and lexical competition, with lower-level auditory grouping mechanisms. If high-level contextual information is shown to play a role in the sequential

integration of speech sounds, it may have implications for discussing higher-level processes not only as abstractions from the signal, but also as altering the perceptual grouping of the signal.

Two experiments were designed in order to investigate the relationship of contextual information and sequential integration and the influence of contextual information as it relates specifically to perceptual segregation or integration.¹

¹ Two pilot experiments were run in order to determine appropriate manipulations for the current paradigm, both in terms of how perceptual segregation is elicited and in terms of what contextual information is altered. The first of these experiment utilized nonwords with embedded nonwords (“slaz”) and words with embedded words (“slam”), as well as a lateralization continuum similar to that of Experiment 1, but with a “6th” step that was fully dichotic. The second of these experiments used words with embedded nonwords (“slant”), nonwords with embedded words (“sland”), and illegal nonwords with embedded words (“sran”), also using a lateralization continuum similar to that of Experiment 1. The results of these experiments were used in order to refine conditions of contextual information and select appropriate maxima for the continuum of perceived spatial location presented in Experiment 1.

Chapter 2: Experiment 1

The purpose of Experiment 1 is to demonstrate the influence of different levels of linguistic information on sequential integration, and their contribution to integration and segregation. The current study uses a novel paradigm in order to examine schematic processing's contributions to the sequential integration of speech sounds. Sequential integration in speech is remarkably robust, making it necessary to use stimuli where segregation can be easily induced. /s/-initial words and nonwords, such as "slim" or "slib", are used where the /s/ is manipulated to segregate from the rest of the stimulus, or base. /s/ onsets are preferable for the current study because /s/ will likely segregate from other phonemes. This is because /s/ is a high-frequency hiss noise, which is spectrally dissimilar from the voiced phonemes that follow it in the base. This spectral dissimilarity is what contributes to a more easily segregated stimulus. Also, formant transitions in /s/ are limited, providing fewer bottom-up cues to sequentially integrate /s/ with the voiced phonemes of the base. In order to vary the strength of cues to segregation, the lateralization of the /s/ and the base are gradually changed on a continuum. At the first step of the continuum, all of the speech is lateralized at the center of the head. As each step increases, /s/ is lateralized to one side while the base is lateralized to the opposite side, providing a way of inducing segregation. A participant's task during the experiment is to indicate whether they hear an item as beginning with /s/ (an integrated percept), or as not beginning with /s/ (a segregated percept).

Condition	Abbreviation	Example / Embedding
Words with embedded words	WW	slam / s-lamb
Words with embedded nonwords	WN	slant / s-lant
Nonwords with embedded words	NW	sland / -s-land
Nonwords with embedded nonwords	NN	slaz / s-laz
Illegal nonwords with embedded words	INW	sran / s-ran

Table 1. Conditions of contextual information in Experiment 1

Five conditions are utilized in order to test integration and segregation effects (listed in Table 1). Four comparisons between conditions are made in order to demonstrate different effects of contextual information on the sequential integration of speech (listed in Table 2). First, the effect of lexical status on sequential integration is tested. The effect of lexical status is the most well-attested effect of contextual information on sequential integration from the previous literature. If effects of lexical status are not observed, it would show that the paradigm used in this study is ineffective for examining the sequential integration of speech. In order to test this effect, WN items (“slant”) are compared with NN items (“slaz”). As /s/ is lateralized further from the base, words should be more likely to remain integrated than nonwords. A difference in the observed segregation between WN and NN items would demonstrate this effect. The lexical status of the carrier, the full item with /s/ and the base, would be acting to

Comparison	Effect	Example
WN vs. NN	Lexical Integration	"slant" (s-lant) vs. "slaz" (s-laz)
NW vs. NN	Lexical Segregation	"sland" (s-land) vs. "slaz" (s-laz)
WW vs. WN	Lexical Competition	"slam" (s-lamb) vs. "slant" (s-lant)
NW vs. INW	Phonological Segregation	"sland" (s-land) vs. "sran" (s-ran)

Table 2. Specific comparisons made across contextual conditions in Experiment 1

integrate the /s/ with the base. Such an effect would show that lexical information can act as a “confirmatory integrator”, increasing integration within an acoustic context that otherwise has evidence to segregate.

In contrast, the ability for lexical information to promote segregation of the /s/ from the base is examined by comparing NN and NW items. In NW items (“sland”), the lexical information of the base may act as a cue that the /s/ should not be sequentially integrated with the base. In other words, the lexical information promotes the perception of the base as a foreground percept, while the /s/ becomes background. Similar effects were observed in Shoaf & Pitt (2002) when it was found that words are more often regrouped as their veridical percept than pseudowords, suggesting that lexical information helps to determine a foreground percept. Therefore, it is expected that as the /s/ is lateralized further from the base, NW items (“sland”)

will be more likely to segregate than NN items (“slaz”). If confirmed, this comparison would demonstrate that lexical information can contribute evidence to segregate speech sounds that do not belong with a lexical item. This effect suggests a more intricate role of lexical status in the sequential integration process than merely adding confirmatory evidence to group the constituent phonemes of a lexical entry, as is demonstrated by the first comparison. In addition to integrating the correct phonemes, word recognition requires excluding, perhaps segregating, phonemes that do not belong to a word.

The two previous comparisons examine an overall effect from the presence or absence of lexical information contained in either the carrier (/s/ and the base) or base of an item. A lexical candidate is activated, and increased integration of the lexical candidate’s constituent phonemes is observed. However, the activation of multiple lexical candidates, rather than a single candidate, may also contribute to how speech sounds are ultimately grouped. In WW items, a segregated percept can yield one of two valid lexical percepts. For example in “slam”, the carrier “slam” and base “lamb” are both words, and both are thus valid percepts in an acoustic context that is perceptually unstable with respect to sequential integration. As cues to segregate the /s/ from the base increase, these two words may compete to be recognized. Lexical competition occurs when multiple words are activated in response to a stimulus (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). In order to determine which word is heard, these words “compete” until one word is chosen as the recognized percept. Competition is influenced by characteristics of the individual words that are competing, such as word frequency and

neighborhood size (Luce & Pisoni, 1998). WW items are compared with WN items in order to test effects of lexical competition. In WN items (“slant”), /s/ should be more likely to integrate with the base because lexical information is acting to integrate the speech into a familiar percept. In WW items (“slam”), as /s/ is manipulated to segregate from the base (“lamb”), the carrier and base may compete to be the recognized percept. As /s/ is manipulated to segregate from the base, the lexical information of the base may make WW items more likely to segregate than WN items, which would indicate that the lexical status of the carrier and base in WW items influences which grouping of the speech is ultimately perceived. Like the comparison of NN and NW items, this comparison examines the capacity for contextual information to promote segregation. However, in this instance the increased segregation would be explicitly due to the high-level process of lexical competition. Such a result would therefore demonstrate that seemingly abstract processes of spoken word recognition are in fact tightly bound with the perceptual grouping processes from which they arise.

The final comparison adds to the current experiment by introducing contextual influences from the phonological level. Phonological information’s contribution to sequential integration is examined by comparing legal and illegal nonwords with embedded words, NW (“sland”) and INW (“sran”) items. In order to integrate INW items, /s/ has to be integrated with phonemes that are rarely heard in conjunction with /s/, in this experiment /r/ and /v/. /sr/ and /sv/ are phonotactically-illegal word onsets. Like the inability of English listeners to treat click consonants as speech (Best & Avery, 1999), phonotactics may play a large role in what

combinations of sounds can be quickly integrated into a speech stream. Combinations of sounds that are illegal may in that sense produce increased segregation relative to combinations of sounds that are legal in a listener's language. If this relationship between phonotactics and sequential integration is true, participants should be less likely to integrate /s/ with the base in INW items, such as "sran", than in legal NW items like "sland".

A number of predictions are drawn from the stimulus conditions in this experiment, each showing a specific way in which contextual information may influence the sequential integration of speech. By observing an effect of contextual information on sequential integration, the current experiment may offer valuable insights into processes underlying spoken word recognition which have not received much attention in the literature. Although intuitively the ability of listeners to group speech sounds over time should play an important role in speech perception, there are many open questions in how higher-level process could relate to perceptual grouping. The two experiments presented in the current study address certain vital questions in order to begin uncovering the role of high-level contextual information in sequential integration. Particularly, the current research speaks to the extent to which contextual information can be expected to influence grouping and different processes of integration and segregation that may be entailed in the influence of context on sequential integration. If the preceding predictions are confirmed, it would suggest that schematic processes play a more intricate role in auditory grouping than previously thought, provide specific examples of contributions of contextual information to sequential integration, and serve

as a necessary starting point for exploring top-down mechanisms in the sequential integration of speech.

2.1 Method

2.1.1 Participants

Twenty-two undergraduates from The Ohio State University participated in this experiment, receiving credit toward a requirement for a class. Listeners were native speakers of American English with self-reported normal hearing in both ears.

2.1.2 Stimuli

Items were /s/-initial words and nonwords that either had a word or nonword base following the /s/. Items were constructed to be in one of five conditions, summarized in Table 1: (1) word with an embedded word (WW), (2) word with an embedded nonword (WN), (3) nonword with an embedded word (NW), (4) nonword with an embedded nonword (NN), or (5) phonotactically illegal nonword with an embedded word (INW). In phonotactically legal items, the initial /s/ was followed either by /l/ or /w/, while the initial /s/ was followed by /r/ or /v/ in the INW condition to produce phonotactically illegal onset consonant clusters. Items were constructed in sets across conditions to minimize phonetic differences between conditions. An example of one such set is “slam”, “slant”, “sland”, “slaz” and “sran”. Items were selected so that word frequency differences between the WW and WN conditions were minimized as much as possible, with means of 12.625 and 4.5 respectively (CMU Pronouncing Dictionary; CMU

Speech Group, 2007), and furthermore so that lexical items in any condition, whether in the base or the carrier, were within a similar range of word frequency.

In order to induce segregation, the current experiment used cues to spatial localization, specifically the interaural time difference (ITD) and interaural level difference (ILD). As mentioned in the introduction, sequential integration enables a listener to track a sound source across time. Localization cues can be powerful in helping a listener to follow a sound source over time. The dominant cue to localization in the horizontal azimuth is the interaural time difference (Darwin, 2008). It is the slight delay across ears resulting from the waves emanating from a sound source arriving at one ear earlier than the other. For example, if the attended sound source is at a person's left side, the sound waves will arrive at the person's left ear a short interval (say, 0.6 milliseconds) before it arrives at the right ear. Calculating this difference as the signal from the two ears is integrated allows a listener to infer where a source is, at least in terms of how far to the left or right the source is. Above 1600 Hz, however, the dominant cue to localization is the interaural level difference, or the difference in intensity of a sound across ears (Darwin, 2008). This is due in part to ITD's becoming unreliable at higher frequencies and intensity differences becoming more easily detectable at higher frequencies. The current study used a combination of ITD and ILD in order to induce perceptual segregation of /s/ from the base.

Stimuli were recorded by a male speaker of American English. Each stimulus was processed using Praat speech analysis software (Boersma, 2001) in order to produce a 5-step

continuum of perceived spatial location, or how lateralized a stimulus was perceived to be. Stimuli were initially normalized to 70 dB SPL. At step 1, the initial /s/ and the base were all presented at the center of the head. Each step introduced an increase of 0.6 milliseconds of ITD and 3 dB of ILD. As steps increased, /s/ increased in lateralization to the left or right and the base became lateralized in the opposite direction. At step 5, the ITD was 2.4 milliseconds (beyond the physiological range of approximately 0.7 milliseconds) and the ILD was 12 dB, maxima that were selected in order to achieve strong effects of perceived spatial location. Each item was presented with the /s/ lateralized to the left and the base to the right as well as with the /s/ lateralized to the right and the base to the left. This yielded 10 presentations of each item during the test session. With the exception of items having a /w/ following the initial /s/, all stimuli had the same /s/ token spliced in place of their original /s/ to reduce acoustic variability between items. Items having a /w/ following /s/ used the same /s/ token.

2.1.3 Procedure

Participants were tested up to four at a time in sound-attenuated booths. Participants were presented stimuli that were played via a computer soundcard with 16-bit resolution at a sampling rate of 48 kHz. Stimuli were presented binaurally through Sony MDR-V900HD headphones. Participants were asked to respond to whether the item in each trial began with /s/ or not using a button board. Participants were given a practice session in order to become familiar with the task as well as a brief auditory introduction to the spatial continuum, listening to a nonword not used during the test session played from step 1 to step 5 of the continuum, in

order to ensure that they were able to perceive the difference in localization. The response of “/s/-initial” or “not /s/-initial” is assumed to indicate differences in segregation or integration of the initial /s/ with the. Five continuum steps, two lateralizations, five conditions of contextual information, and eight items per condition yielded 400 trials for participants to complete, with a break in the middle of the test session, which took approximately 40 minutes. Trials in the test session were blocked such that within each block of 40 trials, the ordering of items was random and each continuum step in each lateralization was played only once. Across ten blocks, items were rotated so that each item was played in each continuum step and lateralization.

2.2 Results

Displayed in Figure 2 as a function of the spatial continuum on the x-axis are the mean proportion of “not /s/-initial” responses on the y-axis, with each line in the display representing one of the five contextual conditions from Table 1. A few trends emerge, one of which is that segregation of /s/ from the base increased along the spatial continuum across conditions, showing larger increases from step 1 to step 3, and smaller increases from step 3 to step 5. If the results were ideal, segregation would not occur at step 1 because within step 1, the entire stimulus was presented at center. However, it is clear from Figure 2 that participants experienced the /s/ as segregating from the base even at step 1. Also, segregation never reached 100% in this stimulus set, even in conditions that should be the most disposed toward segregation. In addition to differences across the continuum, the various conditions of contextual information influenced how often the /s/ integrated with the base. The ordering of

Figure 2. Results of Experiment 1, showing mean proportions of responding “Not /s/-initial” by spatial continuum step and condition of contextual information

conditions in terms of segregation patterned in accordance with the contextual cues to segregate and integrate, with the exception of WW and WN conditions, which were similar across the continuum.

A logistic mixed effects model (using the lmer package in R Statistics) was fit to test the overall effects of spatial localization, lateralization (base presented to the left or right), and the effect of each condition of contextual information. Second, contrast tests were run within the

Fixed effects:	Estimate	Standard Error	z value	p value
(Intercept)	-1.94984	0.77541	-2.515	0.011917
NN	-1.92179	0.37379	-5.141	2.73E-07
NW	-1.3225	0.3245	-4.075	4.59E-05
WN	-3.04042	0.44595	-6.818	9.24E-12
WW	-3.05667	0.43271	-7.064	1.62E-12
Step	2.16469	0.51987	4.164	3.13E-05
Step ²	-0.24924	0.06534	-3.814	0.000136
Lateralization	-0.22544	0.05945	-3.792	0.00015

Table 3. Fixed effects tests for the logistic mixed effects model in Experiment 1

same model to test the four hypotheses introduced earlier. The model had fixed effects for contextual condition, lateralization, and spatial continuum, which was modeled as a quadratic function. The model had random effects for contextual condition and spatial continuum within participant and a random coefficient for items. A chi squared goodness of fit test compared the final model with a simpler model lacking a term for lateralization of the base relative to the /s/, $\chi^2(1) = 14.203$, $p < 0.001$, showing that the final model, presented here, fit the data best.

Results of fixed effects tests from the final model are displayed in Table 3. The model demonstrates a few important things. First, the spatial continuum successfully elicited

increasing segregation of /s/ from the base, indicating that the paradigm is successful in eliciting varying amounts of integrated and segregated percepts. Second, the different conditions of contextual information predicted how likely a given stimulus was to segregate. Finally, segregation was more likely to occur when the base was presented to the right ear, demonstrating a modest “right-ear advantage” for segregating the base from the /s/.

Four contrast tests were used within the model in order to make the comparisons listed in Table 2. First, WN and NN items were compared (“slant” vs. “slaz”) in order to determine whether lexical information helped to integrate /s/ with the base. A consistent difference between the two conditions is readily apparent in Figure 2. The contrast test returned significance, $\beta = 1.1191$, $p < 0.001$, indicating that WN items were significantly more likely to integrate than NN items. This comparison demonstrates that the lexical information available to listeners allowed them to integrate speech information in the presence of cues that might otherwise cause /s/ and the base to segregate.

The second test compared NN and NW items (“slaz” vs. “sland”) in order to assess whether lexical information in the base of a stimulus can act as a perceptual foreground, increasing segregation of /s/ and the base. In Figure 2, one can note that the mean of NW items is always higher relative to NN items. The contrast test returned significance, $\beta = 0.5992$, $p < 0.05$, indicating that /s/ was more likely to segregate from the base in NW items. This result confirms that lexical information can promote segregation of sequential speech sounds in addition to integration. It is interesting to note, however, that the increased segregation due to

lexical information does not appear to be as powerful an effect as the previous effect for increased integration due to lexical information.

The third test compared WN and WW items (“slant” vs. “slam”). In WW items, it was predicted that the lexical information in the base would compete with the lexical information of the carrier when perception was unstable, resulting in increased segregation of WW items relative to WN items. From Figure 2, it is apparent that these two conditions were experienced as highly similar to participants. The contrast test failed to return significance, $\beta = -0.01625$, $p = 0.96$, and in fact showed that the two conditions were remarkably similar. The reason for this non-significant comparison is unclear. It appears likely that the current paradigm’s use of segregating a word-onset may be the reason. Lexical processing is strongly influenced by initial phoneme activation (Marslen-Wilson & Zwitserlood, 1989), which may have inhibited lexical information in the base from impacting word recognition, even in the presence of localization cues to segregate the /s/ from the base.

The final contrast test compared NW and INW items (“sland” vs. “sran”) in order to show whether phonological information, specifically phonotactics, could influence the sequential integration of /s/ with the base. In Figure 2, this comparison shows the largest difference, a mean difference of 22.3%, with INW items quickly separating from NW items as each step increases. Unsurprisingly, then, the contrast test returned significance, $\beta = 1.3226$, $p < 0.001$, indicating that the phonotactically-legal NW items were more likely to integrate than INW items. This result demonstrates that phonological information is used in order to help

determine sequential integration of speech sounds. In the absence of certain phonological cues to integrate sequential speech sounds, as in the illegal onset consonant cluster of the INW items (“sran”), segregation of /s/ from the base becomes more likely.

2.3 Discussion

The results of Experiment 1 demonstrate a number of things. The data suggest that the current paradigm is an effective means of testing sequential integration differences due to contextual influences. Localization cues influence the likelihood of integrating the /s/ with the base, showing that they are an effective way of inducing bottom-up segregation of sequential speech sounds. Also, the different conditions of contextual information predict different likelihoods that a given item will segregate. Contrasts between contextual conditions tested specific hypotheses about contextual influences on grouping. It was found that lexical information of the carrier can increase integration of /s/ with the base, lexical information in the base can increase segregation of /s/ from the base, and illegal phonotactics give evidence to segregate /s/ from the base. Finally, the comparison of WW and WN items was not significant, suggesting that there was no effect of lexical competition from the base in WW items in the current experiment.

Overall, the results of this experiment demonstrate that high-level contextual information can influence the sequential integration of speech. At the lexical and phonological levels, cues to integrate or segregate speech sounds resulted in changes in the sequential

integration of speech. These results are non-trivial when discussing the interface of perceptual grouping and spoken word recognition. Though spoken word recognition is understood to depend on a lower-level process of perceptual organization, it is not often studied as it relates to the mechanisms of perceptual grouping. The current experiment demonstrates that it may be useful to understand certain processes at higher levels such as the lexical and phonological levels in terms of perceptual integration and segregation. For example, the large effect of phonotactics on the sequential integration of /s/ with the base observed in this experiment is a clear example of stored knowledge influencing the perceptual grouping of sounds. The listeners' experience with the English language resulted in differing sequential integration for familiar (e.g. /sl/) and unfamiliar (e.g. /sr/) sequences of sounds, suggesting that phonotactic information may function perceptually in a way that corresponds with Bregman's (1990) conception of schematic grouping.

The current experiment's results may also provide a way of thinking about role of the contextual information in terms of contribution to differing grouping processes. For example, the effect of integrating the /s/ with the base and the effect of segregating the /s/ from the base entail different influences of lexical status on sequential integration. In the integrative effect, lexical information groups information across sounds that have bottom-up cues to segregate. In the segregation effect, lexical information selects the base as a foreground, preventing its integration with the preceding /s/. It could be the case that lexical information is differentially able to influence segregation and integration. For example from Figure 2, it appears as though

integration due to lexical information (WN vs. NN) is somewhat stronger than the additive effect of segregation due to lexical information (NW vs. NN). This possibility is discussed in greater depth in the General Discussion.

One assumption that was made in carrying out this experiment is that a participant's response of "/s/-initial" or "not /s/-initial" to a given stimulus reflected a sequential integration process. However, such a response is an indirect measure of grouping. The inclusion or exclusion of /s/ from a participant's reported percept across conditions of contextual information could change due to other factors, such as response strategies. Experiment 2 was designed in order to address this issue of interpretation.

Chapter 3: Experiment 2

The results of Experiment 1 are promising in that different conditions of contextual information appear to elicit changes in the sequential integration of speech. However, interpreting the results of Experiment 1 in terms of sequential integration rests critically on the assumption that the response participants made of whether each stimulus began with an /s/ implied a change in sequential integration. For example, responding that a stimulus does not begin with /s/ reflects a segregated percept. Experiment 2 is intended to demonstrate that the responses made in Experiment 1 reflect perceived changes in grouping, and not merely a bias resulting from the changes in contextual information.

In order to validate the interpretation of results from Experiment 1, the response of whether subjects heard a stimulus as /s/-initial should be used in conjunction with a task that is known to measure perceptual grouping. An important property of stream formation that allows segregation to be inferred is the ability of the listener to track temporal information within and across perceptual streams (Bregman, 1990). Within a stream, temporal information such as the relative timing of events is easier to detect than across perceptual streams. A number of tasks involving temporal judgments have been developed in which either segregation or integration promote better performance (Micheyl & Oxenham, 2010), one of which is gap detection. In gap detection, a participant tries to identify whether a silence is present within a stimulus, responding either that there is or is not a gap in the stimulus, or indicating whether the size of the gap changes. Because temporal judgments are impaired by stream segregation, sensitivity

to the gap decreases in segregated relative to integrated percepts (Woods, Sorkin, & Boggs, 1979). In addition to the benefit of using gap detection in order to gauge perceptual streaming, the task also allows discrimination to be measured apart from response bias.

In order to show that responses from Experiment 1 reflect changes in sequential integration, each trial in the experiment used a gap detection response followed by the response to whether the stimulus was /s/-initial. For example, a participant might hear a stimulus and respond that they heard a gap and that the stimulus was “not /s/-initial”. By having participants make both responses to each stimulus, the differences in gap sensitivity aggregated across the response used in Experiment 1 reflect a trial-by-trial measure of subjective changes to sequential integration. Two predictions were drawn from the design. First, stimuli categorized as “/s/-initial” will show higher gap sensitivity than stimuli categorized as “not /s/-initial”, indicating a change in sequential integration. Second, NN items will show lower gap sensitivity relative to WW items, due to the lack of lexical information integrating /s/ with the base. Confirming these predictions would validate that the effects observed in Experiment 1 are indicative of contextual influences in the sequential integration of speech.

3.1 Method

3.1.1 Participants

Thirty undergraduates from The Ohio State University participated in this experiment, receiving credit toward a requirement for a class. Thirteen participants failed to meet the

criterion for gap detection performance (described in Procedure), did not complete the experiment, and were excluded from analyses. Listeners were native speakers of American English with self-reported normal hearing in both ears.

3.1.2 Stimuli

Two conditions of contextual information were taken from Experiment 1: Ten words with word bases (WW; e.g. “slab”) and ten nonwords with nonword bases (NN; e.g. “slav”). In each condition, four items were drawn directly from Experiment 1. WW and NN items were constructed in pairs that differed only by the coda of each item. All items had an /sl/ onset, which was used because the closure of /s/ before /l/ is less pronounced, making a gap within the onset consonant cluster easy to detect. All items used the same token of /s/spliced in place of each original /s/ onset to reduce acoustic variability.

Stimuli were recorded by a male speaker of American English. Each stimulus was processed using Praat speech analysis software (Boersma, 2001). All stimuli were presented at step 5 of the perceived spatial location continuum from Experiment 1, meaning each stimulus had an ITD of 2.4 ms and an ILD of 12 dB. The decision to present all stimuli at step 5 was based on a few factors. First, to get a reliable gap sensitivity measure, temporal information around the gap needed to be held as constant as possible. Second, the localization cues at step 5, based on the data seen in Figure 2, provide the best chance of inducing segregated and integrated perception. Relatedly, for signal detection analysis to be carried out, a sufficient number of

trials is needed in “/s/-initial” and “not /s/-initial” response categories, which step 5 of the spatial appeared most likely to yield. Only one lateralization was used in which the /s/ was lateralized to the left and the base was lateralized to the right. Gap trials were constructed by inserting a silence after the /s/ and preceding the following /l/. Gap duration varies in the calibration phase of the experiment according to participant performance, so stimuli were created with silences following the /s/ onset ranging from 2.5 ms to 87.5 ms. 5 ms cosine ramps were used around the silence in order to ensure no detectable artifacts were created by inserting a gap into the stimulus.

3.1.3 Procedure

Participants were tested individually in sound-attenuated booths. Participants were presented stimuli played via a computer soundcard with 16-bit resolution at a sampling rate of 48 kHz. Stimuli were presented binaurally through Sony MDR-V900HD headphones. There were two phases in this experiment, a calibration phase and a test phase.

The purpose of the calibration phase was to estimate a gap duration that a participant identified accurately in 70% of trials. Participants’ task was to respond “gap” or “no gap” in regard to whether they heard a silence following the /s/ in each stimulus. For each response, participants received visual feedback on a computer monitor about their accuracy. Participants were given a brief practice session to ensure understanding of the task. The test session began with the gap duration set at 50 ms. As participants responded, the gap duration adjusted based

on a rolling average over twenty trials of participants' accuracy until a gap duration was reached at which they responded accurately in 70% of trials. After this value for gap duration stabilized within a range of 5 ms for ten trials, the test procedure terminated, or the procedure would terminate after a maximum of 400 trials. Presentation of stimuli was randomized, with the ten items in each condition being presented five times with a gap and five times without a gap, yielding 200 trials. If participants exceeded 200 trials, the ordering of the first 200 trials was repeated, helping to ensure an even distribution of item presentations throughout the calibration phase. In order to continue to the test phase, the estimated gap duration of the calibration phase had to be no more than 50 ms. This cut-off was selected a priori based on gap detection thresholds reported in dichotic presentation in the literature (Moore, Shailer, & Black, 1992; Roberts & Lister, 2004).

During the test phase of the experiment, gap duration was fixed at the duration estimated from the procedure in the calibration phase. Visual feedback for participants' gap detection response was no longer provided. In addition to the first response, participants made a secondary response indicating whether each item was "/s/-initial" or "not /s/-initial", the same response that participants made in Experiment 1. A brief practice session was given to ensure understanding of the task. The same items from the calibration phase were used in the test phase. Each of the ten items in the WW and NN conditions was presented five times without a gap and five times with a gap, yielding 200 trials. The ordering of stimuli was randomized for each participant. The entire experiment took approximately 40 minutes to complete.

3.2 Results

The gap durations estimated in the calibration phase varied across participants, having a mean of 27.02 ms and a standard deviation of 14.67 ms. d' values for the test phase were aggregated by participant, having a mean of 0.60 with a standard deviation of 0.56, which indicates a wide range of performance. The overall low d' indicates that most participants had difficulty with the task in the test phase.

The primary result of interest was whether changes in the responses in Experiment 1 reflected a change in perceptual streaming. In order to evaluate this, gap detection responses in the test phase were aggregated by responses of whether or not the stimulus began with /s/. An observed difference in gap detection between “/s/-initial” and “not /s/-initial” responses would suggest a change in sequential integration across these responses, specifically that “not /s/-initial” responses indicate segregation of the /s/ from the base. d' values were obtained for each participant in each response category. The mean participant d' for “/s/-initial” responses was 0.62, compared with 0.36 for “not /s/-initial” responses. A within-participants t test performed on the participant d' values showed a statistically significant difference, $t(16) = 3.6579$, $p < 0.01$. Changes in response to whether a stimulus was “/s/-initial” or not corresponded to changes in gap sensitivity, from which a change in sequential integration is inferred. A t test for β values was also performed, $t(16) = -3.5707$, $p < 0.01$, showing that subjects were significantly more likely to report hearing a gap when they heard a stimulus as “not /s/-initial”. This change in response bias may be due to the inserted silence in gap trials

increasing the likelihood of hearing the /s/ as streamed, since the temporal separation of /s/ from the base likely increases segregation.

A secondary result of interest is changes in gap sensitivity resulting from changes in lexical information, the comparison of NN and WW items. A significant comparison of these conditions would lend further credence to the conclusions drawn from Experiment 1 by directly demonstrating that changes in contextual information, in this case lexical information, result in differences in sequential integration (roughly comparable to the conclusion drawn from the contrast test conducted on WN and NN items in Experiment 1). Mean participant d' values were 0.50 and 0.66 for NN and WW conditions. A within-participants t test on these d' values approached significance, $t(16) = -1.7489$, $p = 0.10$. The non-significant difference in gap sensitivity across NN and WW conditions could be interpreted as showing that lexical status has a weak, if any, effect on gap sensitivity, however some issues with interpreting this result are raised in the discussion section. A t test on β values, $t(16) = 2.8687$, $p = 0.01$, shows that participants were more likely to say there was a gap in NN rather than WW stimuli. The change in response bias could be partially indicative of streaming, with the less coherent nonwords producing more perceived separation of /s/ from the base, and thus an inferred gap, regardless of whether a gap was present.

3.3 Discussion

The primary focus of Experiment 2 was to assess whether the responses given in Experiment 1 reflected sequential integration or not. Listeners' gap detection can be used to assess changes in sequential integration, since gap sensitivity suffers in segregated percepts. By aggregating gap detection responses by whether a stimulus was perceived as /s/-initial, the current results demonstrate that changes in the response to whether an item begins with /s/ reflect changes in sequential integration. In other words, Experiment 2 lends credibility to the argument that the influence of contextual information observed in Experiment 1 reflects mechanisms of perceptual grouping.

The non-significance of the comparison of gap sensitivity between NN and WW items is an interesting result. Assuming that the observed relationships of gap sensitivity to the secondary response in Experiment 2 and contextual information to this same response in Experiment 1 are both valid, one would expect to see gap sensitivity be significantly better in the WW condition compared with NN condition. Given the difference between these conditions in Figure 2 (averaging 16% at step 5), however, it is clear that this comparison did not provide the strongest possible test of changes in sequential integration due to contextual information. Rather, it provided a conservative test of differences in gap sensitivity and was the logical comparison to make, given that a majority of studies of top-down processes in perceptual grouping involved a change in lexical status as a cue to integrate.

As touched on briefly in the results section, participants' gap detection performance was highly variable. The estimates of gap duration obtained in the calibration phase varied greatly,

ranging from 2.5 - 50 ms, as well as participants' overall d' values in the test phase, ranging from -0.18 to 2.17. Participants' sensitivity across phases likely suffered due to feedback not being provided in the test phase. In addition, while most participants performed as well or worse in the second half of trials in the test phase, six participants showed an increase in d' averaging 0.33, perhaps suggesting a learning effect for these participants that the calibration phase did not eliminate. The other eleven participants showed an average decrease in d' of 0.14.

Participants were highly variable in how they performed in response to Experiment 2, but it is clear that Experiment 2 was effective in establishing the relationship between the response used in Experiment 1 and sequential integration. Future improvements on this experiment could focus on reducing inter-participant variability, perhaps by increasing the length of the calibration phase to eliminate learning effects and keeping visual feedback in the test session. Additionally, when aggregating by /s/-initial response and lexical status, there were too few responses per cell to explore whether these variables interacted. The interaction of lexical status and "/s/-initial response" could offer further insights into how contextual information influences sequential integration. Thus, later experiments should address this problem with an extended test phase.

Chapter 4: General Discussion

The two experiments in this study provide a novel, efficient way of examining contextual influences in the sequential integration of speech. The results of these experiments provide evidence that schematic processing influences the perceived grouping of speech sounds. Experiment 1 demonstrates that stored knowledge at the lexical and phonological levels can be used to influence the sequential integration of speech. Different conditions manipulating contextual information elicited different response patterns in terms of perceived grouping of /s/ with the following base (Figure 2). Comparisons between these conditions provide evidence that schematic processing can influence sequential integration of speech in multiple ways. First, the comparison of WN and NN conditions shows that lexical information can be used to promote integration in cases that are likely to segregate. This result echoes effects of lexical status on grouping that have been reported in several methodological paradigms, as reviewed in the Introduction. Second, the comparison of NN and NW items suggests that, in addition to promoting integration, lexical information can be used in order to select out a foreground from a speech stream, effectively promoting segregation. Finally, the comparison of NW and INW items revealed that phonotactically-illegal sequences are more likely to segregate relative to legal phoneme sequences. Phonotactically-illegal phoneme sequences lack the cues to integrate compared with phoneme sequences that are predictable from the listeners' experience with language. This lack of predictability, perhaps because it operates on a shorter time scale than the lexical effects, has a powerful influence on the sequential integration of speech sounds.

Overall, Experiment 1 was successful in uncovering changes in sequential integration of speech due to differences in contextual information of the stimulus.

Experiment 2 strengthened the interpretations of Experiment 1 by using gap detection, which is a more direct method of assessing changes in sequential integration. Aggregating gap detection responses by whether a stimulus was perceived as “/s/-initial”, the measure used in Experiment 1, Experiment 2 revealed that participants were less sensitive to gaps in stimuli reported as “not /s/-initial”. This reduction in sensitivity to a short silence is indicative of perceptual streaming because analyzing temporal information becomes more difficult across perceptual streams. As a result, Experiment 2 vindicated the interpretation of Experiment 1’s results as reflecting changes in the sequential integration of speech sounds.

The results drawn from this paradigm have important implications for research in spoken word recognition. The processing of contextual information in speech is often examined as though it is solely an abstraction from the auditory percept, not interacting with perception of lower-level grouping of the speech stimulus. For example, the TRACE model of spoken word recognition (McClelland & Elmer, 1986) uses as its input a series of feature-level specifications (e.g. voicing) used to identify different phonemes. The perceptual grouping of these features at one moment in time and subsequent integration of sounds into a single stream across time is assumed in how the model is implemented. This architectural decision is, of course, justified, since the perceptual grouping of speech sounds tends to be remarkably robust (Remez, 2005). However, studying spoken word recognition in this manner, with perceptual grouping being

assumed to be resolved, may undermine the importance of high-level speech representations in the perceptual grouping process. As the current study demonstrates, contextual information in the stimulus can influence the sequential integration of speech. This result suggests that the perceived grouping of speech is not somehow fixed once the linguistic information in the signal is processed. Indeed, the results of the many different paradigms reviewed in the Introduction- duplex perception, sinewave speech, VTE, migration- indicate just the opposite. Top-down processes, likely acting in accordance with Bregman's (1990) description of schematic grouping, do influence the perceptual grouping of speech.

One problem with the result, which motivated the current study, is that there is a limited amount of research that has been conducted in order to investigate how processes of spoken word recognition can influence the low-level grouping of speech. Seemingly the largest body of literature that speaks to this problem focuses on issues of domain specificity- whether the processing of speech is subject to the same low-level processes as general auditory sounds (Remez et al 1994; Nygaard, 1993). Interpretations of results in this literature arguing for "speech-specificity" in perceptual organization would conflict with the ideas being put forth by the current research, specifically that top-down processing should occur in accordance with something similar to Bregman's (1990) schematic grouping process. What is clear, then, given the disputes in the literature and the lack of available research investigating this specific problem, is that the results of the current study can motivate future research to uncover contextual influences in perceptual grouping in finer detail.

One way in which the current study contributes to how the sequential integration of speech is understood is by explicitly investigating how contextual influences can contribute to both segregation and integration. The distinction between schematic grouping's effects on these processes may be of paramount importance to understanding how speech schemas are used by the listener to group sounds. It is claimed that the default for the auditory perceptual system is to integrate sounds (Bregman, 1990; Darwin, 2008). Contributions of contextual information to integration therefore act as confirmatory evidence to integrate while contributions to segregation act as positive evidence to segregate. Comparing word carriers (WN, WW) to nonword carriers (NN, NW, INW) in Experiment 1, displayed in Figure 2, shows that lexical schemas can help integrate sounds across time. This result should not be surprising, as schematic grouping is described as a process of organizing evidence from a sound mixture. Perhaps a more striking result is that lexical information can also increase the segregation of sequential speech sounds, as in the comparison of NN and NW conditions in Experiment 1. Constituent phonemes of a lexical schema are perceptually bound together, while the surrounding speech information, in this case a preceding /s/, is excluded from contributing to the ultimately realized percept. Thus lexical information can contribute to integration (of its constituent phonemes) and segregation (of non-constituent phonemes). The results reported in Experiment 1 therefore support treating words as higher-order schemas, not only in the sense that they represent perceptual hypotheses about the identity of the incoming word, but also in the sense that these schemas ultimately contribute to the weighing of evidence to integrate or

segregate sounds. This interpretation of the results of Experiment 1 is bolstered by the d' values presented in Experiment 2 that were aggregated across the “/s/-initial” response. Gap sensitivity decreased in “streamed” trials, supporting the idea that the influences of lexical information observed in Experiment 1 are due to its role in sequential integration. The observed influence of lexical information on sequential integration shows how schemas operate as “confirmatory integrators” in some contexts and act as a source of positive evidence for segregation in others.

It is also informative to consider not only whether contextual information can contribute to both integration and segregation, but also to what extent contextual information is able to contribute to either process. Bregman (1990) argued that schematic grouping could be distinguished from pre-attentive, primitive scene analysis in part by the asymmetry of segregation and integration. In other words, the reliance on attention for schematic grouping results in differing contributions to one’s ability to integrate or segregate sounds. This contrasts with primitive scene analysis, in which whether one goes from a segregated to integrated percept or vice versa, the change from one percept to another will occur in the same place, for example at a certain frequency distance between alternating tones. Therefore, the contribution of a certain type of contextual information to either integration or segregation could be an important characteristic within the schematic grouping process. As an example, in Figure 2 one can see that the effect of lexical information on integration (WN vs. NN) is somewhat greater than the effect of lexical information on segregation (NN vs. NW). The differential effects of

lexical information on segregation and integration are perhaps indicative of what would be predicted by conceptualizing lexical information as contributing to an attention-driven, schematic grouping process. Different types of contextual information may have different asymmetries in their contribution to integration and segregation. For example, phonotactics appear to more strongly influence segregation relative to lexical information, as demonstrated by the comparison of NW and INW conditions in Figure 2 (Mean difference of 22.3%) relative to the comparison of NN and NW conditions (Mean difference of 10.4%). Investigating the various segregation/integration asymmetries of different types of contextual information could lend a great deal of insight into how the perceptual system appeals to linguistic knowledge in the perceptual organization of speech.

The current study provides a novel paradigm for studying schematic processing in the sequential integration of speech. There are several advantages to this experimental paradigm. First, it is an efficient way of comparing the influences of different types of contextual information, as demonstrated in Experiment 1. Several lexical and phonological effects were tested within a relatively short session, meaning that the paradigm is not constrained by the need for thousands of trials or skilled listeners. Second, cues that promote segregation, specifically lateralization cues, are directly manipulated. Several paradigms for studying sequential integration in speech, such as VTE and migration paradigms, rely on the emergence of perceptual phenomena, a new foreground or a migration of a sound across ears. These methods make perceptual grouping difficult to quantify. Though the responses in Experiment 1

never reach clear integrated and segregated endpoints, the manipulation of cues to segregate produces predictable proportions of integrated and segregated responses, making the effect of contextual information on sequential integration more easily quantified.

There are certain limitations to the current research. In order to induce segregation, localization cues are used. The use of these cues is the likely cause of why participants' never unambiguously heard stimuli as being segregated. In other words, it is why none of the conditions presented in Figure 2 reached 100% segregation. Localization cues are not as powerful of cues to auditory grouping as more typical cues such as frequency separation (Darwin, 2008). Therefore, the current paradigm is limited in how strongly primitive segregation (segregation that is independent of attention) can be elicited. It is further pertinent to note that Experiment 1 yielded no significant interactions of contextual information and localization cues to segregate. This result would imply that contextual information and localization cues to segregation had an additive effect, which can be contrasted with the interactive effects of context information and cues to segregate, such as lexical status and onset asynchrony in Experiment 3 of Nygaard (1993). Perhaps this apparent difference in how cues interact is due to the current research's focus on sequential as opposed to simultaneous integration. In other words, because the effects in the current paradigm occur over time, rather than in a single moment, sequential integration may reflect a process of weighing evidence to integrate or segregate sounds, as opposed to the presumably attention-independent interactive effects

observed in simultaneous integration. This process would, perhaps, produce the pattern of additive effects of context and localization cues observed in Experiment 1.

The current study serves as a starting point for relating mechanisms of sequential integration with higher-level representations in spoken word recognition. The success of the current study in eliciting changes in sequential integration through the manipulation of high-level contextual information demonstrates that further work may yield interesting and theoretically important results. Future experiments based on this work will focus on exploring the relationship of contextual influences at specific levels (lexical, phonological) on integration and segregation mechanisms. The current study provides an overview of what effects one might expect to find. Experiments designed to focus on specific levels of linguistic information and their influences on specific processes, segregation or integration, would allow for the influence of contextual information on sequential integration to be better understood. One such study could focus on changes in lexical information and its contributions to segregation and integration. Finer steps of the spatial continuum and more items could be used to produce functions that more precisely reflect the contribution of lexical information to either integration or segregation, and how these influences change as bottom-up cues to segregate increase. A number of experiments like this applied to different types of linguistic information could elucidate how, overall, the perceptual system incorporates schematic influences when weighing evidence to integrate or segregate. Furthermore, there are specific processes at different levels of linguistic information, such as word segmentation at the lexical level, where it may be

possible to test their relationship to sequential integration by extending the current paradigm. Processes such as word segmentation and regressive place assimilation could be tested under conditions that explicitly manipulate cues to sequential integration, using stimulus manipulations similar to those of the current study. Studies such as these could be invaluable in uncovering how spoken word recognition can be conceptualized in terms of the perceptual grouping process on which it rests.

References

- Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, 10(1), 65-70.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA: MIT Press.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244-249.
- Carnegie Mellon University. (1995). Cmu dict: The cmu pronouncing dictionary 0.6. Retrieved from <ftp://ftp.cs.cmu.edu/project/speech/dict>
- Cutting, J. E. (1975). Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance*, 104(2), 105-120.
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Science*, 1(9), 327-333.
- Darwin, C.J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B*, 363, 1011-1021.

- Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, 11, 411-412.
- Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, 122(4), 2340-2353.
- Dorman, M. F., Cutting, J. E., & Raphael, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 121-129.
- Kolinsky, R., & Morais, J. (1992). Représentations intermédiaires dans la reconnaissance de la parole: Apports de la technique de création de mots illusoires. *Proceedings of the 19th journées d'étude sur la parole. SFA, ULB, Brussels*, 129-133.
- Kolinsky, R., & Morais, J. (1996). Migrations in speech recognition. *Language and Cognitive Processes*, 11, 611-619.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing*, 19, 1-35.
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31, 1-24.

- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 576-585.
- Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of memory and language*, 36, 87-116.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- Micheyl, C., & Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *Journal of the Association for Research in Otolaryngology*, 11(4), 709-724.
- Moore, B. C. J., Shailer, M. J., & Black, M. J. (1992). Dichotic interference effects in gap detection. *Journal of the Acoustical Society of America*, 93(4), 2130-2133.
- Nygaard, L. C. (1993). Phonetic coherence in duplex perception: Effects of acoustic differences and lexical status. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 268-286.
- Remez, R. E. (2005). Perceptual organization of speech. In D. Pisoni & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 28-50). Oxford, England: Blackwell.

- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, 101(1), 129-156.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212(4497), 947-950.
- Roberts, R. A., & Lister, J. J. (2004). Effects of age and hearing loss on gap detection and the precedence effect: Broadband stimuli. *Journal of Speech, Language, and Hearing Research*, 47, 965-978.
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474-494.
- Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes. *Journal of Experimental Psychology: General*, 125(1), 28-51.
- Shoaf, L., & Pitt, M.A. (2002). Does node stability underlie the verbal transformation effect? A test of node structure theory. *Perception & Psychophysics*, 64(5), 795-803.
- Warren, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.
- Warren, R. M., & Obusek, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, 9(3B), 358-362.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence:
Confusion of patterns other than speech or music. *Science*, 164, 586-587.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*,
233, 30-36.

Woods, D. D., Sorkin, R. D., & Boggs, G. J. (1979). Stimulus context and duration discrimination.
Perception & Psychophysics, 26(2), 127-132.