Investigating how perceptual organization and linguistic memory processes interact to promote spoken word perception

A DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy, in the Graduate School of The Ohio State University.

By

Marjorie Freggens, M.A.

Graduate Program in Psychology

The Ohio State University

2023

Dissertation Committee:

Dr. Mark Pitt, Advisor

Dr. Julie Golomb

Dr. Cynthia Clopper

Abstract

Intro. A full account of speech perception requires explaining how listeners organize the acoustic signal into speech objects (perceptual organization) and how listeners use their memory for language to impart meaning to the speech objects (linguistic memory). Traditionally these mechanisms have been investigated separately, and thus theorized as two independent, sequentially- applied mechanisms. The prominent view of perceptual organization defines it as an early, low-level process, one that occurs prior to linguistic processing of speech. However, recent studies have found influences of organizational cues on linguistic percepts, indicating that the two mechanisms might not be sequential and independent. This dissertation attempts to explicate how perceptual organization and linguistic memory (lexical memory for words and sentential memory for context) interact when organizing and perceiving speech.

Methods. I presented participants with complex, un-organized speech scenes, with a stream of [s]s occurring in one ear and a sentence in the other. I asked participants to listen for the last word and report whether it started with a voiceless (p/t) or voiced (b/d) sound. Due to the phonotactic properties of English, the perceived identity of the target word changes from voiced to voiceless if the listener organizes it with a [s] from the s-stream ("[s]+base" perceived as "space"). By manipulating the grouping strength of the [s] stream and the linguistic biases embedded in the sentence (lexical and sentential), I tested whether and how perceptual organization and linguistic memory mechanisms interact during speech perception. The response time and cue weighting strategies of participants were also examined, to give more in depth information about the processing ease and cue tradeoffs.

Results. Experiment 1 compared categorization of target words with lexical biases (voiced, unbiased, voiceless) to s-streams with grouping strengths (strong, weak) and found independent

effects of both manipulations on perception. Cue weighting analysis found a consistent trade-off in cue reliance by participant, with most but not all relying more on memory cues than organizational cues. The response time data indicated that the speed of response was dependent on the cued organization: it was easier to make responses that agreed with the organization implied by the s-streams. Experiments 2-4 tested the interaction of organization and sentential processing (sentences biased by context) and found an effect of grouping strength that varied (but was never eliminated) based on the presence of sentence. Cue weighting data showed that sentential context cues were almost universally relied upon for categorization, while response time data supported the findings of Experiment 1, that response speed was influenced primarily by organization cues.

Conclusions. Taken together, my results imply that that linguistic memory and organizational information are both applied in an overlapping fashion, with an interactive juncture at the level of lexical processing and a much weaker interactive juncture at the sentential level of processing. In all experiments, both linguistic and organizational cues influenced the perceived target word, clarifying the organization of the two overlapping auditory streams. The combined results of the dissertation experiments bring together the two literatures on auditory perception: linguistic and perceptual organization cues are integrated together to inform the speech percept.

Acknowledgments

This paper is dedicated to my wife, Zoe. I could not have done this without you. I love you so much. I thank my friends and family for their support and love through this seven-year journey. I appreciate you all more than you know. Finally, I thank my advisor for his mentorship over the years.

Vita

B. A. Psychology, Georgia State University……..……2015

M. A. Psychology, The Ohio State University……..….2018

Graduate Researcher, Psychology Department, Ohio State University……..…..2016-2022

Graduate Teaching Assistant Psychology Department, Ohio State University……2017-2018

Research Experience Coordinator Psychology Department, Ohio State University……2019-2021

Publications

Freggens, M., Thomas, A., & Pitt, M. A. (2019). A test of linguistic influences in the perceptual organization of speech. *Attention, Perception, & Psychophysics*, *81*(4), 1065–1075. https://doi.org/10.3758/s13414-019-01699-3

Daltrozzo, J., Emerson, S.N., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L. & Conway, C.M. (2017). Visual statistical learning is related to natural language processing ability in adults: An ERP Study. *Brain and Language, 166*, 40-51.

Fields of Study

Major Field: Psychology

Specialization: Psycholinguistics and Auditory Perception

Contents

List of Tables

List of Figures

Introduction

Spoken language perception is the process by which listeners transform a mixture of incoming acoustics into speech objects bound with conceptual meaning. Listeners take in all the speech sounds occurring in the environment, combining all the acoustic properties from various talkers and background noise into a single, averaged signal hitting each ear. To make sense of the signal, the listener must decode it by grouping together the acoustics that came from the same talker and separating the irrelevant acoustics (a process of perceptual organization). Even once separated from irrelevant sounds, speech still needs to be organized further to promote accurate understanding. Continuous speech does not have clear acoustic boundaries for words: the listener must apply their memory for language to impose order on the speech (Cutler, 2012). For example, the spaces between words in this written sentence are not present in the acoustic signal for spoken speech (as they are perceived to be) but must be constructed by combining several acoustic and linguistic cues with listener knowledge (Remez, 2021). In doing this, the perceived utterance is constructed by the listener; it is not inherently evident in the acoustic signal. Organization imposed by the listener is necessary for accurate spoken language perception. The dissertation examines the interface of perceptual organization and linguistic memory in understanding spoken language.

Spoken Language Processing

Psycholinguists have focused their investigations on the process whereby the listener converts the relevant portions of the acoustic signal into linguistic understanding (Cutler, 2012). Current theories emphasize the incremental and combinatorial nature of speech processing: speech understanding occurs as speech arrives, not buffered or gated until the end of the utterance (McQueen, 2007). The raw acoustic signal must be analyzed for acoustic and linguistic cues to

identify speech sounds, which themselves are integrated to make words that access concepts, and then grouped into multiple words (typically, a sentence). Larger units are built from smaller units in real time, with linguistic knowledge adding competition and feedback processes to ensure rapid and accurate recognition and interpretation of the message (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Johnsrude & Buchsbaum, 2016).

The details differ among researchers (Weber & Scharenborg, 2012), but the major levels of linguistic processing are broadly agreed upon: sub-lexical (involving the sub-parts of words), lexical (words), and sentential (multiword combinations). Sub-lexical processing involves analyzing the incoming acoustics and comparing them to stored mental categories to determine which speech sound(s) the acoustics likely match (Raphael, 2021). When enough sub-lexical portions of speech are combined and processed, this allows for lexical processing to occur (Dahan & Magnusen, 2006). For example, the word "cash" is made up of the sequence of three sub-lexical phonemes ("cash" = /k/, /æ/, /ʃ/). What distinguishes lexical units from the sub-lexical level is that the units (words) connect to conceptual meaning (only when speech sounds combine into the word "cash" does the mental concept of the real-word object become activated). Evidence supporting and opposing the lexical activation of the word is summed together, and the lexical candidates compete for recognition (Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Vitevitch & Luce, 2016). In this way, lexical processing deals with determining the identity of individual words, independent of context. The most abstract level of linguistic processing is sentential: colloquially referred to as "context", this involves any combined or nuanced meaning gathered from multiple word strings or any understanding of how lexical units relate to one another (Elman, 2009; Magnuson, 2016). Only with words before or after "cash" can the listeners form a conceptual interpretation about that object (*Who is holding or receiving the*

cash?). The listener gains a conceptual understanding of the message by combining lexical items (e.g., *cash in the picture* conveys a different mental scene than *cash in my hand*). In this way, sentential processing refines and constrains the mental picture conveyed by the speech.

It is generally agreed that successful spoken language perception requires the completion of multiple linguistic processes, applied to the signal in a sequential but overlapping fashion: first the acoustic signal is analyzed by sub-lexical processing, then the resulting sub-word pieces are grouped and analyzed by lexical processing, and finally multiple words are combined for conceptual interpretation by sentential processing (Davis & Johnsrude, 2007; McClelland, Mirman, & Holt, 2006). Evidence from neuro-imaging studies suggests that these processes are applied in a hierarchical and interactive way, such that information from each level is fed to the next-highest process to interpret, and in some cases, abstract lexical or sentential information can provide feedback to earlier levels of processing (Davis & Johnsrude, 2003; de Heer et al., 2017). Under this linguistic framework, organization of speech acoustics is applied primarily by linguistic memory, using the listener's experience with language to group the acoustics into speech segments and match those segments with remembered words (Foucart, Ruiz-Tada, & Costa, 2015). In this way, memory is widely considered to be deeply involved in linguistic processing.

Perceptual Organization of Speech

In addition to processing speech as linguistic objects, successful spoken language perception requires a way to identify, isolate, and group the relevant incoming acoustics from background noise or competing talkers (Bregman, 1990, Remez, 2021). In normal listening environments, speech is not presented or heard in isolation: there is always some degree of background noise that the speech is embedded in. For listeners to accurately perceive speech, they must engage in

an organizational process to be able to focus on the relevant sounds. Researchers have long been interested in how the auditory system distinguishes multiple talkers given an acoustic mixture of sounds (Cherry, 1953). The auditory system uses multiple cues in the acoustic signal to determine which acoustics belong to the same talker and which acoustics should be grouped with separate talkers (Shinn-Cunningham, Best, & Lee, 2017). By researching the impact of multiple grouping cues on perception, researchers have proposed a general auditory system that can handle both the automatic grouping of unfamiliar sounds and the application of memory for previously encountered sounds to organize and interpret the signal (Bregman, 1990; Ciocca, 2008; Remez, 2021). This perceptual organization framework asserts that grouping based on organizational heuristics (rules applied to unfamiliar sounds) is applied to the acoustics before the memory system engages, asserting early acoustic influences on organization (auditory grouping) and later influences of memory on perception (memory-based grouping).

Auditory grouping, the first and most basic organizational process, depends on acoustic similarity, continuity, and repetition to impose order onto the signal (Bregman, 1990; Ciocca, 2008; Cutting, 1975). The auditory system makes sense out of the acoustic mixture the ears receive by either collecting acoustics together into a single perceptual object (like a person talking) or separating into multiple objects (like two separate conversations, or background noise from a talker). This allows the irrelevant sounds to be separated from the speech, for easier speech processing. Memory-based grouping, the second and higher-level organizational process, depends on the listener's memory for previously encountered sounds to further process important information in the signal (Darwin, 2008). It is theorized that after the auditory grouping has finalized, previously encountered and remembered speech sounds from the listener's memory are matched to the acoustic signal, and these matches are extracted for further processing (Davis &

Johnsrude, 2007). In this way, understanding speech produced in the real world relies upon both early application of auditory organization and later application of listener knowledge.

Applying this viewpoint to speech processing, most listeners can easily distinguish and focus on one talker in conversation, even though real-world speech is always embedded in a background of noise (speech-in-noise studies: Culling & Stone, 2017) or other talkers (Cocktail Party studies: Shinn-Cunningham, Best, & Lee, 2017). To achieve linguistic understanding, listeners must isolate and separate one signal (the target talker) from an acoustically similar mixture (background noise and/or other talkers). It is widely believed that only after this initial auditory organization can speech processing begin (Ciocca, 2008; Remez, 2021). Experiments examining this phenomenon typically present multiple utterances simultaneously, with organizational cues to separate them (by location or talker identity) and ask participants to report back the speech of the target talker. Researchers have found that errors in understanding target speech are often due to listener inability to separate simultaneous utterances and isolate the target utterance, either because the background noise masks the target (Speech-in-Noise experiments) or because the conflicting speech is not distinct from background speech (Cocktail Party experiments). These experiments show that organizing the acoustics into perceptual objects is a necessary first step to understanding the intended utterance (Culling & Stone, 2017; Shinn-Cunningham, Best, & Lee, 2017).

Typically, researchers consider this problem of speech isolation and grouping a separate issue from speech understanding, dividing this research into separate lines of investigation. The interesting question from the standpoint of linguistic processing is how the listener uses the speech acoustics to extract meaning (spoken language perception), and for this reason, research on linguistic processing often involves speech that has been pre-organized (i.e., presented in

isolation), unlike what listeners hear in the real world. Contrarily, the interesting question from the perspective of perceptual organization is how the listener groups and separates the acoustic signal into auditory objects. From this perspective, speech is considered another example of memory-based stimuli, like music or other known sounds, all of which are initially subject to mandatory, basic organizational processing to create auditory objects (Remez, 2021). Consequently, research on perceptual organization has prioritized exploring the contributions of various organizational cues on grouping perceptions, while the potential contribution of linguistic memory is acknowledged but rarely elaborated upon. The explanation of how linguistic memory works to further organize speech remains unanswered in this framework.

Because researchers have used similar stimuli to study different problems (perceptual organization versus linguistic memory), they have explored and elaborated on different processes necessary for successful auditory perception. Though both perspectives invoke memory-based effects to explain speech processing, the perceptual organization framework elaborates on the effects of organizational cues on initial auditory grouping, leaving the potential effects of memory on grouping unstudied, while the linguistic processing framework elaborates on memory effects on speech perception, without including potential effects from auditory organization. As a result, this problem of how listeners group speech together and separate it from irrelevant sounds has been understudied: for example, how do organization processes interface with lexical and sentential memory?

A complete model of human language comprehension requires understanding how these two processes (perceptual organization and linguistic memory) interface when organizing and perceiving speech. But currently, there are no models of spoken language perception that consider organization as part of the process. It is challenging to integrate the two perspectives as

the literature currently stands, though the simplest explanation might categorize auditory grouping as a separate, early, and independent process from linguistic processing, and linguistic processing as a part of secondary, memory-based grouping, along with other higher-order processes that make use of listener knowledge. This formulation would imply that when perceiving speech, the auditory grouping cues are integrated and finalized before memory-based grouping begins (linguistic cues integrated as a second step). However, the literature suggests that the auditory system's integration of the two cue types (organizational and linguistic) is more complicated than that.

Evidence of Integration

Sub-Lexical

How do listeners integrate both sources of auditory information (auditory organization and linguistic memory) together to organize and form speech percepts? It makes sense for listeners to integrate organizational cues into their sub-lexical processing, because sub-lexical memory representations are derived directly from experience with speech acoustics. Indeed, many classic studies have successfully documented an interaction between auditory grouping and sub-lexical processing, as low-level organizational cues like acoustic continuity and similarity have affected the sub-lexical units that listeners perceived (Darwin, 2008 for a review). For example, Pastore Szczesiul, and Rosenblum (1984; using a paradigm from Dorman, Raphael, & Liberman, 1979) manipulated the length of silence between concatenated recordings of 's' and 'day', asking participants to listen to the sequence and identify the word they heard. They found that listeners identification of the stop phoneme reliably changed (from /t/ to /d/) when enough silence was added between the recordings. The organizational cue of acoustic discontinuity (the silent

interval) informed listeners' sub-lexical processing, influencing the phoneme categorization process.

Researchers have demonstrated that listeners were sensitive to continuity and similarity in various aspects of the acoustic speech signal, like spatial location (Darwin & Hukin, 1999), silence-length (Pastore, Szczesiul, & Rosenblum, 1984; Repp et al., 1978), and pitch (Culling & Darwin 1993; Darwin & Gardner, 1986; Meyer & Barry, 1999) when engaged in classifying phonemes. In all these studies, cues for organization influenced speech processing, such that the identification of sub-lexical units (phoneme identity) changed based on the auditory grouping. Based on the number of studies showing the integration of organizational and linguistic cues for sub-lexical perception, it is unlikely that the two cue types (perceptual organization and linguistic memory) are integrated in a completely independent or fully sequential manner, rather that they are both integrated into the percept. Because linguistic processing has been demonstrated to be incremental and cumulative, these influences of auditory grouping on sub-lexical units could percolate up through the system, impacting lexical processing as well.

Lexical

As linguistic memories become more abstracted from the acoustics (lexical and sentential), the corresponding linguistic processes should have less ability to integrate organizational information for perception. By studying stimuli that have been pre-organized, researchers implicitly assume that successful extraction of the linguistic meaning requires speech to be in a perceptually stable, organized form at some point in speech processing: whether this is completed by the initiation of lexical or sentential processing is unclear. Results from different paradigms show that organizational cues have different weights on perception, influencing word perception differently depending on the context in which the grouping cues are presented.

Lexical processing and perceptual organization have been studied using both short, single presentation paradigms (Cutting, 1975; Cutting & Day, 1975; Day, 1968; Morais, 1996; Mattys & Melhorn, 2005; Poltrock & Hunt, 1977; Sexton & Geffen, 1981) and longer, repetition paradigms (Billig et al., 2013; Freggens, Thomas, & Pitt, 2019; Pitt & Shoaf, 2002; Warren, 1968).

Single presentation paradigms consist of short trials, where two words are presented dichotically to listeners simultaneously over headphones (Cutting, 1975; Cutting & Day, 1975; Day, 1968; Morais, 1996; Mattys & Melhorn, 2005; Poltrock & Hunt, 1977; Sexton & Geffen, 1981). Researchers typically find that despite organizational cues to separate the words, participants perceive a single, fused word across ears (Cutting & Day, 1975). For example, two words like "back" and "lack" are presented to opposing ears aligned by their acoustic onset, and participants are asked to report back what they hear. Instead of reporting back the two words, participants perceptually combine them together into one word ("black"). The organizational cues of spatial location and onset similarity were not enough to prevent the lexically driven fusing of speech percepts. In addition, Cutting (1975) found that other, more informative organizational cues, like acoustic dissimilarity in pitch, loudness, or onset timing also had little to no effect on fusion rates. These results indicate that organization cue weights are smaller when applied to speech of a short duration (<500ms), failing to influence lexical perception. Lexical knowledge overpowers auditory organization in determining the percept, preventing the organizational cues from informing word perception.

In contrast, the repetition paradigms present participants with long trials (many seconds in duration) consisting of repeating stimuli, where the listener indicates how their percept changes over time (Billig et al., 2013; Freggens, Thomas, & Pitt, 2019; Pitt & Shoaf, 2002;

Warren, 1968). Using the repetition paradigm, researchers have demonstrated that fast repetitions of the same word for many seconds can influence the lexical percept experienced by the listener (e.g., "write" eventually heard as "rye", "try", and "trite"). Initially, the lexical percept is veridical (listener hears "write" repeating), but after several repetitions the percept becomes reorganized by its acoustic properties (the continuous portion of the word ([rɑi]) perceptually separating from the stop consonant ([t]), such that other similar words involving those sounds ("rye", "try", "trite") are perceived by the listener (Pitt & Shoaf, 2002). This indicates that given adequate time for auditory cues to accumulate (i.e., a speech context of many seconds), low-level organization cues can cause a reorganization of speech sounds into various acoustically-similar words. This means that auditory grouping is influencing not only how the words are organized but also which words are perceived by the listener. In contrast to the single presentation studies, these repetition studies demonstrate the organizational cue weighting strengthening over time, eventually increasing enough to inform the listener's word perception. The outcomes of studies from the single presentation and repetition paradigms demonstrate the range of possible interaction between auditory organization and lexical processing: the weighting of both cues in determining the lexical percept varies depending on the buildup of both forces.

Sentential

If any linguistic percept should be insulated against the influence of auditory grouping cues, it would probably be sentential, because these representations are completely abstracted from the acoustics (Magnuson, 2016). Indeed, there is very sparse evidence for auditory grouping affecting processing of sentences or sentential processing impacting speech organization, as barely any studies have investigated these questions. Uddin and colleagues (2018) investigated the ability of listeners to understand sentences using mixed lexical and non-speech

representations. They presented participants with sentences over headphones ("The [bleat] went out to the pasture to eat grass."), wherein a non-speech sound (ex: an acoustic recording of a sheep bleat) was substituted for its lexical component (the word "sheep"). Participants were tasked with evaluating the mental scene evoked by the sentence in terms of whether it was sensible. They found that participants showed little difference in accuracy or response time when the word was replaced with its acoustic counterpart. Though organizational heuristics should prevent grouping disparate acoustic sounds and speech together (like their mixed sentences), Uddin et al.'s experiment indicates that mixed sentence integration occurs with similar ease as natural sentence processing. This result implies that sentential cues could be weighted heavier than auditory grouping cues, overpowering them in sentential perception, or that different rules apply in these contexts.

To summarize the current state of knowledge, the process of organizing speech has not often been studied alongside language comprehension, and thus has often been left out of conceptualizations of spoken language perception. Organization of speech components is assumed to start and end swiftly, before much linguistic processing has begun. Evidential support for auditory grouping cues being integrated into linguistic perception has been shown primarily at the sub-lexical level. Results from different paradigms imply that auditory grouping can have a range of interactions with lexical processing, depending on the build-up of cues. In addition, no studies have directly investigated the potential of auditory grouping to interact with sentential processing, mostly because such abstract language processing is assumed to be immune to non-linguistic forces (Davis & Johnsrude, 2007). A reasonable assumption is that by the time sentential processing has started, the speech signal should be in a perceptually stable, organized form, but there is not enough evidence to fully support that idea. This means that a

potential influential force on language processing has been overlooked, and conceptualizations of spoken language perception could be missing a source of information about how listeners organize and understand incoming speech.

Purpose Statement

This dissertation attempts to synthesize the two disparate frameworks into a cohesive account of the architecture underlying spoken language processing, by examining how perceptual organization and linguistic memory interface when organizing and perceiving speech. Though the two frameworks have been assumed to be independent and sequentially applied to speech (acoustic grouping occurring first, followed by linguistic analysis), evidence from select studies has hinted that mechanisms within these frameworks could interact in a more parallel or overlapping fashion. If this is true, then auditory grouping and linguistic processing might be more integrated than previously assumed, requiring a revision to the proposed speech processing architecture.

This dissertation evaluates how organizational cues and lexical/sentential knowledge are integrated during speech perception, by embedding linguistic and organizational cues in the speech to create conflicting percepts and recording the resulting percept. For example, when lexical bias indicates grouping incoming speech together into the word 'sponge,' but auditory grouping cues indicates that the [s] does not belong with the speech, how does the system settle on the linguistic percept ('sponge' or 'bonge')? Responses can indicate how the two sources of information are applied and weighted when determining the intended linguistic meaning, allowing for the potential trade-offs between cues. If responses indicate that both linguistic memory (lexical or sentential biases) and auditory grouping have an equivalent effect on the percept, then this places doubt on the idea that auditory grouping is completed and applied to the

signal before memory-based processing begins, implying a parallel application of both mechanisms. If auditory grouping and linguistic memory cues are unequal in their influence on speech perception, then this implies that the auditory system is more interactive and hierarchical in application of organization and memory processes. My experimental design allows for the potential trade-offs between organizational and memory contributions on speech perception to be explicitly measured. In this way, the dissertation helps clarify the potential architectural structure underlying speech perception, by integrating two frameworks of auditory processing: perceptual organization and linguistic memory.

Experimental Design

In exploring these questions, I used a hybrid paradigm, which combines the fusion responses of the single presentation studies (Cutting, 1975; Cutting & Day, 1975; Day, 1968; Morais, 1996; Mattys & Melhorn, 2005; Poltrock & Hunt, 1977; Sexton & Geffen, 1981) with longer scenes of the repetition paradigm (Billig et al., 2013; Freggens, Thomas, & Pitt, 2019; Pitt & Shoaf, 2002; Warren, 1968). This hybrid paradigm allows effects of auditory grouping to sufficiently build throughout the trial, culminating in linguistic fusion. Freggens and Pitt (2023, under review) developed a precursor to this paradigm (based on Pastore et al., 1984) to examine auditory grouping on lexical perception. Their paradigm involves presenting an [s]+stop initial word ("spring") split by location ([s] to one ear and word base [bring] to the other ear) and gathering the resulting identification of the stop consonant ("Did you hear P or B ?"). Due to the phonotactic properties of English, voiceless stops ([p]) produced after fricative consonants ([s]) are perceived as voiced versions of the stop ([b] for [p]) when that consonant is removed (spring – [s] = bring). Therefore, whether the reported speech sound is a voiced stop ("b") or a voiceless stop ("p") indicates whether the [s] was perceptually organized with the word base. Freggens and

Pitt have found that listeners overwhelmingly group the isolated [s] with the base (reporting voiceless “p”), despite the organizational cue to separate the two (maximum spatial distance). But when given additional lexical context connecting the isolated [s] to another word (i.e., simultaneously presenting “spring” on left and “start” on right), listeners grouped the [s] away from the target word (reporting voiced “b”). The paradigm modulates the perceptual force pulling the [s] away from the word, by adding or removing lexical context to the [s]. The hybrid paradigm in this study elongates the stimuli of the paradigm Freggens and Pitt developed, to allow organizational and linguistic cues to build over time.

The hybrid paradigm can demonstrate auditory grouping interacting with lexical and sentential processing in a context that approximates natural speech (complete sentences). This is important, as evidence from single presentation and repetition paradigms imply that the integration of auditory grouping and lexical memory processes varies depending on the stimuli length. The effects found in those studies could be underreported or exaggerated due to the extremely short and long contexts in which the cues were presented, compromising the external validity of their results. This hybrid paradigm allows for the combined contributions of auditory grouping and linguistic memory to be measured directly. In addition, the use of sentences in the paradigm allows for direct comparison between levels of linguistic processing: lexical and sentential. The results of this comparison will have implications for the interface of auditory grouping and linguistic processing hierarchy: at which junctures of speech processing can auditory grouping interact with linguistic processing to influence the percept? Using this paradigm, I can systematically explore the effects of perceptual organization and linguistic memory on spoken language perception, informing the potential interface between the two frameworks of speech processing.

To achieve enough organizational flexibility to test the integration of organizational cues into linguistic percepts, the auditory scene must contain enough acoustic ambiguity to be able to support multiple organizations. I set up a complex auditory scene containing both linguistic and auditory streams, ending in the target phoneme identification task described above (see Figure 1, below). In this scene, listeners heard a sequence of linguistic objects (a predictive or non-predictive sentence) in one ear and an s-stream (a series of identical [s]'s repeating isochronously) in the opposing ear. At the end of each trial, the linguistic stream contained an [s]-sized gap (silence) before the last word, and a target [s] was embedded in the s-stream, temporally aligned to the linguistic gap. The listeners' task was to report the phoneme they heard (choosing either "p" or "b") in the last word, and, in so doing, signifying how they organized the target [s]: integrated or segregated from the final word.

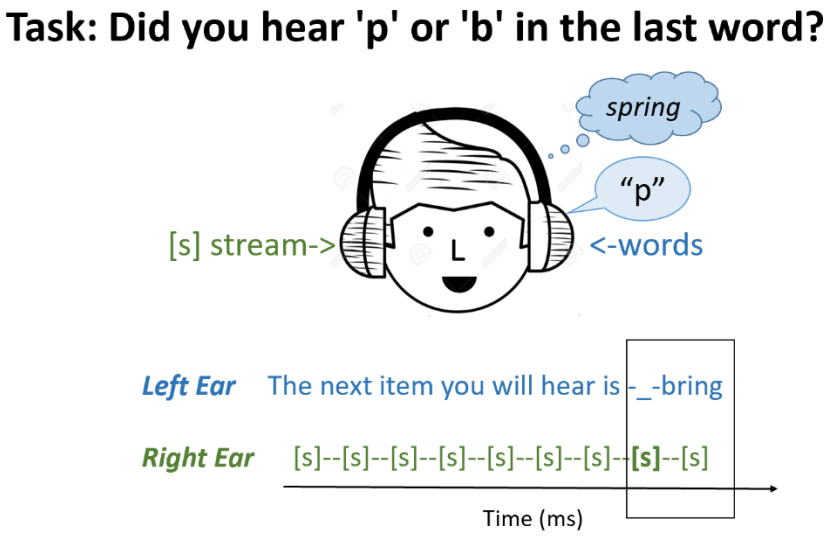

Figure 1: *A listener in this setup hears sounds over headphones and picks which phoneme they heard out of two choices. A non-predictive, neutral sentence occurs in the left ear and a string of [s]s simultaneously occurs in the right ear, with a target [s] presented just before the target word (bring).*

If participants responded with “b”, then that indicates that the target [s] was not organized into the linguistic stream. In order to hear “b”, listeners must have perceived two objects: a background s-sequence containing the target [s] separated from a foreground word “bring”. If listeners indicated “p”, then that means that the target [s] was organized into the linguistic stream: listeners perceived the [s] as part of a combined linguistic object, “spring”. Because the target [s] was acoustically identical to the adjacent [s]s in the s-stream, organizational cues should insulate the target [s] against being integrated into the speech stream. At the same time, the sentence imposes linguistic bias to integrate the target [s] into the speech stream. In this way, the paradigm allows for a clear understanding of the listeners’ organization and perception of the [s] acoustics, indicating how auditory grouping and linguistic memory interact during speech processing.

Manipulations and Predictions

To evaluate how auditory grouping interacts with linguistic processes to influence speech perception, I manipulated the strength of the organizational cues in the scene’s s-stream and measured the change in the reported word. Varying the number and rate of the non-target [s]s changes the perceptual pull of the s-stream on the target [s], changing its influence on the linguistic percept (see Figure 2, green text). Using a string of identical, fast-repeating [s]’s sets up the expectation for the auditory system that the same pattern will continue in that ear for the entire trial (Strong grouping, Figure 2B). Since the target [s] is identical to the repeating pattern that came before it, the perceptual system is likely to keep the embedded [s] integrated with the competitor s-stream, preventing it from contributing to the word (causing a lexical percept of “bring”, a voiced response). For the weak condition, the s-stream will not build up before the target, as the repetitions will be too few and far apart to cohere into a perceptual stream (Figure

2A, green text). This means that the s-stream will not have a strong cohesion among the [s]s, allowing the speech stream to integrate the [s] instead (lexical percept of "spring", a voiceless response).

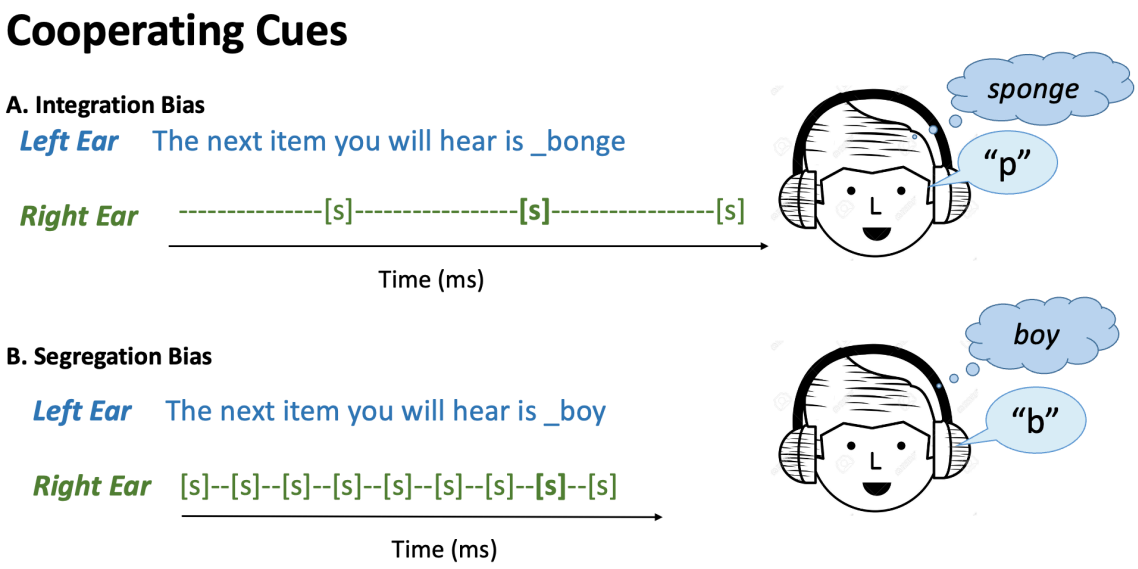

Figure 2: *This image shows the trial setup for endpoint conditions in Experiment 1. The text in blue is the linguistic stream (non-predictive sentence ending in a word or nonword) and the text in green is the s-stream (fast or slow repeating [s]s). Conditions 2A and 2B serve as endpoints, demonstrating that the linguistic and organizational forces are exerting appropriate influence over the resulting percept.*

To measure the influence of linguistic memory, I included conditions where the linguistic streams contained linguistic biases: lexicality bias (Experiment 1), the tendency for listeners to perceive words rather than non-words, and sentential bias (Experiments 2-4), the tendency to use prior words to predict upcoming words (Connine & Clifton Jr., 1987). These biases were intended to either enhance or prevent the target [s] from binding with the target word, by providing a linguistic reason to integrate or segregate the target [s], reliant on listener knowledge of English words and their relations. In the lexical bias conditions, the last word in the sentence was either only a valid word with the [s] ("sponge" vs "bonge", Figure 2A) which biases voiceless responses, or only a valid word without the [s] ("boy" vs "spoy", Figure 2B) which

biases voiced responses. In comparison, the sentential bias changed the content of the pre-target words in relation to a lexically unbiased target word, either with context biasing the voiced target ("Sam muzzled his dog so it wouldn't bark") or the voiceless target ("The wood on the fire created that spark").

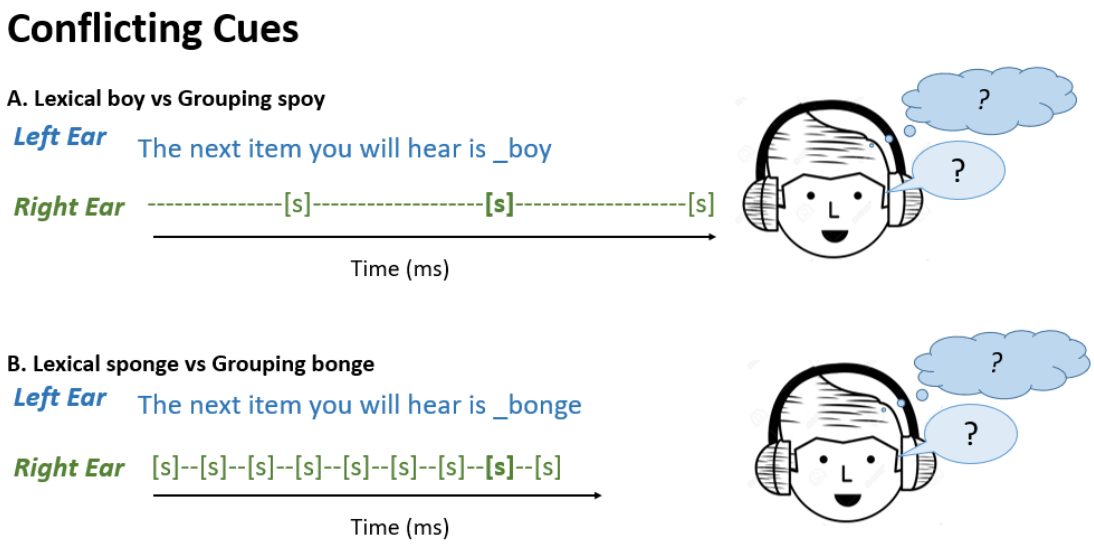

Figure 3: *This image shows the trial setup for the critical conditions in Experiment 1. Responses to conditions 3A and 3B are uncertain and will demonstrate the relative strength for linguistic and organizational forces.*

The combination of manipulations will allow me to measure the influence of auditory grouping (s-stream strength) in the context of linguistic biases (lexical and sentential), demonstrating the extent to which perceptual organization and linguistic memory interface to influence speech perception and organization. Figures 3A and 3B demonstrate these conditions: in both, the s-stream and the lexical bias do not predict the same percept. In one case (Figure 3A), the linguistic cues bias [s] segregation (word "boy" instead of non-word "spoy") while the auditory grouping biases the opposite outcome (weak s-stream allows [s] integration: "spoy"). Contrarily, in the other condition (Figure 3B), the linguistic forces bias [s] integration (word

"sponge" instead of non-word "bonge"), whereas the auditory grouping cues bias [s] segregation (strong s-stream prevents [s] integration: "bonge"). Responses to both conditions can indicate which process exerted a stronger bias for word perception, and thus elucidate how the auditory system integrates organizational and memory-based processing.

The predictions are the same for both lexical and sentential biases: they are based on the evidence for interaction between the organization and memory processes. If perceptual organization and linguistic memory mechanisms are completely independent and sequentially activated (first grouping then linguistic knowledge), as proposed by the organization framework (and assumed by the linguistic framework), then we should find only an effect of linguistic memory, without an accompanying grouping effect. This result implies that the organization mechanism initiated and finalized processing the stimuli before the higher-order memory mechanism initialized. This outcome is unlikely for lexical processing, considering that research has already found interactions of varying strength between organization and lexical memory (Billig et al., 2013; Cutting, 1975; Freggens, Thomas, & Pitt, 2019). But this could be the case for sentential processing, as it operates on more abstract, higher-level information than speech acoustics: both the linguistic and organization frameworks support the idea that sentential processing occurs after organizational processing is finalized.

In contrast, finding effects of both organization and memory cues can give more detailed information about the interaction between the two mechanisms. If the organization and memory mechanisms are engaged in parallel (or with some overlap) during speech perception, then I should find strong effects of both grouping strength and linguistic memory, neither of which change in magnitude when embedded in conflicting conditions. This result would not distinguish between two mechanisms operating independently on a parallel timescale and two

interdependent processes interacting via an architectural juncture. However, If I find an interaction between the manipulations (an effect of grouping strength which lessens based on the linguistic bias), then this gives definitive evidence for an interactive connection between organizational and memory mechanisms. In this case, speech processing likely involves both organization and memory processes operating interactively and in tandem to create the resulting speech percept. Finding evidence for parallel processing or interactivity between perceptual organization and linguistic memory mechanisms would contradict the assumption that speech organization must settle into a stable form before linguistic processes begin to be applied to speech. Rather, that both organization and meaning-extraction are working together to organize and understand the signal, as the speech is accumulating.

Upcoming Sections

The structure of the dissertation is as follows: Experiment 1 demonstrates the interaction of auditory grouping and lexical processing during word perception, replicating and extending earlier studies with natural stimuli. Experiment 2 will be the first to directly explore the interaction of sentential biases and auditory grouping cues on word perception. Experiments 3 and 4 replicate the effects of Experiment 2 and strengthen the evidence for a gradient effect of auditory grouping on speech perception, demonstrating the additive nature of the cue weighting. By the end of this dissertation, I will provide direct evidence that organization and memory mechanisms work together in promoting linguistic understanding.

Experiment 1: Lexical Bias

Introduction

Past evidence of interaction between lexical memory and organization mechanisms when perceiving speech has varied in strength, likely due to the length of the stimuli used in the task. Single presentation (Cutting, 1975; Poltrock & Hunt, 1977; Sexton & Geffen, 1981) studies, involving trials with single word stimuli, typically find that lexical knowledge can insulate the percept from effects of organizational cues (a lack of interaction). In contrast, repetition paradigms (Billig et al., 2013; Freggens, Thomas, & Pitt, 2019), with stimuli lasting over several seconds, find that organizational cues can build in strength to interrupt or influence lexical perception. This experiment serves as a first test of the hybrid paradigm, to reveal interactions between memory and organization in a context that approximates natural speech (complete sentences without pre-organization). The results of this experiment should more accurately reflect how organization and lexical memory impact word perception.

The purpose of Experiment 1 is to explore how perceptual organization and linguistic memory (specifically lexical memory for words) interact to promote speech perception. Both the grouping and lexical bias manipulations were applied to the scene to measure the effect of perceptual organization and lexical knowledge on word formation. In each trial, participants listened to a scene with a non-predictive sentence ("The next item you will hear will be…") ending in a word or non-word in one ear, while a simultaneous s-stream ("[s]-[s]-[s]…") occurred in the other ear (see Table 1 for condition overview). The s-stream either consisted of a fast-repeating sequence of many [s]s (strong condition) or a slow sequence of few [s]s (weak condition), influencing the strength of the grouping effect. Lexicality was manipulated by the choice of word at the end of the sentence: whether the target was a word only with the [s] (voiceless bias: "sponge"-"bonge"), only without the [s] (voiced bias: "spoy"-"boy"), or both

with and without the [s] (unbiased: “spring”-“bring”), as demonstrated in Table 1 (below). The unbiased condition provided a neutral baseline against which to measure the effects of voiced and voiceless bias. By directing participants to listen to the sentence and report the type of stop consonant they heard (voiceless or voiced), I can identify how the two mechanisms (organization heuristics and linguistic memory) are applied when determining the intended linguistic meaning. The combination of these manipulations (specifically the conditions where the two cue types are in opposition) will have implications for the theorized structure of the auditory system when organizing and perceiving speech.

If perceptual organization and linguistic memory are sequentially applied (grouping processes completing before application of lexical knowledge), as proposed by the organization framework (and assumed by the linguistic framework), then I should find a singular effect of linguistic memory, without an accompanying grouping effect. Given previous studies on lexical processing and organization, this outcome is unlikely. If the organization and memory mechanisms are engaged in parallel (or with some overlap) during speech perception, then I should find strong effects of both grouping strength and linguistic memory, neither of which change in magnitude when embedded in conflicting conditions. If I find an interaction between the manipulations (an effect of grouping strength which lessens based on the linguistic bias), then this gives strong evidence for an interactive (rather than independent) connection between organizational and memory processes.

Method

Participants

The participants consisted of 60 undergraduate Ohio State University students from various ‘Introduction to Psychology’ courses. They were given course credit for completing the

experiment. All participants in this study self-reported being native speakers of English and having no hearing loss. Three participants' data were excluded due to poor foil performance, leaving 57 participants. This dissertation study was approved by the Ohio State Institutional Review Board (study number: 2009B0066).

Stimuli

The stimuli were designed to create ambiguity in the auditory organization upon which lexical and grouping forces could act. Two simultaneous auditory streams were created at different spatial locations, one with a spoken sentence ending in a target word and the other with a sequence of isochronously repeating [s] sounds. The target word split the [s] from the base, presenting the [s] in the other stream, embedded within the sequence of identical [s]s. Calibration pilots were conducted to ensure that the scene was ambiguous enough to allow for organization and linguistic cues to change the target percept. In a few pilot experiments, the number of [s]s (1, 3, 7, 11), repetition rate (0, 100, 300, 500, 900, 1300ms), and spatial separation (90 and 150 degrees) of the s-stream were varied, to find the combination that allowed strong and weak conditions to reach 75% and 25% voiced responses. The best configuration for the strong s-stream condition consisted of 11 [s]s with a 0ms repetition rate, and the weak s-stream consisted of 3 [s]s with a 900ms repetition rate (see Table 1 for a visual representation). Surprisingly, spatial separation did not influence the results (no difference in voiced reports for 90 compared to 150 degrees), so 90 degrees was used for the experiments.

	Strong Grouping	**Weak Grouping**
Voiced Bias	Left: The next item you will hear is **_boy** Right: [s][s][s][s][s][s][s][s][s]**[s]**[s][s]	Left: The next item you will hear is **_boy** Right: [s]—900ms-----**[s]** —900ms-- [s]
Unbiased	Left: The next item you will hear is **_bring** Right: [s][s][s][s][s][s][s][s][s]**[s]**[s][s]	Left: The next item you will hear is **_bring** Right: [s]—900ms-----**[s]** —900ms-- [s]
Voiceless Bias	Left: The next item you will hear is **_bort** Right: [s][s][s][s][s][s][s][s][s]**[s]**[s][s]	Left: The next item you will hear is **_bort** Right: [s]—900ms-----**[s]** —900ms-- [s]

Table 1: *Conditions and examples for Experiment 1: Lexicality Bias. The bold items indicate the targets in each stream. The grouping strength is reflected in the s-stream, while the lexical bias is reflected in the sentence-final word.*

Target words were chosen to engage a lexical bias: voiced bias ("boy" vs. "spoy"), voiceless bias ("sponge" vs. "bunge"), and unbiased ("spring" vs. "bring"). For each of the three lexical conditions, 32 monosyllabic items were chosen: half of those (16) were /sp/ initial while the other half were /st/ initial. The vowels following the initial consonant were chosen to span the range of vowel space in English (see Appendix A for a list of items). Each item (both the integrated and segregated version: "sponge"-"bonge") was reviewed for frequency in the spoken section of the Corpus of Contemporary American English (COCA: Davies, 2008-). The average spoken frequency for the items by condition was evaluated on a base 10 logarithmic scale, to match the psychological representation of frequency effects (Whaley, 1978). The words in the unbiased condition had similar log frequency for both their integrated ("spring"; M=2.82, SD=1.03) and segregated ("bring"; M=3.16, SD=1.18) versions. Biased voiced stimuli had a high average frequency for segregated versions ("boy"; M=3.6, SD=0.8). Biased voiceless stimuli had the opposite pattern, a higher frequency for integrated versions ("sponge"; M=2.86, SD=1.12). The average log frequency of the voiced bias was significantly stronger than the voiceless bias ($t(31)=2.82$, $p=.006$), showing that the lexical bias to segregate the [s] was significantly stronger than the lexical bias to integrate it. For this reason, the logarithmic difference in word frequency was included as a covariate in the logistic models for both the categorization and RT data. To give the grouping cues a chance to build and to equate the sentence and s-streams in length, three non-predictive sentence frames were created to precede the target words (e.g., "The next item in the sentence is…"). The frames all had similar wording and timing (M=1,421ms). The frames and lexical targets were recorded separately, then combined using a custom python script.

A female native-English speaker recorded the stimuli on a Tascam HD-P2 audio recorder in a sound-attenuated room, stored as WAV files with a sampling rate of 44.1kHz. The stimuli

were isolated and normalized in volume to 70dB using a custom Praat script (Boersma & Weenink, 2022). They were then lateralized to give the perception that they were spatially separated by 90 degrees (+45 and –45 degrees from center) using a custom python script, using Head-Related Transfer Functions (Kayser et al., 2009). For each target word, the [s] was isolated and lateralized to appear to be 90 degrees spatially separated from the base (e.g., left ear [s] and right ear "boy", for the target "spoy"). Two versions of each lateralization were made, one with the [s] on each side. The stimuli randomly switched sides of presentation throughout the experiment, such that participants could not predict which side the sentence would be presented on.

To create the s-stream, the isolated [s] from each target was lateralized and repeated after a specific amount of silence (see Table 1). For the weak grouping condition, only three [s]s, one pre-target and one post-target, were present in the s-stream, each separated by 900ms of silence, to create a weakly connected stream of [s]s. Table 1 shows the target [s] (bolded, in the middle of the s-stream), which was temporally aligned to occur just before the target word (using the natural gap timing between the [s] and the stop phoneme from the recording of the word). In the strong grouping condition, eleven [s]s were presented in an isochronous sequence: eight preceding the target [s] and two following the target [s]. For the strong grouping condition, there was no period of silence between the [s]s (0ms gap).

Foil trials were created to identify participants whose responses reflected conscious strategy instead of listening to the stop phonemes. There were two strategies that the foils were designed to identify: any participants who categorized the stop based on only lexical knowledge (only categorizing stops in a way that would make a word, regardless of what they sounded like) and any participants who categorized the stop based on how many [s]s they heard (all weak s-

streams as voiceless and all strong s-streams as voiced). The lexical strategy was addressed by hyper-articulating the stop (p/b/t/d) and adding an [s] before the nonword ([s]+blit). The correct response would reflect the stop identity (voiced stop for b/d-initial words and voiceless stop for p/t-initial words), regardless of whether the item was lexical with or without the [s]. The foil trials were 20 non-words beginning with either [b p d t] (e.g., “blit”, “pall”, “druck”, “tive”). The same female native-English speaker recorded the foil stimuli to emphasize the stop identity (hyper-articulation), and then had an [s] from one of the target words spliced into the foil word ([s]+tive). The foil words were then attached to one of the three sentence frames and added to spatially separated streams via a custom-generated python script.

The voiceless foils (“pall” and “tive”) were paired with the strong grouping condition, and the voiced foils (“blit” and “druck”) were paired with the weak grouping condition. This was intended to discourage the strategy to report all strong s-streams with voiced stops and all weak s-streams with voiceless stops, to allow the identification and exclusion of participants engaging in that strategy. The voiceless foils should elicit no voiced responses (0%), while the voiced foils should have only voiced responses (100%). In total, participants were tested on 242 trials, with 40 foils (20 voiceless and 20 voiced), 10 practice trials, and 192 test trials (6 conditions x 32 words).

Procedure

The experiment was presented through a custom-built, browser-based platform, accessible from a URL link. Participants used their own computer and headphones to complete the experiment. The only restrictions were that participants had to use a computer or laptop (not a phone or tablet) and access the link via either Firefox or Chrome web browser.

Before starting the experiment, participants had to complete three pretests. The first was a sound-level test, where participants clicked on a button to play a sound file (a male, native-English speaker saying "spark") and adjusted the volume on their machine to a comfortable level. The second test was a stereo test to ensure that participants were listening over headphones or earbuds in stereo. In this test, participants heard a 500Hz tone, with an interaural timing difference (ITD) adjustment to appear to be presented slightly more to the left or right headphone. On each trial of the pretest, participants had to choose where the sound was coming from (left or right). After correctly getting a sequence of five tones correct, the participant passed the test and moved onto the last pretest. If they didn't get the entire tone sequence correct, they repeated the 5-trial test once before being politely removed from the experiment. The last pretest was an electronically presented consent form. To proceed to the experiment, the participant had to press a button to confirm their consent to participate.

In the experiment, the participants first saw an instructions page, with a visual depiction of the task and written instructions. On each trial, they would hear a sentence in their left or right headphone with "background noise" (the s-stream) in the other ear. Their task was to use the keyboard to indicate the sound that they heard in the last word of the sentence. Participants had 5000ms from the start of the audio to make their response (the long response time range was due to the varying lengths of time for the sentences to complete), although they were encouraged to respond as soon as they heard the last word. The response choices were between [p] and [b] in one block and [t] and [d] in another block (counterbalanced between lists). Response choices (voiceless p/t or voiced b/d) and their corresponding keyboard keys (F: voiceless and J: voiced) were presented visually on the screen during the trial. The trial ended as soon as a response was made. After the trial, feedback was presented based on their responses and response time. If

participants made no response or responded with a key outside of the two response options, they saw “Respond Faster! (F or J buttons)” in red text on the screen for 1000ms before continuing to the next trial. Otherwise, they saw no feedback. The inter-trial interval was jittered such that on any trial the wait could be 500, 1000, or 1500ms before the next trial.

Participants had a practice block of five trials to familiarize them with the task. There were two test blocks (p/b and t/d), with self-paced breaks scheduled every 58 trials. Before the second block, there was an updated instruction screen and five new practice trials. Participants completed 242 trials in total, taking 20-30 minutes on average. Block order and presentation side (whether the speech was in the left or right ear) was counterbalanced across lists. When participants finished the experiment, they saw a conclusion screen with a link to the experiment survey. The survey took 5 minutes on average and asked demographic and participation questions (see Appendix A for survey questions).

Results & Discussion

Exclusion and Analysis

Participants who failed to respond to more than 25% of the target trials were excluded from the study (n=2). Any participant scoring below 75% correct on foil trials was also excluded from the study (n=1). Any trials where the response time was before the onset of the target phoneme was removed (3.5%). To analyze the results, I used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical software (v4.1.2; R Core Team 2021) to employ logistic mixed modelling on the voiced responses with grouping strength, and lexical bias as fixed effects and participant and item as random effects. Due to the unequal word frequency biases for voiced and voiceless items, the logarithmic adjusted-difference in word frequency was included as a covariate in the models. Including word frequency difference as a covariate significantly

improved model fit ($X^2(16)=8.63$, $p<.001$). Comparison of the null ("response ~ word_frequency + (bias|pcode) + (grouping|item)"), main effect ("response ~ bias + grouping + word_frequency + (bias|pcode) + (grouping|item)"), and interaction ("response ~ bias * grouping + bias + grouping + word_frequency + (bias|pcode) + (grouping|item)") models was performed using a likelihood ratio test to determine which relationship fit the dataset best.

Categorization Data

Figure 4 shows the proportion of voiced responses ("d" or "b") on the y-axis, averaged over lexical conditions (x-axis: voiced, unbiased, voiceless) and grouping strength conditions (color: strong with blue and weak with red) for each participant. Each participant has an average voiced response proportion for each of the six conditions, represented by a dot on the figure. The spread of dots in each condition shows the similarity or dissimilarity in the percept among participants, while the boxplots show the concentration of voiced proportions for each condition: a large spread indicates that the participants perceived the stimuli differently.

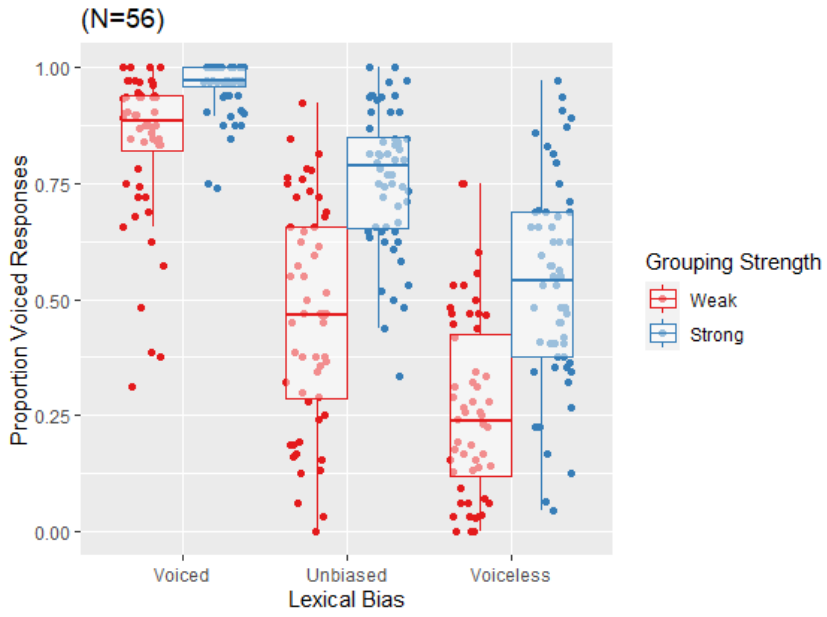

Figure 4: *Boxplots of the proportion of voiced responses by lexical bias and grouping strength conditions. Lexical bias conditions are shown on the x-axis (voiced, unbiased, voiceless), and grouping strength is distinguished by color (weak: red, strong: blue). Each dot represents a participant's average.*

Based on the biasing conditions, we should see a couple patterns in the figure. Lexical bias, or the tendency for listeners to perceive ambiguous speech as words, should cause the proportion of voiced responses to be highest for voiced bias items, less for unbiased, and lowest for voiceless bias (Cutler, 2012; Vitevitch & Luce, 2016). Due to organizational heuristics preferencing the integration of acoustically similar and repetitive sounds, participant's proportion of voiced responses should be higher for strong grouping (blue) than weak grouping (red) conditions (Ciocca, 2008; Darwin, 2008). Crucially for the dissertation, the presence or absence of an interaction between the two manipulations will demonstrate how the perceptual organization and linguistic memory mechanisms interface: a consistent effect of grouping and lexical manipulations across the conflicting conditions implies that the auditory system applies both mechanisms in a parallel manner, whereas varying effect sizes would indicate that the system applies organizational and memory-based processing in a parallel and interactive fashion.

The data in Figure 4 shows effects of lexical bias (voiced > unbiased > voiceless) and grouping strength (strong > weak), both in the predicted directions. The effect sizes for both manipulations are large, indicating each manipulation had a strong effect on perception: averaged voiced proportions for lexical bias conditions differed by 50% (voiced *M*=.89, *SE*=.005 and voiceless *M*=.40, *SE*=.008), and for grouping strength conditions differed by 23% (strong *M*=.75, *SE*=.007, weak *M*=.52, *SE*=.006). Though strong grouping conditions (blue) had more voiced responses than weak grouping (red) for each lexical bias condition, the grouping strength effect was much smaller for the voiced bias (.13), compared to the unbiased (.31) and voiceless bias (.28) conditions. Though the ceiling effect found in the voiced bias conditions render it less useful in interpretation, the similarity in the grouping effect size for unbiased and voiceless bias conditions lends evidence against an interaction. This indicates that both grouping cues (strong

vs weak) and lexical cues (voiced vs voiceless bias) were influential in determining the perceived word.

To assess the reliability of these results, I employed logistic mixed modelling on the voiced responses with grouping strength and lexical bias as fixed effects and participant and item as random effects. Comparison of the null, main effect, and interaction models using a likelihood ratio test indicated that the main effect model significantly improved the fit over the null model ($X^2(11)=232.35$, $p<.001$), which included just the random effects and the word frequency. The interaction model did not improve the fit compared to the main effect model ($X^2(14)=3.93$, $p=.14$), indicating that the lexical and grouping manipulations had consistent and independent effects on perception. This suggests that the smaller grouping effect for the voiced bias condition is due to the distribution of voiced responses: they are clustered at the ceiling of the possible response proportions (>.75), preventing differences between grouping conditions from fully appearing. If the responses for voiced bias were not at ceiling, it is likely that the same magnitude of difference to grouping strength conditions would be found as in the other two conditions.

This equivalent influence of both manipulations indicates that grouping and lexical cues are both integrated when perceiving spoken words. Both cue types have an impact on the percept, but not in an additive way, as can be seen by the independent effects of both manipulations. The lexical manipulation has a consistent effect on the word the participants hear, with bias causing a 50% change in the average reported percept. Fixed contrasts support these findings, as voiced biased words were 5.18 times more likely to be categorized voiced as unbiased words ($Z=5.22$, $p<.001$), and unbiased words were 1.71 times more likely to be categorized voiced as voiceless biased words ($Z=1.8$, $p=.19$). But at the same time, the grouping

manipulation also has a consistent effect on the organization of the [s] with the sentence, with s-stream strength changing the average reported percept by 23%. Fixed contrasts support these findings, as strong items had 5.56 higher odds to be categorized voiced, compared to weak items (Z=26.64, p<.001). Both manipulations had independent effects on perception, suggesting that perceptual organization processes overlap at least slightly with lexical memory processes (evidence of parallel processing). The lexical bias, though strongly influential for perception, did not insulate the percept against the grouping manipulation.

This experiment has implications for the underlying architecture of human language comprehension, demonstrating that both perceptual organization and lexical memory are active during lexical processing. This dual influence is contrary to the strict hierarchy of the linguistic framework, which assumes that acoustic organization processes start and end before the influence of linguistic memory when processing speech. This strict hierarchy is supported by studies of word segmentation, which show that linguistic memory (both lexical and sentential) can override acoustic organizational cues (Mattys & Melhorn, 2007; Mattys, White & Melhorn, 2005). However, this experiment shows that during the traditionally linguistic process of word identification, acoustic grouping cues can influence perception. And even when these two cues conflict (lexical bias and grouping strength), lexical memory cannot fully insulate incoming speech from being influenced by perceptual organization. For this to be true, there must be an interactive juncture between lexical processing and perceptual organization, such that auditory grouping processes are continuing during linguistic processing. This implies that perceptual organization processing overlaps with lexical processing, indicating that both mechanisms work together to choose lexical percepts.

Cue Weighting

This dissertation attempts to explicate the architectural structure of the auditory system: specifically, the relationship of organizational and memory mechanisms during spoken language perception. This phrasing of the research question leads to the assumption that this relationship would work in the same manner for most listeners, but the large variation in categorization responses suggests that this might not be the case. The stimuli in this experiment are multi-dimensional: they have multiple auditory cues tied to them, both indicating the content of the speech (sub-lexical and lexical cues in the sentence) and how the complex scene should be organized (grouping strength and spatial separation of the s-stream from the sentence). Researchers have shown that when categorizing complex or ambiguous speech sounds, listeners differ in the attention paid to specific cues (Holt & Lotto, 2006; Repp, Liberman, Eccardt, & Pesetsky, 1978; Mattys, White, & Melhorn, 2005; Sanders & Nevillle, 2000). Thus, it is possible that listeners in this experiment relied on some cues more than others, either paying more attention to the lexical bias (linguistic cue) or the grouping strength (perceptual organization cue) to influence their categorizations.

Figure 4 shows vast individual differences in response by participant, such that variability among participants accounted for 72% of the variance in the statistical model, which is more than double the variability explained by items (27%). The large variation among participants indicates that participants' experience with the stimuli might not be uniform; instead, the variation in perception may indicate categorically different cue weighting strategies or habits. Some listeners could preferentially use lexical information rather than grouping information to identify the word, while others rely more heavily on grouping cues when responding. This should be reflected in response pattern differences across participants. To investigate this

possibility, I need to know if participants are similar in their responses across conditions: if not, then the variation is likely due to factors outside experimental control and are possibly meaningless.

I first visually inspected individual data patterns for each participant, to investigate whether any obvious similarities or groupings existed between participants' data. When examining participant's individual effects of lexical bias (memory effect) and grouping strength (grouping effect), I observed three distinct patterns among participants. Most participants showed a larger memory effect than grouping effect, some showed equivalent effects, and very few participants showed a larger grouping than memory effect. To further examine and quantify this pattern, I generated effect sizes for both the memory effect (averaged voiced response for voiced – voiceless lexical bias conditions) and grouping effect (averaged voiced response for strong – weak grouping conditions). Then I comparatively graphed these effect sizes for each participant in a scatterplot (Figure 5) to see the distribution, with memory effect sizes on the y-axis and grouping effect sizes on the x-axis. Participants in the upper left corner of the graph have a large memory effect and small grouping effect, whereas participants in the bottom right have a large grouping effect and small memory effect. Participants in the center, along the black diagonal line, have roughly equivalent memory and grouping effects.

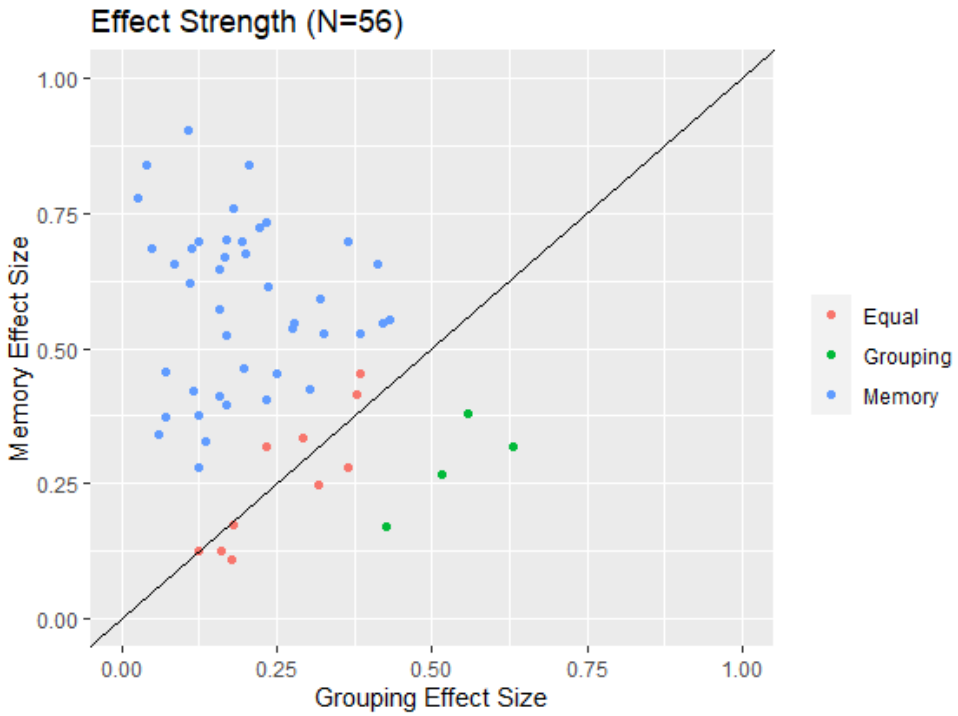

Figure 5: *Scatter plot of participant effect size by cue (memory on the y-axis and grouping on the x-axis). Each participant is represented by a dot. The colors refer to participant groupings, given based on the comparative influence of the two effects.*

As can be seen in Figure 5, memory effect sizes range from .11 to .90 (y-axis) and grouping effect sizes range from .05 to .62 (x-axis). Most participants (n=41, 71%) show a pronounced memory effect compared to their grouping effect (top left of graph, blue dots). Then a smaller group of participants (n=11, 19%) show an equivalent influence for the two cues, defined by having less than .2 difference (.1 in either direction) in effect sizes (red dots). There was also a minority of participants (n=5, 9%) who had a larger grouping influence than memory (lower right, green). These results suggest that most participants primarily relied upon their linguistic memory (specifically, lexical knowledge) more than organizational cues (indicated by grouping strength) to inform their categorization decision. But there is a contingent of participants that disregarded lexical memory to some degree, in favor of organizational cues.

Most participants who relied heavily on lexical knowledge also ignored grouping cues when categorizing the phonemes. For example, the participants who showed the largest memory effects (>.75) also had the smallest grouping effects (<.25). In contrast, those participants who relied heavily on grouping cues (>.5), also showed smaller memory effects (<.35). In fact, though there is a group of four participants who showed weak reliance on either cue (<.25 for both effects, bottom-left of graph), no participant demonstrated strong reliance on both cues (top-right of graph). This negative correlation (r= -0.25) is statistically significant (t(54)= -1.87, p<.034), indicating a potential trade-off between participants' reliance on memory versus organization information when perceiving speech. Individuals who weight memory cues more also weight grouping cues less for perception. This implies that there is a finite pool of resources to integrate memory and grouping cues, and that relying on one cue for word perception (lexical processing) allows for less resources available to use the other cue. For most (but not all) listeners, lexical memory cues outweighed organizational heuristics when categorizing the phoneme.

It is important to note that all groups were affected to some degree by both cues, though a few participants in the memory group (blue) were able to largely ignore grouping cues (blue dots to the extreme left on the x-axis in Figure 5: grouping effect less than .10). Given that all participants were self-reported native speakers of English, differences in relative effect strength should not be due to differences in linguistic familiarity. All included participants were able to answer the majority (>75%) of foil trials correctly (clear phoneme identity with opposing grouping strength), which shows they were not using a conscious strategy to categorize based solely on the properties of the s-stream or the lexical bias. Rather, participants show a trade-off

in their memory and organizational cue influence on speech perception, specifically phoneme categorization.

Response Time Data

In addition to categorization data, time taken to respond can provide additional information about the interaction of the two mechanisms, for example the processing time or mental difficulty in completing the task. During the trials, the organizational cue of grouping strength is building in prominence (as the s-stream is consistently presented during the sentence presentation), whereas the lexical cue is only present at the end of the trial. In fact, listeners are prompted to categorize the initial stop as soon as they heard the relevant speech sound, before the relevant lexical information has been presented (the lexicality of "bun" vs "bonge" depends on the phonemes after the [b]). Investigating the timing of responses by response and condition can inform the proposed time-course of lexical and grouping processes. In addition, response time has been shown to correlate with task difficulty, such that responses that take longer occur in difficult trials, whereas easy trials evoke faster responses (Luce, 1986). In Figure 6, the mean time to respond (y-axis) for each lexical (x-axis) and grouping (color) condition is shown, taken from the target phoneme onset time, and split by response categorization (voiced "b" responses on the left graph and voiceless "p" responses on the right graph). By splitting the response times according to participant response (either voiced or voiceless), we can see the time cost of conflicting (and time boost of concurring) grouping and lexical cues on the formation of the word percept.

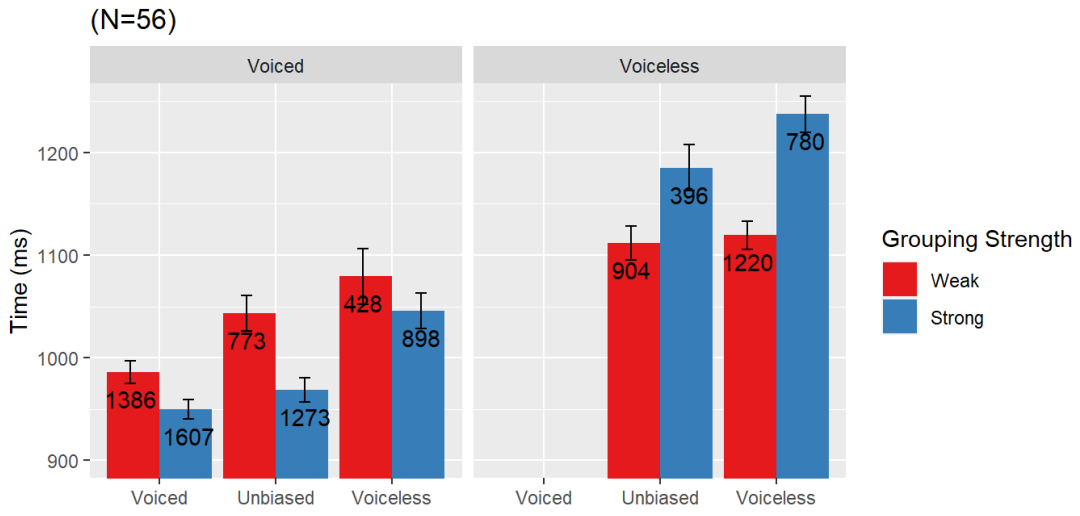

Figure 6: *Bar graph of time taken to respond (y-axis) averaged over participants and separated by sentential bias (x-axis) and grouping strength (color). The graphs are divided by response type: voiced ("b/d") on the Left and voiceless ("p/t") on the Right. Error bars are standard error from the mean and number correspond to number of observations.*

If lexical bias and grouping strength both affected response times, the quickest responses should those where the lexical bias and grouping strength conditions both signal the same response (e.g., a voiced response to the voiced-strong condition, and a voiceless response to the voiceless-weak condition). And in conjunction, the slowest responses should be for conditions where the response opposes the lexical and grouping cues (e.g., a voiced response to voiceless-weak condition, and a voiceless response to a voiced-strong condition). This is only true for the voiced responses, whereas the voiceless responses do not show the expected pattern. The time to respond seemed to depend most strongly on what the response was, rather than the lexical bias: voiced responses occurred 100-150ms faster than voiceless responses across all three lexical conditions. This data pattern implies that participants were quicker to categorize the stop phonemes as voiced throughout the experiment, regardless of experimental manipulations. Logistic modelling indicated that this data pattern was not due to any word frequency bias in the stimuli (including word frequency as a covariate in the model did not improve fit, $X^2(16)=0.96$,

p=.33). This suggests that the speeded response times for voiced responses is due to some other factor.

To assess the reliability of these results, I employed linear mixed modelling on the reaction time data with response, grouping strength, and lexical bias as fixed effects and participant and item as random effects. Any trials with responses that occurred before or within 150ms of the target onset were removed from analysis (3.5%). Comparison of the null, main effect, and interaction models indicated that the interaction model, specified as "responseTime ~ grouping * bias * response + (1|pcode) + (1|item)", significantly improved the fit over the main effect (X^2(11)=86.19, p<.001) and null (X^2(14)=222.73, p<.001) models. The fixed effects support the descriptive results: the strongest predictor of reaction time, categorization response, estimated longer reaction times for voiceless responses than voiced responses, in four of the six comparisons (weak.voiced 125ms, Z=4.56, p<.001; strong.voiced 267ms, Z=5.32, p<.001; strong.unbiased 237ms, Z=10, p<.001; strong.voiceless 195ms, Z=9.61, p<.001).

Like the categorization data, the response time data show an effect of both grouping strength and lexical bias manipulations. If the lexical bias manipulation influenced the response time, responses should be fastest in conditions where the participants' response agree with the lexical bias (voiced responses to voiced biased items: "boy"), and slowest in conditions where the participants' response oppose the lexical bias (voiced responses to voiceless biased items: "spoy"). Instead, the results were the same regardless of response, with faster responses to voiced items ("boy") than voiceless ("sponge") items (by 150ms). This is the case even when the participant response and the lexical bias were both voiceless: responding "p" to "sponge" took longer than responding "p" to "spoy". The fixed effects support the descriptive results, as the lexical bias manipulation did not cause a significant change in response time for any conditions.

This result is unexpected and implies that the lexical bias manipulation did not influence response time as much as it did categorization responses.

What we should see if the grouping strength manipulation influenced response time is a consistent difference in response time over grouping strength conditions, such that responses that agree with grouping strength (voiced responses to strong s-streams) occur faster than those that oppose grouping strength (voiced responses to weak s-streams). This is shown in both response graphs: responding voiced ("b/d") to weak grouping items took longer (25-80ms) than responding voiced to strong grouping items. And in concert, responding voiceless ("p/t") to weak grouping items was faster (50-100ms) than responding voiceless to strong grouping items. The fixed effects support the descriptive results: the change in grouping strength from strong to weak led to reliable reaction time differences depending on lexical bias and categorization response (VLbias.VLresp 118ms, Z=6.35, p<.001; Unbias.VLresp 91ms, Z=3.76, p<.004; Unbias.Vresp -82ms, Z=-4.47, p<.001). Decision time was impacted prominently and consistently by grouping strength cues: responses that agreed with the grouping cues were made faster than those in violation of the grouping cues. This implies that while lexical knowledge had a stronger impact on word formation (for most participants), grouping strength had a stronger impact on the ease of that decision.

Another interpretation of the grouping effect is that it conveys the organizational ambiguity of the entire scene. Responses are fastest when there is less intrusion from the competing s-stream into the speech stream. For voiced responses, the strong s-stream clearly conveys acoustic cues that the [s] should be segregated (excluded from) the speech stream, giving an unambiguous preferred organization of the scene: two separate streams with the target [s] embedded in and belonging to the s-stream only. For voiceless responses, the weak s-stream

conveys that the [s] likely does not belong with the s-stream, and therefore can be integrated into the speech stream. This organization is more ambiguous, because, though the weak stream does not convey strong evidence that the s-stream is a separate stream, it could still plausibly be organized that way, which may explain why all voiceless responses took longer to make than voiced responses. Responses that oppose the grouping cues took more time to process, indicating that grouping heuristics of acoustic similarity and repetition were used to infer the 'correct' organization of the target [s] within the two streams.

Conclusion

In summary, this experiment showed that both lexical knowledge and perceptual organization can influence word formation, indicating a somewhat parallel time-course of organization and lexical processing. Further, the equivalent effects of both lexical and grouping strength conditions implies that both manipulations had an independent effect on voiced responses: that neither lexical knowledge nor grouping strength cues completely outweighed the other for phoneme categorization. This result implies the presence of a connection between auditory grouping and linguistic memory when perceiving speech, specifically located at the lexical processing stage. The reaction time data showed that grouping strength had a consistent and predictable effect on processing speed, such that responses that agreed with the grouping strength manipulation (voiceless and weak) were faster than those that opposed the manipulation (voiceless and strong). This makes sense, as perceptual organization primarily serves as an automatic process that groups and separates sound in auditory scenes: responses that agree with the preferred/cued organization should be faster than those in opposition. Interestingly, participants varied in the degree to which they weighted each cue, demonstrating a trade-off between lexical and grouping cue weights, with lexical bias outweighing grouping strength cues

for most participants. This pattern indicates that the integration of memory and perceptual organization cues could be expressed or experienced differently among listeners. Taken together, results imply that that lexical knowledge had more influence on word formation (for most participants), whereas grouping strength had more influence on the ease of that decision. In Experiment 2, I will continue to explore the auditory system's integration of perceptual organization and memory-based cues for speech perception by using sentence context biases to replicate and extend the findings.

Experiment 2: Sentence Context

Introduction

Both behavioral and neuro-imaging studies have demonstrated the hierarchical nature of spoken language processing. Fmri and MEG studies have compared neural responses to acoustic stimuli at various linguistic levels, with pre-lexical stimuli activating neurons in regions responsible for basic auditory processing (de Heer et al., 2017; Humphries et al., 2014), and lexical/sentential processing activating more distributed cortical regions (Davis & Johnsrude, 2003; Scott & Johnsrude, 2003; Sheng et al., 2019). These neural studies conclude that linguistic processing of speech is hierarchical, occurring across many timescales in a semi-parallel manner. Behavioral studies support these assertions, with researchers comparing acoustic, lexical and sentential cues during both word segmentation (Kim, Stevens, & Pitt, 2012; Mattys & Melhorn, 2007; Sanders & Neville, 2000) and auditory word recognition (Connine, 1987; Connine, Blasko, & Hall, 1991; Connine, Blasko, & Wang, 1994). In these studies, listeners preferentially used the highest level of linguistic cue available for speech understanding, such that sentential context overpowered lexical bias, which in turn overpowered pre-lexical cues. These behavioral and neural studies imply that sentential information, processed on a higher level of the linguistic hierarchy, could interact with perceptual grouping differently than lexical information did in Experiment 1.

To further examine the auditory system's integration of linguistic memory and perceptual organization cues when perceiving spoken language, I examined the effects of sentential context bias (using past words to predict future words in an utterance) and perceptual grouping (using basic acoustic heuristics to organize the auditory scene) when organizing and forming words. The design of this experiment is almost identical to Experiment 1, with the following changes listed. In each trial, participants listened to a scene with a semantically predictive sentence ("Sam muzzled his dog so it wouldn't...") ending in a word (lexical both with and without the [s]:

"spark"-"bark"). A sentential bias was applied to the pre-target words in the sentence: the sentence either predicted a word starting in [s] (voiceless bias: "Putting new wood on the fire created that spark"), predicted a word starting without [s] (voiced bias: "Sam muzzled his dog so it wouldn't bark"), or did not predict a word (unbiased: "The next item you hear is…"). An effect of sentential context would be indicated by having more voiced responses to voiced biased sentences ("Sam muzzled his dog so that it would not bark") than voiceless biased sentences ("Putting new wood on the fire created that bark").

The predictions are similar to Experiment 1: if responses are made based on sentential context rather than grouping strength in conflicting conditions (i.e., if biased conditions show no effect of grouping strength), this will indicate the lack of an architectural connection between perceptual organization and linguistic processing at the level of sentential context. If an architectural juncture exists between perceptual organization and sentential linguistic memory when processing speech, we should see some effect of both grouping and sentential cues. If the effects are equivalent, such that neither cue overpowers the other in conflicting conditions, then this would indicate that grouping and sentential information are integrated in an independent manner, like in Experiment 1. An interaction between grouping and sentential effects would indicate the presence of an interactive juncture connecting the two processes. Comparing the effect of grouping and linguistic manipulations on phoneme categorization can tell us more about how linguistic memory and perceptual organization work together to facilitate word perception and formation.

Method

Participants

A separate group of 60 Ohio State University students participated in this experiment. Four participants were excluded using the same criteria as in Experiment 1 (poor foil performance, missing target responses, or fast response times), leaving 56 participants' results.

Stimuli

The purpose of the scene setup was the same as Experiment 1, to create ambiguity in the auditory organization upon which linguistic and organizational cues could act and was achieved in the same way as Experiment 1. In this experiment, the lexical status of the target word was kept constant: both versions of the targets were words (like the Experiment 1 unbiased condition: "spay"-"bay"). The linguistic cue examined in this experiment was the buildup of sentence context, or the multi-word expectation for predicting the final word in a sentence (sentential bias manipulation). The grouping manipulation was the same as Experiment 1: an s-stream with either strong or weak grouping cues. The structure and setup of the trials was the same as in Experiment 1, but the content of the linguistic stream was changed (see Table 2 for conditions and examples).

For each target word, a pair of sentences was created, one with sentence context predicting the integrated word ("Putting new wood on the fire created that spark") and another with context predicting the segregated word ("Sam muzzled his dog so that it would not bark"). Forty-two target words were used for each condition, 21 starting with /sp/ and 21 starting with /st/. For each target word, a pair of sentences was created (84 sentences total), with one context predicting the voiced version ("bark") and the other predicting the voiceless version ("spark"). Each pair of sentences was matched by number of syllables, ranging from 9-13 syllables

(*M*=1828ms, *SD*=263ms). Refer to the target sentences in Appendix B for reference. The process of recording the stimuli was almost identical to Experiment 1, with one change: the sentences and the target words were recorded together ("Sam muzzled his dog so that it would not bark"), to enhance the naturalness of the sentence recordings.

Weak Grouping			
Voiced Bias	L:	Sam muzzled his dog so it would not _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Unbiased	L:	The next item you will hear is _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Voiceless Bias	L:	Putting new wood on the fire created that _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Strong Grouping			
Voiced Bias	L:	Sam muzzled his dog so it would not _**bark**	
	R:	[s][s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)
Unbiased	L:	The next item you will hear is _**bark**	
	R:	[s][s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)
Voiceless Bias	L:	Putting new wood on the fire created that _**bark**	
	R:	[s][s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)

Table 2: *Conditions and examples for Experiment 2: Sentential Bias, with context conditions and grouping strength across rows. The bold words indicate the target items.*

To provide continuity with Experiment 1, an unbiased condition was created, which used a single non-predictive sentence frame preceding the target word ("The next item you will hear is..."). This condition had trials created in the same way as in Experiment 1, appending an isolated target word recording ("spark") to a sentence frame (see Table 2 for an example). The process of creating the opposing s-stream was the same as in Experiment 1: the number of [s]s and repetition rate of the s-stream was the same as Experiment 1.

Foil sentences were created to ensure that participants were categorizing the stop phonemes based on how they sounded, rather than relying only on the number of [s]s in the s-stream or only the sentence context in the speech stream. 60 foil sentences were created by pairing one of 30 target words with a non-predictive but plausible sentence context (see

Appendix B for the foil list). The full sentences (a voiced and voiceless version for each of the 30 targets) were recorded by the same female Native-English speaker as in Experiment 1. To ensure the clarity of the stop and naturalness of the foil trials, the target word was recorded with the sentence, not spliced in after the recording. Similar to the target trials, an s-stream was added to the sentence, but unlike the target trials, the s-stream was temporally misaligned, such that the target [s] overlapped with the stop. This was supposed to prevent the [s] from being integrated into the target word, ensuring that participants would report back the stop, uninfluenced by cross-ear integration. The foil s-stream had slightly different properties from both the strong and weak conditions: there were three [s]s total, each separated by 500ms of silence. By adding these foil trials, I can identify and exclude participants who are using conscious response strategies instead of performing the task as instructed.

Pilot Experiments

Two pilot studies were conducted to confirm that the sentence contexts were creating the intended expectations for the target word. In the first pilot (N=7), participants read the sentence text presented visually on the screen (When a pipe is clogged the water stays in the _____) and chose which of the two targets fit it better (strain or drain). Both context conditions indicated the expected response patterns: voiced bias sentences (context predicts "drain") elicited 99% voiced target responses and voiceless bias sentences (context predicts "strain") elicited 1% voiced target responses. This pattern was identical for all seven respondents. This demonstrates that the sentence contexts were eliciting the intended predictions (biasing one target word over another).

In the second pilot (N=11), participants were visually presented with a completed sentence (When a pipe is clogged the water stays in the drain) and rated how well the last word fit into the sentence on a scale from 1 ("no fit") to 7 ("best fit"). For each sentence context, there

were two sentence variants, one presented with the voiced target (When a pipe is clogged the water stays in the drain) and one with the voiceless target (When a pipe is clogged the water stays in the strain). The version matching the intended target (e.g., voiced context and voiced target) should have a higher fit rating than the sentence with the non-matched target. Almost all participants (10 of 11= 90.9%) exhibited the correct pattern across all items, indicating that the sentence contexts fit better with the target (M=6.1, SD=.31) than its alternative (M=2.1, SD=.36). Any sentences with too similar fit ratings for both versions (within one fit rating of another) were revised (5 sentences total).

Procedure

The procedure was almost identical to Experiment 1, with the differences listed as follows. The stimulus lists did not repeat items within the list, due to the memorability of the sentence stimuli (to minimize learning effects). Instead, each list included 21 target words per condition (126 total), with each target word having a unique sentence for all three sentential bias conditions, and all three sentences paired with one of the three grouping conditions. Participants went through 196 trials (10 practice, 126 targets, and 60 foils), with self-paced breaks introduced every 31 trials, so that the experiment took about 15-20 minutes on average to complete. For this experiment, in addition to showing feedback after the trial when responses were too slow, feedback was also given if the participant responded before the target word onset ("Wait for the last word to respond!" in red for 1000ms).

Results & Discussion

Exclusion and Analysis

Participants who failed to respond to more than 25% of the target trials were excluded from the study (n=2). Any participant below 75% correct on foil trials was excluded from the study (n=2).

Any trials that had a response time faster than 150ms (from the onset of the target word) were excluded (2%.). To analyze the results, I used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical software (v4.1.2; R Core Team 2021) to employ logistic mixed modelling on the voiced responses with grouping strength and sentential bias as fixed effects and participant and item as random effects. Comparison of the null ("response ~ (bias|pcode) + (grouping|item)"), main effect ("response ~ bias + grouping + (bias|pcode) + (grouping|item)"), and interaction ("response ~ bias * grouping + bias + grouping + (bias|pcode) + (grouping|item)") models was performed using a likelihood ratio test to determine which relationship fit the dataset best.

Categorization Data

The proportion of voiced responses ("d" or "b") was averaged over sentential bias and grouping strength conditions for each participant (Figure 7). Data are graphed like Experiment 1, except with sentential bias on the x-axis. Based on the biasing conditions, we should see a couple patterns in the data: an effect of sentential bias (voiced > unbiased > voiceless) and an effect of grouping strength (strong > weak). Most importantly, the presence or absence of an interaction between the two manipulations will demonstrate how the perceptual organization and linguistic memory mechanisms interface. Finding an interaction between these effects of the organization and linguistic manipulations, specifically such that sentential bias outweighs grouping strength, would reinforce the proposed hierarchical structure behind auditory speech perception, that the system integrates cues from both sources in an ordered fashion (perceptual organization then sentential memory). However, if a consistent effect of both manipulations is found, like in Experiment 1, this implies that the auditory system is less strictly hierarchical than previously

assumed, with interactive junctures between perceptual organization and linguistic memory occurring at multiple levels of linguistic processing.

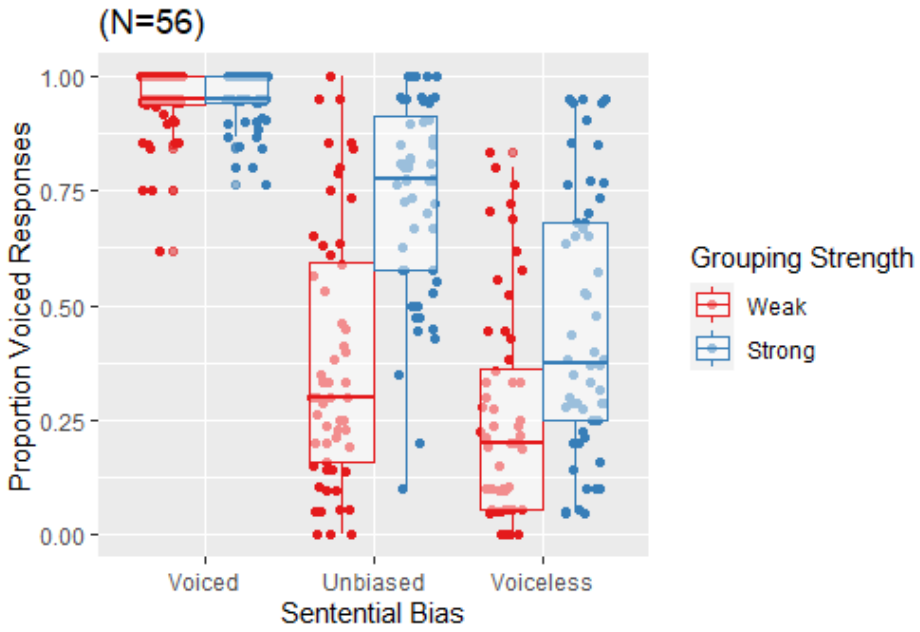

Figure 7: *Boxplots of the proportion of voiced responses by sentential bias and grouping strength. Sentence bias conditions are shown on the x-axis (voiced, unbiased, voiceless), and grouping strength is distinguished by color (weak: red, strong: Blue). Each dot represents a participant's average.*

Like in Experiment 1, the linguistic manipulation affected participant phoneme classification, with voiced biased sentences eliciting the highest proportion of voiced responses (M=.95, SE=.007) on average, followed by unbiased (M=.56, SE=.013), with voiceless biased sentences eliciting the lowest proportion of voiced responses (M=.35, SE=.14). However, unlike Experiment 1, the effect of the grouping strength manipulation differed based on the sentential bias, ranging from .01 to .48, indicating a potential interaction between grouping strength and sentential bias. Indeed, logistic mixed modelling indicated that the interaction model significantly improved the fit over both the main effect model (X^2(16)=81.42, p<.001) and the null model (X^2(10)=608.48, p<.001), using the same procedure as in Experiment 1. Paired

contrasts indicated that a quadratic function fit the difference between grouping conditions best (strong-weak=3.58, Z=7.65, p<.001), indicating that the effect of grouping strength differed by each sentential bias condition, peaking at the unbiased context condition. With an unbiased sentence, the log-odds of the strong grouping strength being categorized as voiced are 2.38 units higher than the weak grouping strength (Z=18.21, p<.001). With a voiceless context, the log odds of strong grouping being categorized as voiced are 1.46 units higher than weak grouping (Z=11.48, p<.001). The voiced contrast was smallest and nonsignificant, with the log odds of strong grouping being categorized as voiced only 0.13 units higher than weak grouping (Z=0.6 p=.97). This means that the effect of the grouping manipulation was strongest when presented without any sentential cues, and significantly smaller with sentential bias (but not nullified completely). When sentential cues are present in the speech, the auditory system preferences those cues in determining the percept. This makes sense, as sentential context is highly relevant for speech perception.

This experiment has implications for the underlying architecture of human language comprehension, demonstrating that perceptual organization likely has a weak interface with sentential memory during speech processing. This result partially supports the hierarchical conception of auditory processing discussed in the Experiment 2 introduction. As in word segmentation (Kim, Stevens, & Pitt, 2012; Mattys & Melhorn, 2007; Sanders & Neville, 2000) and auditory word recognition (Connine, 1987; Connine, Blasko, & Hall, 1991; Connine, Blasko, & Wang, 1994) studies, listeners preferentially used the highest-level cues available to respond. For this experiment, linguistic memory, specifically sentential context cues, was relied upon for phoneme identification, to the detriment of perceptual organization cues. Notably, when sentential context cues were absent from the stimuli (unbiased condition), listeners relied upon

grouping strength to complete the task. This gives evidence for an interactive juncture between sentential linguistic processing and perceptual organization.

Direct comparison across experiments can strengthen the claim that perceptual organization processes interact with linguistic memory across linguistic hierarchy of processing (Experiment 1: Lexical, Experiment 2: Sentential). In Experiment 1, the grouping strength effect was more consistent across linguistic conditions (voiced .12, unbiased .30, voiceless .28), with all lexical conditions demonstrating a significant difference between strong and weak grouping conditions. In contrast, Experiment 2 demonstrates a varying effect of grouping strength cues: largest for the unbiased (.38) condition, and absent (voiced, .01) or halved (voiceless, .21) with conflicting sentential context. The difference between lexically unbiased and voiceless items in Experiment 1 was much smaller compared to Experiment 2 (.02 and .17, respectively), and the grouping strength effect for the voiced lexical condition was much larger than in Experiment 2 (.12 and .01, respectively). This indicates that sentence context partially constrained organization of the speech in a way that lexical bias failed to do. These experiments, taken together imply an integration of linguistic memory and perceptual organization that differs across the linguistic processing hierarchy: perceptual organization interacts strongly with lexical processing, but weakly with sentential processing.

Cue Weighting

The categorization responses indicated that participant differences explained 64% of the variance in the model, compared to the 36% explained variance from item differences. This is roughly the same pattern as in Experiment 1 (participants responsible for more variance than items), though unsurprisingly, full sentences in this experiment caused more item variance than the frame and word combinations from Experiment 1. To further investigate these participant differences in cue

weighting, I graphed these effects in a scatterplot (Figure 8), using the same procedure as in Experiment 1. I generated effect sizes for both the memory effect (averaged voiced response for voiced – voiceless bias conditions) and grouping effect (averaged voiced response for strong – weak grouping conditions).

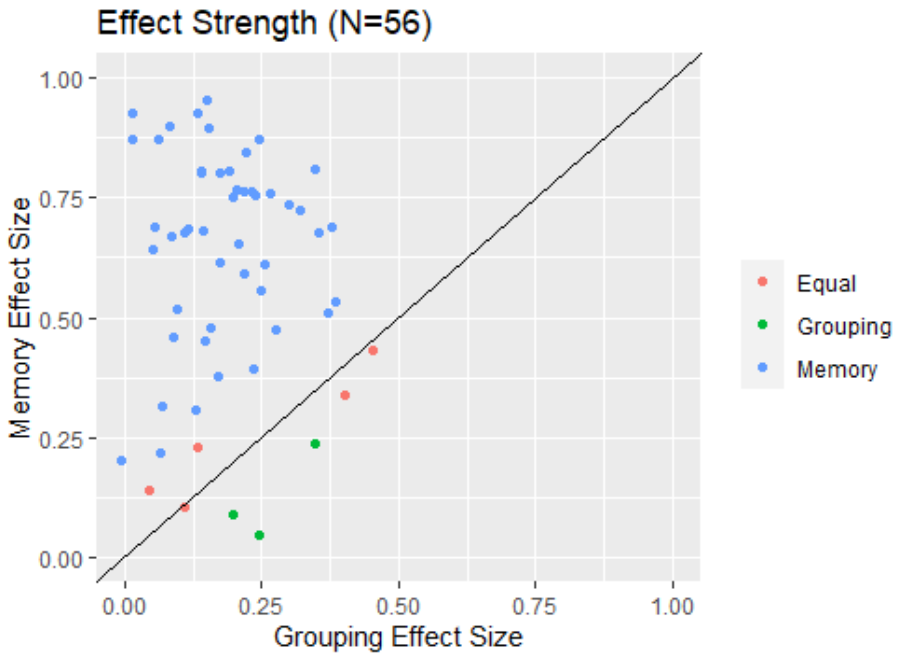

Figure 8: *Scatter plot of participant effect size by cue (sentential bias on the y-axis and grouping strength on the x-axis). Each participant is represented by a dot and a label. The colors refer to participant groupings based on the comparative size of the two effects.*

Interestingly, when the grouping strength effect size is averaged over all the sentential bias conditions, the results are vastly different from Experiment 1. In Experiment 1, there was a trade-off in cue strength, such that participants with a larger lexical memory effect also had a smaller grouping effect, and those with a small lexical memory effect had a larger grouping effect. Figure 8 shows that the effect of grouping strength in this experiment is miniscule for all participants, such that no participant has a grouping effect size larger than .5, whereas the memory effect is just as strong as in Experiment 1, with a couple of participants even reaching an

effect size of .97. Almost all participants (n=48, 86%) show a pronounced memory effect compared to their grouping effect (compared to 70% in Exp1). The minority of participants had an equivalent influence for the two cues (n=5, 9%), or a larger grouping influence than memory (n=3, 5%). This difference in cue influence was not due to an increased grouping effect size relative to other participants (like in Experiment 1), but rather a decrease in their memory effect size. These results suggest that most participants primarily relied upon their linguistic memory (specifically, sentential context) more than organizational cues to inform their categorization decision.

As categorization response data indicated an interaction between grouping strength and sentential bias in this experiment, the greatest influence of grouping strength should occur at the unbiased condition (without sentential context). To investigate whether this lack of grouping effect exhibited in Figure 8 is due to any change in the effectiveness of the grouping manipulation, I measured the grouping effect at the unbiased condition, instead of averaging the grouping effect over all sentential bias conditions (see Figure 9, below).

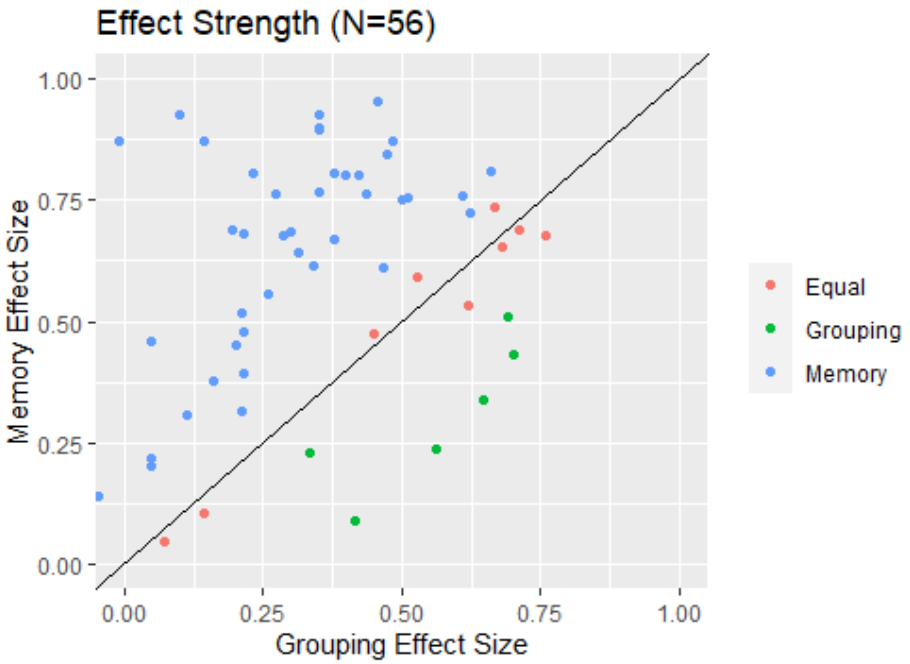

Figure 9: *The same as Figure 8, except that the grouping strength effect is taken from the unbiased sentence condition.*

The results change drastically when the grouping effect size is measured without the conflicting cue of sentential context. The memory effect size remains the same (.05 to .95), but the grouping strength effect size increases (0 to .80). The data now show a positive correlation ($r=.28$), such that memory and grouping effect sizes increase together ($t(55)=2.18$, $p<.017$). Compared to Figure 8, a smaller majority of participants (N=41, 73%) show a pronounced memory effect, whereas a slightly larger group of participants show an equivalent effect (n=9, 16%) or a larger grouping effect (n=6, 11%). In fact, the participant groupings are almost identical to those in Experiment 1 (Memory=73%, Equal=18%, Grouping=9%) when the grouping effect is measured this way. This shows that even though the grouping strength manipulation was still a strong indicator of categorization response when presented in isolation, sentential memory cues overpowered the grouping cues when in conflict.

The large grouping effects seen in Figure 9 are due to the ambiguity of the sentence context at the points where grouping cue effect size was measured. These effects are greatly diminished when opposing sentence context is present in the trial (Figure 8). This gives more evidence that sentence context is a more influential cue when pitted against perceptual organization, more so than lexical knowledge. In fact, the effect of lexical bias on categorization in Experiment 1 was equivalent to the effect of an unbiased sentence in Experiment 2. Since the lexically biased words were embedded in context-absent frame sentences ("The next item you will hear is…"), this demonstrates that lexical knowledge had less impact on word perception than sentence context. This makes sense, as the lexical bias was applied to the speech stream after the target began (a word initial target) and had no chance to build over the course of the trial, unlike sentential bias in Experiment 2.

Response Time Data

In Figure 10, the average time to respond for each condition is shown, taken from the target onset time, and split by categorization (voiced "b" responses on the left and voiceless "p" responses on the right). Any conditions with fewer than 100 responses were not included in the analysis (Voiceless responses for Voiced sentences). By splitting the response time data according to participant response (either voiced "b/d" or voiceless "p/t"), we can see the time cost of conflicting (and time boost of concurring) grouping and context conditions on the formation of the word percept. Like in Experiment 1, comparison of the null, main effect, and interaction models using a likelihood ratio test indicated that the interaction model significantly improved the fit over the main effect model ($X^2(18)=119.08$, $p<.001$) and the null model ($X^2(14)=46.5$, $p<.001$).

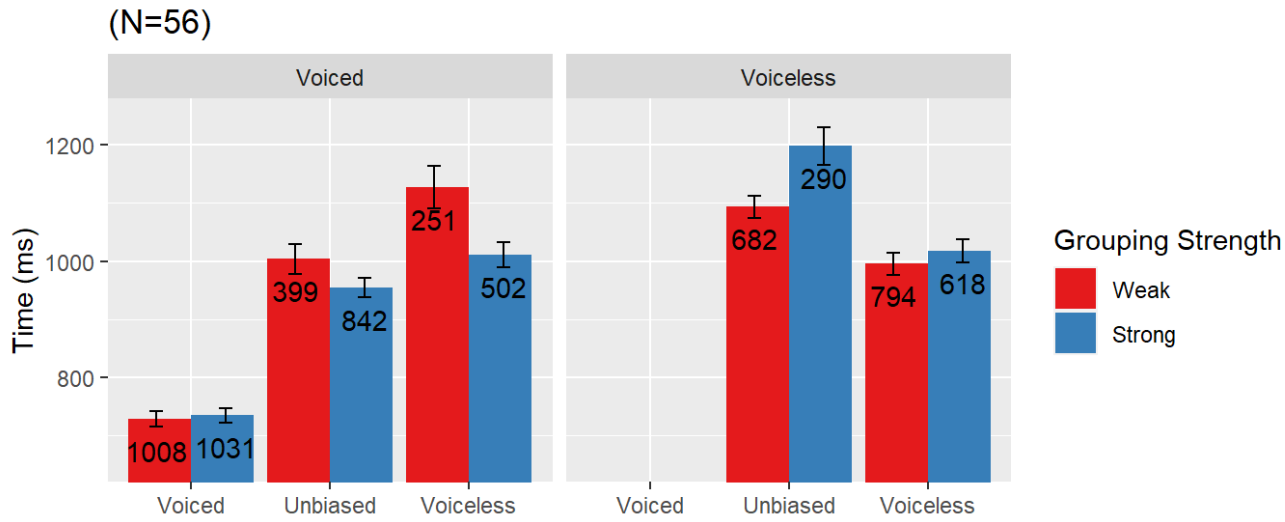

Figure 10: *Bar graph of time taken to respond (y-axis) averaged over participants and separated by sentential bias (x-axis) and grouping strength (color). The graphs are divided by response type: voiced ("b/d") on the Left and voiceless ("p/t") on the Right. Error bars are standard error from the mean and number correspond to number of observations.*

What we should see, if sentential bias is dominating the decision speed, is an inverted "u-shape" across the two graphs, such that its quickest to make a voiced response to the voiced context items (congruent) than to the voiceless context items (incongruent), and vice versa for voiceless responses. Figure 10 partially shows this, as an asymmetrical pattern: voiced responses fit the predicted pattern, and voiceless responses only partially fit the pattern. For voiced responses, there is an increase in response time by sentence context, such that responding voiced to voiced contexts occurs 300ms faster than unbiased sentences, and the same response to unbiased sentences are also faster than voiceless contexts (100ms). Fixed comparisons support this result, as voiced responses were estimated to occur on average 285ms faster for voiced than unbiased sentences (Z=2.93, p<.004), with no reliable difference between unbiased and voiceless sentences (20ms, Z=0.35, p=.73). This shows that only the voiced responses are facilitated by congruency of response and condition. For voiceless responses, participant response time is less varied by context condition: responses to the unbiased sentences took longer (100ms) than

responses to the voiceless sentences. For voiceless responses only, the lack of sentence context caused longer response delays than conflicting cues.

If perceptual organization is dominating the decision speed, we should see a consistent difference in response times by grouping strength: voiced responses should occur faster to strong s-streams and voiceless responses should occur fastest to weak s-streams. We partially see this pattern in Figure 10. Responding voiced to targets embedded in weak s-streams took longer (50-100ms) than responding voiced to targets embedded in strong s-streams, in the unbiased and voiced conditions. Indeed, responding voiceless to targets in weak s-streams was faster (100ms) than responding to targets in strong s-streams, but only for the unbiased condition. Fixed effects show that responses were 264ms faster on average between strong and weak items ($Z=2.71$, $p<.007$). This corroborates the categorization data, as the largest effect of grouping strength was when sentence context was neutral. Decision time seemed to be impacted less consistently by grouping strength cues in this experiment than in Experiment 1: responses that agreed with the grouping cues were made faster than those in violation of the grouping cues, but only in some of the conditions (unbiased, voiced and voiceless response).

Interpretation of the reaction time data is less clear in this experiment. Response time patterns differed by participant response in unexpected ways. Voiced responses to voiced biased items gave a time boost of 300ms, whereas the same effect for voiceless responses to voiceless context was not found. The grouping strength effect also differed based on participant response: for voiced responses, the strong s-stream benefit was largest for the conflicting sentence context (voiceless), whereas the strongest benefit for voiceless responses was for sentences without biasing context.

Conclusion

In summary, the interaction between the grouping strength and sentential bias conditions implies that the auditory system likely has a strong connective junction between sentential memory and perceptual organization processing: sentence context processing and perceptual organization are co-occurring and influencing each other. Cue weighting data found that individual differences in participant performance showed uniform reliance on sentence context, more than grouping strength, except in the unbiased condition. This pattern indicates that unlike with lexical knowledge (Experiment 1), listeners rely on sentence context cues more than organization cues when forming and perceiving words. The response time data was different based on the participant's categorization of the stop, but overall, the effect of grouping strength on processing speed was less consistent than in Experiment 1. Taken together, results imply that that linguistic memory across the linguistic hierarchy interacts with perceptual organization processes, though higher-level linguistic information, specifically context-based predictions, is preferentially integrated into the percept. In Experiment 3, I will continue to explore the same perceptual organization and memory-based interaction for speech perception by adding a medium grouping strength condition to replicate and extend the findings.

Experiment 3: Gradient Organization

Introduction

The purpose of Experiment 3 is to replicate Experiment 2, and to provide stronger evidence that perceptual organization can influence word perception in various sentence contexts, by demonstrating the gradient nature of the effect. This experiment is almost identical to Experiment 2, with one major change: the grouping strength manipulation includes a third s-stream condition. All context conditions remained the same, but a third, medium strength grouping condition was created, to allow for gradient effects of perceptual organization to surface. This condition imposes less organizational force than the strong condition and more force than the weak condition, by having seven pre-target [s]s and one post-target [s], separated by 200ms of silence (see Table 3). Including the medium strength condition allows the grouping strength effect to be compared across three levels, like the sentential effect. We should see a gradual increase in voiced responses as s-stream strength increases (weak < medium < strong), demonstrating how word formation can be incrementally affected by perceptual organization.

In addition to strengthening the argument for perceptual organization influencing word perception, this experiment should replicate the main finding of Experiment 2: an interaction between the grouping strength and sentential bias manipulations. This would strengthen the conclusions made in Experiment 2, that memory and organization cues both influence word formation and perception, to different degrees. We should find that the preceding sentence context has a larger effect on the final percept than the grouping manipulation (when in conflict), again demonstrating that sentential cues are given more weight than perceptual organization cues during speech processing, somewhat insulating incoming speech against being affected by organizational forces. But when the context is absent (when memory has less influence on the

speech), perceptual organization cues are relied upon to influence the identified word, in addition to organizing speech objects.

Methods

Participants

The participants consisted of a new batch of 90 Ohio State University students, using a larger group of participants to compensate for the increased number of conditions in this experiment (causing less reliable estimates for each item). Four participants were excluded using the same criteria as in previous experiments (poor foil performance, missing target responses, or fast response times), leaving 86 participants' results.

Stimuli

Pilot testing (N=20) was done on a representative subset of items, to find the combination of features (number of [s]s and repetition rate in s-stream) that would achieve an intermediate proportion of voiced responses, compared to weak and strong grouping conditions. Repetition rate varied from 100-500ms in 100ms increments, and number of [s]s varied between 3, 6, and 8 total. I found that the best combination of features was 200ms repetition rate and 8 [s]s total (6 pre-target and 1 post-target). This configuration consisted of more [s]s with a faster repetition rate compared to the weak condition, and less [s]s with a slower repetition rate compared to the strong condition (see Table 3 for a visual depiction). When tested on 20 pilot participants, this configuration of the s-stream elicited the most consistent intermediate voiced proportions. Since there were more conditions in this experiment than in the previous ones (9 compared to 6), stimuli were distributed slightly differently across the lists, such that each condition contained fewer target words (14 instead of 21). Otherwise, the structure, setup, and creation of the trials was the same as in Experiment 2.

Weak Grouping			
Voiced Bias	L:	Sam muzzled his dog so it would not _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Unbiased	L:	The next item you will hear is _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Voiceless Bias	L:	Putting new wood on the fire created that _**bark**	
	R:	[s]----------**[s]**----------[s]	(900ms)
Medium Grouping			
Voiced Bias	L:	Sam muzzled his dog so it would not _**bark**	
	R:	[s]--[s]--[s]--[s]--[s]--[s]--**[s]**--[s]	(200ms)
Unbiased	L:	The next item you will hear is _**bark**	
	R:	[s]--[s]--[s]--[s]--[s]--[s]--**[s]**--[s]	(200ms)
Voiceless Bias	L:	Putting new wood on the fire created that _**bark**	
	R:	[s]--[s]--[s]--[s]--[s]--[s]--**[s]**--[s]	(200ms)
Strong Grouping			
Voiced Bias	L:	Sam muzzled his dog so it would not _**bark**	
	R:	[s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)
Unbiased	L:	The next item you will hear is _**bark**	
	R:	[s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)
Voiceless Bias	L:	Putting new wood on the fire created that _**bark**	
	R:	[s][s][s][s][s][s][s][s]**[s]**[s][s]	(0ms)

Table 3: *Conditions and examples for Experiment 3: Gradient Organization, with context conditions across rows, separated by grouping conditions. The bold words indicate the target items. The number in parentheses represents the repetition rate of the s-stream.*

Procedure

The procedure was identical to Experiment 2, except that the items were split between three grouping conditions instead of two (14 Targets * 3 Grouping Conditions * 3 Context Conditions = 126 test trials).

Results & Discussion

Categorization Data

Figure 11 shows the proportion of voiced responses (y-axis) averaged over sentential (x-axis) and grouping (color) conditions for each participant. If grouping strength impacts word perception in an accumulative fashion, we should see a difference in voiced responses per grouping condition, such that medium strength s-streams (blue boxes) elicit more voiced responses than weak s-streams (red boxes) and less voiced responses than strong s-streams

(green boxes). Otherwise, this data should show roughly the same patterns as in Experiment 2: effects of sentential bias (voiced > unbiased > voiceless) and grouping strength (strong > medium > weak). In addition, the size of the grouping effect (strong - weak) should be largest in the unbiased condition, compared to the others. The sentential bias patterned as predicted, with the most organizational segregation (voiced responses) on average for voiced (*M*=.94), then unbiased (*M*=.49), and lastly voiceless (*M*=.19) sentences. Grouping strength also affected voiced responses as predicted, with more voiced responses with strong (M=.62, green boxes) than medium (M=.53, blue boxes), and lastly weak (M=.47, red boxes) s-streams. Implications of this effect will be discussed in-depth later in this section.

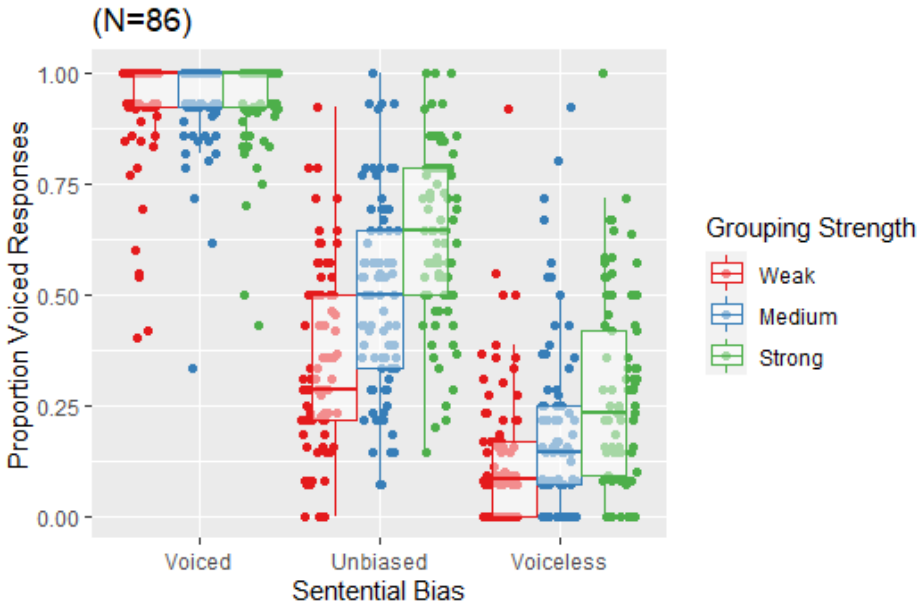

Figure 11: *Boxplots of the proportion of voiced responses by sentential bias and grouping strength conditions. Each dot represents a participant's average.*

As in Experiment 2, the results demonstrate an interaction between grouping and sentential effects, as grouping strength had a different effect size depending on the sentential bias. Though strong grouping conditions almost always elicited more voiced responses than

weak grouping, the grouping strength effect was absent for the voiced bias (.01), largest for unbiased (.30), and half as strong for the voiceless bias (.14). These effect sizes are similar to Experiment 2 (.01, .38, and .21, respectively) and distinct from Experiment 1 (.12, .30, and .28, respectively). Comparison of logistic models using a likelihood ratio test confirmed that the interaction model significantly improved the fit over both the main effect model ($X^2(17)=36.76$, $p<.001$) and the null model ($X^2(13)=457.9$, $p<.001$). This replication strengthens the conclusions of Experiment 2, that grouping strength and sentential memory are independently integrated into the speech percept.

Paired contrasts indicate that the effect of grouping strength differed by context as follows. With an unbiased sentence, the log-odds of strong grouping eliciting voiced responses was significantly higher than both medium ($\beta=.89$, $Z=2{,}309.13$, $p<.001$) and weak ($\beta=1.67$, $Z=3{,}048.25$, $p<.001$) grouping conditions. With a voiceless context, the log odds of strong grouping eliciting voiced responses are again higher than both medium ($\beta=.82$, $Z=1{,}487.26$, $p<.001$) and weak ($\beta=1.29$, $Z=1{,}662.30$, $p<.001$) grouping, but to a lesser extent. The voiced contrast was smallest, with the log odds of strong grouping eliciting voiced responses only slightly more than both medium ($\beta=.16$, $Z=291.46$, $p<.001$) and weak ($\beta=.21$, $Z=265.58$, $p<.001$) grouping conditions. This indicates that the effect of predictive context was constrained by the organizational strength of the s-stream. When the organizational cues connecting the s-stream were weak, a stronger effect of sentential bias emerged, which was lessened by each increase in grouping strength. The auditory system's integration of memory and organizational cues traded-off based on the relative strength of those cues.

Unsurprisingly, grouping strength affected stop identification the most when the sentence carried no context information (unbiased sentences), likely because there was no conflicting

linguistic information for the auditory system to preferentially integrate into the word percept (Mattys, White, & Melhorn, 2005). In the sentential biased conditions, the organization of the auditory scene (and resulting target phoneme perception) was constrained by the context of the preceding speech. This replication strengthens the evidence that when conflicting cues are present in the signal, linguistic memory, specifically predictive context, is weighted heavier for perception than basic grouping heuristics. But when memory-based cues are ambiguous or absent, other cues (like perceptual organization) can influence word formation.

The cumulative effect of grouping strength was demonstrated in two of the three sentence contexts: unbiased (weak: *M*=.34, medium: *M*=.48, strong: *M*=.64) and voiceless sentences (weak: *M*=.13, medium: *M*=.18, strong: *M*=.27). This incremental increase in voiced responses demonstrates that the strength of the organizational cues (grouping strength manipulation) influenced how likely the target [s] was to be perceptually included with the sentence and influencing word formation. The weak s-stream was not strong enough to prevent the target [s] from integrating with the target word (causing less voiced responses), but each increase in s-stream grouping strength (medium and strong conditions) made it more likely that the target [s] would segregate from the target word (decreasing voiced responses incrementally). As in Experiment 2, no differences in voiced reports by grouping strength were found for the voiced context, likely due to responses being at ceiling.

The graded effect of the grouping strength manipulation on target word perception gives evidence that the effect of perceptual organization on speech perception is not all or nothing, that organizational ambiguity can occur for perception. And when organizational ambiguity is present, the auditory system accumulates cues for segregation in an incremental way, with plausible organizations competing for prominence (Ciocca, 2008). If evidence for a segregated

organization passes a threshold, then that organization becomes perceptually dominant, with the listener hearing more than one stream of sounds. With strong and weak s-streams, the organizational ambiguity should be resolved quicker due to the clear evidence in the [s] sequence, either to separate the acoustics into two streams (in strong conditions) or to integrate the [s]s with speech (in weak conditions). But with medium strength s-streams, the evidence for organizational integration or segregation is less clear: in some cases, the evidence in the medium s-stream passes the segregation threshold (perceiving a voiced word), but in others the [s] is integrated with the target word (perceiving a voiceless word). What is the reason why the threshold is surpassed in some cases but not others? The large individual differences found in all experiments thus far suggest that listeners can have different thresholds for segregation, such that conditions or cues that are sufficient for one listener to segregate the sounds might not be sufficient for another (cue weighting differing between participants).

Cue Weighting

Unlike in previous experiments, the response variance explained by participant differences (53%) was almost equal to that explained by item differences (47%). This equivalence between item and participant variation was not found for Experiments 1 or 2, indicating that the addition of the medium grouping condition might have unified participant's cue weighting patterns. To examine participant performance, I performed the same grouping procedure described in Experiment 2 (graphing the grouping effect at the unbiased condition). Graphing the effect in this way allows one to compare the maximum influence of memory and organization cues on perception.

Figure 12 shows the distribution of effect sizes for both manipulations: the effect sizes of the conditions range from 0 to 1 for sentence context and from 0 to .7 for grouping strength.

Interestingly, almost all participants had a larger memory effect than grouping effect (n=78, 92%), with only few participants having a relatively equivalent effect size (n=5, 6%). Only 2 participants showed a larger effect of grouping strength on voiced responses, which is fewer than in previous experiments (Exp3: 2%, Exp2: 9%, Exp1: 9%). It seems that adding a third grouping strength condition caused fewer participants to rely more on grouping than memory cues, perhaps because of the additional uncertainty caused by including an intermediate s-stream strength.

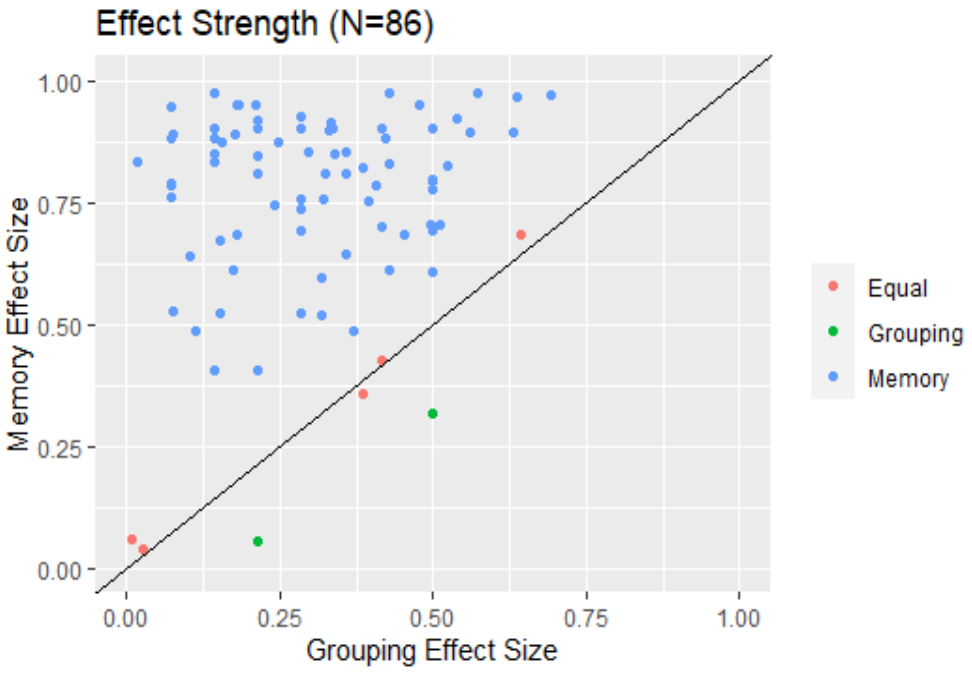

Figure 12: *Scatterplot of participant effect size by cue (memory effect on the y-axis and grouping effect on the x-axis). Each participant is represented by a dot. The colors refer to participant groups, given based on the comparative influence of the two effects.*

The only difference between this experiment and Experiment 2 is the inclusion of the medium grouping condition, which implies that the reduction in participant variance is probably due to this change. The medium condition adds ambiguity to the scene organization, as it does not strongly promote integration (like the weak s-stream) or segregation (like the strong s-

stream). Perhaps this additional ambiguity in the grouping conditions caused participants to shift their weighting of the cues in the experiment, such that memory cues were used to identify the target phoneme consistently more than grouping cues (or, similarly, more participants engaged in that strategy).

Though the categorization data indicate that stimulus items were responsible for more variance than in previous experiments, an examination of responses by items revealed no interesting differences. The variance was due to the effectiveness of the sentence context's bias for the last word more than differences in how each item was affected by grouping strength. The largest variation between voiced responses occurred for sentences in the unbiased condition, which is sensible given that the target word was preceded by a frame sentence designed to be uninformative.

Response Time Data

In Figure 13, the average time to respond for each condition is shown, split by categorization choice. By separating response times according to participant response, we can see the time cost of conflicting grouping and sentential manipulations on the formation of the word percept. As in previous experiments, any condition with less than 100 observations was removed from the analysis. Longer response times indicate that participants required more time to resolve the auditory scene. Response time averages should be fast for easy decisions and slow for harder decisions: more ambiguous conditions (unbiased sentences and medium s-streams) should have longer response times than non-ambiguous conditions (Cutler & Norris, 1979).

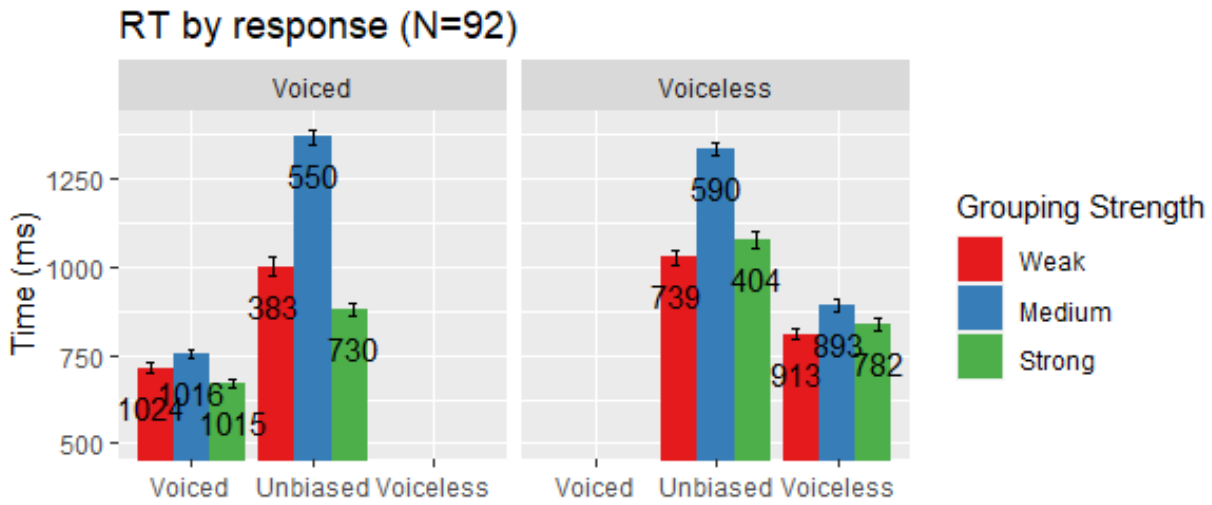

Figure 13: *Bar graph of average time (in ms) taken to respond (y-axis) averaged over participants and separated by context condition (x-axis) and grouping condition (color). The graphs are divided by response type: voiced ("b/d") on the Left and voiceless ("p/t") on the Right. The numbers on the bars represent the number of observations per condition, and the brackets show the standard error from the mean.*

Response decisions were faster (and likely easier) when the sentential bias and the grouping strength conditions were congruent (e.g., voiced bias and strong s-stream) than when they were in opposition (e.g., voiced bias and weak s-stream). Comparison of the linear mixed models supported this, indicating that the interaction model significantly improved the fit over the main effect ($X^2(19)=395.15$, $p<.001$) and null ($X^2(14)=442.09$, $p<.001$) models. For voiced responses (left graph), participants took less time to respond to strong than weak s-streams (Unbiased: 138ms, $Z=5.38$, $p<.001$; Voiced: 45ms, $Z=2.54$, $p<.02$), with the opposite occurring for voiceless responses (right graph; Unbiased: 51ms, $Z=2.05$, $p<.041$; Voiceless: 46ms, $Z=2.34$, $p<.02$). The grouping conditions that led to organizations that were congruent with the response were quickly processed. Those responses that were incongruent to the organization implied by the grouping strength manipulation took longer to process and resolve than congruent responses. This implies that listeners were relying on both cues to some degree when identifying the last word in the sentence, which is consistent with categorization responses.

Immediately upon viewing Figure 13, the very long response times for the medium grouping conditions are conspicuous. Regardless of sentential bias, sentences presented with medium strength s-streams took 50-500ms longer to generate a response for both strong s-streams (Vresp: 213ms, Z=15.83, p<.001; VLresp 125ms, Z=7.97, p<.001) and weak s-streams (Vresp: 128ms, Z=8.86, p<.001; VLresp: 176ms, Z=12.18, p<.001). This implies that the medium strength s-stream was more organizationally ambiguous than the weak or strong condition s-streams, requiring more processing time to resolve. In addition, responding to the unbiased sentences took longer (200-400ms) on average than either of the biased contexts (Vresp: 197ms, Z=5.32, p<.001; VLresp: 290ms, Z=8.74, p<.001). This means that a non-predictive sentence context ("The next item you will hear is…") was less helpful to listeners in identifying the stop, and thus required more time to elicit a response. As cue weighting analysis suggests that memory cues outweighed organizational cues, this longer response time with unbiased sentences makes sense. With only the grouping strength cue being informative, participants required more time to resolve the organization and respond.

Conclusion

In summary, this experiment replicated and extended the finding that both linguistic memory and perceptual organization cues interact in word formation. The replicated interaction between the grouping strength and sentential bias strengthens the conclusion that sentential memory constrains perceptual organization in word perception. This experiment also demonstrated that perceptual organization influences speech perception in a graded, accumulative fashion. Individual differences in participant performance were the most consistent of the three experiments thus far, with participants relying on sentence context more than grouping strength to inform their phoneme categorization. This pattern indicates that the addition of an ambiguous

organization condition (medium s-strength) caused the grouping cue to be less reliable, and likely decreased its cue weighting among participants. The RT data supported the categorization biases, showing that the medium s-stream condition was the most difficult to process, regardless of response or sentence context. Taken together, results imply that that contextual knowledge had more influence on word formation than perceptual organization, though both cues exhibited influence proportionate to their cue strength. In Experiment 4, I will continue to explore the same perceptual organization and memory-based interaction for speech perception by making the unbiased sentences more ambiguous and unique.

Experiment 4: Ambiguous Context

Introduction

The purpose of Experiment 4 is to provide evidence that the grouping effect found in previous experiments relies upon the predictive ambiguity in the sentences, rather than being an artifact of specific properties or peculiarities of the frame sentence. Two common ways to reduce or eliminate the effect of sentential bias are by using either frame sentences or neutral sentences. Frame sentences do not convey any information related to the target word ("The next word you will hear is spark"), while neutral sentences convey information that equally predicts the target alternatives ("Maria jumped after being startled by a spark/bark"). These operationalizations of unbiased sentences are assumed to have the same effect: reducing the buildup of context information available prior to the target word. But the two types of unbiased sentence have subtle, potentially important, differences: frame sentences eliminate sentence context as a variable, such that other factors must be responsible for any effects found, whereas neutral sentences refine sentence context such that it could only be applied to a subset of items (typically the alternative targets being compared). In all previous experiments, frame sentences were used in the unbiased condition, and, in fact, the critical interaction relied upon the large grouping effect in this condition. In this experiment, I expect that the grouping effect found in Experiments 2 and 3 does not change or disappear when the non-predictive sentence frame is replaced with a more natural, ambiguous sentence context, one that equally predicts both segregated and integrated target words. I should find the same interaction as in previous experiments.

Methods

Participants

The participants consisted of 100 Ohio State University students. Eight participants were excluded using the same criteria as in previous experiments (poor foil performance, missing target responses, or fast response times), leaving 92 participants' results. The large number of participants (N=92) is to compensate for having fewer trials per condition in the experiment (each participant's responses are averaged over 15 items for each of the 9 conditions).

Stimuli

In contrast to previous experiments, where all the sentences in the unbiased condition were one of three uninformative sentence frames ("The next item you will hear is…."), I created unique sentences in which the context equally predicted both variants of the target word. For example, the following sentence preceded the word spark: "Maria jumped after being startled by a…." Both word variants were required to fit syntactically and semantically at the end of the sentence. Other than this change to the unbiased condition, the experiment stimuli are identical to Experiment 3.

Due to the constraints involved in creating ambiguous sentences that adequately predicted both target variants, fewer stimuli were used in this experiment, compared to the previous experiments. In total, 15 ambiguous sentences were created from a subset of the 42 target words in Experiment 3 (see Appendix C for a list). The voiced and voiceless biased sentences were a subset of those from Experiment 3, chosen to match the ambiguous targets (45 target sentences total). A preliminary pilot test was conducted to ensure that the newly created ambiguous sentences were similar to normal sentences that participants might hear outside of the lab. On each trial, seven participants read the sentences presented visually with one of both target

endings (voiced “bark” or voiceless “spark”) and rated them in terms of how natural they sounded (1: “Very Unnatural” to 4: “Very Natural”). On average, all sentences were rated as highly natural by all seven participants (*M*= 3.0, *SD*= 0.6), with no difference between the voiced (*M*=3.03, *SD*=0.61) and voiceless (*M*=2.97, *SD*=0.59) versions of the sentences. When comparing the sentence ratings by target word (voiced vs. voiceless: spark/bark), I removed or changed any sentences where one target was rated 1 rating different from the other (7 sentences total).

After the sentences were finalized, I conducted a second pilot experiment to ensure that these neutral sentences predicted both target variants equivalently. In this pilot, participants (N=11) read a sentence presented visually on the screen and rated how well the last word fit in the sentence, on a scale from 1 (“no fit”) to 7 (“best fit”). Each sentence was presented twice, paired with both variants of the target word (e.g., spark/bark). If the sentences were ambiguous (i.e., the context predicted both variants equally) then both versions should elicit the same (high) fit ratings. This was true on average, with both the voiced (*M*=4.9, *SE*=.01) and voiceless (*M*=4.7, *SE*=.01) variants receiving the same score. This pattern was supported by 10 of the 11 participants (90.1%).

Procedure

Other than the shorter length, this experiment’s procedure was identical to Experiment 3. Because there were only 105 trials total (45 targets and 60 foils), the experiment only took about 5-10 minutes to complete.

Results & Discussion

Exclusions

Like in previous experiments, the same exclusion criteria was used to ensure the data quality. Any participants who scored less than 75% correct on the foil trials were excluded (n=3). Participants who did not answer more than one-third of the trials were excluded (n=2). And finally, participants who consistently responded too fast (before the presentation of the target word) were removed from analysis (n=3).

Categorization Data

The focus of this experiment was to replicate the findings of previous Experiments (2-3), while using a sentence context that was neutral (equally predictive of both target versions) rather than one that was completely non-predictive (as sentences in Experiments 2-3 carried no context information). This will ensure that the critical grouping effect found in previous experiments is not due to a peculiarity of the unbiased sentence's structure or repetitions. Figure 14 shows the phoneme categorization results of Experiment 4, averaged over participants, using the same graphing conventions as in previous experiments. The results look more stratified because fewer items were in each condition, compared to previous experiments (15 rather than 40). If the large grouping strength effect for unbiased sentences is due to the differential cue weighting of memory and organizational cues on perception rather than the structure of the unbiased sentence, then the ambiguous sentences should show a similar effect as in previous experiments.

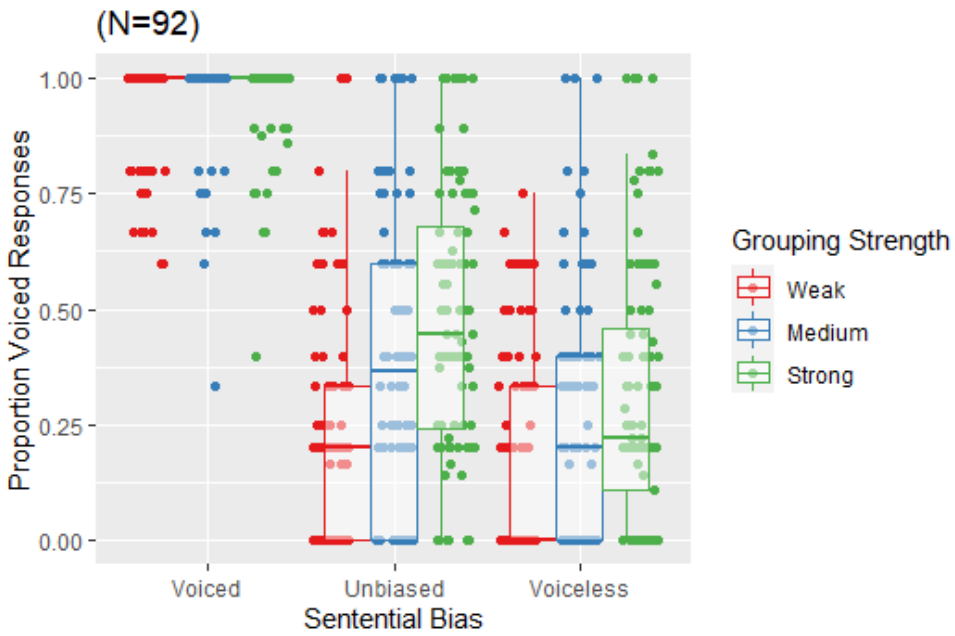

Figure 14: *Boxplots of the proportion of voiced responses by sentence context (x-axis) and grouping strength conditions (color). Each dot represents a participant's average.*

Figure 14 shows the same pattern as in previous Experiments (2-3), in that the effect of grouping strength differs based on the context of the sentence and is largest for the unbiased sentences. The grouping strength manipulation had no effect when the sentence context was biased toward voiced (strong – weak = .02), was largest when context was ambiguous (.36), and was halved for the voiceless bias (.15). This grouping strength effect size is almost identical to Experiment 3 (voiced .01, unbiased .30, voiceless .14) and similar to Experiment 2 (voiced .01, unbiased .38, voiceless .24). The grouping effects are reliably similar to Experiment 3, ($t(4)=0.36$, $p=0.74$), as are the sentential effects ($t(4)=0.05$, $p=0.96$). This implies that the frame sentences in previous experiments were processed similarly to the ambiguous sentences in this experiment. Again, the effect of perceptual organization on word identification seems to be somewhat constrained, but not eliminated, by the sentence context. This solidifies the conclusion from Experiment 3, that memory cues dominate organizational cues when forming and

perceiving words. The cue weighting and RT analyses were the identical to the previous experiment, and thus were excluded from the results section.

Frame sentences are unnatural (likely only heard in lab-environments), repetitive, and context-free, while neutral sentences are natural (similar to those heard in the real-world), unique, and context-ambiguous. Do the subtle differences in frame and neutral sentences cause any differences in perception? The sentence perception literature does not directly address this question, but the word segmentation experiments of Kim et al. (2012) shed some light on this question. Kim and colleagues investigated how listeners differentiate between one- and two-word phrases that have similar acoustics (‘a door’ vs ‘adore’). They presented these ambiguous targets to participants either isolated or embedded in biased or neutral sentences and asked participants to rate their confidence in the percept (1: confident one-word to 7: confident two-words). They found participant ratings of 3-4 (indicating uncertainty) for both isolated targets (‘a door’ vs ‘adore’) and neutral sentences (“The servant came to [A DOOR - ADORE]…”). Participant responses to isolated targets and targets embedded in neutral sentences were uncertain, indicating that they were unable to reliably differentiate between the two segmentation alternatives. Only when the context biased one segmentation over another (“The lovers came to adore…”) did participants indicate that they were confident in their response. These results demonstrate that ambiguous word-phrases in neutral sentences are perceived similarly when presented in isolation, implying that frame and neutral sentence processing would not differ. The results of these dissertation experiments support this supposition, as the effect of frame sentences in Experiments 2-3 gave similar results as the effect of neutral sentences in this experiment. The differences found were due to participants, not items. This experiment is the first to explicitly

demonstrate that context-free (frame) and context-ambiguous (neutral) sentences are processed in the same way.

Conclusion

In summary, this experiment replicated and extended the conclusion that the auditory system integrates both memory and organizational cues while forming the speech percept. The replicated interaction between the grouping strength and context conditions strengthens the conclusion for the presence of an interactive juncture between sentential memory and perceptual organization. Cue weighting and response time analysis yielded no differences from experiment 3, but will not be discussed further. This experiment demonstrated that context-neutral sentences (ones that equally predicted both target versions) were processed in the same way as context-free sentences from earlier experiments, when averaged over items, not participants. This strengthens the conclusions made from Experiments 2-3, that sentence context interacts with perceptual organization when forming words in a hierarchical manner, such that differences in grouping strength mainly arose in conditions without sentential bias.

General Discussion

Summary of Problem

A complete model of human language comprehension requires an understanding of how perceptual organization and linguistic memory interface when organizing and perceiving speech. Researchers have previously considered speech organization to be a separate problem from speech perception, and thus have elaborated on different processes necessary for successful auditory perception. Though both perspectives invoke memory-based effects to explain speech processing, the perceptual organization framework elaborates on the effects of organizational cues on initial auditory grouping, while the linguistic processing framework elaborates on memory effects on speech perception, with no clear understanding of how the two systems connect and interact. The simplest (and classically theorized) architectural explanation would categorize auditory grouping as a separate, early, and completely independent process from linguistic processing, and linguistic processing as a part of secondary, memory-based grouping, along with other higher-order processes that make use of listener knowledge (Bregman, 1990; Ciocca, 2008; Davis & Johnsrude, 2003). This formulation would imply that when perceiving speech, the auditory grouping cues are integrated and finalized before memory-based grouping begins (linguistic cues integrated as a second step). However, the literature suggests that the auditory system might contain more interactions between initial organization processes and later linguistic processing than initially theorized.

This dissertation systematically explores how auditory grouping and linguistic memory interact to promote speech perception, by comparing how organizational cues and linguistic knowledge (lexical and sentential) are integrated during speech perception. My experimental design allows for the potential trade-offs between organizational and memory contributions on speech perception to be explicitly measured. In this way, the dissertation can clarify the likely

architecture underlying successful speech perception, from auditory organization of acoustic input to memory-based processing, resulting in conceptual understanding. For both lexical (Exp 1) and sentential (Exp 2-4) processing, the presence or absence of an interaction between organizational and linguistic manipulations can inform the likely architecture of the auditory system. Finding a consistent effect of grouping cues alongside lexical or sentential biases places doubt on the idea that auditory grouping is completed and applied to the signal before high-level (conceptual) memory-based processing begins. This result would require the redefining of perceptual organization beyond an early, basic process, integrating it with memory-based processing. However, finding a non-existent effect of the grouping manipulation in the presence of conflicting higher-level linguistic knowledge (lexical or sentential) implies that organizational processing has no influence on the percept and likely no overlap with memory-based processes, potentially finalizing before linguistic memory is applied to the signal. This would lend evidence to the traditional view of speech processing, one that assumes little concurrent influence of organizational processes and memory-based processing, instead favoring a sequential application.

The Lexical Juncture

Experiment 1 tested the integration of organizational grouping and lexical knowledge during word perception, with results indicating a potential lexical interactive juncture connecting the linguistic and auditory frameworks. The lexical status of the target word and the grouping strength of the accompanying s-stream were both manipulated, such that participants received conflicting information about the percept: responses demonstrated participants either integrating the s-stream into words (“sponge”) or segregating the s-stream into non-lexical speech (“bunge”). Categorization results indicated that both manipulations had a consistent and

independent effect on perception, such that both lexical knowledge and auditory organization influenced the resulting percept. Critically for the theorized architecture behind auditory processing, neither lexical nor grouping cues insulated the percept against the influences of the other. This gives direct evidence that auditory grouping and lexical processing interact when processing speech, such that perceptual organization could be working with linguistic memory at the lexical level to parse and understand the speech input.

That the initial phoneme categorizations were affected by lexical status is particularly impressive when one considers the time course of the trials. In all cases, lexical status of the stimuli was not certain until 100-300ms after the stop phoneme (e.g., "bun" is a word, but "bunge" is not), and yet listeners perception of the initial stop (voiceless "p" or voiced "b") was partially dependent on lexical status. Participants were instructed to categorize phonemes, a task which can be completed independently of the words they are embedded in, but the data shows that their decisions were heavily influenced by their lexical knowledge, applied after the presentation of the critical phoneme. This implies that the system buffers its phonemic perception when presented with continuous speech, waiting to formalize the speech percept until information matches with higher-level, lexical knowledge. Thus, the strict linguistic processing hierarchy might be an oversimplification of a complex system, one where multiple levels of processing are occurring concurrently and interactively. One potential explanation of this data is that each level of processing passes along multiple potential interpretations of the speech (sub-lexical processing indicating an equal likelihood of "p" or "b" identification), which is then refined by higher-level processing (lexical in this case).

Given that categorization data indicate perceptual organization is engaged for a longer timescale than previously assumed (lexical as well as sub-lexical), examining the time course of

participant responses gives more information about the timescale of organizational and linguistic processing. In Experiment 1, the response time data showed that grouping strength had a consistent and predictable effect on processing speed, such that responses that agreed with the grouping strength manipulation were faster than those that opposed the manipulation. This result is sensible in the context of perceptual organization, as an automatic process that groups and separates sound in auditory scenes: responses that agree with the cued organization should be easier to make, and thus faster, than those in opposition (Bregman, 1990; Ciocca, 2008; Remez, 2021). Participant responses were still made that opposed the grouping cue (in favor of the conflicting lexical cue), but these responses took significantly longer to carry out. This implies that, as theorized, organizational processes began early, forming a preferred or likely organization (given the currently available information), which was then integrated with the lexical information presented at the end of the trial. If the lexical bias agreed with the cued organization, responses were made much faster than if they opposed the cued organization. This suggests that the organizational processing was still occurring in an overlapping fashion with the higher-level lexical processing, influencing the resulting word perception.

Experiment 1 demonstrates that the auditory system integrates both lexical and grouping cues into decisions about speech organization and perception. This implies that organizational and linguistic processing are more interconnected than previously assumed. This is the first experiment to find these results with a hybrid, sentence streaming paradigm, which more closely resembles how listeners experience speech in the real world, compared to repetition and single word paradigms (Billig et al., 2013; Cutting, 1975; Cutting & Day, 1975; Day, 1968; Freggens, Thomas, & Pitt, 2019; Morais, 1996; Mattys & Melhorn, 2005; Pitt & Shoaf, 2002; Poltrock & Hunt, 1977; Sexton & Geffen, 1981; Warren, 1968). In addition, the combination of

categorization, cue weighting, and response timing data provides a fuller picture of the interaction than suggested by previous studies. This first experiment provided additional context on the potential architecture underlying the interaction of perceptual organization and linguistic memory on speech processing, with specific implications for the existence of an interactive juncture between organization and lexical processing. It provides the backdrop for the next series of Experiments (2-4), which are the first to examine whether sentential memory is integrated with organizational information, when perceiving speech.

The Sentential Juncture

Experiments 2-4 demonstrated the integration of sentential memory (i.e., predictive context) and grouping strength on the organization and perception of speech, investigating the potential for an interactive juncture between perceptual organization and linguistic memory at the sentential level of processing. Using a target word that was lexically valid both with and without the [s] (e.g., "spring" and "bring"), the preceding sentence context was varied, such that the context either predicted the version of the target that integrated the [s] ("Putting new wood on the fire created that spark") or the version that segregated the [s] ("Sam muzzled his dog so it wouldn't bark"). Contrary to Experiment 1, I found a reliable interaction between the grouping strength and sentential bias conditions, such that the grouping effect was strongest in conditions without sentential bias ("The next item in the sentence is …"), and greatly reduced when predictive context was included in the sentence (halved or nullified). This implies that the interactive juncture between organization and memory mechanisms theorized for lexical processing is likely different or weaker at the sentential processing level. This result would fit with the hierarchical conceptualization of the linguistic framework, purporting that as linguistic processes operate on more abstract representations (sub-lexical -> lexical -> sentential), they

become impervious to the influence of low-level, organizational processes (Davis & Johnsrude, 2003; Davis & Johnsrude, 2007; de Heer et al., 2017; McClelland, Mirman, & Holt, 2006).

This interaction between sentential and organizational processing demonstrates the hierarchical relationship between the two processing frameworks (perceptual organization and linguistic memory). The preceding sentence context influenced listeners' prediction of the upcoming target word, and, by doing so, biased the organization of [s] with the final word. This prominent sentential influence on word perception has been found directly in previous auditory perception studies (Connine, 1987; Connine, Blasko, & Hall, 1991; Connine, Blasko, & Wang, 1994), as well as studies investigating different speech perception phenomena, like word segmentation (Kim, Stevens, & Pitt, 2012; Mattys & Melhorn, 2007; Mattys, White, & Melhorn, 2005), and linguistic memory effects on ambiguous phonemes (Cutting, 1975; Ganong, 1980; Getz & Toscano, 2019; Samuel, 2001).

Most directly applicable are the studies by Connine (1987), who studied the interaction of sentence context and acoustic clarity on word recognition. She presented listeners with biased sentences ending in an acoustically ambiguous final word (e.g., dent - tent), whose initial phoneme varied in clarity. She found that sentence context had the largest influence when the initial phoneme was acoustically unclear, and that in these cases, listeners identified the phoneme that agreed with sentence context. She concluded that listeners used sentence context information to resolve acoustic ambiguity. From this perspective, it makes sense that the effect of grouping strength varied by context condition in Experiments 2-4. The perceptual organization cues are less abstract (dealing with acoustic similarity and continuity) than sentential context, and more likely to be relied on when the memory-based cues are either unavailable or unreliable.

My experiments show that listeners use sentence context information to resolve organizational ambiguity as well as acoustic ambiguity (Connine, 1987; Connine, Blasko, & Hall, 1991; Connine, Blasko, & Wang, 1994). The grouping strength effect did not completely disappear under conflicting conditions, indicating that there could be some overlap in processing between perceptual organization and sentential processing. This partially reinforces the processing hierarchy theorized by linguistic framework of speech perception: more abstract and high-level processes (e.g., sentential) have more influence on the final speech percept than low-level processing stages (Davis & Johnsrude, 2003; 2006; de Heer et al., 2017; McClelland, Mirman, & Holt, 2006). Certainly, the sentential processing had a qualitatively different interaction with grouping strength than lexical processing did in Experiment 1.

This study is also one of the first to directly compare sentential memory and grouping cue influence on the initial organization of speech, by determining the comparative influence of each competing source of information on [s] integration into the following word. Word segmentation studies involve similar investigations of speech organization within a single speech stream, in which researchers are interested in how a continuous stream of speech is perceptually separated into discrete words (Kim, Stevens, & Pitt, 2012; Mattys & Melhorn, 2007; Mattys, White, & Melhorn, 2005). The experiments by Kim and colleagues (2012) are of most relevance: they investigated how listeners disambiguate words that have multiple lexically plausible segmentations (“adore” vs “a door”). The experimenters presented participants with recordings of these items as one-word or two-word productions, both in isolation and embedded within sentences. Participants were instructed to indicate their perception (one word or two words) and certainty (confident or uncertain). The researchers found that when the word-phrase targets were presented in isolation, listeners were uncertain about whether they heard the item as one-word or

two-words. This uncertainty was replicated when the word-phrase targets were embedded in a sentence with a neutral precursor ("The servant came to…" ADORE – A DOOR). But when the sentence precursor included biasing sentential context for one-word ("Lovers are meant to adore…") or two-word ("The hallway leads to a door…") versions, responses became more certain and uniform. Simply the addition of plausible words as a precursor was not enough to confidently segment the word-phrase target. Only when the context was meaningful and disambiguating did participant perceptions become stable. Kim et al.'s data pattern is similar to that of my sentential experiments: the sentential biased conditions influenced speech organization between the speech and [s] streams, while the unbiased condition remained organizationally ambiguous. Thus, the unbiased condition showed the largest effect of grouping, compared to the sentential biased conditions.

The time course of participant responses gives more information about how organizational cues are integrated with memory-based cues. When grouping cues were presented alongside sentential biases (Experiments 2-4), responses that agreed with the grouping cues were made faster than those in opposition to the grouping cues. This adds evidence to support the idea that perceptual organization is active over a longer timescale than previously assumed, as it affected the speed that the percept was resolved into a word, even when presented alongside a biasing sentence context. Interestingly, the effect of these response time differences was much smaller than in Experiment 1, especially in conditions where sentential bias was present. This likely reflected the increased strength of the sentential cues, compared to the lexical cues: the categorization responses demonstrated that the sentential cues were able to partially insulate the speech stream against intrusions from the s-stream. Though organizational cues had less impact on processing speed when pitted against sentential memory than lexical memory, the sustained

presence of the response time differences imply that perceptual organization processes are active even as high-level memory-based processing is occurring. This is further evidence that both processes work in parallel to organize the speech scene.

The inclusion of Experiment 3 was intended to confirm the graded nature of the grouping cues, by adding a medium grouping strength condition, but also had implications for the timing of perceptual organization. This medium s-stream was designed to be more ambiguous regarding the preferred organization of the scene: indeed, the categorization results indicated that it had more influence on responses than the weak s-stream, but less influence than the strong stream. The response time analysis revealed that participants took much longer to respond to the words presented with the medium s-stream than either the strong or weak conditions, regardless of both categorization response and sentential bias. This suggests that the strong and weak grouping conditions were effective in creating the expectation for a preferred organization of [s] and target word. When that information was made to be less informative (as in the medium s-stream), participant's response speed was hampered, regardless of the more influential sentential biasing condition. This implies that one of the functions of perceptual organization is to speed the ease of speech processing, by preferencing the likely organization of complex scenes, influencing which sounds are included in the perceived speech.

Cue Weighting

Across all experiments, participants showed large individual differences in their relative reliance on memory and grouping cues, adding to the literature supporting varying cue weighting strategies or biases during human language comprehension (Giovannone & Theodore, 2021, 2023; Kapnoula et al., 2017; Kaufeld et al., 2020; Li et al., 2019). This cue reliance variance between listeners was particularly pronounced in Experiment 1, likely because of the hierarchical

nature of linguistic processing (lexical information was less constraining than sentential information). To compare cue weights within participants, I measured the comparative effect sizes of the memory and grouping effects. Though most participants in that experiment exhibited a larger memory effect than grouping effect, a small proportion of participants showed the opposite pattern, preferencing grouping cues over lexical cues. In fact, a significant cue weighting trade-off was found, such that larger reliance on one cue was reliably associated with smaller reliance on the other cue. Participants who responded based primarily on their linguistic knowledge also placed less emphasis on organizational information, and participants who relied primarily on grouping heuristics to categorize the stop placed less emphasis on their lexical memory. This cue weighting trade-off suggests a pattern of cue reliance that stays consistent throughout the experiment. These differences between participants imply that auditory perception is not necessarily identical between listeners: cue weighting patterns differentiate listeners.

Ishida, Samuel, and Arai (2016) noted a similar pattern when listeners reported words with embedded temporal reversals (50-200ms segments of the word were presented reversed in time). Though the average data showed that words more than pseudo-words remained intelligible despite reversals, they found that participants differed in the strength of this effect. Some participants relied on lexical knowledge more than others when reporting back the reversed words (i.e., they were much better at accurately reporting reversed words than pseudo-words), while others relied on the acoustic cues more (their reporting was similar for words and pseudo-words). Further, they demonstrated that this individual difference in lexical reliance was stable when applied to a phonemic restoration task: that participants who heavily relied on lexical knowledge for one task also relied on lexical knowledge for a different task, suggesting a general

pattern of cue reliance. This result of a stable, lexical dependence trait has been supported by similar experiments (Giovannone & Theodore, 2021, 2023). I conjecture that these same "lexically sensitive" individuals from Ishida et al.'s experiment would perform similarly to the memory group in Experiment 1. This suggests that something specific to the individuals can differentially affect cue weighting during word perception. The reason for the difference in cue weighting in both cases is unknown currently, but it provides evidence that the integration of organization and memory information varies between individuals.

Freggens and Pitt (under review) found similar individual differences in cue weighting when investigating selective attention to short, simultaneously presented speech. They used the same stop categorization task as this dissertation with word-length stimuli, varying how much linguistic information was presented before and after the target word (less linguistic: "start" and "suh" presented to the opposite ear vs more linguistic: "start" and "such"). Much like in Experiment 1, individuals responded in patterns that indicated a preference for acoustic, organizational cues (always separating the speech) or a reliance on linguistic cues (only separating the speech when enough linguistic information was present). These results had implications for participants' auditory selective attention abilities, such that participants who relied less on organizational cues demonstrated weak selective attention ability: they were unable to use organization cues alone to focus on the target word, requiring additional linguistic information. In contrast, those with strong selective attention ability were able to focus on the target word in all conditions, based solely on the acoustic cues to organization. These results imply that cue weighting preferences between organizational and memory-based cues could have an effect on selective attentional abilities. Applied to this dissertation, participants that showed a

reliance on lexical cues rather than grouping strength cues might have weaker selective attention ability.

In contrast to Experiment 1, individual differences in participant performance for Experiments 2-4 were reduced, such that all listeners showed relatively uniform reliance on sentence context to predict the final word, more than grouping strength. This pattern indicates that unlike with lexical knowledge (Experiment 1), listeners unilaterally rely on sentence context cues rather than organizational cues when perceiving speech. Mattys, White, and Melhorn (2005) found a similar pattern of results when investigating the interaction of cue reliance during word segmentation. They contrasted the effects of a sub-lexical cue (word stress) with an opposing sentence context cue in segmenting ambiguous, sentence-final words. The experimenters presented participants with a spoken sentence ending in a two-syllable word (with an embedded word inside: "cre-MATE") and then measured the priming effect of each word version ("cremate" and "mate") in a visual lexical decision task. The sentence context favored the full word ("An alternative to traditional burial is to cremate the dead"), while the word stress (sub-lexical cue) favored the embedded word ("mate" is the strongly stressed syllable in "cremate"). They found that under normal listening conditions, the priming effect of the full word was almost double that of the embedded word (140 vs 80ms). However, as the acoustic quality of the sentence deteriorated (by adding increasing levels of noise to the sentence), this effect diminished (20ms priming difference) and eventually reversed (cremate 30ms vs mate 55ms). They concluded from a series of experiments with similar outcomes that cue reliance for word segmentation seems to be hierarchical and graded, such that listeners prefer using more abstract, linguistic cues of lexical and sentential context until the signal becomes degraded. As the

preceding context becomes acoustically less clear, reliance on high-level, context cues decreases, in favor of low-level, sub-lexical cues, like stress.

Interestingly, when the medium s-stream condition in Experiment 3 introduced more variance in the grouping strength cues, individual differences in cue weighting strategies severely decreased. Whereas in previous experiments, the memory group was the majority (Exp1: 70%, Exp 2: 86%), almost all participants in Experiment 3 preferentially weighted sentential memory cues as more informative for word perception (92%), with almost no participants relying more on grouping cues (2%). Holt and Lotto (2006) demonstrated that participant cue weights for acoustic information could shift due to the distribution of trials in an experiment. Though their participants were trained to discriminate sine waves based on two separate acoustic cues (center frequency and modulation frequency), participants exhibited initial biases in which cue they used for their responses. Like in these dissertation experiments, participants showed cue preferences despite both cues being informative and discriminable. And like the reduction of individual differences found in Experiment 3, participants' response bias shifted when the variance of one cue was increased between experiments. Where participants had previously depended primarily on central frequency for their response, they became dependent on modulation frequency. In my Experiment 3, adding the medium grouping condition likely caused increased variance among the grouping cues, making the cue less reliable for categorization. Thus, more participants preferentially used sentence context information, rather than organizational information.

The graded effect of the grouping strength manipulation on target word perception gives evidence that the effect of perceptual organization on speech perception is not all or nothing, that organizational ambiguity can occur for perception. And when organizational ambiguity is

present, the auditory system accumulates cues for segregation in an incremental way, with plausible organizations competing for prominence (Remez, 2021). If evidence for a segregated organization passes a threshold, then that organization becomes perceptually dominant, with the listener hearing more than one stream of sounds. With strong and weak s-streams, the organizational ambiguity should be resolved quicker due to the clear evidence in the [s] sequence, either to separate the acoustics into two streams (in strong conditions) or to integrate the [s]s with speech (in weak conditions). But with medium strength s-streams, the evidence for organizational integration or segregation is less clear: in some cases, the evidence in the medium s-stream passes the segregation threshold (perceiving a voiced word), but in others the [s] is integrated with the target word (perceiving a voiceless word). What is the reason why the threshold is surpassed in some cases but not others, even with strong grouping and sentential cues? The large individual differences found in all experiments suggest that listeners can have different thresholds for segregation, such that conditions or cues that are sufficient for one listener to segregate the sounds might not be sufficient for another (causing cue weighting differing between participants). This implies that the auditory system has the capability to integrate both organizational and memory sources of information, but this is not realized in the same way for all listeners.

Architecture Summary

One assumption of both the current perceptual organization and speech perception frameworks is that organizational cues are applied to the acoustic signal only before higher-level processing begins (Darwin, 2008; Shinn-Cunningham, Best, & Lee, 2017). In fact, perceptual organization is often referred to as a "basic" or "primitive" process for this reason (Bregman, 1990, Ciocca, 2006; Remez, 2021). This implies that organizational cues should not be able to

interact with high-level perceptions, like word identification. However, categorization responses in the dissertation experiments demonstrated that supposedly primitive organizational cues were co-occurring with memory processes to create the listener's word perception. In addition, my experiments also show that linguistic memory (both lexical and sentential) affected the organization of the speech signal (whether the [s] was integrated into or segregated from the target word). Both results imply that auditory organizational processing does not finalize before memory-based processing begins. This evidence contradicts the assumption that perceptual organization is finalized and applied to the signal before memory-based processing initiates.

The RT results show that perceptual organization information begins being taken in early, influencing the ease of processing for ambiguous speech or complex scenes. The grouping cues impacted how quickly participants resolved the percept, causing delays when responses reflected a non-cued organization and speeding responses when they reflected cued organizations. Without informative cues (Experiment 3's medium grouping condition), participant responses were delayed. This indicates that though perceptual organization processes initiate early, as assumed in the literature, they do not finalize before memory-based processing of the speech occurs. This implies that perceptual organization is not likely composed of a serial two-step process, whereby acoustic grouping cues are first analyzed to create the organization of the scene, and later memory-based processing is applied. The response times imply that perceptual organization might be a process that starts early and engages continuously in the background, presenting live updates on the likely organization of complex auditory scenes.

Taken together, my results imply that that linguistic memory and organizational information are both applied in a partially overlapping fashion, with an interactive juncture at the level of lexical processing and a much weaker interactive juncture at the sentential level of

processing. In all experiments, both linguistic and organizational cues influenced the perceived target word, clarifying the organization of the two overlapping auditory streams. The combined results of the dissertation experiments bring together the two literatures on auditory perception: linguistic and perceptual organization cues are integrated together to inform the speech percept. The degree of interaction between organization and linguistic memory differs based on the hierarchical level of linguistic processing: sentential biases, crucial for listeners' nuanced understanding of speech, interacted less with organization cues than lexical processing. This reinforces the proposed hierarchical structure of linguistic memory: lexical information was not strong enough to prevent the influence of organizational cues in the same way that sentential information did.

Limitations

In hindsight, I should have replicated Experiment 1 with a different sample of participants, like I did with the sentential bias experiments (2-4). The reason I did not replicate Experiment 1 originally was that it was very similar to other experiments that had been done in the lexical perception literature. My hybrid paradigm was a combination of the single presentation studies (Cutting, 1975; Cutting & Day, 1975; Day, 1968; Morais, 1996; Mattys & Melhorn, 2005; Poltrock & Hunt, 1977; Sexton & Geffen, 1981) and the repetition paradigm (Billig et al., 2013; Freggens, Thomas, & Pitt, 2019; Pitt & Shoaf, 2002; Warren, 1968). For this reason, I considered Experiment 1 a replication in itself, as well as a proof of concept for the experimental manipulations. I didn't expect to find anything surprising or different in my results (like the cue weighting data). Replicating Experiment 1 would strengthen my conclusion in the individual differences found for cue weighting, specifically the trade-off in cue reliance.

In addition, differences between lexical and sentential bias strength could be explained by the difference in cue build-up over the trials, instead of the hierarchical relationship between the two memory processes. In Experiments 2-4, sentential bias accumulated throughout the trial, like grouping strength: the listener could use both the grouping and context evidence presented throughout the trial to categorize the target phoneme. In contrast, the lexical bias was not experienced by the listener until after the target phoneme was presented (as lexicality judgements occurred for the word that the target phoneme was embedded in e.g., bun and bunge). Maybe the lexical influence on speech perception and organization would have been equivalent to the sentential influence, if it were given adequate time to build.

Future Research

Given the limitations listed in the above section, future research should conduct additional experiments using the lexical bias, with a version of the hybrid paradigm that allows lexical information to build before the target phoneme. Perhaps lexical bias, when given the chance to accumulate, could exhibit the same prohibitive influence on the effect of grouping cues that sentential bias did. One option for testing this is by contrasting voiced perceptions of words with an [s]+stop in the beginning ("spellbinding"), middle ("conspiracy") or end of a longer word ("counterspy"). You would expect that the effect of lexical bias would increase as the [s]+stop location was moved later in the word. This is because it would have a chance to build up the evidence for organization similar to the organizational cues. However, an issue with this method is that there are few, if any, multi-syllabic words that would create lexically valid words both with the [sp] and the [b], the way that single syllable, s-initial words do (e.g., "spring", "bring").

To strengthen conclusions about the consistency of individual cue weighting biases, I would recruit the same group of individuals to participate in both the lexical and sentential

experiments. Results from the same individuals across these two experiments would be informative about whether individual cue weights are consistent across different types of linguistic information. Based on a similar lexical study from Ishida, Samuel, and Arai (2016), which demonstrated that lexical reliance carried through to multiple tasks, I would expect that listener cue weights would remain the same or similar across experiments. Those individuals that relied on lexical information would also rely on sentential cues more than grouping cues. And those few individuals who relied more on grouping cues in the sentential experiments would also rely more on grouping cues for the lexical experiment.

Conclusions

This dissertation is the first to examine the integrated architecture of organizational and memory mechanisms when processing speech, with results suggesting implications for the structure and time frame of auditory processing. Results from the dissertation experiments show that the auditory system integrates cues from both organizational and memory-based sources of information in an overlapping manner. In addition, the relative reliance on either cue strongly differed by individuals: though stable patterns emerged in the aggregate, groups of listeners relied on the same cues to different degrees. This implies that the auditory system has the capability to integrate both organizational and memory sources of information equivalently, but this is not realized in the same way for all listeners. Lastly, response time results demonstrated that organizational processes are not necessarily finalized before high-level, memory-based processing initiates. This defies the assumption that perceptual organization is a system relegated to early time points in auditory scene processing.

The dissertation results suggest a different route of auditory processing than the traditional view, which is that perceptual organization of the speech stream is formed and

finalized first, and only then is that information is passed to the next process. Instead, my results suggest perceptual organization is influencing the outcome of perception simultaneously with memory processes, even when processing abstract speech sounds. It is possible that perceptual organization is continuously engaging in the background throughout the stages of speech perception, never finalizing completely or passing off the information to a different system. In this way, perceptual organization could have a supportive role, by providing continuous information about the probable organization of the scene, and in so doing, speeding the processing time of speech that occurs in complex acoustic environments. These dissertation results suggest that perceptual organization and memory-based processing occur in tandem, at least partially overlapping when perceiving speech. This means that these two processes are more connected than previously assumed.

References

Arbogast, T. L., Mason, C. R., & Kidd Jr, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2169-2180. https://doi.org/10.1121/1.1861598

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Billig, A. J., Davis, M. H., Deeks, J. M., Monstrey, J., & Carlyon, R. P. (2013). Lexical Influences on Auditory Streaming. *Current Biology*, *23*(16), 1585–1589. https://doi.org/10.1016/j.cub.2013.06.042

Bregman, A. S. (1990). *Auditory Scene Analysis: The perceptual organization of Sound*. MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.12, retrieved 17 April 2022 from http://www.praat.org/ .

Ciocca, V. (2008). The auditory organization of complex sounds. *Frontiers in bioscience, 13, 148-69.* https://doi.org/10.2741/2666

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, *26*(5), 527–538. https://doi.org/10.1016/0749-596X(87)90138-0

Connine, C.M., Blasko, D.G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234-250. https://doi.org/10.1016/0749-596X(91)90005-5

Connine, C.M., Blasko, D.G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics, 56* (6), 624-636. https://doi.org/10.3758/BF03208356

Culling, J. F., & Darwin, C. J. (1993). Perceptual separation of simultaneous vowels: Within and across-formant grouping by F_0. *The Journal of the Acoustical Society of America*, *93*(6), 3454–3467. https://doi.org/10.1121/1.405675

Culling, J. F. & Stone, M. A. (2017). Energetic masking and masking release. In Middlebrooks et al. (Eds.). *The auditory system at the cocktail party.* Springer International Publishing. ISBN: 9783319516622

Cutler, A. (2012). What is spoken language like? In *Native listening: Language experience and the recognition of spoken words.* MIT Press. https://doi.org/10.7551/mitpress/9012.003.0003

Cutler, A., & Norris, D. (1979). Monitoring Sentence Comprehension. in W.E. Cooper & E.C.T. Walker (eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett*, pp. 113-134. https://hdl.handle.net/2066/15691

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83*(2), 114–140. https://doi.org/10.1037/0033-295X.83.2.114

Cutting, J. E., & Day, R. S. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, *3*(2), 99–113. https://doi.org/10.1016/S0095-4470(19)31353-1

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. In *Handbook of psycholinguistics* (pp. 249-283). Academic Press.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5-6), 507-534. https://doi.org/10.1080/01690960143000074

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1011–1021. https://doi.org/10.1098/rstb.2007.2156

Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *The Journal of the Acoustical Society of America*, *79*(3), 838–845. https://doi.org/10.1121/1.393474

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: the role of interaural time differences. *Journal of experimental psychology. Human perception and performance*, *25*(3), 617–629. https://doi.org/10.1037//0096-1523.25.3.617

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, *23*(8), 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Day, R. S. (1968). Fusion in dichotic listening. Doctoral dissertation, Stanford University (Psychology). *Dissertation Abstracts (1969), 29*, 2649B, (University Microfilms No. 69-211, Ann Arbor, Michigan.)

Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x

Giovannone, N., & Theodore, R. M. (2021). Individual Differences in the Use of Acoustic-Phonetic Versus Lexical Cues for Speech Perception. Frontiers in Communication, 6. https://www.frontiersin.org/articles/10.3389/fcomm.2021.691225

Giovannone, N., & Theodore, R. M. (2023). Do individual differences in lexical reliance reflect states or traits? *Cognition, 232*. https://doi.org/10.1016/j.cognition.2022.105320

Ezzatian, P., Li, L., Pichora-Fuller, M. K., & Schneider, B. A. (2011). The effect of masker type and word position on immediate sentence recall. *Proceedings of the 24th Annual Meeting of the International Society for Psychophysics, 24*, 209-215. https://proceedings.fechnerday.com/index.php/proceedings/article/view/204

Freggens, M., Thomas, A., & Pitt, M. A. (2019). A test of linguistic influences in the perceptual organization of speech. *Attention, Perception, & Psychophysics*, *81*(4), 1065–1075. https://doi.org/10.3758/s13414-019-01699-3

Freggens, M., & Pitt, M. A. (2023, *under review*). Characterizing individual differences in selective attention to speech. Submitted to *Attention, Perception, & Psychophysics.*

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112-2122. https://doi.org/10.1121/1.1354984

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human perception and performance, 6* (1), 110-125. https://psycnet.apa.org/fulltext/1981-07020-001.pdf

Getz, L.M., & Toscano, J.C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science, 30* (6), 830-841. https://psycnet.apa.org/fulltext/1981-07020-001.pdf

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. *The Journal of Neuroscience*, *37*(27), 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059–3071. https://doi.org/10.1121/1.2188377

Humphries, C., Sabri, M., Lewis, K., & Liebenthal, E. (2014). Hierarchical organization of speech perception in human auditory cortex. *Frontiers in Neuroscience, 8,* 1-12. https://doi.org/10.3389/fnins.2014.00406

Ishida, M., Samuel, A. G., & Arai, T. (2016). Some people are "More Lexical" than others. *Cognition*, *151*, 68–75. https://doi.org/10.1016/j.cognition.2016.03.008

Johnsrude, I. S., & Buchsbaum, B. R. (2016). Representation of speech. in Gaskell, G., & Mirković, J. (Eds.). *Speech perception and spoken word recognition*. Taylor & Francis Group. https://doi.org/10.4324/9781315772110

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural

room impulse responses. *EURASIP Journal on advances in signal processing*, *2009*, 1-10. http://doi.org/ 10.1155/2009/298605

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594. https://doi.org/10.1037/xhp0000410

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 549–562. https://doi.org/10.1037/xlm0000744

Kim, D., Stephens, J. D. W., & Pitt, M. A. (2012). How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. *Journal of Memory and Language*, *66*(4), 509–529. https://doi.org/10.1016/j.jml.2011.12.007

Lenth, R.V. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.3. https://CRAN.R-project.org/package=emmeans

Li, M. Y. C., Braze, D., Kukona, A., Johns, C. L., Tabor, W., Van Dyke, J. A., Mencl, W. E., Shankweiler, D. P., Pugh, K. R., & Magnuson, J. S. (2019). Individual differences in subphonemic sensitivity and phonological skills. *Journal of Memory and Language, 107*, 195–215. https://doi.org/10.1016/j.jml.2019.03.008

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* Oxford Science Publications, ISSN 1362-9972.

Magnuson, J. (2016). Mapping spoken words to meaning. in Gaskell, G., & Mirković, J. (Eds.). *Speech perception and spoken word recognition*. Taylor & Francis Group.

https://www.semanticscholar.org/paper/Mapping-spoken-words-to-meaning-Magnuson/e884a295e113cfddce38949951012e6caf430374

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America*, *122*(1), 554–567. https://doi.org/10.1121/1.2735105

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in cognitive sciences*, *10*(8), 363-369. https://doi.org/10.1016/j.tics.2006.06.007

Meyer, G., & Barry, W. (1999). Continuity based grouping affects the perception of vowel-nasal syllables. In ICPhS 99: proceedings of the 14th International congress of Phonetic sciences, San Francisco, 1-7. Berkeley: University of California. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0203.pdf

Middlebrooks, J. C., Simon, J. Z., Popper, A. N., & Fay, R. R. (Eds.). (2017). *The Auditory System at the Cocktail Party* (Vol. 60). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2

Morais, R. K. J. (1996). Migrations in Speech Recognition. *Language and Cognitive Processes*, *11*(6), 611–620. https://doi.org/10.1080/016909696387015

Pastore, R. E., Szczesiul, R., & Rosenblum, L. (1984). Does silence simply separate speech components? *The Journal of the Acoustical Society of America*, *75*(6), 1904–1907. https://doi.org/10.1121/1.390955

Pitt, M. A., & Shoaf, L. (2002). Linking verbal transformations to their causes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 150–162. https://doi.org/10.1037/0096-1523.28.1.150

Poltrock, S. E., & Hunt, E. (1977). Individual differences in phonological fusion and separation errors. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 62. https://doi.org/10.1037/0096-1523.3.1.62

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing,* Vienna, Austria. URL https://www.R-project.org/.

Raphael, L. J. (2021). Acoustic cues to the perception of segmental phonemes. In Pisoni, D. B. & Remez, R. E. (Eds.). *The Handbook of Speech Perception*. Wiley Blackwell. 603-631. https://arakmu.ac.ir/file/download/page/1601463054-the-handbook-of-speech-perception.pdf#page=196

Remez, R.E. (2021). Perceptual organization of speech. In Pardo, J. S., Nygaard, L. C., Remez, R. E., & Pisoni, D. B. (Eds.). *The Handbook of Speech Perception*. Wiley Blackwell. 722. https://doi.org/10.1002/9781119184096.ch1

Repp, B. H. (1985a). Can linguistic boundaries change the effectiveness of silence as a phonetic cue? *Journal of Phonetics*, *13*(4), 421–431. https://doi.org/10.1016/S0095-4470(19)30787-9

Repp, B. H. (1985b). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 799. https://doi.org/10.1037/0096-1523.11.6.799

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 621. https://doi.org/10.1037/0096-1523.4.4.621

Samuel, A.G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12* (4), 348-351. https://psychology.illinoisstate.edu/jccutti/psych369/F04/Samuel.pdf

Sanders, L.D., & Neville, H.J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research, 43*, 1301-1321. https://doi.org/10.1044/jslhr.4306.1301

Sexton, M. A., & Geffen, G. (1981). Phonological fusion in dichotic monitoring. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 422. https://psycnet.apa.org/record/1982-00353-001

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L.H., Huang, M, Ding, N, Gao, J.H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, *29* (8), 3232–3240, https://doi.org/10.1093/cercor/bhy191

Shinn-Cunningham, B., Best, V., & Lee, A. K. C., (2017). Auditory object formation and selection. in Middlebrooks, J., Simon, J. Z., Popper, A. N., & Fay, R. R. (Eds.). *The auditory system at the cocktail party.* Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-51662-2_2

Uddin, S., Heald, S. L. M., Van Hedger, S. C., Klos, S., & Nusbaum, H. C. (2018). Understanding environmental sounds in sentence context. *Cognition*, *172*, 134–143. https://doi.org/10.1016/j.cognition.2017.12.009

Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, *70*(4), 261–270. https://doi.org/10.1037/h0026275

Whaley, C. P. (1978). Word—Nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X

Appendix A: Experiment 1 Stimuli

Bias Voiced	Unbiased	Bias Voiceless
spack – back	span – ban	sparse – barse
spad – bad	spare – bare/bear	spawn – bon
spall – ball	spark – bark	spew – bew
sparn – barn	spay – bay	spine – bine
spath – bath	speak – beak	splat – blat
spest – best	spear – beer	splay – blay
spid – bid	speed – bead	splice – blice
spirth – birth	spill – bill	split – blit
spomb – bomb	spin – bin	spoke – boke
spone – bone	spoon – boon	sponge – bonge
spox – box	spore – bore	spook – buke
spoy – boy	spot – bot	sport – bort
sprag – brag	spud – bud	spouse – bouse
sprand – brand	spun – bun	sprint – brint
spus – bus	spunk – bunk	sprout – brout
spuuk – book	spy – bye/by	spry – bry
stad – dad	stab – dab	stack – dack
stam – dam/damn	stare – dare	staff – daff
stawn – dawn/don	start – dart	stage – dage
sten – den	stay – day	stake/steak – dake
stesk – desk	steal/steel – deal	stand – dand
stime – dime	steep – deep	star – dar
stip – dip	steer – dear/deer	stoke – doke
stirt – dirt	stew – dew/do	stone – doun
stish – dish	still – dill	stool – dool
stive – dive	sting – ding	stoop – doop
stog – dog	stock – dock	stop – dop
stot – dot	store – door	street – dreet
strag – drag	stub – dub	strike – drike
strink – drink	stun – done/dun	string – dring
stuke – duke	stunk – dunk	strong – drong
stust – dust	stye – dye/die	stuck – druck

Survey Questions

1. What was the first language you learned to speak? *(open-ended)*
2. If not English, at what age did you learn to speak English? *(open-ended)*
3. How hard did you find the experiment? *(choose between "very difficult", "slightly difficult", "neither easy nor difficult", "slightly easy", "very easy")*
4. Did you feel that the instructions were sufficient to know what the task was about? If not, then what could be done to make the instructions better? *(open-ended)*
5. Did you feel that the number of practice items was enough? *(choose between "Yes, that was enough practice", "No I could have used LESS practice", "No I could have used MORE practice")*
6. Did you use both response options equally? *(choose between "Yes my f and j button responses were about equal", "No, I responded with more j responses (B/D)", "No, I responded with more f responses (P/T)")*
7. Which response method would you have preferred? *(choose between "I would prefer to use the keyboard buttons", "I do not have a preference OR I liked the method used here", "I would prefer to type my responses")*
8. What do you think the purpose of this experiment was? *(open-ended)*
9. Did you hear anything other than the words/sentence you responded to? If so, describe what you heard. *(open-ended)*
10. Do you have any other comments about the experiment? *(open-ended)*

Appendix B: Experiment 2 Stimuli

Voiced Bias

The ball player knew she had to get to home **base**
The government just imposed a new travel **ban**
I deposited my money at the **bank**
She went to the zoo to look at the polar **bear**
Carl muzzled his dog so that it would not **bark**
When hitting a baseball you need to use a **bat**
A small inlet near the ocean is a **bay**
I just saw a parrot break a nut with her **beak**
The bar makes its money selling fresh cold **beer**
My most worn necklace is now missing a **bead**
Tom's favorite exercise is riding his new **bike**
After dinner the waiter brings us our **bill**
Make sure to put the recycling in the **bin**
The new dentist's drill only hurts a little **bit**
Another word for favor is a **boon**
The teacher who droned on and on was a real **bore**
You must be excited about the game you just **bot**
My dog is not just a pet but also my **bud**
Her hair was almost long enough to put in a **bun**
The soldier hated to be woken from her **bunk**
When your friends or family leave make sure to say **bye**
When cleaning a stain remember to **dab**
My friends all laughed when I refused the **dare**
Nothing hurts worse than being hit by his **dart**
The sun only comes out during the **day**
The two business partners finally closed the **deal**
The Grand Canyon is both wide and **deep**
The most shy animal in the forest is a **deer**
In the morning outside I see the wet grassy **dew**
I like all pickles but my favorite flavor is **dill**
The bell on our cat's collar will often **ding**
When making a soup don't forget the **stock**
The grocer sold many veggies in the **store**
Lifting heavy objects puts the rope under **strain**
Before getting in the shower I have to **strip**
When riding on a train you need your ticket **stub**
A risk of cave diving is getting **stuck**
The chainsaw cut down the tree at the **stump**
The magician's tricks always amaze and **stun**
After a hard workout Sheila knew that she **stunk**
My room is so dirty it looks like a pig **stye**

Voiceless Bias

The rocket ship successfully launched into **space**
Drinking alcohol shortens your concentration **span**
When children get in trouble some parents **spank**
We had too many candles so I sold the **spare**
Putting new wood on the fire created that **spark**
Their argument became louder halfway through the **spat**
Please help me catch these feral kittens to **spay**
When Mia is mad she refuses to **speak**
In my backyard I practice throwing the sharp **spear**
You won't get a ticket if you drive the right **speed**
The volleyball champion won the game with a **spike**
The unbalanced cup looks like it will **spill**
When I turn in circles my head starts to **spin**
In the summer Randy roasts a pig over a **spit**
The best way to eat soup is with a **spoon**
My allergies are triggered by one pollen **spore**
Her dog's fur was mostly white with only one dark **spot**
My old grandpa calls the potato a **spud**
Miriam twisted the wheel and watched as it **spun**
The courageous teenager had plenty of **spunk**
Adam was secretly a government **spy**
Be careful as even a dull knife can **stab**
Allen could not look away from my **stare**
She moved to the city to get a new **start**
When Max's dog hears the command it will **stay**
The burglars knew the best items to **steal**
The type of tea I drink needs to **steep**
When driving over ice it can be hard to **steer**
Potatoes onions and carrots all go in the **stew**
The water near the boat was not choppy but **still**
Be careful around bees as they can **sting**
I can only access my boat from the **dock**
After work I drive home and open my **door**
When our pipe is clogged water stays in the **drain**
When Sally's hair is wet it tends to **drip**
Cam watched foreign films without an English **dub**
I almost got hit but I managed to **duck**
On Mondays we take the garbage to the **dump**
Jan spends her time at work waiting to be **done**
The basketball player stunned the crowd with their **dunk**
The tee shirt was white and ready for the tie **dye**

Foil Voiced

I won’t make you follow the club's new **ban**.
Lydia always hates going to the **bank**.
Occasionally he would stand and **bark**.
The drink is made with a lemon **base**.
Lisa has always been frightened of **bats**.
Karen works with many types of **bead**.
Some types of fish also have a **beak**.
When walking in the forest I came across a **bear**.
When Rhonda is done she wants a **beer**.
Oscar couldn't decide whether to walk or take the **bike**.
The worst thing about dinner was the **bill**.
At the children's party my wife got **bit**.
These people are impossible to **bore**.
My invention is part man and part **bot**.
The new bar makes the best sandwich **bun**.
I removed the food from my lips with a **dab**.
Hailey convinced her friends to follow the **dare**.
This game requires you to aim and throw a **dart**.
After much arguing they had a **deal**.
The animals who eat my plants are the **deer**.
Chronic illness can make it hard to **dial**.
The chicken is seasoned with rosemary and **dill**.
I wonder what sports my brother will **do**.
Nick did what he had to and then was **done**.
I ran into Meredith around the **door**.
She opened her eyes feeling the **drain**.
The most annoying sound is a **drip**.
The bird I see on the lake is a **duck.**
I found lots of working items at the **dump**.
Club music is a mix of funk, jazz, and **dub**.

Foil Voiceless

Lewis always cooks his soup in a **pan**.
Michael got in trouble for doing a **prank**.
Squirrels and rabbits live in the **park**.
When I get nervous I tend to **pace**.
The teacher gave the student's head a **pat**.
Jess ran to tell her mom she had **peed**.
I won’t blame you if you try to **peek**.
My daughter did not ask to take a bite out of my **pear**.
Every day I go outside on the **pier**.
My favorite fish are tuna, salmon, and **pike**.
The child refuses to eat their **pill**.
Signs were put up so no one fell in the **pit**.
At night I often look at my dirty **pores**.
I cooked potatoes in my brand new **pot**.
I hate it when my friend starts making a **pun**.
My laptop accidentally closed the **tab**.
If dad pulls too hard on the rope it could **tear**.
The dessert I ate was a little too **tart**.
Mom's shirt color is a certain shade of **teal**.
While Harry was happy he still felt a **tear**.
It's so tacky to make the bathroom floor **tile**.
Playing the violin you might have to **trill**.
My sister wants to play with them **too**.
Weight is measured in a gram, pound, or **ton**.
In the scuffle Nancy's new shirt got **torn**.
As an educator my job is to **train**.
Carlton did not enjoy the long **trip**.
She was suspicious and followed their **truck**.
During the game I laid the ace as a **trump**.
The best types of food arrive in a **tub**.

Appendix C: Experiment 4 Stimuli

Voiceless Bias	Unbiased	Voiced Bias
The rocket ship successfully launched into **space**	We tried unsuccessfully to take over the **[s]base**	The ball player knew she had to get to home **base**
We had too many candles so I sold the **spare**	Inside the garage is where Sally keeps the **[s]bare**	She went to the zoo to look at the polar **bear**
Putting new wood on the fire created that **spark**	Maria jumped after being startled by the **[s]bark**	Carl muzzled his dog so that it would not **bark**
In my backyard I practice throwing the sharp **spear**	The unruly and rowdy kids grabbed at the **[s]beer**	The bar makes its money selling fresh cold **beer**
The volleyball champion won the game with a **spike**	Joe stepped in the yard and hurt his foot on the **[s]bike**	Toms favorite exercise is riding his new **bike**
The unbalanced cup looks like it will **spill**	The waiter was expected to clear the **[s]bill**	After dinner the waiter brings us our **bill**
Be careful as even a dull knife can **stab**	The thanksgiving turkey is the thing we **[s]dab**	When cleaning a stain remember to **dab**
Allen could not look away from my **stare**	My best friend Peter gave me a funny **[s]dare**	My friends all laughed when I refused the **dare**
She moved to the city to get a new **start**	The man told them around back was where to **[s]dart**	Nothing hurts worse than being hit by his **dart**
The burglars knew the best items to **steal**	The buildings construction relied on the **[s]deal**	The two business partners finally closed the **deal**
When driving over ice it can be hard to **steer**	They slowly tried to feed a pear to the **[s]deer**	The most shy animal in the forest is a **deer**
Potatoes onions and carrots all go in the **stew**	My shoes were ruined after stepping on the **[s]dew**	In the morning outside I see the wet grassy **dew**
The grocer sold many veggies in the **store**	I hope they will eventually close the **[s]door**	After work I drive home and open my **door**
Lifting heavy objects puts the rope under **strain**	The economy was experiencing a **[s]drain**	When our pipe is clogged water stays in the **drain**
Before getting in the shower I have to **strip**	David’s hairy leg hurt from the hot waxy **[s]drip**	When Sally’s hair is wet it tends to **drip**