HOW PERCEPTION CONSTRAINS STATISTICAL LEARNING ACROSS DEVELOPMENT

by

Alexis K. Black

B.A., The University of Virginia, 2003

M.A., The University of Virginia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Linguistics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

How perception constrains statistical learning across development

submitted by Alexis K. Black in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Linguistics

Examining Committee:

Carla L. Hudson Kam
Supervisor

Douglas Pulleyblank
Supervisory Committee Member

Molly Babel
Supervisory Committee Member

Kathleen Currie Hall
University Examiner

James T. Enns
University Examiner

Abstract

This dissertation seeks to understand the underlying mechanism(s) of statistical learning (SL), defined as the capacity to extract structure from a perceptual stream by relying on the statistical properties of that stream (e.g., Aslin, 2017). I approach this question in two ways: by examining (1) the output representations of statistical learning (i.e., the quality of representations that emerge from a SL experience), and (2) the effect of input representations on SL (i.e., whether and how an individual's prior knowledge filters and shapes SL). I hypothesized that (i) learners' prior knowledge would impact the accessibility of units to SL, and thereby modify the process of learning; (ii) that SL is composed of more than veridical tracking of transitional probabilities between sounds; and (iii) that the interaction of prior knowledge and the underlying mechanisms of SL would relate to differences in learning outcomes across development.

To test these hypotheses, I created a novel testing paradigm of the word segmentation SL task, in which participants' knowledge of trisyllabic nonce words that were embedded in a continuous familiarization stream is probed by manipulating the nature of syllables in particular ordinal positions. Adult subjects were then tested on streams of speech that incurred varying degrees of perceptual load, either via the nature of the phonetic elements, or via an external and unrelated task. Children were similarly exposed to and tested on a stream of familiar sounds; I predicted that their performance should parallel that of adults under conditions of greater perceptual load.

The results of these experiments confirm that underlying perceptual representations impact learners' capacity for SL, and that the output of auditory SL tasks reflects more than the

underlying statistics embedded in a continuous stream. Performance does not rest on underlying phonetic representations alone; rather, differences in executive function skills additionally impact the SL process.

Lay Summary

Children appear to learn language with ease and speed – but how? One learning mechanism that may help with this task is called *statistical learning* (SL): the ability to unconsciously track the statistical properties of a stream of information, and extract structure based on these statistical properties. In this dissertation, I ask whether SL can actually lead to the kinds of representations that we would expect for language learning. In particular, I ask whether we can use SL processes to find words in a continuous stream of sound, and whether this process is impacted by the different perceptual and cognitive constraints of children versus adults. The results suggest that learners do in fact extract word-like chunks – and that these chunks reflect more than just the statistical relationships between sounds. Aspects of perception and cognition impact the learning process; unfamiliarity appears to limit learning, while lower attentional control may improve chunk extraction.

Preface

The experiments described in this dissertation were conceived and designed by Alexis K. Black, with guidance from Drs. Carla L. Hudson Kam, Molly Babel, and Douglas Pulleyblank. Data collection was performed by Alexis K. Black and by research assistants in the Language and Learning Lab and Living Lab, under the supervision of Alexis K. Black (Alannah Turner, Tess Forest, Chantane Yeung, Amane Halicki, Rose Aunaetitrakul, and Geneva Gamble). All data analyses were conducted by Alexis K. Black, with advice from Carla L. Hudson Kam, Molly Babel, and Douglas Pulleyblank. These projects were funded by an NSERC Discovery Grant (Individual) awarded to Carla L. Hudson Kam "Constraints on language acquisition and how they change (or don't) with age," and by a University of British Columbia Arts Graduate Research Award to Alexis K. Black (2014). All experiments reported on were approved by the University of British Columbia's Research Ethics Board (Adult experiments: #H12-0259; Child experiments: #H13-00740).

The following is a list of presentations and publications in which parts of the dissertation were first introduced.

Research from Chapter 2 was presented as a talk at the 41st Boston University Conference on Language Development (Black & Hudson Kam, 2016a), and as a poster at the Acoustical Society of America Conference (Black & Hudson Kam, 2013). Both the talk and poster were written by Alexis K. Black with assistance from Carla L. Hudson Kam. This chapter also makes up part of a manuscript that has been submitted for publication.

The data in Chapter 3 were presented as a talk at the Fifth Implicit Learning Seminar (Black & Hudson Kam, 2016b). This talk was written by Alexis K. Black with assistance from

Carla L. Hudson Kam. A form of the chapter makes up a manuscript that has been submitted for publication.

The results presented in Chapter 4 were presented as a poster at the First Interdisciplinary Advances in Statistical Learning Conference (Black & Hudson Kam, 2015). This poster was created by Alexis K. Black with input from Carla L. Hudson Kam.

Table of Contents

List of Tables

List of Figures

List of Symbols

The following symbols from the International Phonetic Alphabet are used:

p^h: aspirated voiceless bilabial plosive stop

p': ejective voiceless bilabial plosive stop

t^h: aspirated voiceless alveolar plosive stop

t': ejective voiceless aveolar plosive stop

c': ejective voiceless palatal plosive stop

k^h: aspirated voiceless velar plosive stop

k': ejective voiceless velar plosive stop

b̥: voiceless bilabial plosive stop

b: voiced bilabial plosive stop

ɓ: voiced bilabial implosive stop

d̥: voiceless alveolar plosive stop

d: voiced alveolar plosive stop

ʄ: voiced palatal implosive stop

g̊: voiceless velar plosive stop

g: voiced bilabial plosive stop

ɠ: voiced palatal implosive stop

ɹ: voiced alveolar rhotic

l: voiced alveolar liquid

r: voiced alveolar trill

ʀ: voiced uvular trill

i: high front unrounded vowel

y: high front rounded vowel

ɨ: high central unrounded vowel

u: high back rounded vowel

ʊ: high near-back rounded vowel

ɯ: high back unrounded vowel

o: mid back rounded vowel

œ: mid front rounded vowel

a: low back unrounded vowel

ʌ: low-mid back unrounded vowel

ɒ: low back rounded vowel

List of Abbreviations

EL: English language

ERP: Event-related potential

NEL: non-English language

SL: statistical learning

SEL: Semi-English language

TP: transitional probability

2AFC: 2-alternative forced choice

Acknowledgements

There is no way to adequately describe how grateful I am to the people who have helped me along this journey.

First, I am deeply indebted to Carla L. Hudson Kam. Carla, you inspire me to ask deeper questions and make stronger arguments. Your unfailing kindness and empathy have seen me through some of my most difficult moments. You are my model of what it means to be a good mentor. My (academic and personal) life is much richer for having had the opportunity to learn from and work with you these last few years – and I hope for years to come. Thank you.

I would also like to thank my committee members, Douglas Pulleyblank and Molly Babel. Molly and Doug, every change requested, every complaint raised as you have read these original drafts has improved my thinking and writing. Thank you for taking the time and considerable effort to help me think through the theory, the questions, and the logic of these studies, and pushing me to think more deeply about speech perception and phonological representations. Thank you both also for the personal support you have given me over the years of graduate school.

I would also like to thank the numerous mentors and colleagues that have been fundamental to my growth as a scientist. In particular, Dr. Janet F. Werker, who took me under her wing when I first arrived at UBC and has fundamentally shaped my understanding of language acquisition, and all the members of the Infant Studies Centre, Dr. Joseph Stemberger, my collaborators Christina Bergmann, Laura Batterink and Sheri Choi, and the mentors who struck the initial spark, inspiring my love for language sounds: Professors Mark J. Elson and Stephen Dickey. I would also like to thank the numerous research assistants who have helped run, code, recruit, and problem-solve these studies over the years, and all of the participants, in

particular the children and their parents who consented to listen to some boring language sounds in the middle of their exciting excursions to the Science Museum.

Finally, this work would not have been possible without the support and love of my family and friends. Thank you to my parents, for teaching me to think critically, and the importance of perseverance. To my husband, David: you continually force me to be a better version of myself, academically and personally. I love you, and the life we have created together – thank you for suffering through this last year with me. And Jacob – you may not remember much from this time, but I will never forget how much joy and life and love you bring to me and your Abba, every single day. We love you with all our hearts.

for Nonnie

Chapter 1: Introduction

In our daily lives, we are inundated with streams of sights, sounds, smells, and tactile sensations. This experience is guided and streamlined by a set of expectations about how the world works. Yet how do we form these expectations? There is, of course, no single mechanism that can account for learning of all the perceptual categories that constrain this flow of sensory information. Over the last few decades, however, one mechanism has been implicated as a fundamental contributor: statistical learning. Statistical learning (SL) is – roughly – the capacity to induce structure from statistical patterns that are distributed across continuous streams of sensory input (Saffran, Aslin & Newport, 1996; Maye, Werker, & Gerken, 2002; Thompson & Newport, 2007). This capacity has been successfully demonstrated across perceptual domains, (e.g., vision: Kirkham, Slemmer, & Johnson, 2002; audition: Saffran, Johnson, Newport, & Aslin, 1999; touch: Conway & Christiansen, 2005; visuomotor: Hunt & Aslin, 2001), is relatively automatic and robust to sensory interference (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Turk-Browne, Scholl, Chun, & Johnson, 2009; cf. Toro, Sinnet, & Soto-Faraco, 2005), and is operable by the time an infant is born (Teinonen, Feldman, Näätänen, Alku, & Huotilainen, 2009; Bulf, Johnson, & Valenza, 2011; Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011).

While the power and ubiquity of SL has made it a compelling mechanism for theories of perceptual learning generally (Aslin, 2017), there is no domain in which it has had more of a theoretical impact than that of language acquisition (see, e.g., Kuhl, 2004; Aslin & Newport, 2012). Indeed, SL has been hypothesized to contribute to the acquisition of nearly every level of linguistic hierarchy: phonological categories (Maye et al., 2002; Noguchi & Hudson Kam, 2017),

words (Saffran et al., 1996; Graf Estes, Evans, Alibali, & Saffran, 2007), syntactic classes and combinatorial rules (Saffran & Wilson, 2003; Thompson & Newport, 2007; Finn, Lee, Kraus, & Hudson Kam, 2014), and semantic networks (Smith & Yu, 2008; Yurovsky, Yu, & Smith, 2013). Yet from the earliest days of the SL literature, researchers have disagreed about the nature of the computational and perceptual processes that underlie it, asking, for example, whether learners compute the statistical relationships between units (see Saffran & Kirkham, 2018), or if the input is chunked and encoded in a way that has no direct relationship to the underlying statistics, but yields comparable final structures in memory (see Thiessen, 2017).

In this dissertation, I examine the nature of the output of auditory statistical learning (SL) as a means of elucidating the mechanism(s) that underlie it. I propose that SL involves more than (or something other than) tracking the statistical relationships between sounds, and that evidence of these non-statistical learning processes is reflected in the representations that learners form after exposure to a continuous stream of sounds. I further hypothesize that differences in underlying representations (as realized by developmental change, or by altering the language-learning conditions within an age group), will impact this learning process. I argue that the experimental data (1) provide support for learning mechanisms beyond statistics-tracking during a statistical learning task, and (2) reveal nuanced influences of perception and executive function on the learning process.

In the remainder of this chapter, I lay the groundwork for understanding the type of SL that is the focus of the dissertation (Section 1.1), outline what is currently understood about the outcome of auditory SL (Section 1.2), and the impact that underlying representations have on learning outcomes (Section 1.3), and finally will discuss the paradigm that was designed to further probe these two aspects of SL (Section 1.4).

1.1 Background

The idea that learners can use statistical cues in their environment to induce linguistic categories has a long history (e.g., Harris, 1955; Hayes & Clark, 1970); however, it was a seminal study by Saffran, Aslin, and Newport (1996) that brought the idea to the forefront of theories of language acquisition. In this study, the authors addressed the fundamental dilemma of word-segmentation: how do infants learn where word boundaries are when there are no unique, consistent phonetic cues to signal them, either across languages (e.g., Cutler & Carter, 1987), or – even more strikingly – within a single language (e.g., Cole & Jakimik, 1980; Klatt, 1980; Dumay, Content & Frauenfelder, 1999)?[1] The authors propose two hypotheses: (1) sequence transitions within words occur with higher probability than those across word boundaries,[2] and (2) infants can use this information to postulate word boundaries (Saffran et al., 1996).

To test this second hypothesis, Saffran, Aslin, and Newport presented 8-month-old infants with a brief, continuous stream comprised of four trisyllabic nonce words that repeated in a semi-random order. Importantly, the 12 unique syllables that made up these words were

---

[1]There are, of course, a number of cues that infants might recruit for word segmentation (e.g., prosody (Jusczyk, Houston, & Newsome, 1999), phonotactic constraints (Mattys, Jusczyk, Luce, & Morgan, 1999), coarticulation (Juszcyk, Hohne, & Bauman, 1998), or isolated words (Brent & Siskind, 2001)). These cues, however, are inconsistently correlated with word boundaries, require knowledge of at least some words in order to interpret their relationship to word boundaries, and have generally been shown to be relied on by infants after word-learning is known to have begun (see Jusczyk, 1999 for review on timing of acquisition of relevant cues, and Bergelson & Swingley, 2012, Tincoff & Jusczyk, 2012, Bergelson & Aslin, 2017, for evidence of earlier acquisition of words). The SL hypothesis was offered as an initial stepping stone – a pre-linguistic device that would enable infants (in any linguistic environment) to induce a handful of words, that might then promote learning of the myriad phonetic cues that are ultimately more informative of wordhood.

[2]This proposal derived from the earlier work of Harris (1955), Goodsitt, Morgan and Kuhl, (1993), among others.

stripped of any informative prosodic or phonetic cues to the word boundaries; to extract the boundaries, the infant learners had to recruit the differential statistical relationship between syllables within the words as opposed to across them. This statistical relationship was defined as transitional probability (TP), specifically, the frequency that two sounds (in this case syllables) co-occur, as a proportion of the raw frequency of one of them. Within the stream, certain syllables occurred with a very high TP (these were called "words"), while other syllable sequences occurred with a much lower TP. After infants were familiarized to this continuous stream of syllables, they were tested on their knowledge of the underlying TP-defined structure through the Head-turn Preference Procedure (Kemler-Nelson, Jusczyk, Mandel, Myers, Turk & Gerken, 1995), a paradigm in which infants' discrimination of different categories of stimuli can be tested via looking-time preferences.[3]

Two versions of the study tested infants' discrimination of words (the high TP trisyllabic sequences that made up the familiarization stream) versus different types of non-word foils. In the first version, infants heard either a word or a novel trisyllabic combination (called non-word) composed from the same set of 12 syllables. Infants had longer looking times when listening to non-words, indicating that they had distinguished the two types of structures (Saffran et al., 1996). In the second version of the study, infants were tested on discrimination of words versus part-words – trisyllabic combinations that they had encountered during familiarization, but which crossed a word-boundary (and so contained a 0.33 TP, instead of two 1.0 TPs). Again, infants

[3]In this procedure, infants are seated on a caregiver's lap in a small cubicle, and prompted to look to their right or left side by a blinking light. Once the infant attends to the blinking light, he/she hears a recording that plays repeatedly until the infant loses interest. This process is then repeated (alternating sides) until the trial list is exhausted.

attended longer to the foil than to the words (Saffran et al., 1996). The same authors subsequently demonstrated that this performance was due to entrainment specifically to transitional probabilities, as opposed to simple co-occurrence frequencies (a potential confound in the original 1996 study), by staggering the relative frequency of words during familiarization such that infants could only use transitional probabilities to distinguish words from part-words at test (Aslin, Saffran, & Newport, 1998).

Since this time, SL has come to permeate theories of language acquisition and perceptual learning more generally (see Kuhl, 2004; Pierrehumbert, 2003; Fiser, Berkes, Orban, & Lengyel, 2010; Aslin, 2017; Santolin & Saffran, 2018 for reviews). It has been touted as a domain-general learning mechanism that is evidenced across species (cotton-top tamarins: Hauser, Newport, & Aslin, 2001; rats: Toro & Trobalón, 2005; songbirds: Takahasi, Yamada, & Okanoya, 2010) and the developmental span (newborns: Teinonen et al., 2009; Bulf et al., 2011; elderly: Schwab, Schuler, Stillman, Newport, Howard Jr. & Howard, 2016 (mean age 74-years-old)). And, as is reflected in the terminology I've used in the preceding paragraphs – SL is frequently referred to as a *mechanism* of learning (e.g., Saffran, 2003; Kirkham et al., 2002; Santolin & Saffran, 2017). The ability to detect statistically defined structure, however, does not explain how human (or non-human) minds are capable of computing this information (see Thiessen, 2017, for discussion).

Consider the following definition of a TP-based learning mechanism: learners derive the TPs between sounds/syllables and then store these contingencies in memory. It is worth briefly considering what this process might actually entail. Let us assume the original Saffran, Aslin, and Newport (1996) design as a case study. Imagine that you as a learner are presented with the following brief stream of prosodically undifferentiated speech:

1) T U P I K O B I D A K U T U P I K O P A R O T I

What must you encode in order to compute the transitional probabilities? At the very least, you will need to store a memory trace of the syllable in question and its environment. As learners have been shown to be sensitive to both forward and backward probabilities (Perruchet & Peereman, 2004; Perruchet & Desaulty, 2008; Pelucchi, Hay, & Saffran 2009; French, Addyman, & Mareschal, 2011; Tummeltshammer, Amso, French, & Kirkham, 2017), this environment must include both the preceding and following syllables. If you encode the stream listed above sequentially, we might represent these memory traces as follows:

2) T U P I K O B I D A K U T U P I K O P A R O T I

TU.PI *(first syllable + its environment)*

TU.PI PI.KO *(second syllable and its environment)*

PI.KO KO.BI *(third syllable and its environment)*

KO.BI BI.DA …… (cont.)

This list represents a mere 2.67 seconds of input in the original TP design (Saffran et al., 1996). If we assume that each syllable is stored in memory in tandem with its immediate environment as a single entry, across 2 minutes (the exposure time of the traditional experimentation), the learner will lay down 718 memory traces. This suggests that even the simplest hypothesized mechanism underlying SL is non-trivial. An important issue I have thus far ignored, however, is the size of the unit to which the learner attends and which he/she stores

in memory. As a learner, I might choose to store and track the following (note: the following cues are but a subset of the range of possibilities):

3) T U P I K O B I D A K U T U P I K O P A R O T I

[t^h] [t^hu] *(first segment + its environment)*

(Release burst + 60 msec of aspiration]

(Release burst + 60 msec of aspiration; F2: transition from approximately 2100 Hz at voicing onset to 1800 Hz mid-vowel; F1: transition from approximately 240 to 380 Hz)

Indeed, adult speakers are known to track the subphonemic details of their speech environment (e.g., Pisoni, Aslin, Perey, & Hennessy, 1982; Nygaard & Pisoni, 1998; Goldinger, 1996; McMurray, Tanenhaus, & Aslin, 2002; McMurray, Tanenhaus, Aslin, Spivey, & Subik, 2003; Salverda, Dahan, Tanenhaus, Crosswhite, Masharav, & McDonough, 2007; Babel 2012). Moreover, the claim that infants can and do perceive subphonemic sound distinctions fundamentally underpins theories of infant phonological acquisition (e.g., Best, 1993; Werker & Curtin, 2005; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008). Thus, it seems likely that the process of detecting structure in continuous streams of sound involves encoding across a range of acoustic material – thus inflating what *may* already appear to be an overwhelming burden on perception and memory.

It is not only the relative size of the attended unit that incurs potentially exponential costs in memory, however – the period of time over which a learner continuously stores, updates, and

computes the relevant statistics must also be taken into account. Research has shown that traces of statistically defined structures extracted from brief, lab-based exposure to continuous streams can remain for a period of up to 24 hours (Kim, Seitz, Feenstra, & Shams, 2009; Durrant, Taylor, Cairney & Lewis, 2011; Arciuli & Simpson, 2012a). In real world experience, of course, there is no clear time-limit or break in the flow of information, or clear indication of which units will be relevant to patterning with which other units (Qian, Jaeger & Aslin, 2016) – thus it is unclear how learners would delimit this continual encoding, storing, and updating process. Taken together, then, these data prompt the question: are human memories capable of storing and computing statistical relationships over such staggering amounts of information?

Many have suggested that this is, in fact, not a reasonable model of human pattern-learning (see Perruchet, 2005; Thiessen, 2017, for review). As an alternative, Perruchet and Vinter (1998) proposed that, based on previous experience, perceptual primitives, and attentional resources, learners will automatically perceive an input string as dissociable chunks, rather than a continuous stream of primitives. In the case of word-segmentation, chunks that actually form a word, or a part of a word, will be repeated; this reinforces their memory trace (or representation). Chunks that were incorrect hypotheses, on the other hand, are less likely to be repeated, and will therefore fade from memory. This proposal was instantiated in a computational model, PARSER, and has been successfully applied to a range of linguistic data based on SL paradigms (Perruchet & Vintner, 1998; Perruchet & Peereman, 2004; Giroux & Rey, 2009; Perruchet, Vinter, Pacteau & Gallego, 2010; cf. Frank, Goldwater, Griffiths, & Tenenbaum, 2010). Other chunking models with similar assumptions have likewise found success, in both the auditory (e.g., TRACX2, Mareschal & French, 2017) and visual (Orbán, Fiser, & Aslin, 2008) domains. Further, attempts to pit different models against one another have frequently found support for chunking models

over sequential statistics-tracking models (Giroux & Rey, 2009; Frank, et al., 2010; Perruchet & Tillmann, 2010; Perruchet, Poulin-Charronnat, Tillmann, & Peereman, 2014).

There is as yet, however, no single theoretical model that can account for the full range of puzzles raised by the existing literature. For instance, learning across continuous streams of input is generally restricted to adjacent elements (Newport & Aslin, 2004; Creel, Newport, & Aslin, 2004), but there are unexpected exceptions (Peña, Bonatti, Nespor, & Mehler, 2002; Gebhart, Newport, & Aslin, 2009; Vuong, Meyer, & Christiansen, 2011). And, while learning of statistical relationships appears to be domain-general under some conditions (e.g., see Altmann, Dienes, & Goode, 1995 for evidence of transfer across domains), there appear to be domain-specific constraints (e.g., Conway & Christiansen, 2005; Emberson, Conway, & Christiansen, 2011) – which has led to proposals that the mechanisms themselves may differ by modality (Frost, Armstrong, Siegelman, & Christiansen, 2015). The extent to which SL correlates with linguistic knowledge or linguistic aptitude is also debated. For example, studies have demonstrated relationships between SL and verbal working memory (Misyak & Christiansen, 2012), sensitivity to syntactic structures (Kidd, 2012), or syntactic comprehension (Kidd & Arciuli, 2016), and vocabulary (Evans, Saffran & Robe-Torres, 2009). Yet, others have failed to replicate these relationships (e.g., Siegelman & Frost, 2015, find no relationship between SL and verbal working memory, syntactic comprehension, or rapid automatized naming (a correlate of vocabulary)). Finally, a recent meta-analysis of the SL word-segmentation literature has revealed theoretically unexpected relationships between the nature of the stimuli and infant performance (namely, only stimuli created by synthetic means reliably produced learning of the TP structure; Black & Bergmann, 2017).

Understanding the underlying mechanism(s) of SL will help to illuminate these current puzzles. Decoding the SL phenomenon is important both for theory-driven reasons (e.g., understanding the question of the extent to which language acquisition is driven by low-level perceptual mechanisms versus higher-order rule-based abstraction) and practical ones (e.g., understanding whether SL can be used as a diagnostic tool for communication disorders, or even as a training tool to boost implicit learning). In this dissertation, I address two questions that arise from the literature and I believe will contribute to our understanding of the underlying mechanisms: (1) what do we *learn* from a word-segmentation SL task? And (2) does a change in underlying representations lead to different (e.g., more/less abstract) learning outcomes? In the following paragraphs, I provide a brief review of what is known about the representations that are extracted from word-segmentation SL paradigms. I then turn to the impact that *input* representations may have on this process and the resultant learning.

1.2 What is the output of word segmentation via SL?

The original Saffran, Aslin, and Newport findings (1996) were taken as evidence that infants could use conditional statistics to extract words from continuous speech. It is worth pausing to consider what is meant by "word" in this context. Though often not discussed explicitly in the acquisition literature, the term "word" typically refers to a chunk of phonetic material that is maintained in memory and is associated with some constellation of semantic features/contexts. In other words – if an infant has extracted the phonetic chunk /da/, and recognizes that chunk as being associated with a particular context, it is sufficiently word-like to be considered a word. This chunk may not reflect the adult target – it might be either reduced (e.g., "da" for *dog*), or too large (e.g., "allgone"), consist of several morphemes or one (e.g.,

"singing"), and match some, all, or none of the adult target sounds (e.g. "bo" for *sun*). Its most salient feature is simply that it is a stable (but not static) acoustic form, recognized as a singular chunk by the infant. Typically, however, this form is paired with some consistent (even if low-level/underspecified) meaning. This is clearly not the case in the word-segmentation SL paradigm: unless there is training on sound-object pairings post familiarization, there is no obvious semantics for a learner to associate with acoustic structures "extracted" from continuous speech. Thus, it would seem that the SL literature posits that the outcome of SL is a stable acoustic form, available for association with semantics.

Several studies have demonstrated the viability of this definition. For example, after exposure to a continuous, TP-defined stream, infants learn semantic associations with high TP units more proficiently than with low TP units (Graf Estes et al., 2007; Hay, Pelucchi, Graf Estes & Saffran, 2011). Infants have also been shown to more readily incorporate high TP units learned from a continuous stream into fluent native language speech (Saffran, 2001). Finally, high TP sequences are better primes for pushing infants to establish new categories as opposed to low TP sequences (Erickson, Thiessen & Graf Estes, 2014). Yet these facts are consistent with two possible SL processes: either learners extract a particular structure that is then established in memory as an independent chunk (e.g., a word), or learners entrain to the veridical TP-structure, but do not extract independent chunks sans association with some additional cue (e.g., semantics, or a cue that is itself associated with boundaries, such as silence).

1.2.1 SL yields TP relationships

There are a number of reasons to suspect that the latter process is the case. For example: one puzzle in SL tasks is that participants rarely perform at ceiling in standard SL paradigms

(Siegelman, Bogaerts, & Frost, 2017), despite the fact that the optimal segmentation of the stream is clearly defined (e.g., four trisyllabic words of 1.0 TPs in the original Saffran et al., 1996, paper). If learners set word *boundaries* around the TP-defined word edges, we might expect that once even a single word has been extracted the others should soon follow (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Dahan & Brent, 1999). This does not appear to be the case. Not only are learners rarely aware of or particularly successful at explicitly identifying the underlying structure post-exposure, but even giving learners one of the high TP "words" in advance of exposure has been shown to have no facilitatory effect on learning (Finn & Hudson Kam, 2008). And, while there is some evidence that the presence of a familiar word enhances infants' ability to parse a continuous SL stream (Mersad & Nazzi, 2012), infants are surprisingly sensitive to the consistency of the underlying TP structure, and generally fail when the embedded words are of different syllable lengths (Johnson & Tyler, 2010; Mersad & Nazzi, 2012; cf. Erickson et al., 2014). Thus, while both children and adults appear to treat TP-defined nonce words as viable word candidates (Saffran, 2001; Graf Estes et al., 2007; Hay et al., 2011; Erickson et al., 2014), their failure to fully parse the stream suggests that learning consists of veridical TP tracking, as opposed to independent segmentation of multi-syllabic chunks.

Finally, one feature that characterizes the word-forms stored in adult lexicons is knowledge not just of the sequential nature of the embedded sounds, but also the relative positions of (at least some of) those sounds (MacKay, 1970; Marslen-Wilson & Zwitserlood, 1989; Swingley, Pinto & Fernald, 1999; Brown & McNeill, 1966; Allopenna, Magnuson, & Tanenhaus, 1998). In other words, the adult representation of the word "dog" consists both of the fact that /d/ is followed by /ɑ/, but also that /d/ is the initial sound in the word – a position that it shares with a large number of other possible words (e.g. "doll"). Work that has looked for

position-based encoding under SL conditions has met with largely negative results (Peña et al., 2002; Endress & Bonatti, 2007; Endress & Mehler, 2009a). For instance, in Peña et al. (2002), learners attended to a stream of trisyllabic sequences of the type A_1XC_1, in which A_1 was entirely predictive of C_1, but syllable X varied. Learners successfully used this long-distance dependency to segment the speech stream; however, they failed to generalize the relationship between A and C to novel X combinations. In fact, the longer the familiarization, the more likely participants were to choose low TP (i.e., non-word) trisyllabic sequences that they had encountered (e.g., $C_2\#A_1X$), as opposed to A_1XC_1 combinations with novel X syllables. When participants were given pre-segmented words (i.e., 'words' were flanked by brief pauses), however, they quickly extracted the necessary generalization, and picked forms that followed the A_1XC_1 rule, irrespective of adjacent TPs. Endress and Bonatti (2007) and Endress and Mehler (2009b) extended this work to show that learners can induce classes of syllables that belong in edges (the first or last syllable of multisyllabic words), but fail to do so with internal constituents – and, once again, can only do so when the words are bracketed with a prosodic cue (i.e., subliminal pauses between words, or final syllable lengthening). Taken together, these studies paint a picture of SL as a mechanism that involves primarily (or solely) the extraction of TPs between syllables, or adjacent to locally non-adjacent segments (Newport & Aslin, 2004; see Creel et al., 2004 for parallel results with pure tone stimuli).

1.2.2 SL yields independent chunks

And yet - there is additional evidence that the representations that emerge from SL bear independent, chunk-like features. It has been shown in studies of visual perception that once we perceive a whole – though this whole is constructed from a series of lower-level features –

conscious recognition of and memory for the lower-level features, both in on-line processing and in short-term memory, decreases (e.g., Poljac, de-Wit, Wagemans, 2012). Parallels to this gestalt-like phenomenon have been noted in both the auditory and visual SL literatures. In a study in which participants were trained on a continuous stream composed of di- and tri-syllabic words, Giroux and Rey (2009) found that performance on partial-word recognition (i.e., disyllables extracted from tri-syllable full words) suffered in comparison to full-word recognition after 10 minutes of exposure. After only 2 minutes of exposure, on the other hand, performance on these two types of stimuli was equivalent. These results suggest that statistically coherent "words" are reinforced in memory differently than are the sequences of the features from which they are built.

Fiser and Aslin (2005) similarly demonstrated in visual SL that learners' memories for sub-category features declines in comparison to their memories for the same features of images that are not grouped in a single category. Adult learners were exposed to visual arrays of novel shapes that were grouped into pairs or quadruples. When learners were tested on their discrimination of pairs they had experienced and novel pairs, they only succeeded when the familiar pair had not been embedded in a quadruple structure. Those pairs that had been embedded were indistinguishable from novel, unfamiliar pairs. More recently, Zhao and Yu (2016) demonstrated that adult learners' perception of the number of dots in an array reduces as a function of the statistically-defined embedded pairs. This finding was particularly striking, given that learners failed to distinguish high TP combinations from foils at test; in other words, exposure to the stream had not yet induced robust enough categories to withstand an explicit 2-alternative forced-choice (2AFC) test, but nascent category-level representations were already influencing learners' perception.

Furthermore, though most data show a lack of position-based encoding in words 'extracted' via SL, there are a number of findings that point to asymmetrical encoding of syllables across different locations during SL – that is, that people learn about certain parts of the word better than others. For instance, in both Saffran, Newport, and Aslin (1996) and Saffran, Johnson, Aslin, and Newport (1999) (and as discussed in Johnson, 2012), participants were better able to reject non-words of the structure ABX than of the structure XBC (where ABC represents the three syllables of a nonce word, and X reflects a randomly chosen syllable that did not occur in those sequences). By itself, this result is rather opaque. Perhaps the coherence of medial and final syllables is more strongly encoded than that of initial and medial syllables – this would then lead to better recognition of ABX as violating this coherence, while XBC would not. Alternatively, learners might have encoded the position of word-final syllables (but not/less-so the position of initial syllables), which would lead to easier detection of the word-final illicit syllable.[4]

Regardless of interpretation, however, the finding that encoding is asymmetrical across extracted sequences is echoed in a number of related paradigms. For example, Sanders, Newport, and Neville (2002) found larger N100 event-related-potentials (ERPs) to the initial syllable of embedded trisyllabic nonce words in comparison to both medial and final syllables. This was

[4]Saffran, Johnson, Aslin and Newport (1999) interpret the results as perhaps reflecting that infants calculate forward TPs; however, studies have since demonstrated successful apprehension of both forward and backwards TPs (Perruchet & Desaulty, 2008; Pelucchi, Hay & Saffran, 2009; Tummeltshammer, Amso, French, & Kirkham, 2017), and even some evidence that backwards TPs are *more* relied on than forward TPs (Perruchet & Peereman, 2004), rendering this a less convincing explanation.

replicated with newborn infants (Teinonen et al., 2009;[5] see Kudo et al., 2011 for a similar result, but opposite polarity deflection). Recently, this pattern has been replicated and extended by several studies demonstrating both larger N100 and N400 ERPs to the first syllable; the N400 appears to signify an advanced stage of segmentation (Abla, Katahira, & Okanoya, 2008; Mandikal Vasuki, Sharma, Ibrahim, & Arciuli, 2017). Though not TP-based, the findings from SL of artificial grammars similarly reveal asymmetrical knowledge. For example, learners exposed to a finite-state grammar of pure tones successfully distinguish novel licit from illicit sequences, however, they are differentially sensitive to final versus initial fragments, showing greater awareness of final sequences, at least in the auditory modality (Conway and Christiansen, 2005). Studies have also shown that additional cues used to segment an artificial language, such as vowel lengthening, are more facilitatory to segmentation when placed on final syllables than syllables in other locations (Cunillera, Gomila, & Rodriguez-Fornells, 2008; Tyler & Cutler, 2009), which further indicates position-based effects on SL.

In sum, SL appears to yield representations that can be built upon and transformed into independent chunks; however, it is less clear what these representations look like *before* transformation via association with additional cues. On the one hand, learners appear to be sensitive to a range of varying TPs (Goyet, Nishibayashi, & Nazzi, 2013; Bogaerts, Siegelman & Frost, 2016), and fail to postulate boundaries that would perfectly segment streams composed of very simple TP structures (Siegelman, Bogaerts, & Frost, 2017). And yet, learners' representations also appear – in some circumstances – to involve perceptual grouping of chunks

[5]Note: the negative deflection is temporally later, as would be expected with young infant neural responses; however, the difference between initial and final syllables is the point of interest here.

that are defined by high statistical coherence (Fiser & Aslin, 2005; Zhao & Yu, 2016), or to differ internally in ways that are not easily explained by differences in TPs (Saffran et al., 1999).

In this dissertation, I propose to better understand the underlying mechanism(s) of SL by carefully examining the representations that emerge from a SL experience. I pit two different accounts against each other. On the one hand, if SL is a process of veridically tracking TPs between syllables, output representations should reflect that TP-structure. On the other hand, if SL yields independent, word-like chunks, output representations should reflect a different kind of property – encoding of the position of syllables with respect to word boundaries. To do this, I systematically test learners' knowledge of syllable positions against their knowledge of the TP structure that they were exposed to. This paradigm is described in detail below (Section 1.4). First, however, I will outline the second manipulation that my design is probed to test: whether learners' prior knowledge affects their ability to encode and learn either the TP structure, positional information of syllables, or both.

1.3 Does prior knowledge change the learning process? And if so – how?

In the beginning of this chapter, I invited you to imagine the process of encoding transitional probabilities across a short span of acoustic material. One of the questions that this imaginary case scenario raised could be recapitulated as follows: is the amount of information that SL can operate over unbounded, at either a macro- or micro-scale (i.e., how large and how small are the 'units' that can be tracked)? The evidence suggests that the span for detecting relationships between units in a continuous stream is fairly limited. For example, while learners are capable of tracking the non-adjacent relationships between segments (e.g., the relationship between the two consonants in the sequence "**bid**o"), they do not appear to be able to do so

across syllables (e.g., the first and last syllable in the string "**bi**do**la**"), (Newport & Aslin, 2004), unless the syllables are presented not as continuous speech, but as pre-segmented "words" that are flanked by pauses (Gomez & Gerken, 1999; Gomez, 2002). This finding is reinforced in other statistical learning studies and the broader artificial grammar learning literature: while it is possible for learners to acquire non-adjacent dependencies in a SL experiment, a set of additional constraints limit how and when this can occur (Gomez, 2002; Creel et al., 2004; Onnis, Monaghan, Richmond, & Chater, 2005; Endress, 2010; Vuong, Meyer & Christiansen, 2011). It thus appears that, at the very least, the span over which conditional relationships are tracked is limited. It is less clear, however, to what extent learners are tracking adjacent relationships between syllables, phonemes or sub-phonemic acoustic signals.

For instance, underlying phonotactic knowledge (i.e., rules that govern permissible syllable structures or sound combinations and their positions within words) constrains SL. These constraints can take place at the level of word-forms (e.g., infants who are trained on disyllabic or trisyllabic word lengths prior to exposure to a continuous stream are subsequently limited to extracting same-length structures: Thiessen & Saffran, 2003; Lew-Williams & Saffran, 2012, Johnson & Tyler, 2010, cf. Thiessen, Hill & Saffran, 2005, Mersad & Nazzi, 2012), but are also in evidence at the level of combinations of segments (e.g., adults fail to segment words/morphemes that include non-native onset phoneme sequences when presented in continuous speech: Finn & Hudson Kam, 2008; Finn & Hudson Kam, 2015). The fact that illicit segment sequences inhibit SL, however, does not necessarily mean that SL itself requires tracking of *segments*. In other words – once the learner knows the phonotactic norms, the relevant phonetic cues drive attention. In the case of continuous streams where a phonotactically illicit sequence has been placed within an experimentally defined word boundary, the learner's

prior knowledge of how segments combine (or do not combine) to make words might inhibit perception of that sequence as a single unit, or inhibit tracking of transitional probabilities across those sounds – which in turn would inhibit SL.

On the other hand – the literature beyond the SL word-segmentation paradigm reveals that statistical learning extends to the sub-syllabic level. For example, it has been shown that young infants induce phoneme-like categories by attending to the distributions of tokens (produced as isolate monosyllables) along a continuum (Maye, et al., 2002; Yoshida, Pons, Maye & Werker, 2010). Adult learners can do the same (Escudero, Benders, & Wanrooij, 2011; Wanrooij, Escudero, & Raijmakers, 2013; Escudero & Williams, 2014); moreover, adults can learn allophonic patterns based on contextual distributions of tokens (Noguchi & Hudson Kam, 2017). Importantly – learners exposed to this kind of stimulus are able to extend the newly learned generalization to a novel segmental contrast (Maye, Weiss, & Aslin, 2008) – thus confirming that, whether learners are tracking syllables or sub-syllabic units – they can acquire generalizations at the sub-syllabic level. In other words – learners of all ages attend to and track distributions of signals at a sub-phonemic level; there is no independent evidence (as yet) to suggest that these same signals are not also available for learning of transitional probabilities.

This makes a simple prediction: as infants and children employ phonological representations that differ considerably from the adult targets (see, e.g., Werker & Tees, 1984; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Best, 1993; Rost & McMurray, 2010, and many more), learning that involves tracking of phonemes or phonetic units should differ – in some way(s) – across development. While developmental differences in SL have been noted in the visual SL domain (Bulf et al., 2011; Arciuli & Simpson, 2011), there is surprisingly little account of any such difference in the auditory domain (see Raviv & Arnon, 2017, for review).

There are several reasons this might be the case. First, it is possible that there truly is no difference in performance across development on auditory SL. This would suggest that infants, children, and adults are all tracking and computing statistics across the same perceptual primitives. I will argue in the paragraphs that follow that this scenario is unlikely; however, it is worth noting that it is not impossible. Though children and adults eventually wield representations that can operate at the level of a phoneme, research has suggested that infants are born organizing speech perceptually at the level of syllables (Bertoncini & Mehler, 1981; Bijeljac-Babic, Bertoncini, & Mehler, 1993; Räsänen, Doyle, & Frank, 2018), and that this perception of the speech signal continues as an age-invariant primitive into adulthood (Massaro, 1972; Healy & Cutting, 1976; Mehler, Yves Dommergues, Frauenfelder, & Segui, 1981; Greenberg, 1999). Perhaps, then, the lack of change across development on word-segmentation tasks is due to perceivers' ability to use the same set of perceptual primitives to accomplish the task.

An alternative possibility, however, is that auditory SL does differ as a function of the underlying representations brought to the task, and that we have simply not tested infants, children and adults on sensitive enough measures to compare their learning trajectories or outcomes. There is abundant evidence to suggest that learners' prior knowledge states impact (both negatively and positively) their performance on auditory SL tasks. For example, as discussed previously, infants typically fail to segment languages that are composed of different length words (e.g., di- and trisyllables). They can succeed on mixed-length streams, however, if a familiar word is embedded in the speech stream (Mersad & Nazzi, 2010). On the other hand, learning is impeded when learners are faced with highly unfamiliar sounds. Adults exposed to non-linguistic noises require five times the exposure as that needed for successful segmentation

of streams made from comparatively acoustically simple sine-wave tones or familiar language sounds (Gebhart et al., 2009). Impaired learning can also be seen in individuals' ability to generalize from the patterns extracted during statistical learning: when exposed to a stream of familiar, not acoustically distorted syllables, learners are capable of recognizing the extracted patterns even under severe acoustic distortion at test. When learners are trained on a stream that exhibits that same degree of acoustic distortion, however, they are only able to recognize the distorted pattern; i.e., learners were unable to recognize the same sequences in familiar, non-distorted versions of the syllables (Vouloumanos, Brosseau-Liard, Balaban, & Hager, 2012). Infants similarly show reduced levels of learning when confronted with unfamiliar, complex sounds: 14-month-olds exposed to non-native speech sounds fail to discriminate high- from low-TP items, but successfully discriminate high- from zero-TP items (Graf Estes, Gluck, & Bastos, 2015). Finally, there is also evidence from atypically developing populations that stimulus familiarity changes SL trajectories. In a study with children with specific language impairment, Evans, Saffran and Robes-Torres (2009) found that both typically developing children and children with SLI could successfully segment a language comprised of speech sounds, but only the typically developing children were able to segment a language made of pure tones. The results of this study indicated that children with SLI struggled with SL in general (they required double the amount of exposure to learn the structure of the language-sound stream), but the fact that they struggled even more on the relatively unfamiliar tones reveals the increased difficulty that may be incurred by acoustic novelty.

Results such as these suggest that a reduction in stimulus familiarity leads to a reduction in learning. What exactly does 'reduction' of learning mean, however? Without a clear understanding of the mechanism(s) underlying SL, it is difficult to say. One possibility is that the

sounds encountered are less stable and/or precise in memory, making the TP calculation less precise. This might, then, lead younger learners (or adults learning across novel sounds) to struggle more at discriminating high-TP versus low-TP sequences as compared to high-TP versus zero-TP sequences. The original Saffran, Aslin, and Newport (1996) study failed to find a difference in performance on these two contrast types; however, this effect (if it exists) is likely smaller than the overall effect of learning – that is, we would be unlikely to detect the difference in a single study with a relatively small sample.

Alternatively, reduced learning might refer to the span across which learning can take place. As outlined above, Endress and Mehler (2009a) tested adult learners on non-word foils that had high syllable-adjacent TPs, but zero non-adjacent TPs (called "phantom words"), and found that learners were incapable of distinguishing the foils from actual sequences encountered during familiarization. A subsequent failure to replicate the effect (Perruchet & Poulin-Charronnat, 2012) prompted the authors to propose that the different learning outcomes derived from different underlying representations: while the subjects of the failed replication were listening to native-language sounds (i.e., French speakers hearing French sounds), the subjects in Endress and Mehler (2009a) were not (i.e., Italian speakers hearing French sounds). One interpretation of these two studies is that reduced familiarity with the stimuli constrained learners to adjacent TPs, whereas greater familiarity led to learning of both adjacent and non-adjacent TPs.

This interpretation, however, presumes that SL is a mechanism that tracks TPs. If learning actually involves processes of chunking and associative memory, there are a number of additional possible features that might correlate with reduced learning. I propose that one possibility is that non TP-based features of learning – such as encoding of the positional

information of syllables within a trisyllabic word – may emerge more clearly under conditions of reduced stimulus familiarity. For example, an unintended consequence of the Endress and Mehler (2009a) design is that the phantom-words participants were tested on involved trisyllabic combinations in which all syllable positions were maintained across the sequence, though they had not occurred as a unit in the familiarization stream. Their participants – who were learning from non-native speech – found these items more confusing than the Perruchet, Poulin, and Charronnat (2012) participants, who were learning from native speech.

1.4 The proposal

I propose to elucidate the underlying mechanism(s) of SL by examining (1) the nature of the representations that results from exposure to a continuous, statistically deterministic auditory stream and (2) whether and how manipulating the accessibility of that input auditory stream impacts the learning outcomes. I hypothesize that (1) learners' prior knowledge would impact the accessibility of units to SL, and thereby modify the process of learning; (2) SL involves more than veridical TP-tracking; and (3) the interaction of prior knowledge and the underlying mechanisms of SL will relate to differences in learning outcomes across development.

In Chapter 2, I address the first two of these hypotheses by exposing adult native-English speakers to native-English, semi-English, and non-English sounds in a continuous, TP-structured stream. I systematically tested whether the learners' extracted representations were based primarily on TP-strength, or would be asymmetric across syllable positions within trisyllabic chunks. Based on the findings of the first three experiments, I created a fourth in which I used the same paradigm to test learners on native-English sounds, but taxed their ability to perceive

and encode the speech stream by introducing a secondary, attention-demanding task – watching an engaging, silent cartoon.

In Chapter 3, I propose that multilingualism, musical skill, age, and specific language experience are factors that will impact an individual's ability to efficiently encode the speech stream, and hence alter his/her ability to learn from that stream. To explore this possibility, I re-analyze the data from the first three experiments of Chapter 2 for relationships with these individual difference factors.

Chapter 4 asks the same set of questions and addresses my third hypothesis by testing the same paradigm in a developmental sample. I tested 7- to 13-year olds on their ability to segment a TP-structured stream, and asked whether there are developmental shifts in learning generally, and specifically whether age-based change(s) related to evidence of position-based or TP-based encoding. I also examined the same set of individual difference factors as were explored in Chapter 3.

Finally, in Chapter 5 I review the findings of these experiments and discuss their implications for the future study of statistical learning. It is important to note that the paradigm used in this thesis is not designed to reveal a specific mechanism underpinning SL; rather, it is designed to reveal the *nature* of the underlying mechanism. In other words, a mechanism that tracks TPs across syllables (however that process occurs) is predicted to yield a particular kind of representation, while a mechanism that creates independent chunks (however that process occurs) is predicted to yield a different kind of output representation. Future work will be necessary to take these behavioural results and derive the pattern of performance through specific computational and neurobiological means (e.g., see Schapiro, Turk-Browne, Norman &

Botvinick, 2016 for a recent example of a neurobiologically informed computational model of SL).

Chapter 2: Position-Based Encoding During Statistical Word Segmentation

In this chapter, I seek to shed light on the mechanism(s) of auditory SL by probing the nature of representations that emerge after brief exposure to a continuous stream of speech. In particular, I ask whether the learned/extracted sequences that result from a standard word-segmentation SL task with adult learners actually bear word-like properties (defined below), or are primarily determined by the transitional probabilities between syllables. I predict that, if the embedded trisyllabic structure is chunked from the continuous stream, there will be evidence for non-TP-based knowledge of the position of syllables within the chunk. Moreover, increasing difficulty with encoding the familiarization stimuli will enhance these effects, as fewer resources can be dedicated to veridical encoding of the input stream. Over 4 experiments I manipulated participants' ability to easily perceive or attend to the familiarization language, and tested participants on their knowledge of the position of syllables within TP-defined trisyllabic words. I find evidence that SL induces sensitivity to positional information within trisyllabic chunks in addition to sensitivity to the statistical association between adjacent syllables, and that attention and perceptual familiarity impact the segmentation process in different ways.

2.1 Background

Research across disciplines converges on the idea that certain positions in a word enjoy a privileged status in memory and perception. For example, both the initial (MacKay, 1970; Marslen-Wilson & Zwitserlood, 1989; Swingley, Pinto & Fernald, 1999) and final (Brown & McNeill, 1966; Allopenna, Magnuson, & Tanenhaus, 1998) sounds of words act as an organizing principle in the lexicon, thus suggesting that these positions are represented differently from

word middles (Utman, Blumstein, & Burton, 2000). Position-based differences in processing have further been demonstrated from the early stages of acquisition. Initial sounds of words newly segmented from continuous speech elicit larger ERP deflections in comparison to medial or final sounds in both adult (Sanders et al., 2002) and infant (Teinonen et al., 2009; Kudo et al., 2011) learners, while children's first word productions suggest that final positions within (multi-morphemic) words are particularly maintained in memory (Slobin, 1973), and are used as a tool for segmenting words from streams of speech (Echols & Newport, 1992).[6]

As discussed in Chapter 1, the SL word-segmentation paradigm leads to successful discrimination of high TP sequences from low TP sequences across the developmental span (Teinonen et al., 2009; Evans et al., 2009; Saffran et al., 1996). A number of studies suggest that both children and adults treat TP-defined nonce words as viable word candidates (Saffran, 2001; Graf Estes et al., 2007; Hay et al., 2011); however, it is unclear to what extent these high TP sequences have any word-like properties prior to subsequent association with semantics. Indeed, given the ubiquity of the SL phenomenon across perceptual domains (e.g., Kirkham et al., 2002; Conway & Christiansen, 2005) and species (e.g., Hauser et al., 2001; Toro & Trobalón, 2005), it seems reasonable to assume that the output of auditory SL is relatively general (as opposed to

[6]It should be noted that this discussion of "word" reflects evidence primarily from speakers of Indo-European languages (though Slobin, 1973, canvasses a broader cross-linguistic range). What should constitute a 'word' in the minds of speakers cross-linguistically is an important and contentious topic (see Dixon & Aikhenveld, 2002, and Van Gijn & Zúñiga, 2014, for a review of cross-linguistic and theoretical debates on word-hood); however, as was noted in Chapter 1, the purpose of the present exploration is to determine what native English speakers actually learn from exposure to a continuous, non-meaningful stream of speech. The present discussion, then, is to identify certain properties that might (or might not) be expected from the SL learning experience, but is not meant to determine whether learners' emergent representations are *words*.

limited to a linguistic form that does not exist in vision, touch, or – presumably – for rats). What, then, should we expect the output representations from SL to look like?

As was discussed in Chapter 1, there is evidence to suggest both that the representations that emerge from a SL experience reflect a continuous tracking of underlying TP strength between units (i.e., no 'boundaries' per se: Peña et al., 2002; Endress & Bonatti, 2007; Endress & Mehler, 2009a), and that emergent representations reflect chunks of associated elements that stand independently from the rest of the stream (Giroux & Rey, 2009; Fiser & Aslin, 2005; Zhao & Yu, 2016). Furthermore, there is evidence that representations are influenced by the input conditions faced by the learner, and in particular that this impacts the representations' 'chunk-like' quality (Perruchet & Poulin-Charronnat, 2015). These two types of emergent representations are not mutually exclusive – but suggest different types of learning mechanisms at play (e.g., see Giroux & Rey, 2009). In the current study, therefore, I propose to test for different types of representations in a standard SL word-segmentation paradigm. Specifically, I look for evidence that learners' representations reflect veridical TP-tracking (which I will term the TP-encoding hypothesis), or that representations reflect position-based encoding of syllables within a high-TP trisyllabic unit (which I will term the Position-encoding hypothesis).

In the studies that follow, learners are exposed to artificial languages that consist of four trisyllabic words formed from 12 unique syllables (as in Saffran et al., 1996). I tested for asymmetrical representations across syllable positions within a trisyllabic sequence by comparing performance across distinct test item types that probed position-specific knowledge. Two types of non-word foils were created for this purpose. One type (henceforth *fake-words)* consisted of combinations of syllables from two TP-defined words, with all syllable positions maintained. For example, given the words golabu and padoti, a fake-word with an initial syllable

manipulation could have the form go-doti (see Section 2.2.1.2.1). This was parametrically varied across the three syllable positions, initial, medial, and final. The second non-word foil (*part-words*) consisted of sequences encountered during the familiarization string, but across word boundaries (e.g., doti-go). This type of foil has been previously used in both infant and adult studies (e.g., Saffran, et al., 1996; Thiessen, 2010), and so served as a control comparison for trials involving fake-words. In the first experiment, learners were exposed to an artificial language composed of native English sounds. In Experiments 2 and 3, these sounds were made progressively less English-like. In Experiment 4, learners were again exposed to native English sounds, but simultaneously attended to a distracting, unrelated visual display, thereby dividing their attentional resources.

2.2 Experiment 1

In the first experiment, I exposed adult listeners to a 2-minute auditory stream composed of four trisyllabic words that were formed from 12 unique syllables (as in Saffran et al., 1996). I tested whether participants' representations after familiarization consisted of veridical TP-traces, or knowledge of syllable positions within high-TP sequences. This was done by comparing participants' choice of high(er)-TP versus low(er)-TP items in a 2-alternative forced choice paradigm.

2.2.1 Methods

2.2.1.1 Participants

Forty-four adult native-speakers of English were recruited through the University of British Columbia Psychology Department's paid participants listserv, two of whom were excluded because they did not meet our criterion for English-language exposure (i.e., they did not live in an English-speaking environment until after the age of 3), leaving data from 42 participants. Participants received remuneration of $10 for participating.

2.2.1.2 Materials

Input syllables were digitally recorded at a sampling rate of 44,100 and 16-bit depth with a head-mounted AKG C520 microphone and USB Pre 2 preamp through Audacity 2.0 in a sound-proofed booth. Syllables were produced by a trained phonetician (the author) and recorded in a single session. Syllables that were deemed acoustically clearest and most similar in duration, intonation, and timing were selected (by the author) and manipulated via the Vocal Toolkit plug-in (Corretge, 2012) in Praat (Boersma & Weenink, 2012). Syllable durations were set to 220 milliseconds (as in a number of previous SL paradigms, e.g., Saffran, et al., 1996; French et al., 2011; Mersad & Nazzi, 2012). To achieve this, syllables were first hand-spliced so that their natural intensity contour resembled the shape of a target syllable (*bi*).[7] The proportion of voicelessness/voicing to vowel duration was examined, and either lengthened or shortened by

[7]A base syllable was selected at random in order to apply a similar pitch and intensity contour across all syllables. This method was chosen as opposed to a flat pitch and intensity resynthesis in order to increase the perceived naturalness of the stimuli.

hand to approximate the target syllable. Proportions differed for voiceless and voiced consonant onsets: voiced consonant proportions of the syllable ranged from 14-18%; voiceless consonant proportions ranged from 15-25%. Individual glottal pulses and sections of aspiration were either removed or copied and pasted in order to (respectively) shorten or lengthen consonant proportions. In both cases, sections were selected such that the waveform began and ended at a zero-crossing, to avoid the introduction of acoustic artifacts. The function Change Duration from Vocal Toolkit, which uses the PSOLA resynthesis method, was next applied to create a syllable of exactly 220 msec. Then, the Copy function was employed to ensure similar F0 medians (178 Hz), F0 contours, and intensity contours across syllables (see Figure 2.1).[8] Finally, the Scale Intensity function of base Praat was used to set mean RMS amplitude to 70 dB.

[8]The pitch contour from the model syllable is extracted in a pitch tier; this contour then replaces the contour of the second syllable. The median value is extracted as the 0.50 quantile from the model syllable, which is copied to the second sound. Resynthesis is achieved via PSOLA for both of these functions. For intensity contour, the second sound's intensity is first multiplied by its inverse to flatten it, and then multiplied by the extracted intensity curve of the model syllable.

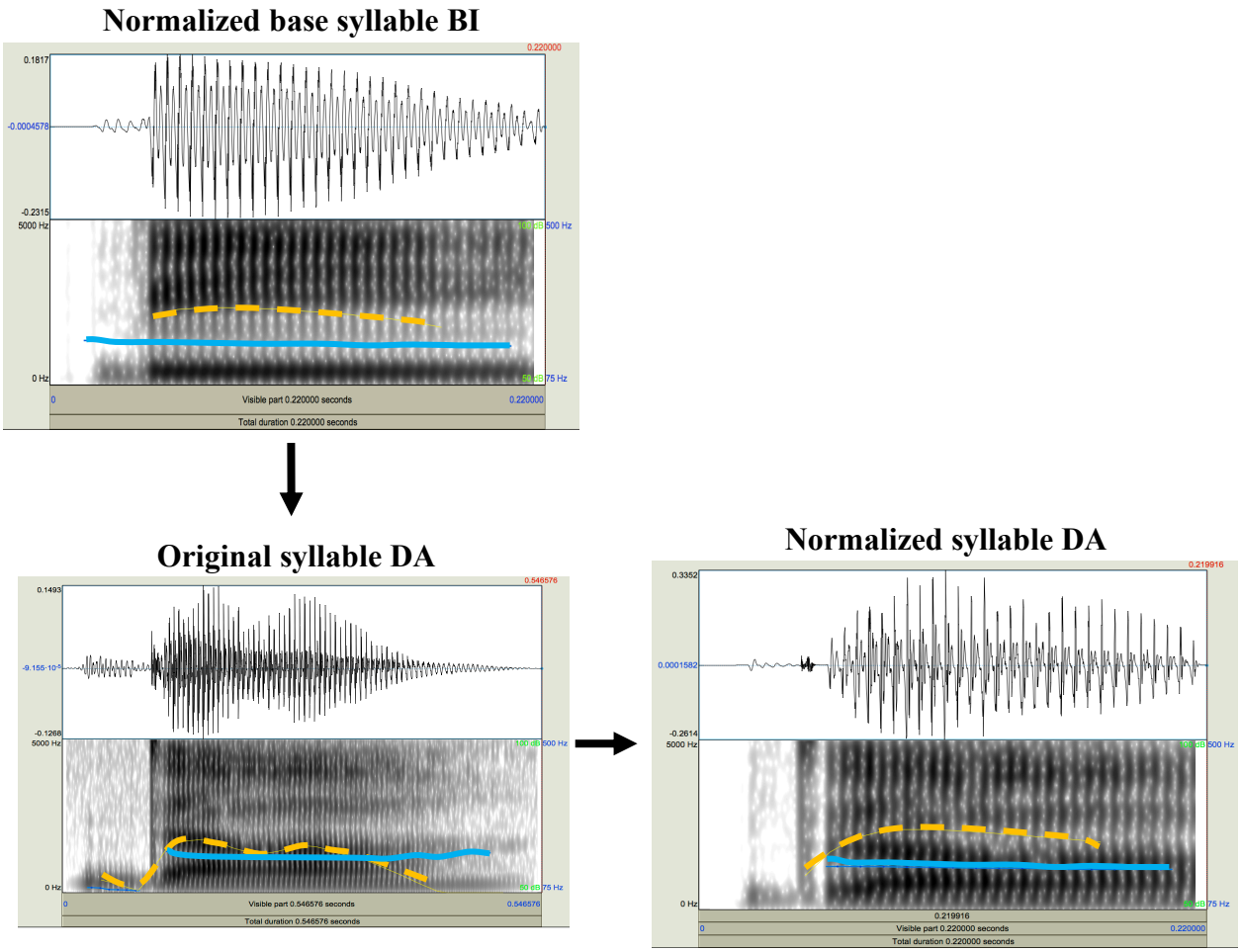

Figure 2.1 Normalization procedure. Syllables were normalized to have the same duration (220 msec), F0 means and contours, and intensity means and contours. The F0 (solid blue line) and intensity (dashed yellow line) contours were copied from a base syllable (BI) to other syllables. The effect of this process is demonstrated for the syllable DA.

Syllables were concatenated into trisyllabic words (see Table 2.1), and words concatenated into two semi-random lists per language. Each word was repeated 48 times and they were interlaced in such a way that every word was followed by the three other words equally often, and never by itself. This created syllable-to-syllable TPs across word boundaries of 0.33, whereas TPs between syllables within a word were 1.0. The resulting familiarization strings were 2 minutes 10 seconds in length. The initial and final 5 seconds ramped up and down in amplitude, respectively (between approximately 32 and 70 dB SPL by multiplying the first

half period of a (1 + cos(x)) / 2 function), to prevent providing participants with a clear cue to word boundaries other than TPs.

CONSONANTS				VOWELS			WORDS	
	BILABIAL	*ALVEOLAR*	*VELAR*		*FRONT*	*BACK*	*LANGUAGE A*	*LANGUAGE B*
ASPIRATED	pʰ	tʰ	kʰ	*HIGH*	i	u	b̥id̥akʰu	d̥atʰub̥i
UNVOICED	b̥	d̥	g̊	*MID*		o	g̊olab̥u	g̊otʰib̥u
APPROXIMANT		ɹ		*LOW*		a	pʰad̥otʰi	ɹokʰula
LATERAL		l					tʰupʰiɹo	pʰid̥opʰa

Table 2.1 Segmental and word inventory from Experiment 1 Segments are displayed by place and manner of articulation, and their respective combinations into four trisyllabic words, presented for both languages A and B.

2.2.1.2.1 Tests

Participants were tested using a 2-AFC paradigm using three types of test items. Item types are described below and presented visually in Figure 2.2.

Words vs. part-words. The first set of items pitted words against part-words. Words are trisyllabic sequences that occurred in the input with perfect TPs between each pair of syllables. Part-words are also syllable sequences that occurred in the familiarization stream, but in these strings one pair of syllables has a high TP (1.0) while the other has a lower TP (.33). These were constructed by taking the final syllable of one word and combining it with the first two syllables of another word, or the final two syllables of a word with the first syllable of another word. If people are sensitive to the strength of the TPs, rather than just whether or not they have heard a particular sequence in their input, they should choose words more frequently than part-words. This is the contrast that has been used most often in statistical word learning studies (e.g.,

Saffran et al., 1996; Thiessen, 2010; Endress & Mehler, 2009a; Peña et al., 2002; Perruchet & Poulin-Charronat, 2012), and so here serves as a within-subject control to demonstrate that people are segmenting the stream as expected.

Words vs. fake-words. In the second type of test item, words were pitted against fake-words. Fake-words are manipulations of words in which the individual syllables remain in their correct ordinal positions, but where one syllable has been replaced by the corresponding syllable of a different word. There were three different kinds of fake-words: Initial-syllable, Medial-syllable, and Final-syllable fake-words. In initial- and final-syllable fake-words, the string comprises one TP of 1.0 and one TP of 0.0. Medial-syllable fake-words have two TPs of 0.0. Overall, participants should prefer words over fake-words, whether they extract the trisyllabic sequences with 1.0 TPs from the speech stream as word-like chunks or simply track and store all relative TPs (given that the fake-words always contain at least one transition that did not occur in the input). The intent behind these items was not just to test the overall preference for words over fake-words, however, it is to test participants' knowledge of the constituent pieces within words. If there is an asymmetry in encoding across syllable positions, as suggested by previous work, some fake-words may be more confusable with words and hence lead to relatively worse discrimination between the two. For example, studies have found that learners struggle to reject combinations like XBC, where X is a novel syllable, but BC are adjacent syllables anchored to the right edge of a high-TP word (Saffran et al., 1996; Saffran et al., 1999). This would predict that our learners might find initial-syllable fake-words particularly confusing. On the other hand, if the output of SL does not involve positional information, performance should be best on words pitted against fake-words with medial syllable manipulations (which have two 0.0 TPs), and better on all fake-word types (which have at least one 0.0 TP) as compared to word versus part-

word (which have one 1.0 and one 0.33 TP) trials. That is, participants should perform better on items where they are choosing between a fake-word and a word than on items where they are choosing between a part-word and a word. There were 8 trials for each sub-type of fake-word (i.e., initial-, medial-, and final-syllable manipulations).

Part-words vs. fake-words. In the third test-item type, part-words were pitted against fake-words. I predicted that participants would prefer part-words across all three syllable position manipulations if participants are merely veridically tracking transitional probabilities, but that they may prefer fake-words (at least some types) if SL yields word-like units with positional information. The reasoning behind this is the following: if participants are extracting word-like units, then fake-words will seem more like the known words than part-words, as fake-words share initial, medial and final syllables with 'real' words, which should lead to something like lexical activation of the novel stored word forms. Again, there were 8 trials for each sub-type of fake-word test item.

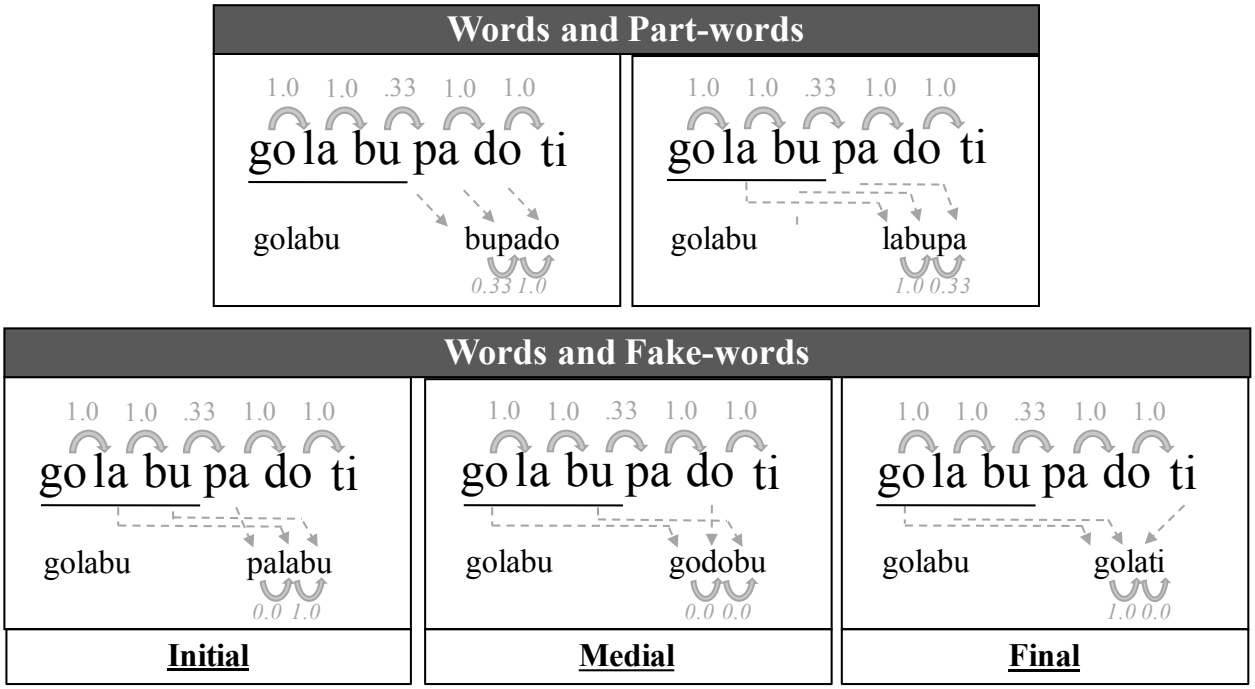

Figure 2.2 Examples of words, part-words, and fake-words and their respective TP-structures. Each panel shows a partial section of the familiarization stream (Language A). The TP between syllables is shown directly above the transitions between syllables. Words, defined as 1.0 TPs between syllables, are underlined. Part-words are syllable sequences that cross word boundaries (one 1.0 TP and one 0.33 TP), and can be found in the top two panels. Fake-words are sequences in which the position of syllables from high TP words is maintained, but are concatenated in novel combinations (creating at least one 0.0 TP). This is done across the three syllable positions, which is shown in the bottom 3 panels.

The 56 trials of the three different types (words vs. part-words (n = 8), words vs. fake-words (n = 24), and part-words vs. fake-words (n = 24) and three syllable manipulations (initial, medial, and final) were randomly presented. Each trial consisted of two trisyllabic tokens presented with a 1,000 msec. ISI (e.g., Newport & Aslin, 2004); participants were given 5,000 msec. to respond (e.g., Toro, Sinnett, & Soto-Faraco, 2005; Newport & Aslin, 2004).[9]

[9]A subset of the participants was accidentally allowed 10,000 msec to respond. Of these, 17 (of 42) participants took longer than 5000 msec on a total of 67 separate trials. All participants were over-limit on 6 or fewer trials (which are spread across the different trial type manipulations), with the exception of one participant who was over-limit on 17 trials. This participant has been excluded from RT analyses (as he is missing half or more of the data for two conditions – the

2.2.1.3 Procedure

Participants were told they would first be listening to some sounds, and then answering some questions about those sounds. They were seated in a sound-attenuated room in front of a computer screen and button box and told to follow the instructions provided by the computer. They were asked to use their two index fingers to provide answers via the two outermost buttons of a button box. The experiment was administered with E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Participants were first led through 4 training trials to ensure understanding of the button box keys: they were asked to listen to two sound files, and indicate the one that sounded like the word "say". After completing these trials, they were asked to please listen quietly to a language called Vesutian. They were prompted to press a button to start, after which the screen turned blank and the familiarization stimuli began playing. After familiarization they were reminded that they would hear two options, and asked to please choose the option that sounded more like a word from the language they had just listened to. At the end of the experiment, a screen thanked the participants and instructed them to see the experimenter; participants then completed an exit interview that assessed meta-linguistic awareness and reactions to the task, and a language background questionnaire (results reported in Chapter 3).

---

two initial syllable manipulation trial types), but retained for all other analyses. The remaining over-limit trials have been individually excluded for the RT analysis (and retained for all other analyses). All non-RT analyses, however, pattern the same when these same over-limit RT trials are excluded.

2.2.1.4 Measures & Analysis

I analyzed participants' proportion choice of words versus part-words, words versus fake-words, and part-words versus fake-words as a means of measuring participants' sensitivity to the statistical structure of the stream. However, I anticipated that evidence for position-based knowledge of the trisyllabic nonce-words might be more subtle than can be easily detected by accuracy scores. I therefore also recorded and analyzed reaction times (RTs) to each trial type, as RTs can reveal processing differences across stimuli that raw accuracy scores do not (e.g., differences in attentional mechanisms, Prinzmetal, McCool, & Park, 2005; developmental shifts in implicit learning, Janacsek, Fiser, & Nemeth, 2012). I hypothesized that slower reaction times would correspond to non-word foils that were more difficult to reject (explicit predictions are detailed in the section that follows). RTs are calculated as the lag between the onset of the second trisyllabic sequence presented in a trial and participant response; as participants might base their 2AFC decisions on the first trisyllabic sequence alone, RTs are considered from the earliest possible responses, and have not been trimmed from the left edge (as is commonly done to prevent the inclusion of false-alarm/unintentional responses). It is also common for RT results to be presented for correct-identification trials only; however, under this paradigm there is no "correct" choice – rather, different choices are hypothesized to reflect different learning mechanisms. Thus, RT data is analyzed using all trials. RT analyses are presented where direct comparisons between trial types are made (e.g., when comparing syllable position manipulations, or direct comparison of major trial types).

Finally, I also present correlations between performance on the various trial types, to look for potential individual differences in segmentation strategies.

All analyses are conducted using R statistical software (Version 3.3.3), using the packages lme4 (Bates, Maechler, Bolker, Walker, 2015), sjPlot (Lüdecke, 2017), and the suite of packages compiled through the tidyverse package (Wickham, 2017). Generalized mixed effects models that predict proportion choice were constructed as follows: I first attempted a fully specified model, which included all fixed main effects and interactions, and in which the random effects structure consisted of interactions, slopes, and intercepts for all within-subject variables grouped by subject intercepts (see Barr, Levy, Scheepers, & Tily, 2013).[10] As it is expected that learning will continue to take place across trials, trial (centered, for the sake of model convergence) was entered as both a fixed and random covariate in all models. If the fully specified model failed to converge, I first increased the number of iterations in the optimization algorithm up to 20,000,000. If the model still did not converge, I progressively eliminated elements from the model beginning with covariance in the random effects structure, followed by random effects interaction terms, the random main effect of trial, and finally fixed effects interaction terms until model convergence was reached. When multiple models were run on the same analysis (i.e., in order to rotate the reference level of a categorical variable), the simplest model structure required for convergence was applied across each model run for consistency. All model results are reported in terms of odds ratios, their 95% confidence intervals (derived via *Wald* tests), and associated *p*-values.

---

[10]I have taken the 'keep it maximal' approach, per Barr et al.'s (2013) recommendations. The current study may be underpowered for this approach (see, e.g., Bates, Kliegl, Vasishth, & Baayen, 2015, and Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). I have retained maximal structure in keeping with much current psycholinguistic research; however, analyses of the data using simplified random effects terms yield similar results.

2.2.1.5 Predictions

Under either the TP-encoding or Position-encoding hypotheses, participants are expected to choose words over part-words. The two hypotheses, however, make different predictions for performance on the other two trial types. Under TP-encoding, I would expect that performance is driven by the TPs. Specifically, if a participant hears an item with 1.0 TPs and an item with 0.0 TPs (fake-words), it should be relatively easy to reject the 0.0 TP item. It may also be easier to reject the 0.0 TP item than it would be a 0.33 TP foil (part-words). Under the position-encoding account, however, some items with 0.0 TPs may in fact be more confusable with the 1.0 TP words because the syllables maintain the correct ordinal positions. This thus predicts that performance on words versus fake-words may be equivalent to or worse than performance on words versus part-words. Finally, the same logic can be applied to the part-word versus fake-word trials. If participants are relying solely on TP strength, then part-words should sound more familiar than the 0.0 TP fake-words. Alternatively, if participants are encoding the positions of syllables within high-TP words, fake-words may be preferable to part-words because they maintain positional information at the expense of TPs. These predictions are graphically depicted in Figure 2.3.

Predictions for RTs follow the same logic. Participants should be fastest at rejecting non-words with 0.0 TPs (i.e., fake-words). Alternatively, if participants are extracting word-like chunks, they may find fake-words (or fake-words of specific syllable-manipulation types) more confusing, and thus be slower to reject them, in comparison to part-words.

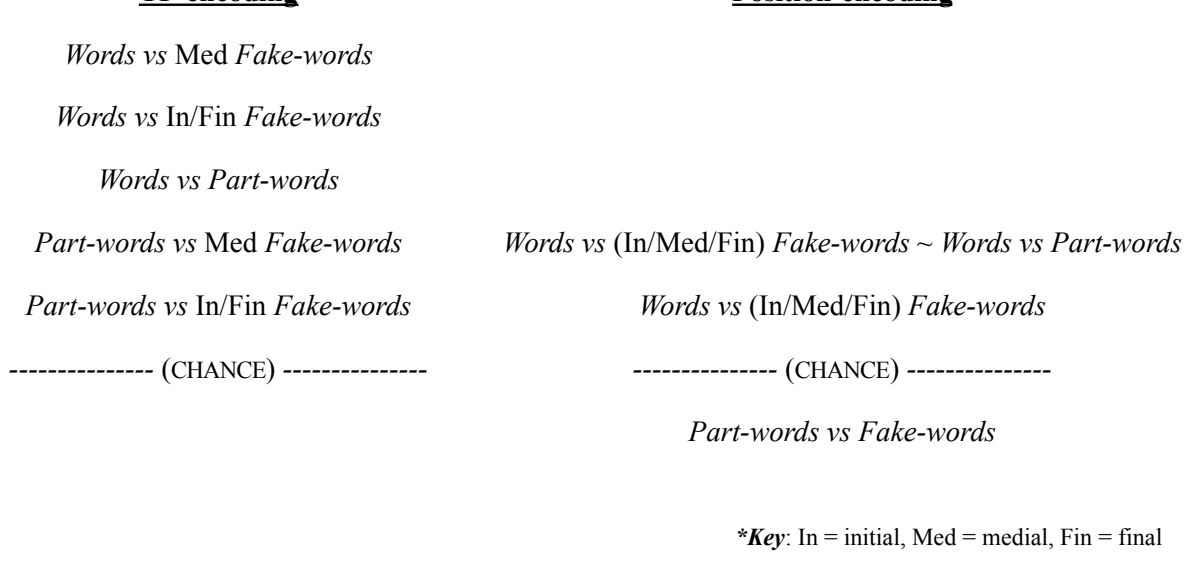

Figure 2.3 Predictions of the TP-encoding and Position-encoding hypotheses. Predicted relative performance by trial type under the TP-encoding hypotheses and Position-encoding hypothesis. Trial types are plotted according to the ordering relationship of relative proportion choice and RT (i.e., trial type A plotted above trial type B means that learners both have stronger judgements on, and are faster to respond to, items of type A than type B). The dotted line reflects equivalent choice (i.e., chance performance). Performance above chance means higher proportion choice of the first sequence type listed (e.g., *words* in the trial type "Words vs. PW"). Performance below chance means higher proportion choice of the second sequence type listed (e.g., *fake-words* in the trial type "Part-words vs Fake-words"). Under the TP-encoding hypothesis, order of performance is predicted by the relative TP comparison (e.g., a Medial Fake-word has 0.0 TPs across both syllable transitions, and should therefore be the easiest type of foil to reject). Under the Position-encoding hypothesis, order of performance is predicted by the maintenance of syllable positionality, not TP strength. *Note*: Items are not plotted with respect to any claim in differences in magnitude.

The position-encoding hypothesis does not make strong predictions with regard to comparisons within the syllable manipulations themselves. Given the linguistic literature on the primacy of edges (Brown & McNeil, 1966; MacKay, 1970; Marslen-Wilson & Zwitserlood, 1989; Echols & Newport, 1992), I might predict that participants will be most accepting of fake-words with medial syllable manipulations – for in these items, the first and final syllable are anchored to each other and each to its correct edge; however, as discussed in Chapter 1, it is unclear if these two positions are encoded in the same way or to the same degree. Given the statistical learning literature (Saffran et al., 1996; Saffran et al., 1999), on the other hand, I would

predict that participants will be most accepting of fake-words of the initial syllable manipulation type (i.e., these are roughly analogous to the trials that pitted ABC versus XBC structures in Saffran et al., 1996 and Saffran et al., 1999, which were found to be more confusing to participants). The predictions of the TP-encoding hypothesis, however, are more straightforward: we would predict that learners will choose words or part-words over fake-words in all syllable manipulations, but that medial fake-words will be easiest to reject, due to the two 0.0 TPs.

Finally, I also examine the correlations between performance on different trial types and syllable manipulations. If segmentation is driven exclusively by TPs, I should find positive correlations between *all* item types (the choice of higher TP items leads to higher proportion choice scores across all trial types), but if segmentation involves position-based encoding, I predict a negative relationship between part-word versus fake-word trials and word versus part-word trials, or that this relationship will hold for certain position-manipulated fake-words (e.g., initial syllable manipulated words), but be uncorrelated for the other positions.

2.2.2 Results

Figure 2.4 shows performance on the 2AFC test for all trial types. Each dot represents one individual participant's proportion choice. Chance performance is reflected by the dotted line at 0.50. Stars and vertical bars represent the group mean plus/minus one standard error.

The same analysis steps conducted below were also undertaken to check for differences between the two languages (Language A, $n = 20$ and Language B, $n = 22$). No significant differences were found, thus the following analyses collapse across them.

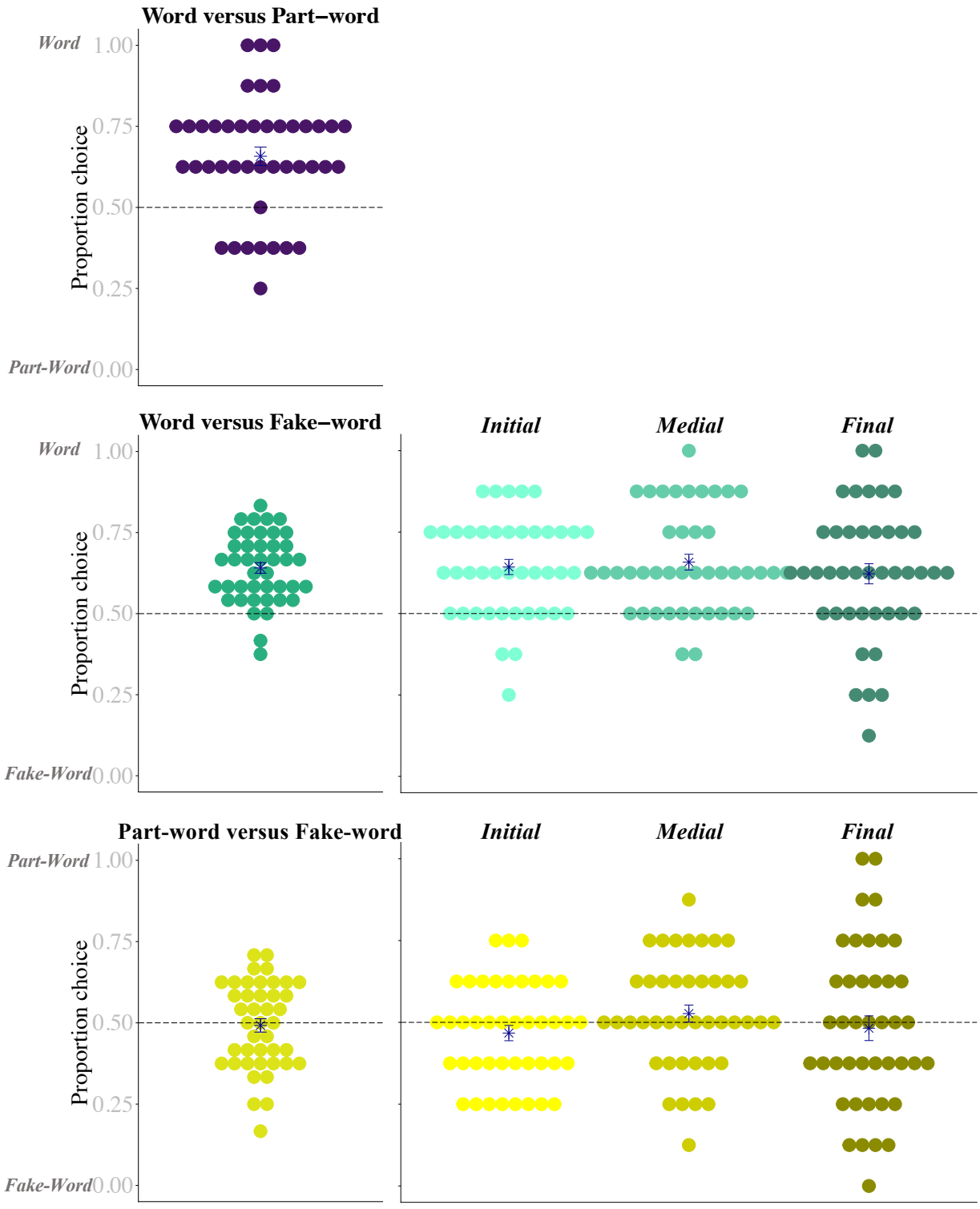

Figure 2.4 Proportion choice across trial types in Experiment 1. Dots reflect individual participant mean scores. Stars reflect mean accuracy scores; error bars are plus/minus 1 standard error. Chance is 0.50.

2.2.2.1 Words versus Part-Words

When asked to choose between words and part-words, participants selected words over part-words at a rate significantly above chance (i.e., above 50%; M = 65.8%, SD = 18.3%, 95% CI = [60%, 71.5%], $t(41) = 5.58$, $p < .0001$, $d = 0.86$; top of Figure 2.4). This finding replicates previous work (e.g., Saffran et al., 1999) and serves as a control comparison for the word versus fake-word trial types (presented below).

2.2.2.2 Words versus Fake-Words

I first report the results for all word versus fake-word trials as a whole, and then break down the results by syllable manipulation type.

2.2.2.2.1 Combined

When asked to choose between words and fake-words, participants selected words at a rate significantly above the 50% chance-level (M = 64.1%, SD = 10.6%, 95% CI = [60.8%, 67.4%], $t(41) = 8.60$, $p < .0001$, $d = 1.33$; see middle of Figure 2.4).

2.2.2.2.2 Syllable Manipulations

Performance by syllable position is displayed in the middle, right-hand panel of Figure 2.4. Participants selected words over fake words across all three syllable manipulations: Initial (M = 64.3%, SD = 15.0%, 95% CI = [59.6%, 69.0%], $t(41) = 6.17$, $p < .0001$, $d = 0.95$), Medial (M = 65.8%, SD = 15.6%, 95% CI = [60.9%, 70.6%], $t(41) = 6.54$, $p < .0001$, $d = 1.01$), and Final (M = 62.2%, SD = 20.0%, 95% CI = [56.0%, 68.4%], $t(41) = 3.95$, $p = .0002$, $d = 0.61$). To probe for differences between the syllable manipulation types, a mixed effects logistic regression

model was fitted to predict item choice across the syllable position manipulations. The fully specified model is as follows:

Choice ~ Syllable position * Trial + (Syllable position * Trial | Subject)

This model, however, failed to converge, even with optimization iterations set to 20,000,000. The model was progressively simplified, beginning by removing the covariance terms in the random effects structure, but did not converge until Trial had been removed from the random effects structure. The final structure was:

Choice ~ Syllable manipulation * Trial + (1 | Subject) + (0 + Syllable manipulation | Subject)

The results of three models (with alternated reference levels for Syllable manipulation) are presented in Table 2.2. There was no effect of syllable position manipulation or trial.

Model Structure:
Choice ~ Syllable manipulation * Trial + (1 | Subject) + (0 + Syllable manipulation | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	1.80	1.44-2.26	**<.001**	1.92	1.53-2.42	**<.001**	1.73	1.31-2.28	**<.001**
Initial Syll				0.94	0.68-1.29	.682	1.04	0.74-1.47	.815
Medial Syll	1.07	0.78-1.47	.682				1.11	0.79-1.57	.540
Final Syll	0.96	0.68-1.35	.815	0.90	0.64-1.27	.540			
Trial	1.00	0.98-1.01	.864	1.00	0.98-1.01	.725	1.01	1.00-1.03	.153
Initial : Trial				1.00	0.98-1.02	.903	0.99	0.97-1.01	.249
Medial : Trial	1.00	0.98-1.02	.903				0.99	0.97-1.01	.198
Final : Trial	1.01	0.99-1.03	.249	1.01	0.98-1.02	.903			
Random Effects									
$\tau_{00,\ Subject}$		0.011			0.012			0.232	
ρ_{01}		1.000			1.000			1.000	
$N_{Subject}$		42			42			42	
$ICC_{Subject}$		0.003			0.004			0.066	
Observations		1008			1008			1008	
Deviance		1281.076			1281.076			1281.077	

Table 2.2 Experiment 1 model of proportion choice word vs fake-words

Participants were numerically faster to respond to fake words with medial syllable manipulations (M = 1690 msec, SD = 535 msec) than initial (M = 1722 msec, SD = 643 msec) or final (M = 1728 msec, SD = 532 msec) syllable manipulations (see Figure 2.5); however, when this data was fitted to a linear mixed effects model this difference did was not significant (Table 2.3).

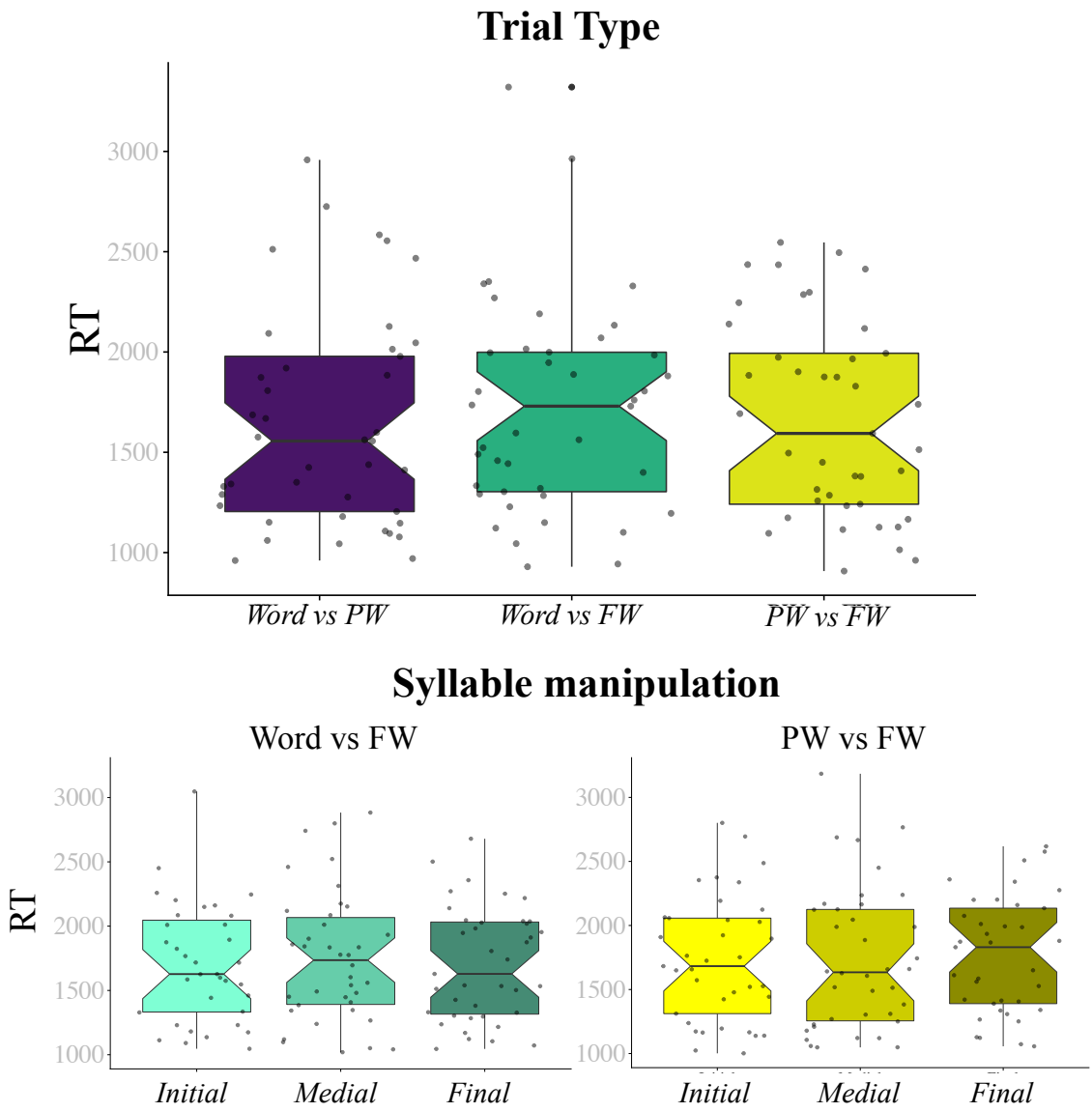

Figure 2.5 Reaction times by trial type and syllable manipulation in Experiment 1 Dots reflect individual participant mean scores. Horizontal lines reflect group medians by condition; boxes cover the 2 middle quartiles, whiskers indicate the range of the top and bottom quartiles.

Model Structure:
RT ~ Syllable position * Trial + (1 | Subject) + (0 + Syllable position | Subject)

	Model 1 Ref level = Initial			**Model 2** Ref level = Medial			**Model 3** Ref level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1714	1520 – 1907	**<.001**	1691	1526 – 1855	**<.001**	1721	1556 – 1886	**<.001**
Initial Syll				23	-99 – 146	.709	-7	-119 – 104	.896
Medial Syll	-23	-145 – 99	.710				-31	-150 – 88	.614
Final Syll	7	-104 – 119	.896	31	-88 – 150	.614			
Trial	-4	-9 – 1	.080	-2	-7 – 3	.393	-4	-9 – 1	.134
Initial : Trial				-2	-9 – 4	.504	-1	-7 – 6	.872
Medial : Trial	2	-4 – 9	.504				1	-5 – 9	.617
Final : Trial	1	-6 – 7	.872	-2	-9 – 5	.617			
Random Effects									
σ^2		463262			463262			463262	
$\tau_{00,\ Subject}$		340373			229898			230396	
ρ_{01}		0.949			0.949			0.999	
$N_{Subject}$		41			41			41	
$ICC_{Subject}$		0.424			0.332			0.332	
Observations		957			957			957	
R^2 / Ω_0^2		.393 / .390			.393 / .390			.393 / .390	

Table 2.3 Experiment 1 model of reaction time to words vs fake-words

2.2.2.3 Word vs PW compared to Word vs FW trials.

If learners are veridically tracking TPs, fake-word foils should be easier to reject than part-words, which were actually encountered during familiarization and therefore consist of non-zero TPs. I therefore compared performance across these two trial types, as in Saffran et al. (1996) and Finn et al. (2014); however, there is no difference between the two trial types (Words vs. Part-words: M = 65.8, SD = 18.3; Words vs. Fake-words: M = 64.1, SD = 10.6; $t(41) = -0.63$, $p = .53$). Reaction times to the two trial types also do not significantly differ, as indicated by a mixed effects linear regression specified for the interaction and main effects of Trial Type and Trial, and the same terms grouped by subject as random effects (Words vs. Part-words: M = 1641 msec, SD = 533 msec; Words vs. Fake-words: M = 1716 msec, SD = 522 msec; $B = 71.4$ +/- 47.9 (standard error), $t(37.6) = 1.49$, $p = .14$). This same model did confirm, however, that participants became faster at word versus part-word trials over the course of the experiment (as was indicated by the model presented in Section 2.2.2.1; results from this model: $B = -6.5$ +/- 2.7 (standard error), $t(41.5) = -2.43$, $p = .02$).

Although this equivalent performance across trial types aligns with the predictions made by the position-encoding hypothesis, it does not provide confirmatory evidence. However, if participants perform worse on certain fake-word types as compared to part-words, this would provide positive evidence in favor of the position-encoding account. I therefore ran mixed effects models with a fixed effects interaction between trial and 2AFC contrast type (a categorical variable with four levels: (1) Reference level: word versus part-word, (2) word versus fake-word initial, (3) word versus fake-word medial, and (4) word versus fake-word final) which showed that neither proportion choice nor response times to each word versus fake-word trial type differ from word versus part-word trials. As in the model comparing the two main trial types (Word vs

PW and Word vs FW), there was an effect of trial in the RT model, showing that participants became a little quicker over time on the word versus part-word trial types (B = -7 +/- 2.5 (standard error), $t(330) = -2.60$, $p = .01$). Full model results (with final, simplified model specifications) can be found in Table 2.4 Panel A for proportion choice, and Panel B for RT.

A. Proportion Choice

Model structure:

Choice ~ Contrast Type * Trial +
(1 | Subject) +
(0 + Contrast Type | Subject) +
(0 + Trial | Subject) +
(0 + Contrast Type : Trial | Subject)

	Odds Ratio	*CI*	*p*
Fixed Effects			
(Intercept)	1.99	1.51-2.62	**<.001**
Initial Syll	0.92	0.65-1.31	.659
Medial Syll	1.00	0.71-1.40	.979
Final Syll	0.90	0.64-1.27	.547
Trial	1.01	0.99-1.02	.292
Trial * Initial	0.99	0.97-1.01	.367
Trial * Medial	0.99	0.97-1.01	.370
Trial * Final	1.00	0.98-1.03	.810
Random Effects			
$\tau_{00,\ Subject}$		0.00	
ρ_{01}			
$N_{Subject}$		42	
$ICC_{Subject}$		0.00	
Observations		1344	
Deviance		1626.98	

B. RT

Model structure:

RT ~ Contrast Type * Trial +
(1 | Subject) +
(0 + Contrast Type | Subject) +
(0 + Trial | Subject)

	B	*CI*	*p*
Fixed Effects			
(Intercept)	1635	1473-1798	**<.001**
Initial Syll	82	-33-198	.168
Medial Syll	56	-55-166	.329
Final Syll	90	-28-209	.144
Trial	-7	-11 - -2	**.010**
Trial * Initial	2	-4-9	.515
Trial * Medial	5	-2-11	.163
Trial * Final	3	-4-10	.366
Random Effects			
σ^2		439975	
$\tau_{00,\ Subject}$		0.001	
ρ_{01}			
$N_{Subject}$		41	
$ICC_{Subject}$		0.000	
Observations		1280	
R^2 / Ω_0^2		.406/.402	

Table 2.4 Experiment 1 model of proportion choice (Panel A) and RT (Panel B) to words versus part-word and words versus fake-words

2.2.2.4 Part-Words versus Fake-Words

Results for both main effects and broken down by syllable position are shown in the bottom panels of Figure 2.4.

2.2.2.4.1 Combined

The TP-encoding hypothesis predicted that participants would be more likely to choose part-words over fake-words, while the position-encoding hypothesis predicted that participants would choose fake-words over part-words (at least in some syllable position manipulations). Participants failed to consistently choose either part-words or fake-words (below 50% performance indicates greater proportion choice of fake-words; above 50% greater proportion choice of part-words: M = 49.2%, SD = 13.3%, 95% CI = [45.0%, 53.4%], $t(41) = -0.39$, $p = .70$, $d = 0.06$). As can be seen in the bottom panel of Figure 2.4, individuals' scores appear to be bimodally distributed. It is possible that the pattern of performance differs by syllable position manipulation (see 2.2.2.4.2); it is also possible that the pattern of performance reflects different learning styles (see correlation analysis, sections under 2.2.2.5).

2.2.2.4.2 Syllable Manipulations

Choice was not significantly different from chance across the three syllable manipulations: Initial (M = 46.7%, SD = 15.1%, 95% CI = [42.0%, 51.4%], $t(41) = -1.40$, $p = .17$, $d = .22$), Medial (M = 52.7%, SD = 16.9%, 95% CI = [47.4%, 57.9%], $t(41) = 1.03$, $p = .31$, $d = .16$), and Final (M = 48.2%, SD = 24.5%, 95% CI = [40.6%, 55.8%], $t(41) = -0.47$, $p = .64$, $d = .08$). These means do not differ, as confirmed by a mixed effects logistic regression model with proportion choice as dependent measure (all p's for main effects > .14; Table 2.5).

Model structure:
Choice ~ Syllable manipulation * Trial + (1 | Subject)

	Model 1 Ref level = Initial			**Model 2** Ref level = Medial			**Model 2** Ref level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	0.88	0.69 – 1.12	.290	1.11	0.87-1.41	.406	0.93	0.73-1.19	.582
Initial Syll				0.79	0.58-1.08	.138	0.94	0.69-1.28	.690
Medial Syll	1.26	0.93-1.72	.138				1.19	0.87-1.61	.278
Final Syll	1.06	0.78-1.45	.690	0.84	0.62-1.15	.278			
Trial	0.99	0.98-1.00	.180	1.01	0.99-1.02	.209	1.00	0.98-1.01	.475
Trial : Initial				0.98	0.96-1.00	.067	1.00	0.98-1.01	.671
Trial : Medial	1.02	1.00-1.04	.067				1.01	0.99-1.03	.160
Trial : Final	1.00	0.99-1.02	.671	0.99	0.97-1.01	.160			
Random Effects									
$\tau_{00,\ Subject}$				0.122					
$N_{Subject}$				42					
$ICC_{Subject}$				0.036					
Observations				1008					
Deviance				1344					

Table 2.5 Experiment 1 model of proportion choice to part-words vs fake-words

A linear mixed effects model fitted to the reaction time data similarly revealed no differences by syllable manipulation (Initial: M = 1655 msec, SD = 516 msec; Medial: M = 1667 msec, SD = 584 msec; Final: M = 1687 msec, SD = 493 msec); results are listed in Table 2.6.

Model structure:
RT ~ Syllable manipulation * Trial + (1 | Subject) + (Syllable manipulation | Subject)

	Model 1 Ref level = Initial			**Model 2** Ref level = Medial			**Model 3** Ref level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1655	1500-1812	**<.001**	1664	1485-1843	**<.001**	1686	1533-1839	**<.001**
Initial Syll					-125 – 107	.876	-31	-142 – 80	.584
Medial Syll	9	-107 – 125	.876				-22	-133 – 90	.702
Final Syll	31	-80 – 142	.584	22	-90 – 133	.702			
Trial	1	-4 – 5	.805	1	-4 – 6	.705	-2	-7 – 3	.375
Trial * Initial				-0	-7 – 6	.910	3	-4 – 9	.422
Trial * Medial	0	-6 – 7	.910				3	-4 – 10	.378
Trial * Final	-3	-9 – 4	.422	-3	-10 – 4	.378			
Random Effects									
σ^2	476422								
$\tau_{00,\,Subject}$	0.00								
ρ_{01}									
$N_{Subject}$	41								
$ICC_{Subject}$	0.00								
Observations	966								
R^2 / Ω_0^2	.349/.346								

Table 2.6 Experiment 1 model of RT to part-words vs fake-words

2.2.2.5 Correlations

I next examined participant-level relationships between performance on the various tests. The TP-encoding hypothesis predicts positive correlations across all three trial types; the position-encoding hypothesis predicts a negative correlation between word versus non-word (both part-word and fake-word) and part-word versus fake-word trials. I first present correlations across the main trial types, and then break the data down by syllable position.

2.2.2.5.1 Main trial types

Participants who chose words over part-words were also more likely to choose words over fake-words ($r(41) = 0.39$, $p = .01$). There was also a significant negative correlation between performance on the words vs. part-words test and the part-words vs. fake-words test: the more successful participants were at choosing words over part-words, the more likely they were to endorse fake-words over part-words: $r(40) = -0.36$, $p = .02$. There was no correlation between performance on word versus fake-word and part-word versus fake-word trials ($r(40) = -0.12$, $p=.45$). This pattern of correlations – i.e., that better learners (as indexed by the Word vs. Part-word trials, which reflect a standard test of successful SL) also preferred positionally-coherent over TP-coherent forms – is more consistent with the position-encoding versus the TP-encoding hypothesis. These relationships are plotted in Figure 2.6.

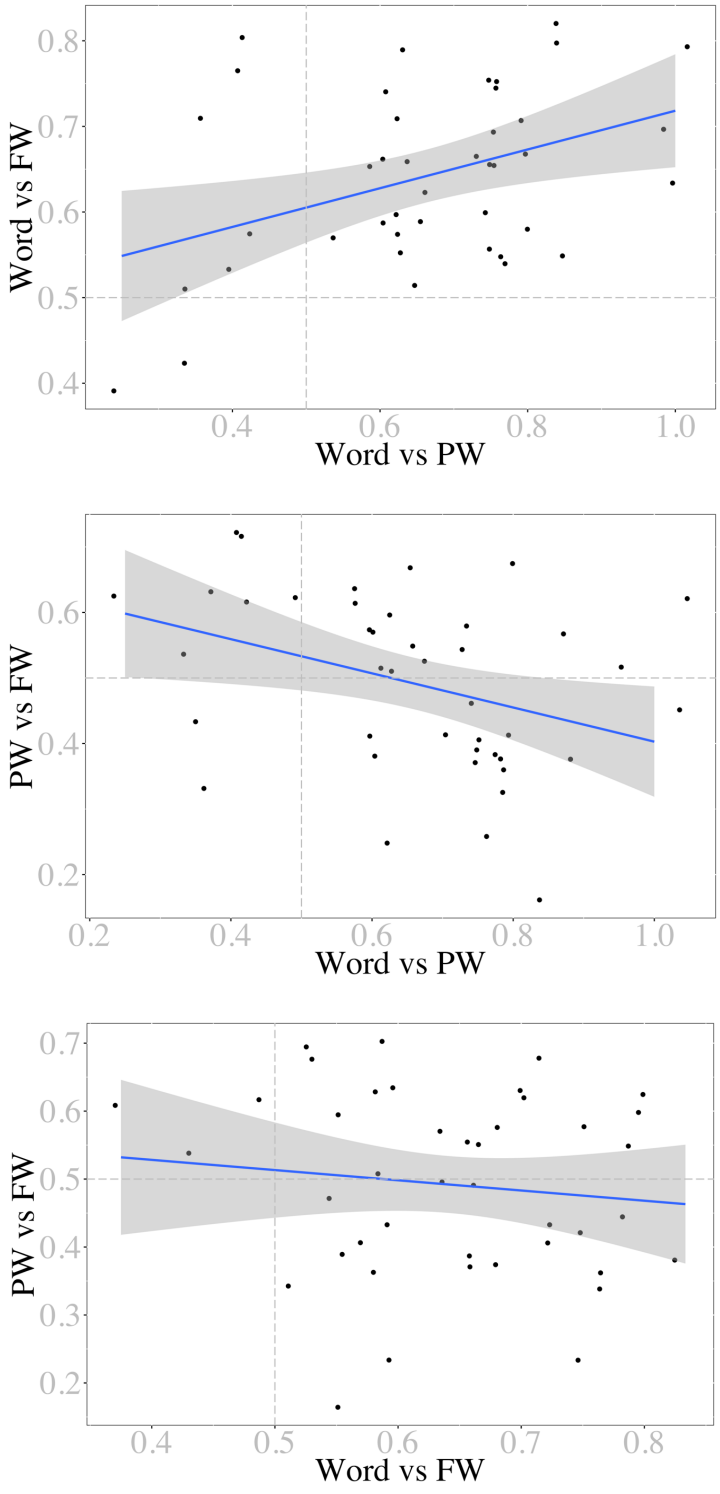

Figure 2.6 Experiment 1 correlations between main trial types Dots represent participant mean performance. The dotted vertical and horizontal lines reflect chance performance in the respective conditions. Thus, dots that fall in the upper right quadrant are above chance on both conditions.

2.2.2.5.2 Syllable manipulations

All correlations by syllable position can be found in Table 2.7. For word vs. fake-word trials, there is no correlation between performance on any of the syllable manipulations, that is, performance on first-syllable manipulations was not related to performance on middle syllable manipulations, etc. There is, however, a relationship between individuals' performance on the standard word segmentation task (word vs. part-word) and word versus final syllable fake-word trials ($r(41) = .39$, $p = .01$). This suggests that learners who successfully recognize words over part-word foils are also more likely to reject fake-words with a 0.0 TP in the final syllable transition.

In part-word versus fake-word trials, performance across syllable manipulation types is positively correlated, but the correlation was only significant between medial and final trial types ($r(41) = .31$, $p = .05$). Finally, the more successfully a participant selected words over part-words, the more likely they were to choose fake-words with final syllable manipulations ($r(41) = -.40$, $p = .008$). In other words – learners who successfully recognize words over part-word foils are also more likely to reject part-words in favor of fake-words with a 0.0 TP in the final syllable transition.

Variable		1	2	3	4	5	6
1. Word vs PW							
2. Word vs FW	*Initial*	.12					
		[-.19, .41]					
3.	*Medial*	.19	-.06				
		[-.12, .47]	[-.36, .25]				
4.	*Final*	**.39***	.14	.14			
		[.09, .62]	[-.17, .43]	[-.17, .43]			
5. PW vs FW	*Initial*	-.14	.03	-.21	-.00		
		[-.43, .17]	[-.28, .33]	[-.48, .10]	[-.31, .30]		
6.	*Medial*	-.14	.13	-.06	.04	.18	
		[-.43, .17]	[-.18, .42]	[-.36, .25]	[-.27, .34]	[-.13, .46]	
7.	*Final*	**-.40****	-.25	-.00	-.09	.19	**.31***
		[-.63, -.11]	[-.52, .06]	[-.31, .30]	[-.38, .22]	[-.12, .47]	[.00, .56]

Table 2.7 Experiment 1 correlations by trial type and syllable position manipulation *Note:* * indicates $p < .05$; ** indicates $p < .01$. Values in square brackets indicate the 95% confidence interval for each correlation.

2.2.3 Discussion

In this study, I presented learners with two minutes of an artificial language composed of four trisyllabic nonce words, which were defined by perfect 1.0 TPs between syllables. I asked whether representations that automatically emerge during SL might share features that characterize words in real-world language acquisition. In particular, I proposed that there would be subtle differences in the nature of the representations across the three syllable positions of a statistically segmented, trisyllabic nonce word. To test this, I asked learners to choose between the nonce words they had been exposed to and two types of non-word foils – those that crossed word boundaries and therefore consisted of one transition with a TP of 1.0 and one transition with a TP of 0.33 (part-words), and those that swapped initial, medial, or final syllables between two 1.0 TP-defined words from the language (fake-words) and so contained at least one transition with a TP of 0.0. I also asked if learners would prefer one type of foil over the other

when they were pitted against each other. I found that learners did not respond in the same way to the different trial types. I also found relationships between segmentation performance and certain position manipulations.

Learners successfully segmented the language, which they demonstrated by endorsing words more frequently than either fake-words or part-words. Mean accuracy across these two trial types was equivalent, which accords with previous work (e.g., Finn et al., 2014), and was moderately correlated overall ($r(38) = 0.46$, $p < .0001$). There was no difference in mean accuracy scores between the various syllable-manipulated fake-words; there were, however, correlational patterns that suggest processing or learning differences across the syllable positions.

For example, successfully choosing words over fake-words of one type (e.g., initial syllable-manipulated) had no bearing on one's performance on other fake-word types (e.g., medial syllable-manipulated). This is surprising, and may suggest that different learners are encoding different parts of the trisyllabic structure. Despite this lack of cohesion between different fake-word trial types, learners who were more likely to reject final syllable manipulated fake-words in favor of words were also better at the classic segmentation task, choosing words over part-words. I did not replicate the Saffran et al. (1996, 1999) findings that participants are more confused by foils with incorrect first syllables but correct final syllable transitions and positions. However, the correlation results in this study appear to align with it. That is, in my task, learners who succeeded on the standard segmentation task were not equally encoding syllable transitions across the trisyllabic word; rather, successful learners were more sensitive to a break in TP between the medial and final syllables than they were between the initial and medial syllables.

When part-words and fake-words were pitted against each other, learners did not prefer either type of non-word foil. This result is difficult to interpret – both the TP-encoding and position-encoding hypotheses predicted a particular direction of choice (i.e., greater proportion choice part-words under TP-encoding, and greater proportion choice fake-words – though possibly not in all syllable manipulations – under the position-encoding account). As in the word versus fake-word condition, however, successful learners behaved differently when faced with fake-words that had final syllable manipulations. In this instance, though, learners (as determined by the word versus part-word task) were more likely to choose the fake-word, as opposed to the relatively higher transitional probability structure they encountered during familiarization (i.e., the part-word). It is worth pausing to unpack what this result might mean. Final-syllable manipulated fake-words break the TP between the medial and final syllable in the trisyllabic chunk. When learners prefer this item over the part-word, it suggests that the medial-to-final syllable transitional probability is a weaker cue to wordhood than is the positionality of the syllables.

The two sets of results relating to final-syllable manipulated fake-words appear to contradict each other. On the one hand, successful learners appear to have homed in on the transition between the last two syllables of a trisyllabic word; on the other, successful learners appear to ignore the coherence of these last two syllables in favor of syllable position. Indeed – there is no correlation between performance on the word versus final-syllable fake-word and part-word versus final-syllable fake-word trials, which may suggest that performance on these two trial types reflects different learning strategies. That these results center on final-syllable manipulations, however, is consistent with other findings of position-based encoding differences in the literature. For example, Conway and Christiansen (2005) found that the statistical

coherence of final sound sequences was more predictive of learning success than that of initial sounds sequences on a grammatical learning task that involved statistical learning (though the paradigm was not the continuous, word segmentation paradigm tested here). Also relevant are studies that show that additional cues used to segment language, such as stress, are more facilitatory to segmentation when placed on final syllables than syllables in other locations (Cunillera et al., 2008). And as mentioned above, children are more likely to extract and produce final syllables than initial or medial syllables when learning words (Slobin, 1973; Echols and Newport, 1992).

And yet the effects uncovered here are quite subtle. There are no significant differences in performance choice across syllable positions, which would provide stronger evidence for differential encoding across syllable positions. The correlational evidence appears to support position-based differences; yet, the significant correlations average around 0.37, with relatively wide confidence intervals. One possibility, then, is that these position-based differences exist, but are very small effects (with greater individual differences) than would be detected under the experimental conditions of previous studies that have looked for their existence, such as in Endress and Mehler (2009b). This is of course not an explanation, however, as to why the positional encoding that appears to emerge from the SL process is of such a weak nature.

The results of Endress and Mehler (2009a) may provide a potential answer. In their study, learners were familiarized to a language structured such that trisyllabic non-word foils could be created that had high adjacent-syllable TPs, but which had never actually been encountered in the speech stream. For instance, syllable A occurred with syllable B frequently, and syllable B occurred with syllable C frequently, but A had never occurred in a trisyllabic sequence with C. Learners chose these non-occurring but high TP sequences as frequently as trisyllabic sequences

they had actually encountered in the stream. This was true even when participants were exposed for eight times the original stimuli duration: participants still failed to distinguish between the two types of items. The only conditions that led to discrimination between encountered and un-encountered high TP items was when prosodic cues signaled the stimuli edges (i.e., small pauses between words, or lengthening of the final syllable vowel durations within the trisyllabic words). Exactly what the results from this study mean is unclear. On the one hand it appears that participants did not encode any non-adjacent TP information (and so, as Endress and Mehler argue, the non-word foils should have been rejected, if participants are encoding the entire trisyllabic sequence in SL tasks); on the other hand, however, the syllables in the high TP foils, though they had never occurred as a unified chunk in the familiarization stream, obeyed positional constraints. It is possible that participants were relying on positional knowledge, and therefore could not distinguish between the two types reliably.

A subsequent study, however, complicates this interpretation (Perruchet & Poulin-Charronat, 2012). In this later study, learners exposed to the same familiarization stream as in Endress and Mehler (2009a) successfully rejected both part-words and the un-encountered but high-TP items when pitted against the trisyllabic chunks they had encountered during familiarization. The authors hypothesized that this discrepancy resulted from a low-level difference in perception: the (Italian-speaking) learners in Endress and Mehler (2009a) may have failed to adequately perceive the unfamiliar (French) speech sounds, whereas their own participants easily encoded their native (French) sounds. Perruchet and Poulin-Charronnat suggest that an inability to accurately encode the encountered sounds impeded the learning mechanism itself. Indeed, lack of familiarity with the stimuli encountered during SL has been shown in other studies to lead to an altered process. For example, Gebhart, Newport and Aslin

(2009) found that adult learners exposed to non-linguistic sounds required increased perceptual salience of those sounds (by adding 150 msec pauses between each sound) and a fivefold increase in exposure. This similarly applies to the segmentation of speech sounds: human infants (14-month-olds) exposed to unfamiliar continuous speech succeed at recognizing non-words as compared to words (0.0 TP vs 1.0 TP), but not when tested on a part-word contrast (0.33 TP) (Graf Estes, Gluck, & Bastos, 2015).

This led me to a new hypothesis: that stronger evidence for position-based effects may be revealed under conditions of reduced phonetic familiarity. To test this, I conducted two additional studies in which adult learners were exposed to two languages that were identical to Study 1 in form, but differed in terms of their relative acoustic familiarity to speakers of North American English. I predicted that the increased difficulty in efficiently perceiving and representing these sounds would lead to greater confusion with or preference for fake-words (see also Morrison & Hudson Kam, 2018, for evidence that unfamiliarity leads to weaker representations and impedes aspects of word learning, and Stager & Werker, 1997, for evidence of a similar effect in infants' word-learning, hypothesized to derive from limited processing efficiency).

2.3 Experiment 2

2.3.1 Methods

2.3.1.1 Participants

Forty-nine adult native-speakers of English were recruited through the University of British Columbia Psychology Department's paid participants listserv. They received $10 for their participation. Five participants were excluded from the analyses: 3 spoke English as a second language, (i.e., were first exposed to English after the age of 3); 1 reported a language disorder; 1 failed to follow instructions.

2.3.1.2 Materials.

Twelve syllables were chosen such that they would structurally parallel the syllables of Experiment 1, but would reflect sounds that are encountered in free variation with a more prototypical form in English, and might not be expected given the syllabic contexts. For example, syllables which in Experiment 1 contained the bilabial sound [p^h] were instead produced with the corresponding ejective consonant [p'] (a p produced with a popping sound that is caused by the release of air compressed between the larynx and oral closure; occasionally heard in conversation in contexts of overemphasis, e.g., if emphasizing the initial or final consonant sounds of the word pop; see Wells, 1982, pg. 261). Syllables in Experiment 1 containing a [b̥] became more prominently pre-voiced versions of /b/ (a free variant of the target

short-lag /b/ of English) in Experiment 2.[11] I will term this language the *Semi-English language*, to reflect that the sounds encountered are English-like, but contain a range of well-known sounds (e.g. [b]) to less familiar ones (e.g. [y]). The entire inventory of sounds and their concatenation in to the 4 trisyllabic words can be found in Table 2.8. They were produced and manipulated in the same way as the materials in Experiment 1.

[11]The original /b/s ranged from short-lag to closures with 1-3 cycles of pre-voicing.

CONSONANTS					VOWELS			WORDS
	BILABIAL	*ALVEOLAR*	*PALATAL*	*VELAR*		*FRONT*	*BACK*	bydʌk'ʊ gœʎʌbʊ t'ʊp'yrœ p'ʌdœt'y
EJECTIVE	p'	t'		k'	*HIGH*	y	ʊ	
PREVOICED	b	d		g	*MID*	œ		
APPROXIMANT			ʎ		*LOW*		ʌ	
TRILL		r						

Table 2.8 Experiment 2 (Semi-English Language) segment and word inventory

2.3.1.3 Analysis

The analysis plan was carried out in the same way as in Experiment 1.

2.3.1.4 Procedure.

The procedure was identical to Experiment 1.

2.3.2 Results

Proportion choice by trial type and syllable manipulations is presented in Figure 2.7; reaction times are presented in Figure 2.8.

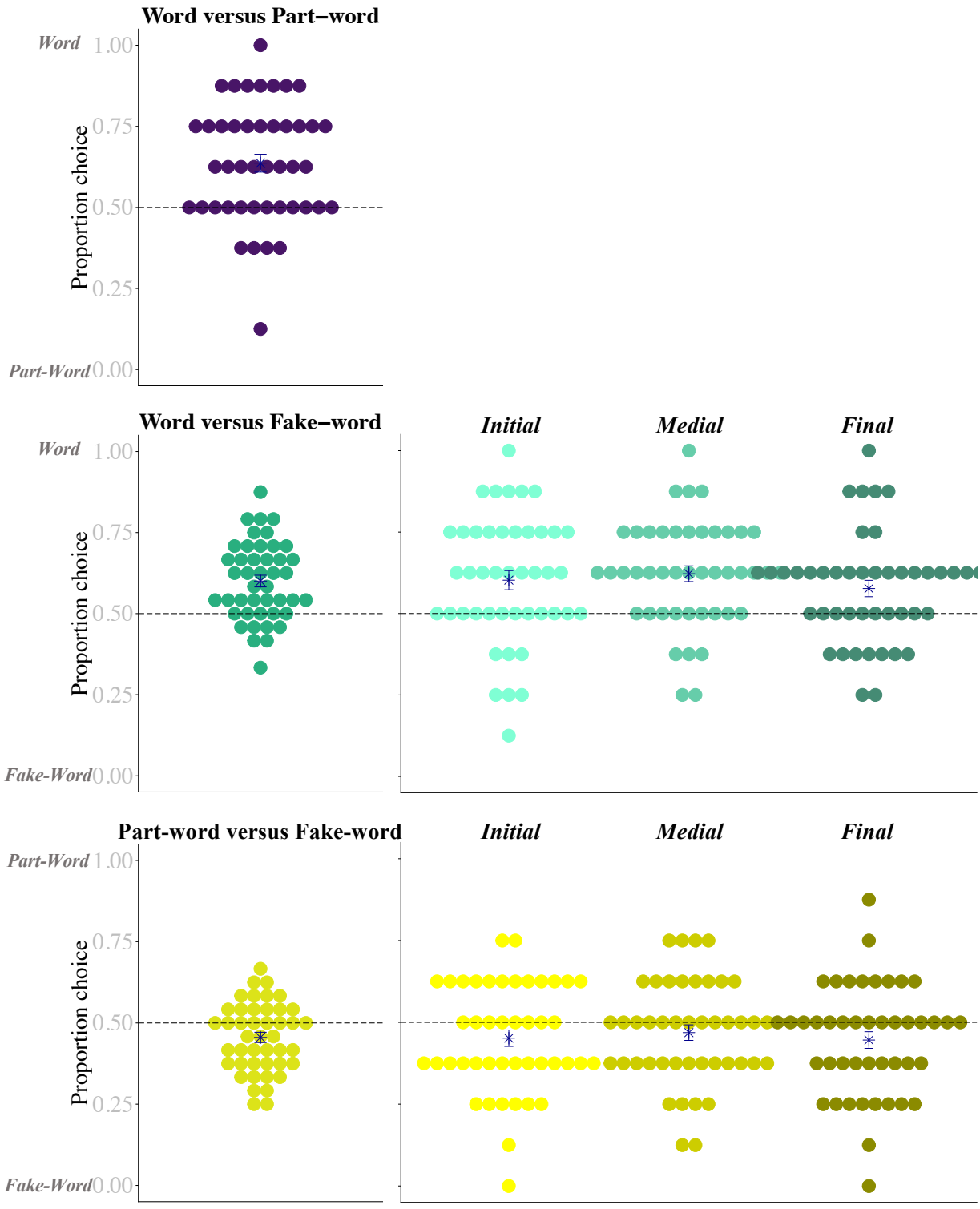

Figure 2.7 Experiment 2 (Semi-English language) proportion choice by trial type and syllable manipulation
Dots reflect individual participant mean scores. Stars reflect mean accuracy scores; error bars are plus/minus 1 standard error. Chance is 0.5 (the dotted line).

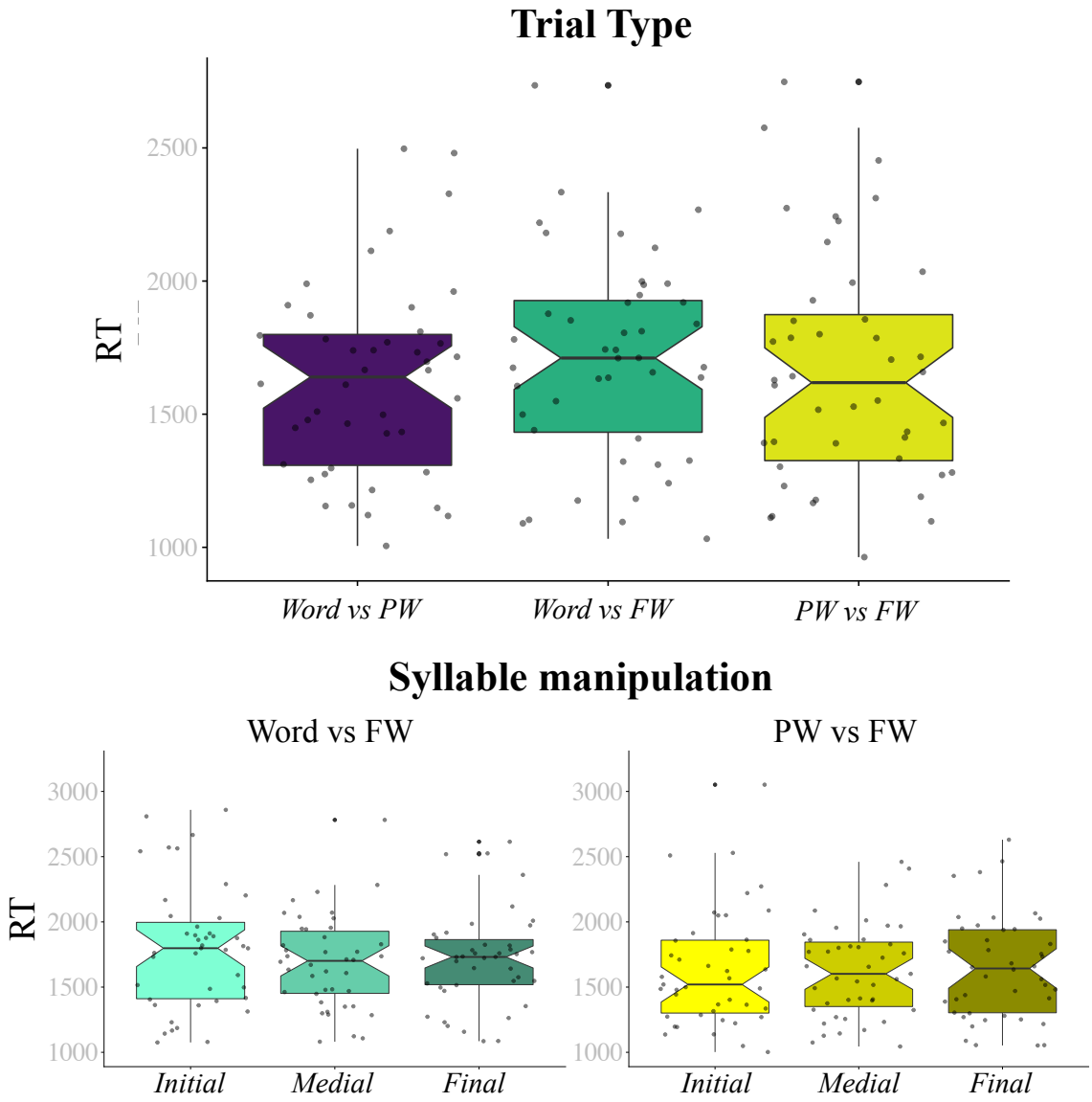

Figure 2.8 Experiment 2 (Semi-English language) reaction times by trial type and syllable manipulation Dots reflect individual participant mean scores. Horizontal lines reflect group medians by condition; boxes cover the 2 middle quartiles, whiskers indicate the range of the top and bottom quartiles.

2.3.2.1 Words versus Part-Words

Participants successfully distinguished words from part-words, as indicated by the fact that they chose words at a rate significantly different from chance (M = 63.6%, SD = 18.2%, 95% CI = [58.1%, 69.2%], $t(43) = 4.96$, $p < .0001$, $d = 0.75$).

2.3.2.2 Words versus Fake-Words

I first report the results for all word versus fake-word trials as a whole, and then break down the results by syllable manipulation type.

2.3.2.2.1 Combined

Overall, participants endorsed words significantly above chance (M = 60.0%, SD = 12.0%, 95% CI = [56.4%, 63.7%], $t(43) = 5.56$, $p < .0001$, $d = 0.83$).

2.3.2.2.2 Syllable Manipulations

Participants chose words significantly more often than fake-words across all syllable positions: Initial (M = 60.2%, SD = 19.5%, 95% CI = [54.3%, 66.2%], $t(43) = 3.48$, $p = .001$, $d = 0.52$), Medial (M = 62.2%, SD = 16.1%, 95% CI = [57.3%, 67.1%], $t(43) = 5.05$, $p < .0001$, $d = 0.76$), and Final (M = 57.7%, SD = 16.7%, 95% CI = [52.6%, 62.7%], $t(43) = 3.05$, $p = .004$, $d = 0.46$). A mixed effects model with interaction and main fixed effects of trial and syllable position, and the same interaction by subject as random effects yields no main effect of syllable position, or of trial. The full model structure and table of results (Table 2.9) can be found below.

Model structure:
Choice ~ Syllable position * Trial + (1 | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	1.52	1.21 – 1.91	**<.001**	1.67	1.32 – 2.11	**<.001**	1.36	1.08 – 1.71	**.008**
Initial Syll				0.91	0.67 – 1.24	.547	1.12	0.83 – 1.52	.470
Medial Syll	1.10	0.81 – 1.49	.547				1.23	0.91 – 1.67	.186
Final Syll	0.89	0.66 – 1.21	.470	0.81	0.60 - 1.10	.186			
Trial	1.00	0.98 - 1.01	.621	0.99	0.98 - 1.00	.204	1.01	1.00 – 1.02	.186
Trial : Initial				1.01	0.99 - 1.02	.572	0.99	0.97 – 1.01	.199
Trial : Medial	0.99	0.98 – 1.01	.572				0.98	0.96 – 1.00	.067
Trial : Final	1.01	0.99 – 1.03	.199	1.02	1.00 - 1.04	.067			
Random Effects									
$\tau_{00,\,Subject}$					0.073				
$N_{Subject}$					44				
$ICC_{Subject}$					0.022				
Observations					1056				
Deviance					1384.879				

Table 2.9 Experiment 2 model for proportion choice words versus fake-words

RT differed by syllable manipulation (Initial: M = 1778 msec, SD = 489 msec; Medial: M = 1657 msec, SD = 379 msec; Final = 1676 msec, SD = 399 msec). This difference was confirmed by mixed effects models, which showed that participants were slower to respond to initial syllable manipulation trials as compared to medial (B = 121, p = .025) syllable manipulation trials (see Table 2.10).

Model structure:
RT ~ Syllable position * Trial + (1 | Subject) + (0 + Syllable position | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1779	1634-1924	**<.001**	1658	1542 – 1774	**<.001**	1674	1556 – 1793	**<.001**
Initial Syll				121	16 – 225	**.025**	104	-3 – 211	.062
Medial Syll	-121	-225 - -16	**.025**				-17	-115 – 81	.737
Final Syll	-104	-211 – 3	.062	17	-81 – 115	.737			
Trial	0	-4 – 5	.861	-2	-6 – 2	.366	2	-3 – 6	.491
Trial : Initial				3	-4 – 8	.447	-1	-7 – 5	.717
Trial : Medial	-2	-8 – 4	.447				-3	-10 – 3	.262
Trial : Final	1	-5 – 7	.717	3	-3 – 10	.262			
Random Effects									
σ^2	434262								
$\tau_{00,\ Subject}$	0.00								
ρ_{01}									
$N_{Subject}$	44								
$ICC_{Subject}$	0.000								
Observations	1056								
R^2 / Ω_0^2	.267/.263								

Table 2.10 Experiment 2 model of reaction time to word versus fake-word trials

2.3.2.3 Word versus Part-Word compared to Word versus Fake-Word

As in Experiment 1, there is no difference in performance between word vs. fake-word and word vs. part-word test trials overall ($t(43) = -1.37$, $p = .18$). RTs are also not significantly different (M = 1661 and 1704 msec, SD = 727 and 750 msec, respectively; a linear mixed effects model yields: $B = -26$ +/- 52 (standard error), $t(41.5) = -0.50$, $p = .62$)[12]. Mixed effects models that compare performance between words versus part-words and each of the syllable manipulation fake-word trial types reveal no differences in proportion choice (Table 2.11, Panel A), but a difference in reaction times: participants were slowest to respond to trials pitting initial-syllable fake words against words (B = 104, $t(1415.6) = 2.09$, $p = .04$; Table 2.11, Panel B).

[12]RT ~ Trial type * Trial + (Trial type * Trial | Subject)

A. Proportion Choice

Model structure:
Choice ~ Contrast type * Trial +
(1 | Subject) +
(0 + Contrast type | Subject) +
(0 + Trial | Subject)

	Odds Ratio	*CI*	*p*
(Intercept)	1.80	1.39 – 2.33	**<.001**
Initial Syll	0.86	0.63 – 1.18	.356
Medial Syll	0.93	0.68 – 1.28	.671
Final Syll	0.76	0.55 – 1.04	.088
Trial	1.00	0.98 – 1.01	.581
Trial * Initial	1.00	0.98 – 1.02	.867
Trial * Medial	0.99	0.98 – 1.01	.619
Trial * Final	1.01	0.99 – 1.03	.200
Random Effects			
$\tau_{00,\ Subject}$	0.000		
$N_{Subject}$	44		
$ICC_{Subject}$	0.00		
Observations	1408		
Deviance	1808		

B. Reaction time

Model structure:
RT ~ Contrast type * Trial +
(1 | Subject) +
(0 + Trial | Subject)

	B	*CI*	*p*
(Intercept)	1668	1545 – 1792	**<.001**
Initial Syll	104	7 – 202	**.037**
Medial Syll	-13	-110 – 84	.792
Final Syll	9	-89 – 107	.861
Trial	-2	-7 – 3	.496
Trial * Initial	2	-4 – 8	.460
Trial * Medial	-0	-6 – 6	.940
Trial * Final	4	-3 – 10	.251
Random Effects			
σ^2	423373		
$\tau_{00,\ Subject}$	119418		
$N_{Subject}$	44		
$ICC_{Subject}$	0.22		
Observations	1408		
R^2 / Ω_0^2	.278/.273		

Table 2.11 Experiment 2 models of proportion choice (Panel A) and reaction time (Panel B) on all word versus non-word trial types

2.3.2.4 Part-Words versus Fake-Words

Results are first reported as main effects, and then broken down by syllable positions.

2.3.2.4.1 Combined

Participants chose fake-words when pitted against part-words at a rate greater than chance (reflected in below performance below 50%; M = 45.5%, SD = 10.5%, 95% CI = [42.4%, 48.7%], $t(43) = -2.81$, $p = .007$, $d = 0.44$; see Figure 2.8).

2.3.2.4.2 Syllable Manipulations

Partipants were significantly more likely to choose fake-words over part-words in the final syllable manipulation, and trended in the same direction of preference across all three syllable manipulations: Initial (M = 45.2%, SD = 16.7%, 95% CI = [40.1%, 50.2%], $t(43) = -1.92$, $p = .06$, $d = 0.29$), Medial (M = 46.9%, SD = 15.7%, 95% CI = [42.1%, 51.7%], $t(43) = -1.32$, $p = .20$, $d = 0.20$), and Final (M = 44.6%, SD = 16.9%, 95% CI = [39.5%, 49.7%], $t(43) = -2.12$, $p = .04$, $d = 0.32$). RT means are similar across the three positions (Initial: M = 1617 msec, SD = 451 msec; Medial: M= 1609 msec, SD = 364 msec; Final: M = 1650 msec, SD = 396 msec). Mixed effects models predicting either proportion correct or reaction time, however, yielded no significant effects (all p's > .3, see Tables 2.12 and 2.13).

Model structure:
Choice ~ Syllable position * Trial + (1 | Subject) + (Trial | Subject)

	Model 1 Ref level = Initial			**Model 2** Ref level = Medial			**Model 3** Ref level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	0.82	0.66 – 1.02	.070	0.88	0.71 – 1.09	.237	0.80	0.65 -1.00	.046
Initial Syll				0.93	0.69 – 1.26	.653	1.02	0.76 – 1.38	.890
Medial Syll	1.07	0.79 – 1.44	.653				1.09	0.81 – 1.47	.556
Final Syll	0.98	0.73 – 1.32	.890	0.91	0.68 – 1.23	.557			
Trial	0.99	0.98 – 1.01	.294	1.00	0.98 – 1.01	.619	1.00	0.99 – 1.01	.950
Trial : Initial				1.00	0.98 – 1.01	.674	0.99	0.97 – 1.01	.472
Trial : Medial	1.00	0.99 – 1.02	.674				1.00	0.98 – 1.02	.757
Trial : Final	1.01	0.99 – 1.03	.472	1.00	0.98 – 1.02	.757			
Random Effects									
$\tau_{00,\ Subject}$					0.000				
ρ_{01}									
$N_{Subject}$					44				
$ICC_{Subject}$					0.000				
Observations					1056				
Deviance					1446				

Table 2.12 Experiment 2 model for proportion choice part-words versus fake-words

Model structure:
RT ~ Syllable position * Trial + (1 | Subject) + (Trial | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1617	1484 – 1751	**<.001**	1608	1496 - 1721	**<.001**	1647	1528 – 1766	**<.001**
Initial Syll				9	-91 – 109	.857	-30	-135 – 76	.585
Medial Syll	-9	-108 – 91	.857				-39	-132 – 54	.412
Final Syll	30	-76 – 135	.585	39	-54 – 132	.412			
Trial	1	-3 – 6	.581	-0	-4 – 4	.864	-2	-6 – 2	.374
Trial : Initial				2	-4 – 8	.606	3	-3 – 9	.313
Trial : Medial	-2	-8 – 4	.606				2	-4 – 7	.609
Trial : Final	-3	-9 – 3	.313	-2	-7 – 4	.609			
Random Effects									
σ^2					378580				
$\tau_{00,\ Subject}$					0.000				
ρ_{01}									
$N_{Subject}$					44				
$ICC_{Subject}$					0.000				
Observations					1056				
R^2 / Ω_0^2					.284/.278				

Table 2.13 Experiment 2 model of reaction time to part-words versus fake-words

2.3.2.5 Correlations

As in previous sections, I first present correlations across the main trial types, and then by syllable position.

2.3.2.5.1 Combined

Similarly to the pattern found in Experiment 1, participants who chose words over part-words were also more likely to choose words over fake-words ($r(42) = 0.39$, $p = .009$), and – though non-significant – the more successful participants were at choosing words over part-words, the more likely they were to endorse fake-words over part-words: $r(42) = -0.26$, $p = .08$. As in Experiment 1, this relationship is attenuated for the comparison of word versus fake-word and fake-word versus part-word trials, but patterns in the expected direction ($r(42) = -0.17$, $p = .28$). These relationships are plotted in Figure 2.9.

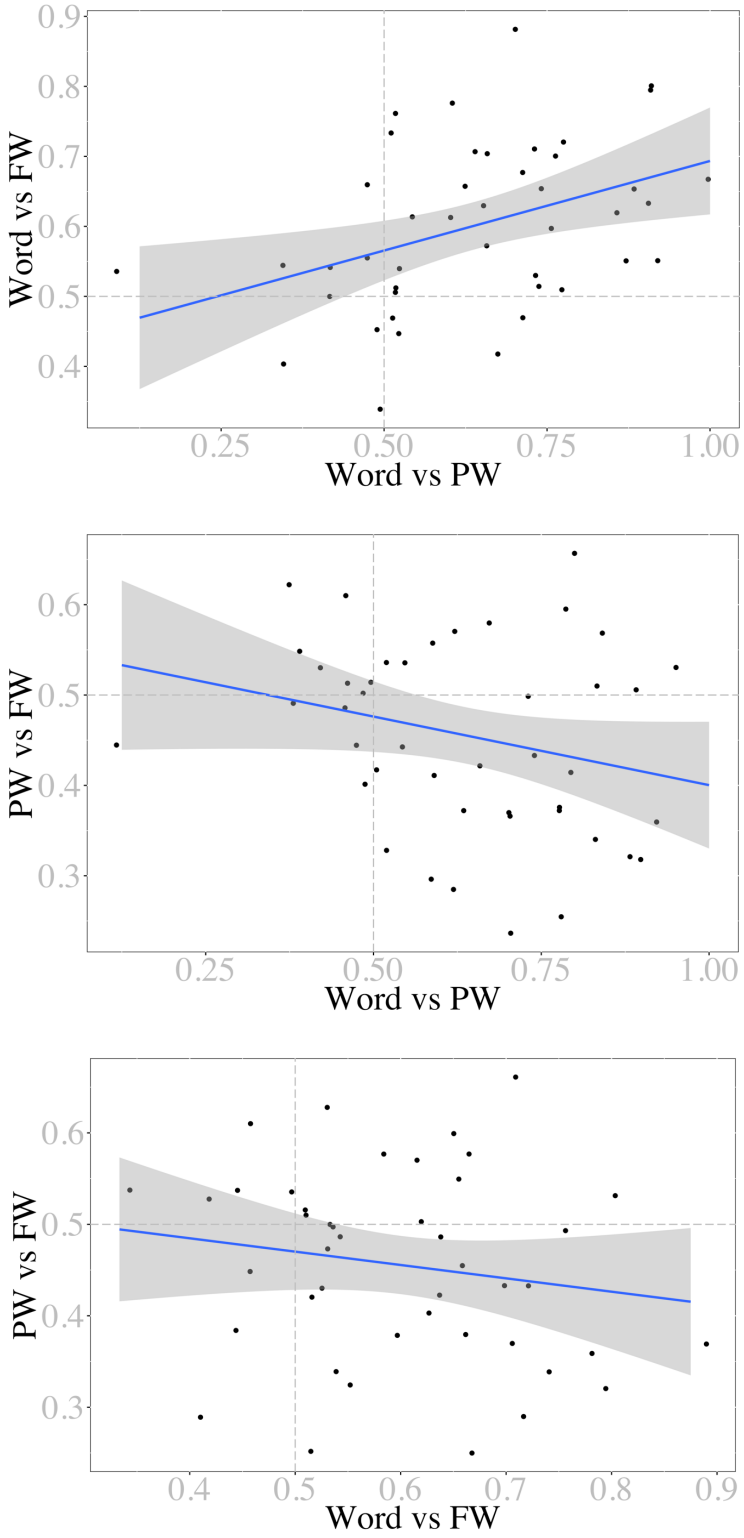

Figure 2.9 Experiment 2 correlations between main trial types Dots represent participant mean performance. The dotted vertical and horizontal lines reflect chance performance in the respective conditions.

2.3.2.5.2 Syllable manipulations.

The full correlation table can be found in Table 2.14. Performance on the standard word segmentation task (words versus part-words) is positively correlated with performance on word versus medial-syllable manipulated fake-word trials, though the same pattern holds across all syllable positions (range r = .23 to .37). As in Experiment 1, there is again a correlation between word versus part-word trials and part-word versus final-syllable manipulated fake-words ($r(42)$ = -0.30, p = .05).

Variable		1	2	3	4	5	6
1. Word vs PW							
2. Word vs FW	*Initial*	.23 [-.07, .49]					
3.	*Medial*	.37* [.09, .60]	.25 [-.05, .51]				
4.	*Final*	.21 [-.09, .48]	.18 [-.13, .45]	.18 [-.12, .46]			
5. PW vs FW	*Initial*	-.28 [-.53, .02]	-.08 [-.37, .22]	-.25 [-.51, .05]	.10 [-.21, .38]		
6.	*Medial*	.09 [-.21, .38]	-.17 [-.44, .14]	-.06 [-.35, .24]	.23 [-.07, .49]	-.11 [-.40, .19]	
7.	*Final*	-.30* [-.55, -.00]	-.26 [-.52, .04]	-.21 [-.47, .10]	.09 [-.22, .37]	.19 [-.11, .46]	.25 [-.05, .51]

Table 2.14 Experiment 2 correlations by trial type and syllable position manipulation *Note.* * indicates $p < .05$; ** indicates $p < .01$. Values in square brackets indicate the 95% confidence interval for each correlation

2.3.3 Discussion

Experiment 1 found some evidence for position-based encoding by demonstrating (1) that learners did not treat lower TP fake-words as easier to reject compared to part-words, (2) that learners do not prefer high TP items (part-words) over lower-TP, but positionally-accurate, items (fake-words), and (3) that learners who better distinguished words from part-words also

preferred fake-words over part-words, in particular those with final syllable manipulations. There was not, however, any difference in mean performance by syllable-position manipulation. I hypothesized that a more taxing listening environment might enhance these positional effects; this hypothesis was confirmed, but in subtle ways. As in Experiment 1, there was no difference in performance between word versus part-word and word versus fake-word trials. Unlike in Experiment 1, however, in Experiment 2 participants showed a slight preference for fake-words over the higher TP part-word counterparts. The correlations and reaction times further point to differences across syllable position manipulations. In particular, there is evidence in Experiment 2 of a special role for initial syllable sequences: participants were slower to reject fake-words with initial syllable manipulations, and trials with initial-manipulated fake-words were negatively correlated with trials pitting part-words against medial and final syllable-manipulated fake-words.

Taken together, the results of Experiment 2 largely replicate the results of Experiment 1, despite the linguistic differences. Though there is correlational and reaction time evidence for position-based encoding differences, I did not find differences in mean performance across syllable positions. It is possible that greater differences were not observed because the SEL sounds may be highly assimilable to existing English speech sound categories (see Best, 1994 and Best, McRoberts, & Goodell, 2001 for relevant models of non-native speech sound assimilation) and are therefore perceived and held in memory much like the familiar English sounds of Experiment 1. Experiment 3 was therefore designed to increase the perceptual distance between the target sounds and native English phonemes.

2.4 Experiment 3

2.4.1 Methods

2.4.1.1 Participants.

Forty-two adult native-speakers of English were recruited through the University of British Columbia Psychology Department's paid participants listserv (22), or the Linguistic Department's subject pool (20). Participants through the Psychology Department listserv were paid $10; participants through the Linguistic Department's subject pool received course credit or $5.[13] Three participants were excluded for the following reasons: 2 spoke English as a second language (i.e. were first exposed to English after the age of 3); 1 failed to follow instructions. The final sample thus consisted of data from 38 participants.

2.4.1.2 Materials

Twelve syllables were chosen such that they would structurally parallel the syllables of Experiment 1, but contained unfamiliar sounds. This included changing the place of articulation for two of the three consonant places of articulation (i.e., alveolar to palatal, and velar to uvular), and the two obstruent manners of articulation (short-lag to implosive, and aspirated to ejective). The vowel system was changed so that rounding – which characterizes high and mid back

[13]These subjects were run at a later time; the norms around amount payed per time spent in the lab differed by the two different subject pools (i.e., Psychology versus Linguistics).

vowels in English – characterized the non-high vowels instead. Given these paradigmatic shifts in place and manner of articulation, it is unlikely that many of these sounds would occur allophonically in English. I term this language the *Non-English Language* (NEL), for easy reference. The full inventory can be found in Table 2.15. Syllables were produced and manipulated in the same way as the materials in Experiment 1.

CONSONANTS				VOWELS			WORDS
	BILABIAL	*PALATAL*	*UVULAR*		*FRONT*	*BACK*	ɓɨʃɒkˈɯ
EJECTIVE	pˈ	cˈ	qˈ	*HIGH*	ɨ	ɯ	ɕœʎɒɓɯ
IMPLOSIVE	ɓ	ʃ	ɠ	*MID*	œ		cˈɯpˈɨʀœ
APPROXIMANT		ʎ		*LOW*		ɒ	pˈɒʃœcˈʏ
TRILL			R				

Table 2.15 Experiment 3 segmental inventory (Non-English Language) The far right column shows how these segments were combined in to the four trisyllabic words of the exposure language.

2.4.1.3 Analysis.

The analysis was conducted in the same way as Experiment 1.

2.4.1.4 Procedure.

The procedure was identical to Experiment 1.

2.4.2 Results

Performance choice across trial types is shown graphically in Figure 2.10; reaction times by trial type are shown in Figure 2.11.

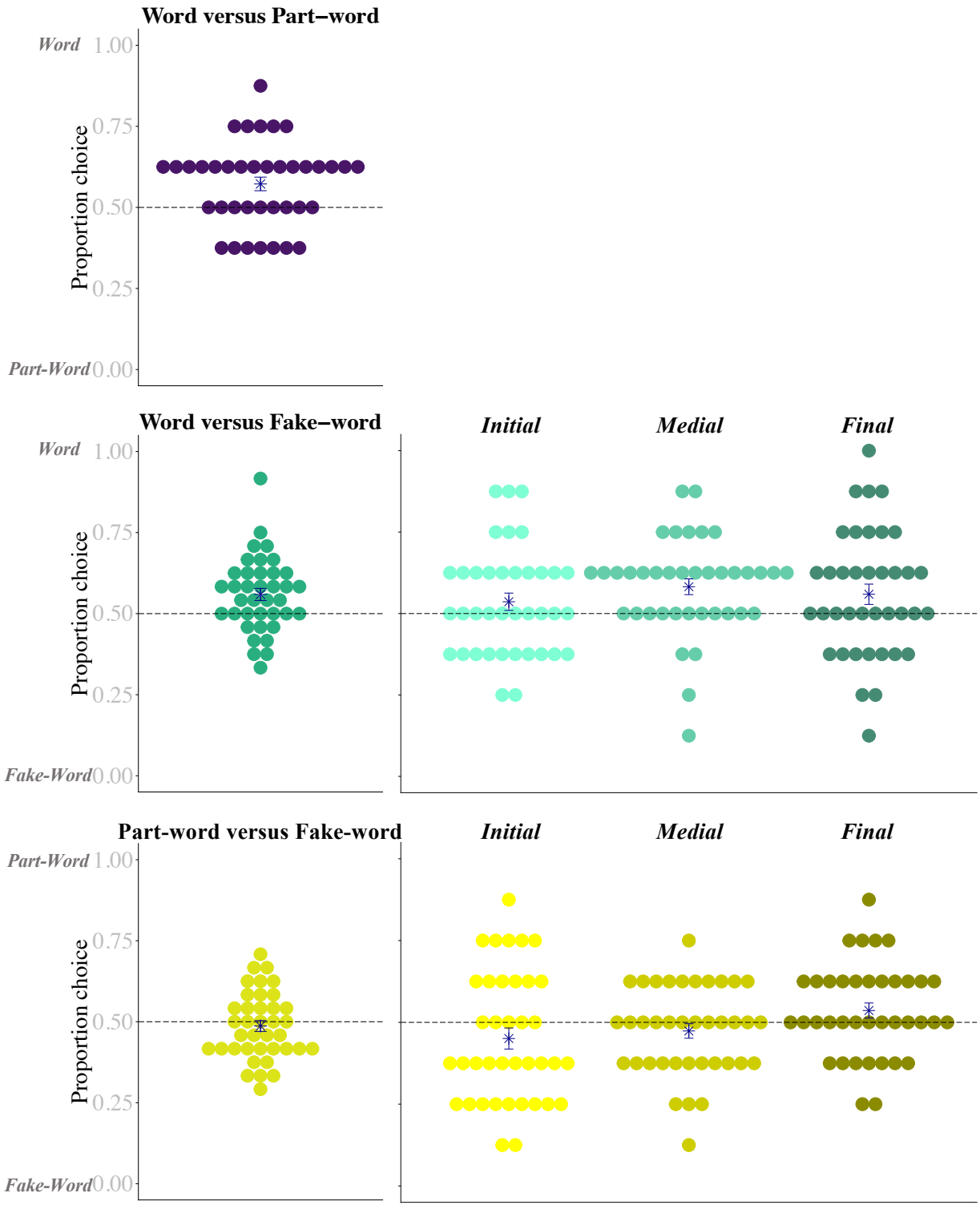

Figure 2.10 Experiment 3 proportion choice across trial types Dots reflect individual participant mean scores. Stars reflect mean accuracy scores; error bars are plus/minus 1 standard error. Chance is 0.5 (the dotted line).

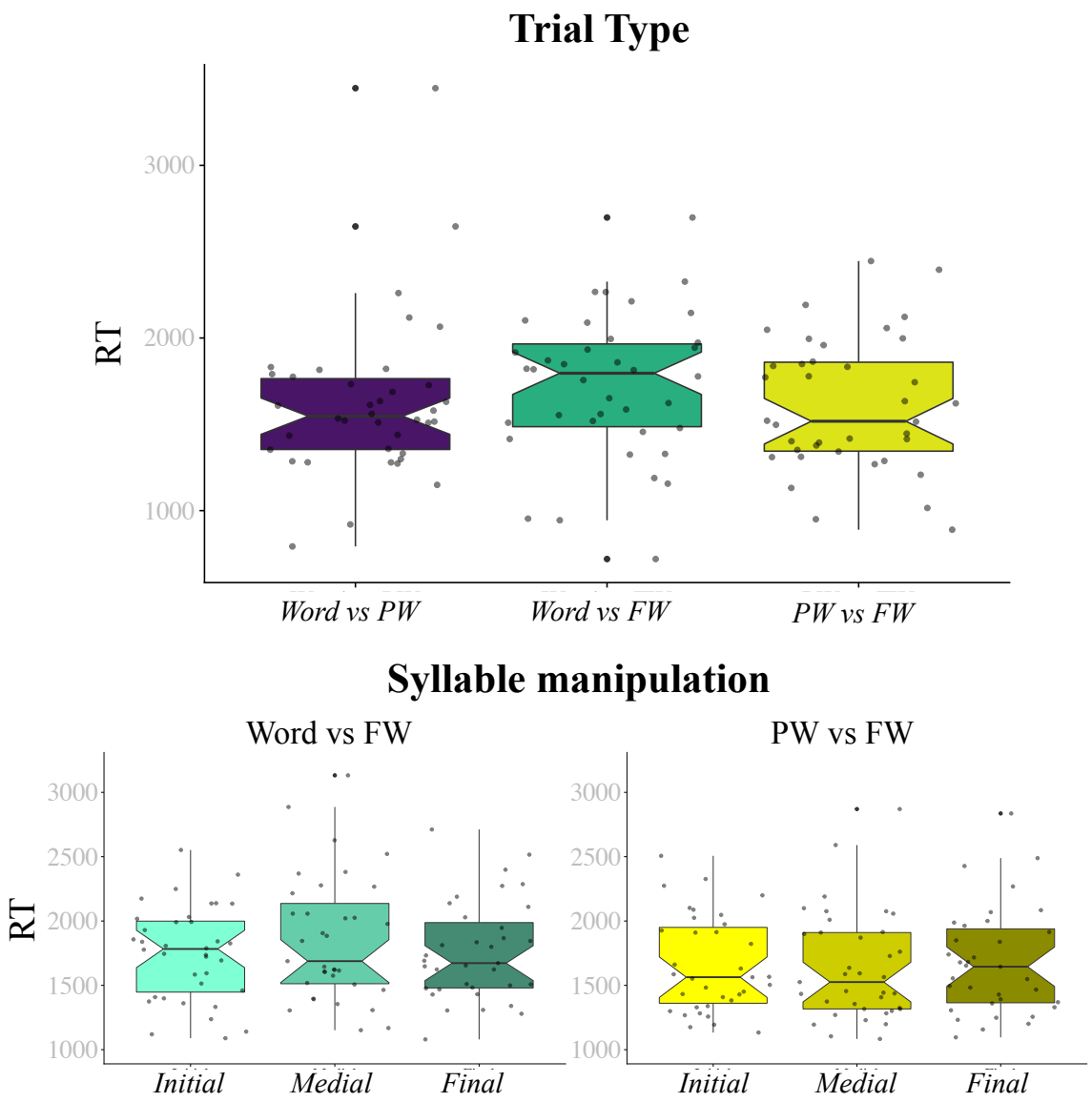

Figure 2.11 Experiment 3 reaction times across trial types and syllable manipulations Dots represent individual's mean RTs; boxes reflect the two middle quartiles; the horizontal line is the median RT; whiskers represent the limits of the bottom and top quartile (excluding outliers).

2.4.2.1 Words versus Part-Words.

Participants chose words at rates significantly above chance (M = 57.2%, SD = 12.9%, 95% CI = [59%, 70%], $t(37) = 3.46$, $p < .001$, $d = 0.56$; see Figure 2.10).

2.4.2.2 Words versus Fake-Words.

Results are first presented across the trial as a whole, and then broken down by syllable position.

2.4.2.2.1 Combined.

Participants chose words at rates significantly above chance (M = 55.9%, SD = 11.4%, 95% CI = [52.2%, 59.7%], $t(37) = 3.19$, $p = .003$, $d = 0.52$).

2.4.2.2.2 Syllable Manipulations.

Not all syllable-manipulation trial types were significantly different from chance: Initial (M = 53.6%, SD = 16.4%, 95% CI = [48.2%, 59.0%], $t(37) = 1.4$, $p = .18$, $d = 0.22$), Medial (M = 58.2%, SD = 14.9%, 95% CI = [53.3%, 63.1%], $t(37) = 3.40$, $p = .002$, $d = 0.55$), and Final (M = 55.9%, SD = 19.2%, 95% CI = [49.6%, 62.2%], $t(37) = 1.90$, $p = .07$, $d = 0.31$). This difference across syllable positions was also reflected numerically in mean response times (Initial: M = 1696 msec, SD = 422; Medial: M = 1780 msec, SD = 536; Final: M = 1687 msec, SD = 443). However, neither proportion choice nor RT means differed significantly from each other in the respective mixed effects models (Table 2.16 for proportion choice; Table 2.17 for RT). Participants became slightly faster on medial- and final-syllable fake-word trials as the task went on (see Figure 2.11).

Model structure:
Choice ~ Syllable position * Trial + (SyllPos | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	1.16	0.91 – 1.47	.235	1.40	1.11 – 1.76	**.004**	1.28	0.99 – 1.66	**.059**
Initial Syll				0.83	0.60 – 1.15	.254	0.90	0.65 – 1.25	.530
Medial Syll	1.21	0.887 – 1.67	.255				1.09	0.78 – 1.52	.624
Final Syll	1.10	0.80 – 1.54	.530	0.92	0.66 – 1.29	.624			
Trial	0.99	0.98 – 1.01	.334	1.00	0.98 – 1.01	.621	1.00	0.98 – 1.01	.491
Trial : Initial				1.00	0.98 – 1.02	.721	1.00	0.98 – 1.02	.835
Trial : Medial	1.00	0.98 – 1.02	.721				1.00	0.98 – 1.02	.883
Trial : Final	1.00	0.98 – 1.02	.835	1.00	0.98 – 1.02	.883			
Random Effects									
$\tau_{00,\ Subject}$					0.000				
ρ_{01}									
$N_{Subject}$					38				
$ICC_{Subject}$					0.000				
Observations					912				
Deviance					1224				

Table 2.16 Experiment 3 model for proportion choice words versus fake-words

Model structure:
RT ~ Syllable position * Trial + (1 | Subject) + (Trial | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1695	1484 – 1751	**<.001**	1772	1629 – 1915	**<.001**	1690	1545 – 1833	**<.001**
Initial Syll				-77	-181 – 22	.150	6	-98 – 110	.914
Medial Syll	77	-28 – 181	.150				82	-22 – 186	.121
Final Syll	-6	-110 – 98	.914	-82	-186 – 22	.121			
Trial	-4	-10 – 1	.119	-6	-12 – -1	**.023**	-9	-14 – -3	**.003**
Trial : Initial				2	-5 – 8	.560	6	-3 – 11	.224
Trial : Medial	-2	-8 – 5	.560				2.59	-4 – 8	.514
Trial : Final	-4	-11 – 2	.224	-2	-8 – 4	.514			
Random Effects									
σ^2					417636				
$\tau_{00,\ Subject}$					148658				
ρ_{01}									
$N_{Subject}$					38				
$ICC_{Subject}$					0.262				
Observations					912				
R^2 / Ω_0^2					.359/.353				

Table 2.17 Experiment 3 model for reaction time to word versus part-word trials

2.4.2.3 Words versus Part-Words compared to Words versus Fake-Words.

There is no difference in proportion choice performance between word versus fake-word and word versus part-word test trials ($t(37) = -0.50$, $p = .62$), but participants were slower to choose on word versus fake-word trials (linear mixed effects model with Trial Type and trial as interactions and main fixed effects, and the same interaction grouped by subject in the random effects structure, $B = 97$ +/- 45 (standard error), $t(106.5) = 2.15$, $p = .03$). Proportion choice and RT mixed effects models were fitted to determine whether word versus fake-word trials of the various syllable manipulations differed from the word versus part-word condition. The logistic regression model results for proportion choice are found in Table 2.18, Panel A, the linear regression model results for reaction time are in Panel B.

A. Proportion Choice

Model structure:
Choice ~ Contrast type * Trial +
(1 | Subject) +
(0 + Contrast type | Subject)

	Odds Ratio	*CI*	*p*
Fixed Effects			
(Intercept)	1.33	1.06 – 1.67	**.014**
Initial Syll	0.87	0.63 – 1.20	.394
Medial Syll	1.05	0.76 – 1.45	.775
Final Syll	0.96	0.69 – 1.35	.832
Trial	1.00	0.99 – 1.02	.523
Trial * Initial	0.99	0.97 – 1.01	.255
Trial * Medial	0.99	0.97 – 1.01	.423
Trial * Final	0.99	0.97 – 1.01	.350
Random Effects			
$\tau_{00,\ Subject}$		0.000	
ρ_{01}			
$N_{Subject}$		38	
$ICC_{Subject}$		0.000	
Observations		1216	
Deviance		1638	

B. RT

Model structure:
RT ~ Contrast type * Trial +
(Trial | Subject)

B	*CI*	*p*
Fixed Effects		
1623	1479 – 1767	**<.001**
71	-33 – 175	.183
155	50 – 259	**.004**
67	-38 – 171	.211
-1	-6 – 4	.736
-4	-10 – 3	.280
-6	-12 – 1	.097
-7	-14 – -1	**.033**
Random Effects		
σ^2	424371	
$\tau_{00,\ Subject}$	151694	
ρ_{01}	-0.296	
$N_{Subject}$	38	
$ICC_{Subject}$	0.263	
Observations	1216	
R^2 / Ω_0^2	.324/.321	

Table 2.18 Experiment 3 models for proportion choice (Panel A) and reaction time (Panel B) to words versus all non-words

While proportion choice did not significantly differ between word versus part-word and any syllable-manipulated fake-word trials, there were RT differences. Participants were slower to respond to word versus medial syllable fake-word trials than they were word versus part-word trials (B = 155 +/- 53.4 (standard error), t(1157.4) = 2.90, p = .004). There was also an interaction such that participants got slightly faster at final-syllable manipulated trials in comparison to words versus part-word trials over the course of the experiment (Final: B = -7 +/- 3.3, t(1152.7) = -2.13, p = .03). Though this same pattern was observed for the medial- and

initial- syllable manipulated trials, these comparisons were not significant (Medial: B = -6, t(1155.3) = -1.66, p = .10; Initial: B = -4, t(1153.1) = -1.08, p = .28). There was no main effect of trial (B = -1, p = .74).

2.4.2.4 Part-Words versus Fake-Words.

I first present results across the main trial types, and then break the data down by syllable position.

2.4.2.4.1 Combined

Participants failed to choose either part-words or fake-words (M = 48.7%, SD = 10.4%, 95% CI = [45.3%, 52.1%], t(37) = -0.78, p = .44, d = 0.13; see Figure 2.10).

2.4.2.4.2 Syllable Manipulations

Performance is at chance across syllable positions: Initial (M = 45.1%, SD = 19.8%, 95% CI = [38.5%, 51.6%], t(37) = -1.32, p = .13, d = 0.25), Medial (M = 47.4%, SD = 13.7%, 95% CI = [42.9%, 51.9%], t(37) = -1.19, p = .24, d = 0.19), and Final (M = 53.6%, SD = 14.2%, 95% CI = [48.9%, 58.3%], t(37) = 1.57, p = .13, d = 0.25). A mixed effects model, however, suggests that performance on initial syllable manipulation trial types consists of greater proportion choice fake-words as compared to final syllable manipulation trial types (OR = 0.72, p = .04) (Table 2.19). RTs, on the other hand, are equivalent across syllable position (all p's > .1, see Table 2.20).

Model structure:
Choice ~ Syllable position * Trial + (Trial | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	0.81	0.65 – 1.02	.077	0.91	0.73 – 1.14	.416	1.14	0.90 – 1.43	.272
Initial Syll				0.90	0.65 – 1.23	.498	0.72	0.52 – 0.99	**.042**
Medial Syll	1.12	0.81 – 1.54	.498				0.80	0.58 – 1.10	.176
Final Syll	1.39	1.01 – 1.92	**.042**	1.25	0.91 – 1.72	.176			
Trial	0.99	0.98 – 1.01	.425	0.99	0.97 – 1.00	.061	0.99	0.97 – 1.00	.058
Trial : Initial				1.01	0.99 – 1.03	.470	1.01	0.99 – 1.03	.447
Trial : Medial	0.99	0.97 – 1.01	.470				1.00	0.98 – 1.02	.964
Trial : Final	0.99	0.97 – 1.01	.447	1.00	0.98 – 1.02	.964			
Random Effects									
$\tau_{00,\ Subject}$					0.00				
ρ_{01}					1.00				
$N_{Subject}$					38				
$ICC_{Subject}$					0.00				
Observations					912				
Deviance					1251				

Table 2.19 Experiment 3 models for proportion choice part-words versus fake-words

Model structure:
RT ~ Syllable position * Trial + (Syllable position | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1600	1470 – 1731	**<.001**	1621	1483 – 1759	**<.001**	1604	1454 – 1753	**<.001**
Initial Syll				-21	-132 – 91	.715	-3	-121 – 114	.955
Medial Syll	21	-91 – 132	.715				17	-90 – 125	.750
Final Syll	3	-114 – 121	.955	-17	-125 – 90	.750			
Trial	-2	-7 – 3	.462	-4	-9 – 1	.094	-6	-10 – -1	**.021**
Trial : Initial				2	-5 – 9	.526	4	-3 – 11	.275
Trial : Medial	-2	-9 – 5	.526				2	-5 – 8	.635
Trial : Final	-4	-11 – 3	.275	-2	-8 – 5	.635			
Random Effects									
σ^2	442016								
$\tau_{00,\ Subject}$	112543								
ρ_{01}	0.106								
$N_{Subject}$	38								
$ICC_{Subject}$	0.203								
Observations	912								
R^2 / Ω_0^2	.274/.269								

Table 2.20 Experiment 3 models of reaction time to part-words versus fake-words

2.4.2.5 Correlations

There are no significant correlations across the full correlation matrix (see Tables 2.21 and 2.22)

Variable	1	2
1. Word vs PW		
2. Word vs FW	.10 [-.22, .41]	
3. PW vs FW	-.07 [-.39, .25]	-.02 [-.34, .30]

Table 2.21 Experiment 3 correlations by trial type *Note:* Values in square brackets indicate the 95% confidence interval for each correlation.

Variable		1	2	3	4	5	6
1. Word vs PW							
2. Word vs FW	*Initial*	-.03 [-.34, .29]					
3.	*Medial*	.08 [-.25, .39]	.24 [-.09, .52]				
4.	*Final*	.15 [-.18, .45]	.29 [-.03, .56]	.03 [-.29, .35]			
5. PW vs FW	*Initial*	-.27 [-.54, .05]	.06 [-.27, .37]	-.04 [-.36, .28]	-.11 [-.41, .22]		
6.	*Medial*	.02 [-.31, .33]	.01 [-.31, .33]	-.04 [-.35, .29]	.13 [-.20, .43]	.22 [-.11, .50]	
7.	*Final*	.20 [-.13, .49]	-.26 [-.53, .07]	.02 [-.31, .33]	.17 [-.16, .46]	.07 [-.26, .38]	.07 [-.25, .38]

Table 2.22 Experiment 3 correlations by trial type and syllable position manipulation *Note:* Values in square brackets indicate the 95% confidence interval for each correlation.

2.4.3 Discussion

While participants successfully distinguished words from non-words (part-words or fake-words) when familiarized to non-English language sounds, several aspects of their performance suggest learning suffered in comparison to learning in the native-English and semi-English sound conditions. First, the average proportion choice was numerically lower (mean proportion choice

of words over part-words of 66% for the native-English language, 64% for the semi-English language, and 57% for the non-English language; $F(2, 121) = 2.76$, $p = .07$) and yielded smaller effect sizes (average Cohen's $d = .4$ in the non-English, as compared to .7 in both the native and semi-English language conditions). Second, unlike in the previous two language conditions, performance is not correlated across trial types.

I had hypothesized that learners' degraded capacity to encode the acoustic signal, as a result of the unfamiliar, non-English sounds, would lead to stronger positional effects. Instead, however, I found that the increased unfamiliarity of the sounds led to reduced learning overall, and a somewhat different pattern with respect to positional information. In the paragraphs that follow, I will break this down first into the ways that the three studies converge, followed by the patterns that diverge.

Under the TP-encoding account, learners' choices should reflect the underlying TPs. That is, a 0.0 TP should be easier to reject than a 0.33 TP. Under the position-encoding account, however, 0.0 TP sequences might be more difficult to reject, because the information coincides with a secondary source of information encoded in the extracted word representation – namely, the position of certain syllables. The three studies each showed that learners did not, overall, find fake-words easier to reject than the higher TP part-words when pitted against the high TP words, and did not clearly endorse the (theoretically) more familiar 0.33 TP part-words over positionally-based 0.0 TP non-words.

The native-English and semi-English learning conditions patterned similarly with respect to the position-based effects. In both, there was a propensity to choose final syllable fake-words over part-words, and this propensity was associated with better word-segmentation performance (as determined by the word versus part-word trials). Additionally, in the semi-native English

language condition, better learners (as determined by the word versus part-word trials) were also more likely to choose initial-syllable fake-words over part-words, performed slightly worse on words versus final-syllable fake-words, and were slower when asked to choose between words and initial syllable fake-words. These patterns were not observed in the non-English language experiment, where there is subtle evidence for medial position effects: participants were better at rejecting, but also slower to respond to, medial-syllable fake-words than they were word versus part-word trials.

The results of the non-English language condition are difficult to interpret. Better performance on the medial-syllable fake-words is consistent with a TP-encoding account; if learners were also better at all word versus fake-word trials as compared to word versus part-word trials, the evidence would further favour this mechanistic explanation. As this expectation was not confirmed (or rejected), we must look to other data for answers. One clear conclusion to draw from this study, however, is that decreased familiarity with the stimuli did not enhance the expected positional learning effects.

In Experiment 4, I continue to ask whether an increase in perceptual load will lead to enhanced position-based effects, but employ a different means of increasing perceptual load. Recent work has suggested that a key component of statistical learning is executive function – in particular, the capacities of attention and inhibition (Toro et al., 2005; Turk-Browne et al., 2009; Finn et al., 2014; Forest, 2017). I hypothesized that this, rather than familiarity with the stimuli, may lead to different learning processes (i.e., position-based encoding versus TP-tracking). To test this hypothesis, I introduced a new manipulation that would tax learners' capacity to attend to the auditory stimuli, but did not alter the perceptual availability of the stimuli, by having participants watch a silent, unrelated cartoon during exposure to familiar language sounds. This

manipulation was chosen as previous work has shown that attention to an unrelated visual display does not impede learning under passive viewing conditions (Toro et al., 2005); that is, it should not simply lead to low learning overall, as in Experiment 3.

2.5 Experiment 4

2.5.1 Methods

2.5.1.1 Participants.

Thirty-nine adult native-speakers of English were recruited through the University of British Columbia Psychology Department's paid participants listserv. Participants were paid $10 for their participation.

2.5.1.2 Materials

The language stimuli were identical to Experiment 1. The cartoon video was a muted clip from the 1969 Russian cartoon version of Winnie the Pooh (Soyuzmultfilm, 1969), timed to coincide with the onset and offset of the language stimuli. This cartoon was chosen because it is sufficiently engaging as to hold participants' attention, and would likely be unfamiliar to most participants [14].

[14]This same video was used in a similar study with young children (not reported on in this thesis).

2.5.1.3 Procedure

The procedure was identical to Experiment 1, with the exception that participants simultaneously watched a 2-minute video of a silent cartoon during familiarization.

2.5.2 Results

Proportion choice and RT scores across all trial types can be found in Figures 2.12 and 2.13.

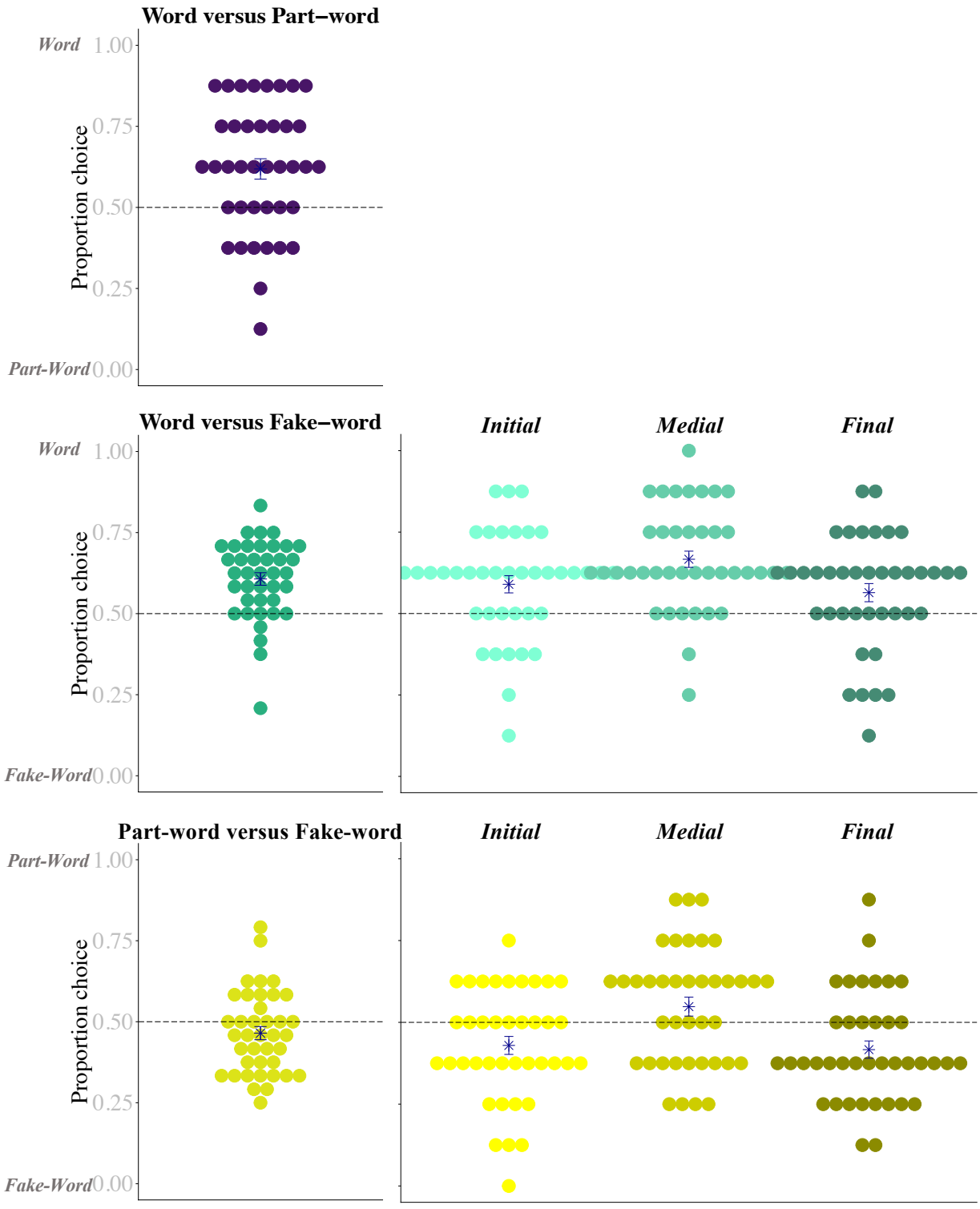

Figure 2.12 Experiment 4 (Video + Native English Language) proportion choice across trial types and syllable manipulations Dots reflect individual participant mean scores. Stars reflect mean accuracy scores; error bars are plus/minus 1 standard error. Chance is 0.5 (the dotted line).

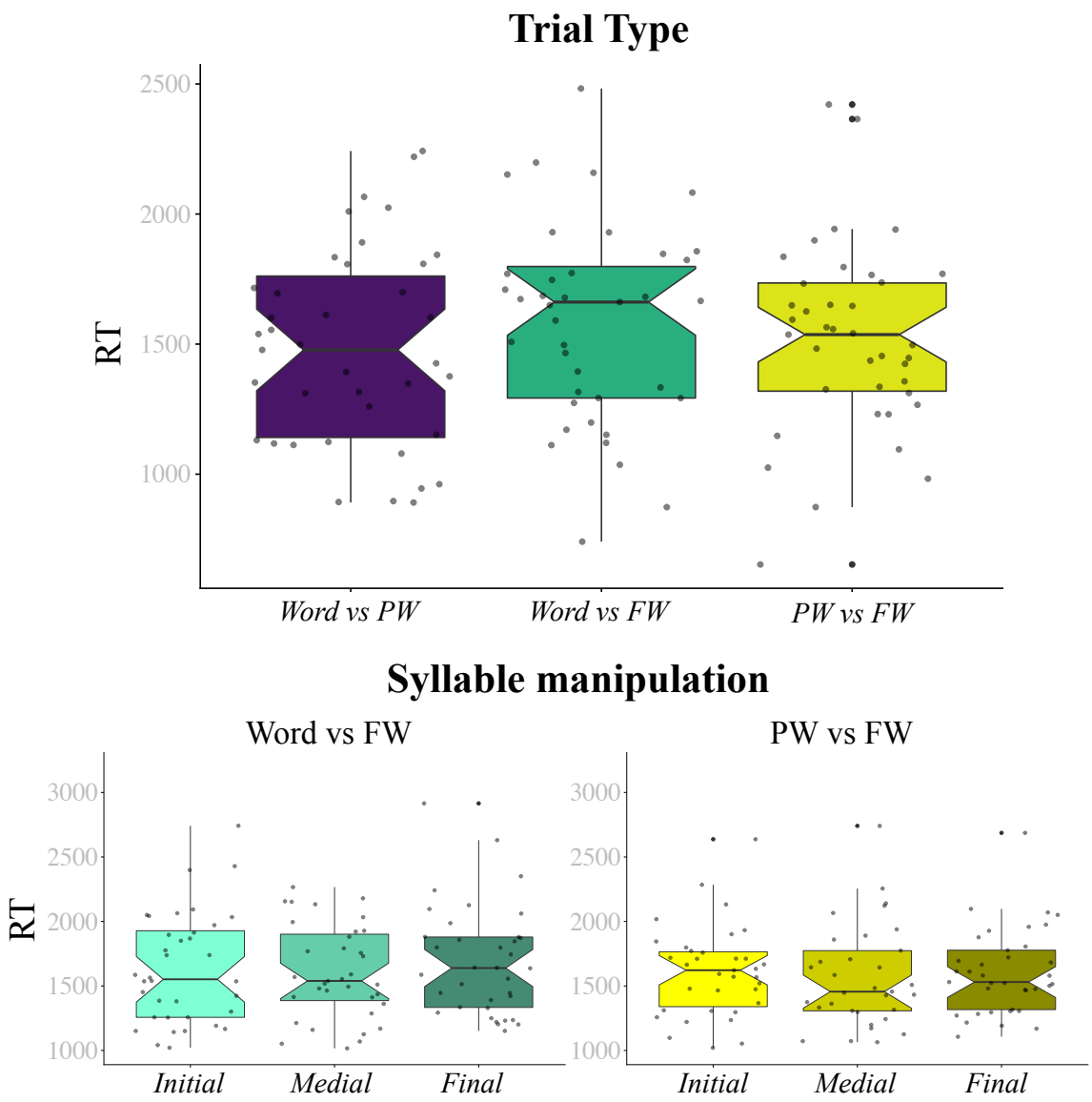

Figure 2.13 Experiment 4 (Video + Native English Language) RT to trial types and syllable manipulations

2.5.2.1 Words versus Part-Words

Participants endorsed words significantly above chance (M = 61.9%, SD = 19.6%, 95% CI = [55%, 68%], $t(38) = 3.77$, $p < .001$, $d = 0.61$).

2.5.2.2 Words versus Fake-Words

I first present the main effects, followed by syllable manipulations.

2.5.2.2.1 Combined

Participants endorsed words significantly above chance (M = 60.7%, SD = 12.0%, 95% CI = [56.8%, 64.6%], $t(38) = 5.54$, $p < .0001$, $d = 0.89$).

2.5.2.2.2 Syllable Manipulations.

Participants successfully chose words over fake-words across syllable positions: Initial (M = 59.0%, SD = 16.5%, 95% CI = [53.6%, 64.3%], $t(38) = 3.40$, $p = .002$, $d = .55$), Medial (M = 66.7%, SD = 15.5%, 95% CI = [61.6%, 71.7%], $t(38) = 6.70$, $p < .0001$, $d = 1.08$), and Final (M = 56.4%, SD = 17.4%, 95% CI = [50.8%, 62.1%], $t(37) = 2.30$, $p = .027$, $d = .37$). Mixed effects models reveal that participants performed better on trials pitting medial-syllable fake-words against words than they did trials involved initial- or final-syllable fake-words (for medial versus initial: $OR = 1.40$, $p = .046$; for medial versus final: $OR = 1.56$, $p = .008$; see Table 2.23). There was no significant difference in mean RTs (Initial: M = 1571 msec, SD = 473; Medial: 1537 msec, SD = 396 msec; Final: 1625 msec, SD = 465 msec; see Table 2.24).

Model structure:
Choice ~ Syllable position * Trial + (Trial | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	1.44	1.12 – 1.85	**.004**	2.02	1.57 – 2.62	**<.001**	1.30	1.01 – 1.66	**.039**
Initial Syll				0.71	.051 – 0.99	**.046**	1.11	0.80 – 1.54	.517
Medial Syll	1.40	1.01 – 1.96	**.046**				1.56	1.12 – 2.18	**.008**
Final Syll	0.90	0.65 – 1.24	.517	0.64	0.46 – 0.89	**.008**			
Trial	1.00	0.98 – 1.01	.933	1.01	0.99 – 1.02	.484	1.00	0.98 – 1.01	.641
Trial : Initial				0.99	0.97 – 1.01	.565	1.00	0.98 – 1.02	.786
Trial : Medial	1.01	0.99 – 1.03	.565				1.01	0.99 – 1.03	.388
Trial : Final	1.00	0.98 – 1.02	.786	0.99	0.97 – 1.01	.484			
Random Effects									
$\tau_{00,\ Subject}$					0.085				
ρ_{01}					-0.821				
$N_{Subject}$					39				
$ICC_{Subject}$					0.028				
Observations					936				
Deviance					1200.178				

Table 2.23 Experiment 4 model of proportion choice words versus fake-words

Model structure:
RT ~ Syllable position * Trial + (Trial | Subject)

	Model 1 Reference level: Initial			**Model 2** Reference level: Medial			**Model 3** Reference level: Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1568	1431 – 1705	**<.001**	1540	1403 – 1677	**<.001**	1625	1488 – 1762	**<.001**
Medial Syll	-28	-139 – 83	.625				-85	-196 – 26	.133
Final Syll	57	-54 – 168	.315	85	-26 – 196	.133			
Trial	-0	-7 – 6	.900	-4	-10 – 2	.239	-5	-11 – 1	.135
Trial : Medial	-3	-10 – 4	.363				1	-6 – 8	.771
Trial : Final	-4	-11 – 3	.231	-1	-8 – 6	.771			
Initial Syll				28	-83 – 139	.625	-57	-168 – 54	.315
Trial : Initial				3	-4 – 10	.363	4	-3 – 11	.231
Random Effects									
σ^2					485729				
$\tau_{00,\ Subject}$					127976				
ρ_{01}					-0.063				
$N_{Subject}$					39				
$ICC_{Subject}$					0.209				
Observations					936				
R^2 / Ω_0^2					.307 / .297				

Table 2.24 Experiment 4 model of reaction time to word versus fake-word trials

2.5.2.3 Word versus Part-words compared to Words versus Fake-words

There was no difference in proportion choice of word versus part-word and word versus fake-word trials ($t(38) = -0.43$, $p = .67$), but participants were slower to respond to word versus fake-word trials as a whole (word versus part-word trials: M = 1483 msec, SD = 378 msec; word versus fake-word trials: M = 1578 msec, SD = 384 msec, $B = 98$ +/- 48 (standard error), $t(86.9) = 2.03$, $p = .045$).[15] None of the syllable manipulations differed by proportion choice from word versus part-word trials (mixed effects model with the interaction of trial and trial type as fixed effects, and trial by subject as random effects; see Table 2.25, panel A). The coefficients of a linear mixed effects model predicting RT by contrast type and trial (Table 2.25, panel B) revealed that participants were slower to respond to trials with final syllable manipulations as compared to word versus part-word trials ($B = 138$ +/- 60.4 (standard error), $t(96.1) = 2.29$, $p = .02$), and that participants became faster over the course of the experiment ($B = -7$ +/-2.5 (standard error), $t(1172.5) = -2.72$, $p = .007$).

[15] RT ~ TrialType * Trial + TrialType * (Trial | Subject)

A. Proportion Choice			
Model structure: Choice ~ Contrast type * Trial + (Trial | Subject)			
	Odds Ratio	*CI*	*p*
Fixed Effects			
(Intercept)	1.67	1.28-2.17	**<.001**
Initial Syll	0.87	0.62-1.21	.401
Medial Syll	1.23	0.87-1.72	.237
Final Syll	0.78	0.56-1.08	.140
Trial	1.01	1.00-1.03	.075
Trial * Initial	0.99	0.97-1.01	.186
Trial * Medial	0.99	0.97-1.01	.402
Trial * Final	0.98	0.96-1.00	.095
Random Effects			
$\tau_{00,\ Subject}$		0.140	
ρ_{01}		-0.804	
$N_{Subject}$		39	
$ICC_{Subject}$		0.041	
Observations		1248	
Deviance		1580	

B. RT		
Model structure: RT ~ Contrast type * Trial + (Contrast type | Subject)		
B	*CI*	*p*
Fixed Effects		
1486	1361 – 1610	**<.001**
85	-34 – 205	.165
53	-62 – 168	.369
138	20 – 257	**.024**
-7	-12 – -2	**.007**
7	-0 – 14	.064
3	-3 – 10	.326
3	-4 – 10	.412
Random Effects		
σ^2	501917	
$\tau_{00,\ Subject}$	94904	
ρ_{01}	0.589	
$N_{Subject}$	39	
$ICC_{Subject}$	0.159	
Observations	1248	
R^2 / Ω_0^2	. 245/.240	

Table 2.25 Experiment 4 models of proportion choice (Panel A) and RT (Panel B) to words versus all non-word types

2.5.2.4 Part-Words versus Fake-Words.

Results are presented first as main effects and then by syllable position type.

2.5.2.4.1 Combined

Participants failed to choose either words or fake-words across all syllable manipulations combined (M = 46.5%, SD = 12.7%, 95% CI = [42.4%, 50.6%], $t(38) = -1.74$, $p = .09$, $d = 0.28$).

2.5.2.4.2 Syllable Manipulations

Performance differed by syllable position: Initial (M = 42.9%, SD = 17.2%, 95% CI = [37.4%, 48.5%], $t(38) = -2.57$, $p = .01$, $d = 0.41$), Medial (M = 54.8%, SD = 18.0%, 95% CI = [49.0%, 60.7%], $t(38) = 1.66$, $p = .10$, $d = 0.27$), and Final (M = 41.7%, SD = 16.6%, 95% CI = [36.3%, 47.0%], $t(38) = -3.14$, $p = .003$, $d = 0.50$). Mixed effects models with the interaction of trial and syllable position as fixed effects and trial by subject as random effects, confirms that initial- and final-syllable trials differ (in the direction of choosing fake-words) from medial-syllable trials (which were in the direction of choosing part-words). These models also reveal that participants increasingly endorsed medial- and initial-syllable fake-words in comparison to final-syllable fake-words over the course of the experiment (initial versus final: $OR = 0.98$, $p = .03$; medial versus final: $OR = 0.97$, $p = .002$). Full results can be found in Table 2.26. Despite numerical differences in mean RT across syllable position (Initial: M = 1534 msec, SD = 401 msec; Medial: M = 1452 msec, SD = 457 msec; Final: M = 1564 msec, SD = 360 msec), they do not significantly differ (all p's > .19; see Table 2.27). Participants became slightly slower to respond to medial syllable manipulations in comparison to initial syllable manipulations over the course of the experiment ($B = 7$, $p = .03$).

Model structure:
Choice ~ Syllable position * Trial + (Trial | Subject)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	0.72	0.56 – 0.93	**.012**	1.30	1.01 – 1.68	**.044**	0.71	0.55 – 0.91	**.007**
Initial Syll				0.56	0.40 – 0.78	**.001**	1.02	0.74 – 1.42	.903
Medial Syll	1.80	1.29 – 2.51	**.001**				1.83	1.32 – 2.55	**<.001**
Final Syll	0.98	0.71 – 1.36	.903	0.55	0.39 – 0.76	**<.001**			
Trial	0.98	0.96 – 0.99	**.005**	0.97	0.95 – 0.98	**<.001**	1.00	0.99 – 1.02	.903
Trial : Initial				1.01	0.99 – 1.03	.366	0.98	0.96 – 1.00	**.029**
Trial : Medial	0.99	0.97 – 1.01	.366				0.97	0.95 – 0.99	**.002**
Trial : Final	1.02	1.00 – 1.05	**.029**	1.03	1.01 – 1.06	**.002**			
Random Effects									
$\tau_{00,\ Subject}$					0.090				
ρ_{01}					0.211				
$N_{Subject}$					39				
$ICC_{Subject}$					0.027				
Observations					936				
Deviance					1202.414				

Table 2.26 Experiment 4 models of proportion choice part-words versus fake-words

Model structure:
RT ~ Syllable position * Trial + (Trial | Subject)

	Model 1 Reference level: Initial			**Model 2** Reference level: Medial			**Model 3** Reference level: Final		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Effects									
(Intercept)	1529	1399 – 1659	<.001	1448	1305 – 1590	**<.001**	1562	1447 – 1681	**<.001**
Initial Syll				81	-31 – 194	.162	-35	-143 – 73	.524
Medial Syll	-81	-194 – 31	.162				-116	-235 – 3	.062
Final Syll	35	-73 – 143	.524	116	-3 – 235	.062			
Trial	-4	-9 – 1	.085	3	-2 – 8	.196	-0	-5 – 4	.865
Trial : Initial				-7	-14 - -1	**.033**	-4	-10 – 3	.276
Trial : Medial	7	1 – 14	**.033**				4	-3 – 10	.305
Trial : Final	4	-3 – 10	.276	-4	-10 – 3	.305			
Random Effects									
σ^2					453159				
$\tau_{00,\ Subject}$					114186				
ρ_{01}					0.247				
$N_{Subject}$					39				
$ICC_{Subject}$					0.201				
Observations					936				
R^2 / Ω_0^2					.246/.240				

Table 2.27 Experiment 4 linear mixed effects regression of reaction time to part-words versus fake-word trials

2.5.2.5 Correlations

Correlations between main trial types is presented first, followed by correlations broken down by syllable position manipulation.

2.5.2.5.1 Main trial types

As in Experiment 1 and 2, participants who were better at choosing words over part-words were also better at choosing words over fake-words ($r(39) = 0.51$, $p < .001$). Participants who were better at choosing words over fake-words were also more likely to choose fake-words over part-words ($r(39) = -0.32$, $p = .04$). This patterned in the same direction for words over part-words, but weakly ($r(39) = -0.16$, $p = .34$). These relationships are plotted in Figure 2.14.

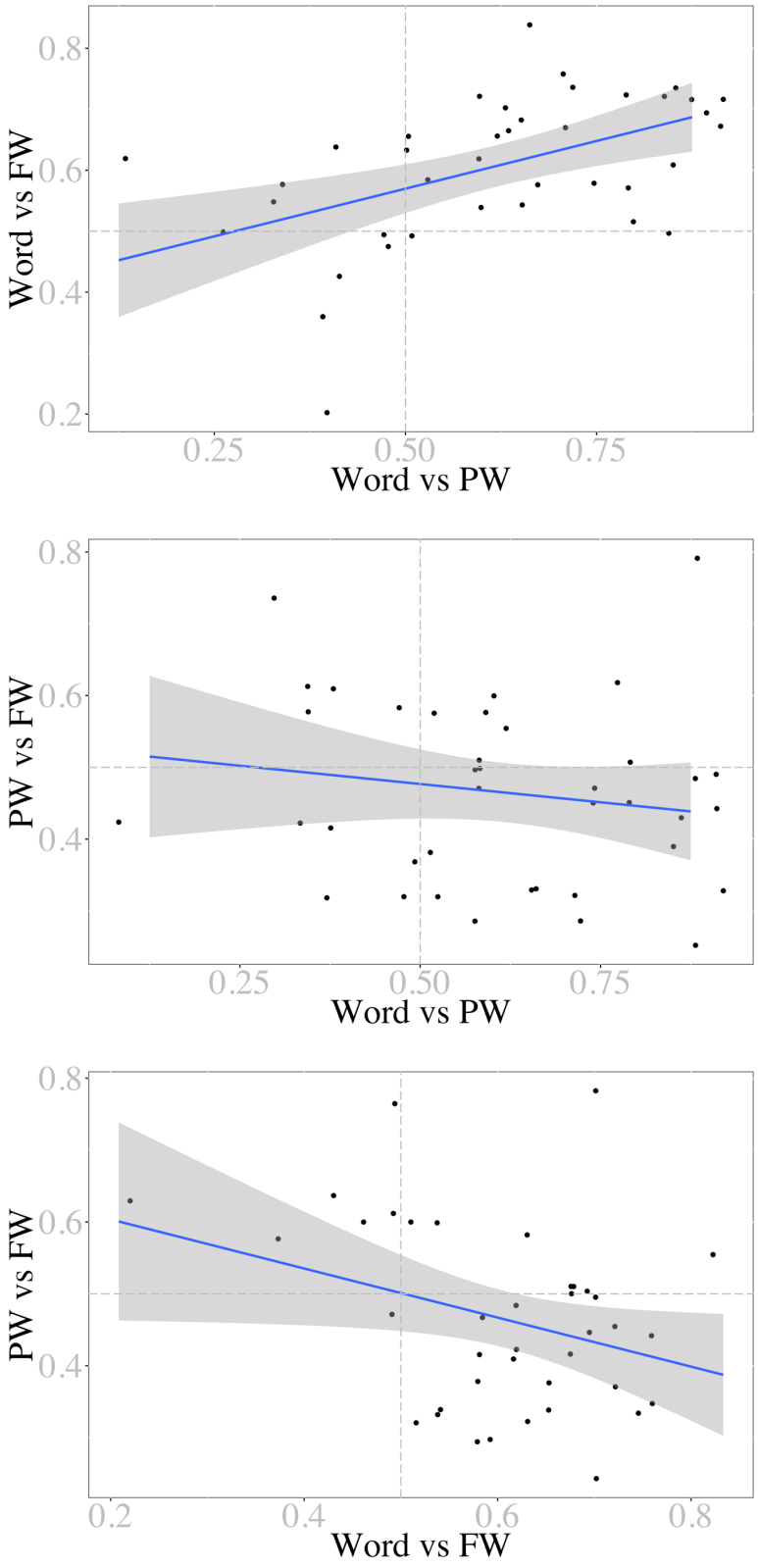

Figure 2.14 Experiment 4 correlations by main trial type

2.5.2.5.2 Syllable positions

Correlations by syllable position manipulation and trial type are presented in Table 2.28. Correlations by syllable positions reveal interesting patterns: the more successful a learner was on the standard segmentation task (word versus part-word), the better they were at rejecting fake-words compared to words – but in particular, those with initial syllable manipulations ($r(39)$ = .51, p = .0008). Unlike in Experiments 1 and 2, the relationships between the different fake-word contrast types (i.e., words versus fake-words and part-words versus fake-words) suggests preferential encoding of initial and final syllable edges. That is, learners who were better at rejecting fake-words with initial and final syllable manipulations in favor of their word counterparts were more likely to choose fake-words with medial syllable manipulations over part-words ($r(39)$ = -0.33, p = .04, and $r(39)$ = -0.40, p = .01, respectively).

Variable		1	2	3	4	5	6
1. Word vs PW							
2. Word vs FW	*Initial*	.51** [.24, .71]					
3.	*Medial*	.33* [.02, .59]	.19 [-.14, .48]				
4.	*Final*	.28 [-.04, .54]	.41** [.11, .64]	.29 [-.02, .56]			
5. PW vs FW	*Initial*	-.14 [-.43, .19]	-.09 [-.39, .23]	-.18 [-.47, .14]	-.18 [-.46, .15]		
6.	*Medial*	-.11 [-.41, .22]	-.33* [-.58, -.01]	-.16 [-.45, .16]	-.40* [-.64, -.10]	.22 [-.10, .50]	
7.	*Final*	-.11 [-.41, .22]	-.07 [-.37, .26]	.14 [-.19, .44]	-.24 [-.51, .08]	.31 [-.01, .57]	.40* [.10, .63]

Table 2.28 Experiment 4 correlations by trial type and syllable position manipulation. *Note:* * indicates p < .05; ** indicates p < .01. Values in square brackets indicate the 95% confidence interval for each correlation.

2.5.3 Discussion

In this experiment, I attempted to elicit greater evidence for position-encoding by taxing learners' attentional resources through a secondary non-auditory task (i.e., watching an unrelated, silent cartoon). As in Experiment 1 (native-English language) and Experiment 2 (semi-English language), participants clearly learned from the language stream, which was evident from their proportion choice of high TP words against part-words or fake-words, and by the fact that performance was correlated across the different trial types (unlike in the non-English language in Experiment 3). Thus, it appears that the simultaneous cartoon did not detract from learning in the same way that the non-English language did. The increase in attentional demands does appear to have shifted the learning curve, however – specifically, there are larger asymmetries of encoding across the different syllable manipulation in comparison to the previous 3 experiments. Moreover, the asymmetrical patterns mirror the position-based effects of all three previous experiments, clarifying the puzzle introduced by the non-English language condition.

When words were pitted against fake-words, participants successfully chose words over fake-words of all three syllable types. However, they were less likely to do so when fake-words had mismatched initial or final syllables. This replicates the results of the non-English language study. As was discussed previously, one interpretation of this result is as support for the TP-encoding account of learning: participants apparently find non-words with two 0.0 TPs easier to reject than non-words with one 1.0 and one 0.0 TP. As in the non-English language experiment, however, the second part of the proposal does not hold true: that is, it is not the case that all items with 0.0 TPs are easier to reject than items with positive TPs (i.e., part-words). I was, therefore, unable to either confirm or reject the TP-encoding hypothesis on the basis of Experiment 3. The remaining results from Experiment 4, however, favor the position-encoding account of learning.

Participants chose fake-words over part-words, specifically when edge syllables were manipulated. Moreover, as the experiment progressed, participants became increasingly likely to choose fake-words over part-words. Finally, participants who chose words over fake-words with initial- and final-syllable manipulations, were also more likely to choose medial-syllable fake-words over part-words – a striking relationship, given that performance at the group level in this condition was in the direction of part-word choice (d = .27). In other words, better segmentation performance was associated with a higher reliance on positional information than on TP-structure.

2.6 General Discussion

In four experiments I examined whether learners encode the positions of syllables that are embedded in trisyllabic words defined solely by transitional probabilities. I hypothesized that if learners are extracting word-like chunks, then trisyllabic sequences that masquerade as words by maintaining the ordinal relationship of syllables, but that create novel syllable transitions, might be more confusable with the statistically defined words. If statistical learning merely involves veridical tracking of TPs, however, I predicted that performance should be consistently dictated by higher TP sequences. The experiments revealed evidence of both mechanisms: participants appear to use TPs in their decision-making processes, but also demonstrate knowledge of positional information from trisyllabic chunks.

The position-encoding hypothesis, in contrast to the TP-encoding hypothesis, predicts that learners will find certain syllable-manipulated fake-words more confusing than others, and more confusing than word versus part-word trials. The TP-encoding hypothesis also predicts an

ordered relationship of performance, but according to TPs. That is, performance under the TP-encoding hypothesis is as follows:

(1) Words vs Medial FW > Words vs Initial/Final FW > Words vs PW

Whereas the position-encoding hypothesis is:

(2) Words vs (some) FW ~ Words vs PW > Words vs (other) FW

Across all 4 experiments, participants performed best (or equivalent to word vs part-word trials) on the word versus medial fake-word trials, which is in line with the TP-encoding hypothesis. While there were no significant differences in proportion choice between word versus part-word and word versus syllable-manipulated fake-word trials in any of these four experiments, there was a consistent ordering relation among these trial types (Words vs Initial FW is excluded, as the pattern of performance was not consistent):

(3) Words vs Medial FW >/= Words vs PW > Words vs Fin FW

This pattern of performance is more consistent with the position-encoding, as opposed to the TP-encoding hypothesis. These relationships can be seen in Figure 2.15 below, where performance on each trial and syllable manipulation type is plotted in order of effect size (Cohen's *d*) by experiment. Contrasts that were significantly different from chance are noted with a *, as are contrasts that were significantly different from one another. Items plotted above the dotted line

reflect greater proportion choice of the higher TP item; items below the dotted line reflect greater proportion choice of the lower TP, but positionally licit item.

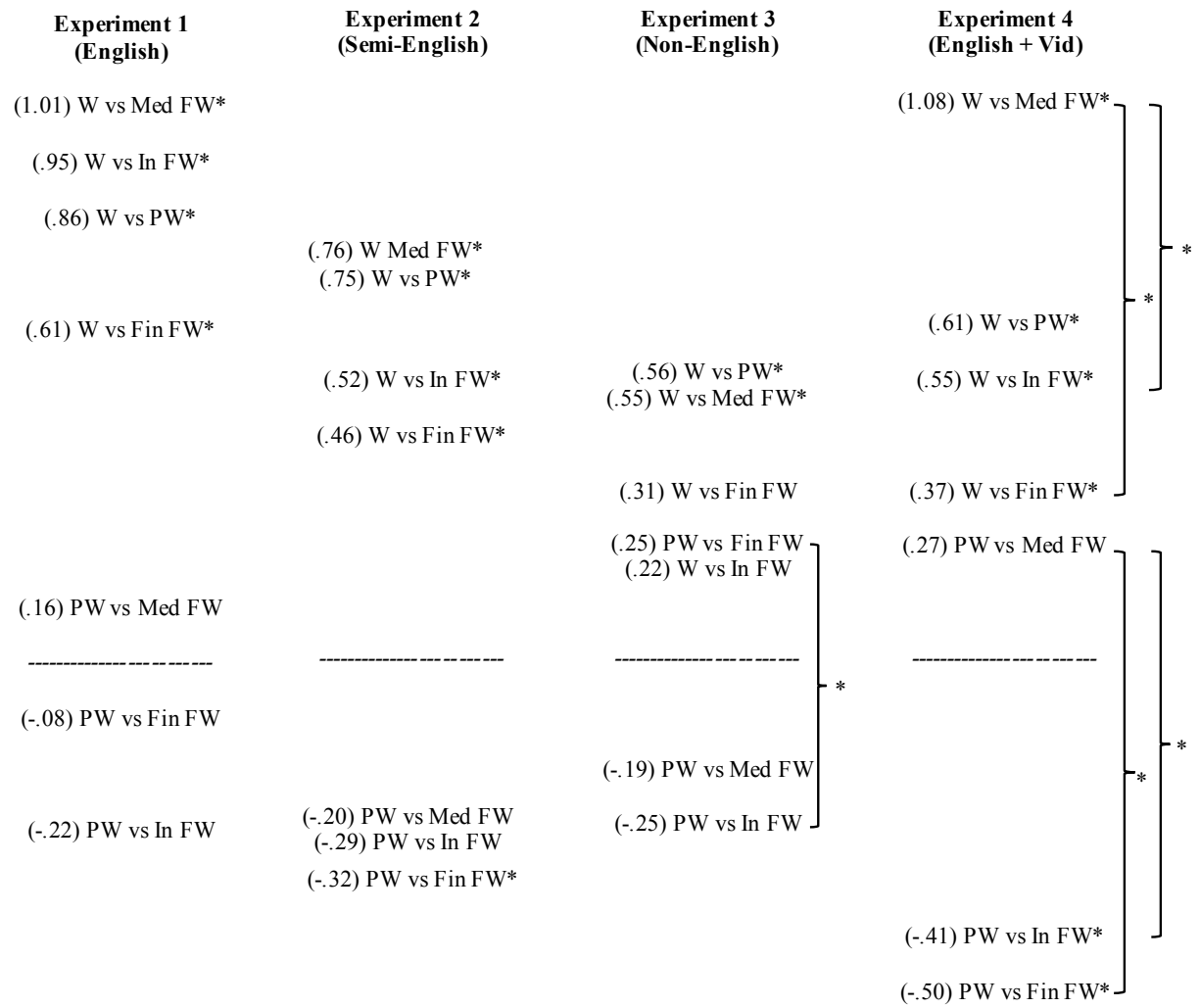

Figure 2.15 Relationship between proportion choices words, part-words, and syllable manipulations by trial type and Experiment. Performance in each experiment is plotted by effect size with respect to proportion choice A over B (where A and B = word, part-word or one of initial-, medial-, or final-syllable fake-words). Cohen's d effect sizes are noted in parentheses to the left of each contrast. Contrasts that were significantly different from chance are noted with a *, as are contrasts that were significantly different from one another. Items plotted above the dotted line reflect greater proportion choice of the higher TP item; items below the dotted line reflect greater proportion choice of the lower TP, positionally licit item.

The second prediction involves how participants treat trials that pit low- versus zero-TP items against each other. The TP- and position-encoding hypothesis made alternate predictions for these items, as follows:

(4) TP-encoding: PW vs Medial FW > PW vs Initial/Final FW > CHANCE

(5) Position-encoding: CHANCE >/= PW vs (some) FW > PW vs (other) FW

Participants did not prefer high TP sequences (part-words) over lower TP sequences with maintained ordinal positions (fake-words) at rates significantly above chance. It is possible, of course, that these items are simply harder to discriminate because the lower TPs are less accessible in memory, in which case the TP-encoding hypothesis would be compatible with at-chance performance across the 3 syllable positions. This was not the case, however: participants under more significant attentional or perceptual demands (i.e., when asked to attend to multiple streams of information at once, or speech sounds that were less familiar to a native English speaker's ear), preferred fake-words that maintained one adjacent 1.0 TP.

The next predictions to consider relate to the relationship between reaction times and each of these contrasts (i.e., words versus part-words, words versus fake-words, and part-words versus fake-words). As RTs reflect ease of decision-making (e.g., Smith, Branscombe, & Bormann, 1988; Tamminen & Gaskell, 2016), I predicted that they might prove a more sensitive measure for detecting position-based versus TP-based decision making processes. By the TP-account, items with larger differences in TP might be easier to distinguish – and therefore RTs should follow the same line of performance as proportion choice (i.e., faster RTs to words versus medial fake-words, getting progressively slower the closer the TP structures become to one another). The position-encoding account, of course, predicts the same underlying process, but that fake-words (at least of certain types) will be harder to distinguish from words, despite their more obvious TP differences. There were few significant differences with respect to RT, but those that exist support the position-encoding hypothesis. In Experiments 3 and 4, participants

were significantly faster to respond to word versus part-word trials as compared to word versus fake-word trials; as can be seen in Table 2.29 below, this pattern was maintained across all four experiments. This suggests that, as a whole, fake-words and words were more difficult to discriminate, as opposed to words and part-words. There were some significant differences reported between syllable position manipulations as well (e.g., participants were slower to respond to initial syllable fake-word trials as compared to medial-syllable trials in Experiment 2), but these did not pattern consistently across the four experiments.

Experiment	W vs PW	W vs FW				PW vs FW			
		total	In	Med	Fin	*total*	In	Med	Fin
Experiment 1 (English)	1641	1713	1722	1690	1728	1668	1655	1667	1687
Experiment 2 (Semi-English)	1661	1704	1778	1657	1676	1625	1617	1609	1650
Experiment 3 (Non-English)	1622	1721	1696	1780	1687	1610	1602	1617	1612
Experiment 4 (English + Video)	1483	1578	1571	1537	1625	1517	1533	1452	1564

Table 2.29 Mean reaction times by trial type and syllable manipulation, by Experiment *Key*: W = Word; PW = Part-word; FW = Fake-word; In = initial; Med = medial; Fin = Final.

Finally, I also predicted that performance on the different trial types would exhibit different correlational patterns according to these two different mechanistic accounts. TP-encoding predicts positive correlations among all trial types, whereas position-encoding predicts a negative relationship between part-word versus fake-word trial types and all word versus non-word trial types. The data evince the latter pattern: in all 4 experiments, performance on part-word versus fake-word trials is negatively correlated with performance on other trial types (Table 2.30).

Experiment	W vs PW W vs FW	W vs PW PW vs FW	W vs FW PW vs FW
1. Experiment 1 (English)	.39*	-.36*	-.12
2. Experiment 2 (Semi-English)	.39*	-.26	-.17
3. Experiment 3 (Non-English)	.10	-.07	-.02
4. Experiment 4 (English + Video)	.51*	-.16	-.32*

Table 2.30 Correlations between main trial types by Experiment *Key*: W = Word; PW = Part-word; FW = Fake-word.

Why might learners automatically encode the positions of syllables during a statistical learning task? The premise of the word-segmentation statistical learning literature has been that learners can use the skill of tracking transitional probabilities to extract coherent chunks from the auditory stream. This process of chunking – if that's what it is – would be useful to language learning. Chunks of linguistic information (such as words), however, bear properties that are not automatically given by pure transitional probability-tracking. Rather, cross-linguistic evidence suggests that positional information – in particular, the edges of linguistic chunks – are particularly salient to memory and processing (e.g., Brown & McNeil, 1966; MacKay, 1970; Marslen-Wilson & Zwitserlood, 1989). For example, languages are much more likely to employ affixes (morphemes that are attached to word bases at the beginning or at the end of the word) than infixes (morphemes that are inserted word internally), though the latter are certainly attested (see Ramscar, 2013, for review). Phonotactic rules (rules that apply to the type or nature of sounds in context in a language) frequently serve to define word-boundaries by limiting the occurrence of certain segments to either word-initial or word-final positions, or the occurrence of certain segment combinations to across word boundaries, or militating against the occurrence of a segment at word edges (Dixon & Aikhenvald, 2002). Stress patterns also highlight the

importance of word edges; according to some accounts, approximately 90% of languages with stress analyzed (ranging from 260 to 306 languages, found in the Hyman, 1977, Gordon, 2002, and Goedemans and van der Hulst, 2011, corpora; as reported by Elordieta, 2014) contained stress patterns defined with respect to the edge of words.

The study described here provides some support for the hypothesis that statistical learning might itself yield these position-based patterns; however, the data also suggest that additional mechanisms are at play. While the SL literature has demonstrated that attention is not necessary for successful SL (Teinonen et al, 2014; Turk-Browne et al., 2009), it does facilitate certain aspects of it. In fact – increased attention has been shown to facilitate adherence to the transitional probability structure of a stream, and so simultaneously impede the acquisition of higher order structure (Finn et al., 2014). These findings accord with the data from the experiments reported on here: only under conditions of increased perceptual load/attentional demand was there clear evidence that learners relied on the positions of syllables in addition to the TP structure. While the interpretation of these results is not entirely straightforward, the pattern argues against one account of SL: namely, that learners track TPs, but can only arrive at positional information with the insertion of additional prosodic cues (Endress & Mehler, 2009b). Rather, the current data is compatible with a mechanism that tracks TPs and position-based information simultaneously, or (possibly) with a mechanism that chunks the input according to non-statistical strategies (e.g., akin to PARSER, Perruchet & Vinter, 1998).

One potential concern with the 2AFC study design is that participants will inevitably learn over the course of the experiment from repeated exposure to the trisyllabic items presented at test. This exposure could entrain position-based encoding, as participants hear word and fake-word test items a combined total of eighty times, whereas they only hear part-words a total of 32

times. I controlled for this possibility by including trial as a factor in all analyses; there was little evidence for change over the course of the experiment except in one case: Experiment 4. In this task, participants simultaneously attended to an engaging cartoon. This finding is interesting, in and of itself: why did these learners' representations undergo a greater shift from exposure to the test items than the shift experienced in other conditions? In other words – if it is simply the case that participants' divided attention leads to impaired learning overall (as highly unfamiliar sounds led to impaired learning), I would expect a similar pattern of results in both Experiments 3 (the non-English language sounds) and 4 (the video condition). Instead, I find that the level of learning in Experiment 4 parallels that of the control case (Experiment 1, native English language sounds), but with an increase in evidence for position-based encoding.

Another factor that may drive performance is the degree of acoustic similarity between fake-word foils and the target word from which the fake-word derived. In other words, a fake word that combines the medial syllable of bidaku and golabu (to yield bilabu) is much more acoustically similar to the target word (i.e., bidaku) than the combination of bidaku and tupiro (which yields bipiku, in comparison to target bidaku). If participants are solely led astray by highly acoustically similar fake-words, we might suspect that participants' choices are based more on processing and memory constraints imposed by the 2AFC task, as opposed to evidence for position-based knowledge. I therefore examined how performance on fake-word foils varied as a function of acoustic distance between the fake-word foil and target word.

Acoustic distance was measured as the difference between normalized mel cepstral coefficients (MFCC; calculated by the Phonological CorpusTools software, Currie Hall, Allen, Fry, Mackie, & McAuliffe, 2015) of target word syllables and fake-word syllable replacements. Lower values reflect more acoustically similar sounds, while higher values reflect more

acoustically distinct sounds. For example, the fake-word bidaBU, which combines bidaKU with golaBU (two acoustically similar syllables), was assigned a score of -1.35, the difference between the normalized MFCC of BU and KU. The fake-word padoRO, which is a combination of the more acoustically distinct final syllables in padoTI and tupiRO, received the inverse score of 1.35. There are two trial types that involve fake-words: words versus fake-words, and part-words versus fake-words. In both cases, I predict that the more acoustically similar a fake-word is to its word counterpart, the more confusable it should be. This would result in a drive to choose fake-words in both conditions, leading to lower performance (or below chance) performance overall. While there were no significant correlations between performance and acoustic distance by trial type across the four experiment sets, they do all pattern in the predicted direction – that is, fake-words that are acoustically closer to the word target are more confusable than fake-words that are more dissimilar (r = [.15 - .45], all p's > .11). The data thus weakly support the hypothesis that acoustic similarity plays a role in participants' decisions (see Figure 2.14). It was not possible, however, to create a trial list that fully balanced the range of acoustic distance given the structure of test items; future research will be needed to determine why and how acoustic similarity influences decisions across different syllable positions.

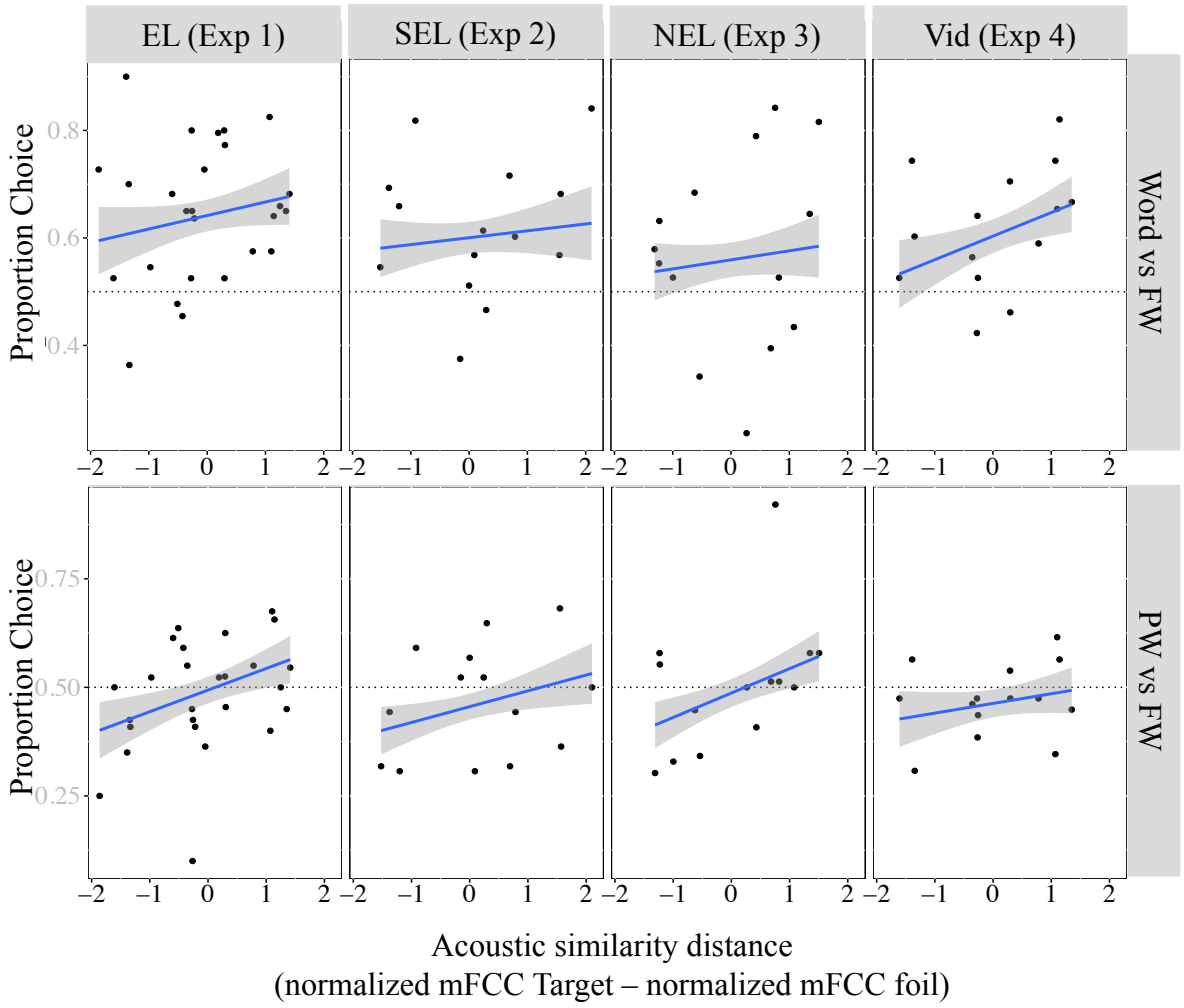

Figure 2.16 The relationship between acoustic similarity of a fake-word to the target word and performance Negative values reflect fake-word syllables that are more acoustically similar to target words (e.g. bidaBU as a fake-word replacement for bidaKU); positive values are more acoustically distance (e.g., bidaTI as a replacement for bidaKU). The top panel shows performance on Word versus Fake-word (FW) trials (performance above 0.50 reflects greater proportion choice words), the bottom panel shows Part-word (PW) versus FW trials (performance above 0.50 reflects greater proportion choice words).

In summary, this chapter explored the representations that emerged from a brief exposure to continuous auditory streams in adult learners. The results indicate that these representations involve more than a simple recording of adjacent TPs – rather, representations are asymmetrically encoded across syllable positions. Moreover, the learning process is altered both by demands on the perceptual and attention systems. In the chapters that follow, I examine (1)

whether SL is impacted by these two factors at the level of individual differences, and (2) how/whether SL is impacted in a population that has relatively lower capacities in both domains (i.e., children).

Chapter 3: Individual Difference Predictors in Statistical Learning

One of the central claims of the statistical learning literature has been that it is a foundational mechanism for (certain aspects of) language acquisition (e.g., Gomez & Gerken, 2000; Pierrehumbert, 2003; Romberg & Saffran, 2010). As such, researchers have sought to tie individual linguistic capacities with statistical learning performance. For example, studies have demonstrated that auditory statistical learning (SL) relates to sentence comprehension in adults (Misyak & Christiansen, 2012), and receptive and expressive vocabulary in children (Evans et al., 2009).[16] Studies demonstrating a connection between SL and linguistic knowledge have not been limited to auditory SL; for example, Arciuli and colleagues have found that visual SL correlates with syntactic knowledge in children (Kidd, 2012; Kidd & Arciuli, 2016), and reading in both adults and children (Arciuli & Simpson, 2012b). Recent work, however, has revealed a decoupling of performance within individuals across differing SL paradigms, as well as varying levels of psychometric validity across different SL tasks (Siegelman & Frost, 2015; Siegelman, Bogaerts, Christiansen, & Frost, 2017). In this chapter, I address these findings by proposing that the outcome of SL crucially relies on the efficiency with which one can encode the stream of sensory stimuli (which I will term the *encoding hypothesis*[17]). Underlying differences in the

---

[16]This latter study also found impaired SL abilities in children with Specific Language Impairment (SLI) (Evans et al., 2009). A decade of work since has largely confirmed the relationship between SLI and SL impairments, but has found little consistent evidence for either heightened or impaired SL abilities in autism or Williams Syndrome (see Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016 for meta-analysis of SL in SLI and autism, and Cashon, Ha, Graf Estes, Saffran, & Mervis, 2016 for work with infants with Williams Syndrome).

[17]Note: Not to be confused with the TP-encoding or Position-encoding hypotheses of chapter 2. "Encoding" is used here to denote the relative efficiency with which one perceives and represents sounds.

learner's knowledge state or experience with the sensory domain will therefore lead to differing capacities for efficient perception and encoding – and hence differing SL outcomes (described in detail below). Should this hypothesis find purchase, it may account for the dissociation in performance on SL tasks across not only different modalities, but different domains within a modality (e.g., lack of correlation between adjacent and non-adjacent SL within a single domain: Siegelman & Frost, 2015).

To test this hypothesis, I examine the relationship between performance on the word segmentation experiments described in Chapter 2 (Studies 1, 2, and 3) and individual-level correlates of auditory skill, which I describe below. The language conditions of Studies 1 (native-(English) language: NL), 2 (semi-English language: SEL), and 3 (non-English language: NEL) were themselves designed as a group-level means for testing the central proposal, namely that one's underlying representations impact the ability to learn from a perceptual stream. I had predicted a linear decline in performance as the familiarity of speech sounds decreased across the three language conditions; however, this was not confirmed by the analysis of the data presented in Chapter 2. Rather, I found that statistical word segmentation of a stream of less familiar, but not entirely unfamiliar, sounds (the semi-English language) was largely indistinguishable from segmentation of a stream composed of native English sounds. Statistical learning from a stream of entirely unfamiliar sounds (non-English language), on the other hand, was – as predicted – negatively impacted: there were reduced rates of learning in the word versus part-word contrast, and no correlation in performance across different contrast types, indicating a lack of internal cohesion to participants' choices. There are a variety of explanations that might account for this non-linear relationship; one possibility that the present analysis serves to exclude/confirm is that the individuals in the SEL were independently higher on potentially relevant auditory skills.

How do we define which 'auditory skills' might be relevant to statistical learning? My hypothesis is that an individual's ability to rapidly encode and store in memory a particular phonetic unit will impact his/her ability to associate (through the computation of TPs or via a process of chunking) that unit with other units. The most direct means of testing this hypothesis would be to test participants on their discrimination of the sound contrasts used in the familiarization stream, and subsequently map continuous measures of performance from the perceptual task to the SL task. This, however, would require exposing participants to the sounds they experience during the SL task either before SL (which might thereby change their capacity to learn), or after SL (which might in turn change their perceptual performance). These concerns are not insurmountable; however, as this analysis was supplementary to the primary research question (i.e., what is the nature of representations formed from SL, explored in Chapter 2), I opted to use a simpler design, and collect relevant self-report data as a proxy.

To assess the effect of general auditory experience and skill I collected information relevant to non-English language experience, musical skill and experience, and age, and examined the relationship of these variables to performance on the SL task. I hypothesized that (1) specific experience with the sounds used in the experimental languages, (2) multilingualism, and (3) advanced musical skill would contribute to an individual's capacity for efficiently encoding speech sounds and therefore enhance SL performance, whereas (4) age would negatively impact that capacity. Results reveal non-linear relationships between the different language conditions and the auditory skill variables. The slightly more difficult or unfamiliar contrasts used in the semi-English language were easier to encode for multilinguals, individuals with advanced musical skill, and older individuals, as compared to monolinguals, people with less musical experience, and younger individuals. These same characteristics, however, had a

negative impact on performance with completely unfamiliar (non-English) sounds. Taken together these results support – with some caveats – the hypothesis that differences in experience with a particular sensory domain result in different statistical learning outcomes.

3.1 Background

Previous work on individual differences in SL has focused primarily on the relationships between SL and linguistic competence, and whether SL is a separable skill from other aspects of cognition, such as executive function (e.g. Misyak & Christiansen, 2012; Miskyak, Christiansen, & Tomblin, 2010; Weiss, Gerfen, & Mitchel, 2010). The intended contribution of the present analysis is to reverse the causal arrow, and look for the influence of specific types of experience on the perception of a continuous stream of sound, and the impact that may have on the outcome of SL itself. This is because I aim to better understand the mechanisms that underlie SL itself – a pursuit that I hope will ultimately guide our understanding of the relationship between SL abilities across different types of SL tasks and other cognitive or linguistic skills. In the following paragraphs, I delineate the individual differences that I predicted would have a direct impact on an individual's ability to efficiently encode (and therefore learn from) continuous auditory streams.

3.1.1 Specific Language Experience

The encoding hypothesis predicts that a participant's previous experience with the sounds encountered in a continuous auditory stream will have demonstrable effects on the learning outcome. Specifically, I predict decreasing performance (i.e., less frequent choice of words over non-words) as the encountered sounds become less familiar. As discussed in Chapters 1 and 2,

there is existing evidence to support this idea. For example, Perruchet and Poulin-Charronat (2012) attribute their failure to replicate a previous SL finding (Endress & Mehler, 2009a) to a familiarity difference between the subjects' experience of the speech stimuli. Endress and Mehler presented Italian learners with a continuous stream of French sounds and found that they (the learners) failed to learn the trisyllabic words; instead, their participants extracted adjacent and non-adjacent bisyllabic combinations. Using the same design, Perruchet et al. found that French students succeeded at extracting the full trisyllabic dependencies – a success that they suggest is at least partially due to the greater parsibility of a French speech stream to French-speaking participants than it was to the Italian speakers.

On the other hand, there is work showing that statistical learning occurs even with novel or unfamiliar stimuli. For example, newborn infants successfully segment sequences of visual shapes, despite their paucity of visual experience (Bulf et al., 2011). Similar findings exist in auditory SL – for example, learners with congenital amusia (a disorder that affects perception of pitch, musical memory, and recognition) are as sensitive to transitional probabilities between tones and syllables as typically developing controls (Omigie & Stewart, 2011). And adults successfully segment a range of unfamiliar sounds: temporally reversed syllables (Vouloumanos et al., 2012), warbles and glides (Hayes & Clark, 1970), and sine-wave tones (Saffran et al., 1999). On the other hand, there is evidence that learning with less/unfamiliar stimuli is more difficult, even when it is possible. For example, adults exposed to non-linguistic, unlabeled noises in a standard segmentation task failed to extract the triadic patterns until familiarized to the stimuli for 100 minutes across three consecutive days – a 5-fold increase over the required exposure for identically constructed tasks with familiar language sounds or tones (Gebhart et al., 2009). And Graf Estes, Gluck and Bastos (2015) found that 14-month old English-speaking

infants only succeeded at segmenting a continuous stream of Mandarin syllables when tested on the embedded trisyllabic sequences against completely novel foils, as opposed to the (potentially) more difficult contrast of trisyllabic sequences encountered in the stream but across word boundaries.

As a whole, therefore, the extant literature supports the idea that lack of stimulus familiarity impedes learning. I thus made two relevant predictions, one at the group-level, and one at the individual differences level. As to the first, I predict that the different language conditions – which were created to be semi-English-like, and entirely non-English-like, would impact our native English learners' capacities. This has already been demonstrated in Chapter 2, which found evidence for reduced learning overall from the non-English language. In that chapter, however, I made no direct comparisons between the different language conditions; in the present analysis, I will be able to compare directly whether performance in the semi-English and/or non-English conditions differ from the English-language sound condition, and whether they differ from each other. In addition to this group-level prediction, however, I also predict that multilingual participants' prior linguistic experiences – if they overlap with the non-English sounds encountered – will facilitate SL performance, in comparison to those who have not had relevant experience.

3.1.2 Multilingualism

I further propose that competency in multiple languages will positively impact an individual's statistical learning capacity above and beyond any specific linguistic experience. There are two possible reasons that multilinguals might have an advantage in SL beyond their specific linguistic experience. First, multilinguals have been characterized as having superior

executive function skills as compared to monolinguals. Executive function has been both directly and indirectly implicated in statistical learning. Weiss, Gerfen, and Mitchel (2010) correlated performance on a segmentation task in which statistical cues and bracketing cues competed for determining the underlying structure. Individuals who scored higher on the Simon task – a non-linguistic cognitive task that taps in to skills such as selective attention and inhibition – were better able to segment the language using either statistical or bracketing cues. There is also less direct evidence for a relationship between executive function and SL. For example, poor sequence learning is correlated with degree of impaired executive function in Parkinson's patients (Price & Shin, 2009). Moreover, children with SLI (also known as Developmental Language Disorder) – a condition that is associated with degraded executive function skill (Wittke, Spaulding & Schechtman, 2013) – are poorer statistical learners (Evans et al., 2009). Thus, the so-called "bilingual advantage" (demonstrated through, e.g., enhanced sensitivity to visual language distinctions, Sebastian-Galles, Albareda-Castellot, Weikum & Werker, 2012, greater inhibitory control, Bialystok & Martin, 2004, Bialystok, Martin, & Viswanathan, 2005, and superior mental shifting skills, Prior & MacWhinney, 2010, cf. Paap & Greenberg, 2013) may therefore further exert itself in the domain of statistical learning.

Second, it may be that bilinguals will enjoy superior SL skills, but not due to a global bilingual advantage. Rather, they may have a specific skill set associated with increased auditory perception skills. That is, early training of the ear to attend to a larger range of sounds than afforded by a single language will result in a general capacity to quickly encode unfamiliar sounds (see Krizman, Skoe, Marian & Kraus, 2014). In the present study we will look for a simple relationship between multilingualism and SL performance; future work would be necessary, however, to discriminate between these two possible sources for such an advantage.

I thus propose that bi/multilingual experience will impact statistical learning capacity, independent of specific language experience. There is, indeed, some existing support for this hypothesis. For instance, Bartolotti, Marian, Schroeder, and Shook (2011) demonstrate that degree of bilingualism *and* inhibitory control contribute to successful SL of unfamiliar acoustic streams (Morse code). Wang and Saffran (2014) found that bilinguals' performance exceeded their monolingual counterparts on a SL task involving novel (to the listeners') tone contrasts – and, in fact, that bilingualism was more predictive of success than previous relevant linguistic experience (also see Potter, Wang, & Saffran, 2017, for similar results with newly trained second-language learners). A similar bilingual advantage for SL has been found in infants as well: 14-month old bilingual infants are able to segment two, statistically distinct streams (Antovich & Graf Estes, 2018), while monolingual infants are not (Antovich & Graf Estes, 2018; Bulgarelli, Benitez, Saffran, Byers-Heinlein, & Weiss, 2017).

3.1.3 Music

As noted above, while it is possible that linguistic experience with multiple sound systems might lead to general auditory expertise, I hypothesize that non-linguistic auditory experience perceiving complex sounds might also translate to an increased ability to perceive and hence encode unfamiliar phonemes. An auditory experience that bears much of the same spectral and temporal complexity that characterizes speech is music. And indeed – musical training has been found to prepare the auditory cortex to more efficiently encode different aspects of complex sound. For example, infants with musical experience show enhanced oscillatory neural entrainment to beat and meter (Cirelli, Spinelli, Nozaradan, & Trainor, 2016), while children with lab-based musical training subsequently show enhanced late event-related

potential signals to musical sounds (Moreno, Lee, Janus, & Bialystok, 2015). Similar effects have been detected in adulthood: musically trained adults have faster and larger magnitude subcortical responses to both music and language (Musacchia, Sams, Skoe, & Kraus, 2007) and more refined audiovisual integration to music and sine-wave speech (Lee & Noppeney, 2014).

A large body of research further supports a direct connection between auditory tuning via musical training and enhanced linguistic perception (e.g., Alexander, Wong & Bradlow, 2005; Wong & Perrachione, 2007; Moreno, Marques, Santos, Santos, Castro, & Besson, 2009; Slater, Skoe, Strait, O'Connell, Thompson, & Kraus, 2015; see Kraus & Chandrasekaran, 2010, for a review). Of particular relevance to the current design, Tierney, Krizman, and Kraus (2015) demonstrated that musical training in adolescence led to enhanced neural responses to sound generally, and, at a behavioural level, improved participants' phonological processing. In the statistical learning literature itself, previous work has found a facilitatory effect of musical expertise on the segmentation of a sung stream of speech (Francois & Schön, 2011), Morse-code sequences (Shook, Marian, Bartolotti, & Schroeder, 2013), and pure tones (Mandikal Vasuki et al., 2017). These findings suggest that musical training can alter the efficiency and accuracy of encoding of language-specific sounds. I hypothesize that this enhanced capacity would positively impact SL.

3.1.4 Age

In the domain of speech perception, aging is commonly associated with high-frequency hearing loss (Agrawal, Platz & Niparko, 2008). This, in turn, can affect adults' ability to encode and keep speech sounds in memory (McCoy, Tun, Cox, Colangelo, Stewart, & Wingfield, 2007). An age-related decline in sensitivity to the acoustic signal, however, appears to extend to cases

even when hearing remains normal; older auditory nerves provide slower and more variable neural encoding of speech sounds (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012). It might, therefore, be expected that auditory SL skill will negatively correlate with age. This hypothesis, however, has failed to find support thus far. In a recent study, Hutson, Palmer and Mattys (2016) found that older and middle-aged adults performed equivalently to younger adults on an auditory SL task, despite decreasing performance on other cognitive tasks (c.f., Penha, 2014). Even more impressively, rate of presentation ("normal" or "slow") was irrelevant to all age groups. Similarly, Neger, Rietveld, and Janse (2014) found that older adults performed equivalently to younger adults on an auditory artificial grammar learning task (i.e., statistical learning, but with non-continuous presentation).

Though the extant research suggests that SL of familiar auditory sounds should remain largely intact across the lifespan, it is less clear what the impact of age on the learning of unfamiliar sounds would be predicted to be. Perceptual adaptation to unfamiliar sounds appears to decrease with age (Neger, Janse, & Rietveld, 2015), which would suggest that older adults will perform less well at segmenting unfamiliar speech streams. I have therefore included this factor in the analysis of SL performance that follows.

To conclude, there are numerous physiological and experience-driven characteristics that can impact an individual's capacity to efficiently encode complex acoustic signals. I propose that efficient encoding of the acoustics of a continuous stream of sounds will impact a learner's ability to extract the statistically-defined chunks embedded in that stream. I therefore examine whether multilingualism, specific linguistic experience, musical skill, and/or age will influence learners' abilities to parse streams of native English, semi-native, or non-native sounds.

3.2 Methods

3.2.1 Participants

The same participants that were reported on in Studies 1, 2, and 3 of Chapter 2 make up the dataset explored here. As this data was reported by study, I repeat the information here for clarity, but collapsing across the entire set. 135 participants were recruited through the University of British Columbia. Participants received $10 or course credit, and gave informed consent prior to the experiment. Their ages ranged from 17 to 50 (mean 23, median 21). Ten participants were excluded for: failure to follow instructions (n = 1), a self-reported hearing or language disorder (n = 2), being a non-native speaker of English (n = 7; all participants listed English as their primary language; however, if a participant did not live in an English-speaking environment by age 3, they were counted as a non-native English speaker). All remaining participants reported no hearing or language disorders. Of these, an additional 18 participants were excluded due to missing questionnaire data (e.g., failure to answer whether they did or did not have music experience), leaving a final sample of 107 participants (81 female). Participants had been randomly assigned to one of 3 language conditions; the final distribution was as follows: English (sounds) language: n = 35; semi-English: n = 37; non-English: n = 35.[18]

3.2.2 Materials

The materials are the same as those used in Experiments 1, 2, and 3 from Chapter 2. The specifications are briefly repeated here for clarity.

---

[18]Participants were assigned to the particular language condition being run in the lab at that time.

3.2.2.1 Stimuli

The native English speech sound inventory was identical to that used in Saffran, Aslin, and Newport (1996; see Table 3.1). The semi-English and non-English language inventories were selected such that they would structurally parallel the syllables of Experiment 1, but would reflect a continuum of sounds that would be more or less familiar to native English speakers. The semi-English language inventory (Table 3.1) included sounds that may occur in allophonic or free variation in English, or have acoustically similar counterparts in English – but that would not be likely given the specific syllabic contexts of the target familiarization language. For example, syllables which in the English-sounds experiment (Experiment 1) contained the bilabial sound /p/ were instead produced with the corresponding ejective consonant (a p produced with a popping sound that is caused by the release of air compressed between the larynx and oral closure; similar to sounds occasionally heard in conversation in contexts of overemphasis, e.g., if emphasizing the final sound of the word pop; see Wells, 1982). The non-English inventory, also listed in Table 3.1, included primarily sounds that would be unlikely to occur in any context in English.

	A: English Language				B: Semi-English Language					C: Non-English Language			
Consonants		*Bilabial*	*Alveolar*	*Velar*		*Bilabial*	*Alveolar*	*Palatal*	*Velar*		*Bilabial*	*Palatal*	*Uvular*
	Aspirated	p^h	t^h	k^h	*Ejective*	p'	t'		k'	*Ejective*	p'	c'	q'
	Unaspirated	b̥	d̥	g̊	*Prevoiced*	b	d		g	*Implosive*	ɓ	ʄ	ɠ
	Approximant		ɹ		*Approximant*			ʎ		*Approximant*		ʎ	
			l		*Trill*		r			*Trill*			ʀ
Vowels		*Front*	*Back*			*Front*	*Back*				*Front*	*Back*	
	High	i	u		*High*	y	ʊ			*High*	ɨ	ɯ	
	Mid		o		*Mid*	œ				*Mid*	œ		
	Low		a		*Low*		ʌ			*Low*		ɒ	

Table 3.1 The consonant and vowel inventories for the English-Language (A), Semi-English Language (B) and Non-English Language (C). (Repeated from Tables 2.1, 2.9, and 2.14)

Input syllables were produced by the author and digitally recorded in a sound-proofed booth. Syllables were matched in duration (220 milliseconds), F0 medians (178 Hz), F0 contours, intensity means (RMS amplitude mean 70 dB) and intensity contours.

Syllables were concatenated into trisyllabic words (Language A: *bidaku, golabu, tupiro, padoti;* Language B (only in the NL): *datubi, gotibu, rokula, pidopa*), and words concatenated into two semi-random lists per language. Each word was repeated 48 times and interlaced in such a way that every word was followed by the three other words equally often, and never by itself. This created syllable-to-syllable TPs across word boundaries of 0.33, whereas TPs between syllables within a word were 1.0. The resulting familiarization strings were 2 minutes 10 seconds in length. The initial and final 5 seconds ramped up and down in amplitude, respectively (accomplished via the Fade function of the Vocal Toolkit plugin), to prevent providing participants with a clear cue to word boundaries.

3.2.2.2 Test items

In the previous chapter, I described participants' proportion choice and reaction times to a variety of different trial types, with the hopes of elucidating the nature of representations extracted from the familiarization materials. In the current analysis, I will examine learning in a more broad-strokes fashion, and collapse the proportion choice measure across all trial types that pit a TP-based word against any other trisyllabic sequence (i.e., word versus part-word and word versus fake-word trials). Given the very small effects observed across the different trial types (see Chapter 2) and the small sample size, I do not expect there to be sufficient power to observe interactions with the different individual difference predictors.

3.2.2.3 Language Background Questionnaire

A questionnaire was designed to determine specific language knowledge, lingualism, and musical skill for each participant. The participant was asked to note what language(s) they and their family members know, each person's proficiency in reading, writing, speaking, and listening in each language, and when the participant began learning that language. It also asked about any musical training (including voice), musical skill level (on a scale of 1 - 4, where 1 = novice, 4 = professional), and the number of years and age span the participant had trained on each instrument. The full questionnaire can be found in Appendix B. Responses to the language and music background questions have been coded for Lingualism, Specific Language Experience (degree of phonetic overlap between languages known by the participant and the test language), and Musical expertise. These were assessed as is described in the Analysis Section below (Section 3.2.4).

3.2.2.4 Exit interview

The exit interview aimed to gauge the participant's affective response to the experience, any strategies they employed, and was a check to ensure that no one in the NL condition was familiar with the original Saffran et al. (1996) language. (This study is taught in several linguistic and psychology courses on campus.) They were asked (1) whether they thought the language was a real language, (2) how confident they felt in their answers, and (3) whether they chose answers more based on what sounded wrong or what sounded right. A full list of questions can be found in Appendix B. This data will not be analyzed in this chapter, however, as these questions are not directly relevant to the encoding hypothesis.

3.2.3 Procedure

The procedure is identical to that described in Chapter 2. As a reminder, the basic procedure was as follows: participants were told they would be first listening to some sounds, and then answering some questions about those sounds. They were seated in a sound-attenuated room in front of a computer screen and button box and told to follow the instructions provided by the computer. They were asked to use their two index fingers to provide answers via the two outermost buttons of a button box. The experiment was administered with E-prime 2.0 (Psychology Software Tools, Pittsburgh, PA). After completing 4 training trials, participants were asked to please listen quietly to a language called Vesutian. They were prompted to press a button to start, after which the screen turned blank and the familiarization stimuli began playing. After familiarization, they were reminded that they would hear two options, and were asked to please choose the option that sounded more like a word from the language they had just listened to. After completion of the experiment, participants were instructed to return to the researcher.

They were first administered the exit interview, and then filled out the language background questionnaire.

3.2.4 Analysis

I analyzed all data through R statistical software (Version 3.3.3), using the packages lme4 and sjPlot. Performance was first evaluated collapsed across all trials that pitted a non-word foil (i.e., part-word or fake-word) against a statistically defined word. Though this obscures any potential patterns in performance across the different non-word manipulations (i.e., part-words versus syllable-manipulated fake-words), these all involve a test of recognition of high TP versus lower TP items, and so can be logically grouped. This then allows us to construct a model that compares the effects of multilingualism, music, and age in the different experimental language conditions (whereas models that contrast the different non-word foils likely have more parameters than the data can support, Matuschek et al., 2017, and are difficult to interpret given three-way interactions with two multi-level categorical variables). The relationship between the covariates and word performance is presented first as correlations, and then analyzed through generalized linear models.

Operationalization of the factors examined is described in the following paragraphs.

3.2.4.1 Lingualism

Of 107 participants, only 10 listed themselves as having some degree of reading, writing, speaking or receptive proficiency in a single language. Forty-five cited knowledge of 2 languages; 52 cited 3 or more (37: 3 languages; 13: 4 languages; 1 each for 5 and 8 languages, see Table 3.2). A simple breakdown based on these numbers of monolingual, bilingual, and

multilingual, however, ignores relevant (and recoverable) information, such as age of acquisition and proficiency across different domains of language use. Therefore, two categories were created: early multilingual experience, and current multilingual proficiency. In both categories, only values associated with speaking and understanding proficiency were considered, as reading and writing are skills that are less directly related to the hypothesized auditory encoding mechanism that is of interest here. For early experience (Early lingual), a composite score was created by (1) taking the *Z*-score (based on the mean and standard deviation) of the summed speaking and understanding scores for a participant's 2nd language, (2) deriving the age of acquisition by taking the *Z*-score of the proportion of time a participant had known their language (i.e., their age of acquisition subtracted from their age, and divided by their age) and (3) summing these two values (see Bartolotti et al., 2011, for a similar composite metric). Current multilingual proficiency (Current lingual) was assessed similarly, but by taking in to account current proficiency across all non-English languages reported. A composite score was therefore created by taking the *Z*-score of the summed speaking and understanding scores for participants' 2nd through 4th languages. In both scales, lower numbers reflect less multilingual experience (with 0 being monolingual), higher numbers reflect higher multilingual ratings. To illustrate – the participant with the highest Current lingual rating had endorsed a 4/4 score on speaking and understanding of 3 languages, and a 3/4 score on speaking and 4/4 score on understanding of a 4th language. To derive her score, I therefore summed these values (31) and transformed that score into a *z*-score (final value = 2.86). This same participant also scored the highest possible rating of Early lingual, as she had been exposed to her second language (Gujarati) from birth, and rated herself as a 4/4 on both speaking and understanding of that language. To derive the Early lingual score, I summed these values (8) and took the proportion of the person's life that

she had spoken the language (27 / 27 = 1), transformed both of these values into z-scores and added the two scores together (final value = 2.09). Individuals' multilingualism scores are plotted across the familiarization language conditions in Figure 3.1 (Panel A: early multilingual experience; Panel B: current multilingual experience). The dotted line reflects the group-wide average.

# Languages Reported	Experimental Condition		
	EL	**SEL**	**NEL**
1	3	3	4
2	15	14	16
3	12	16	9
4	4	4	5
5	1		
8			1

Table 3.2 The number of participants who listed proficiency with between 1 – 8 different languages by experimental language condition. EL stands for English language; SEL is for Semi-English Language; NEL is for non-English Language.

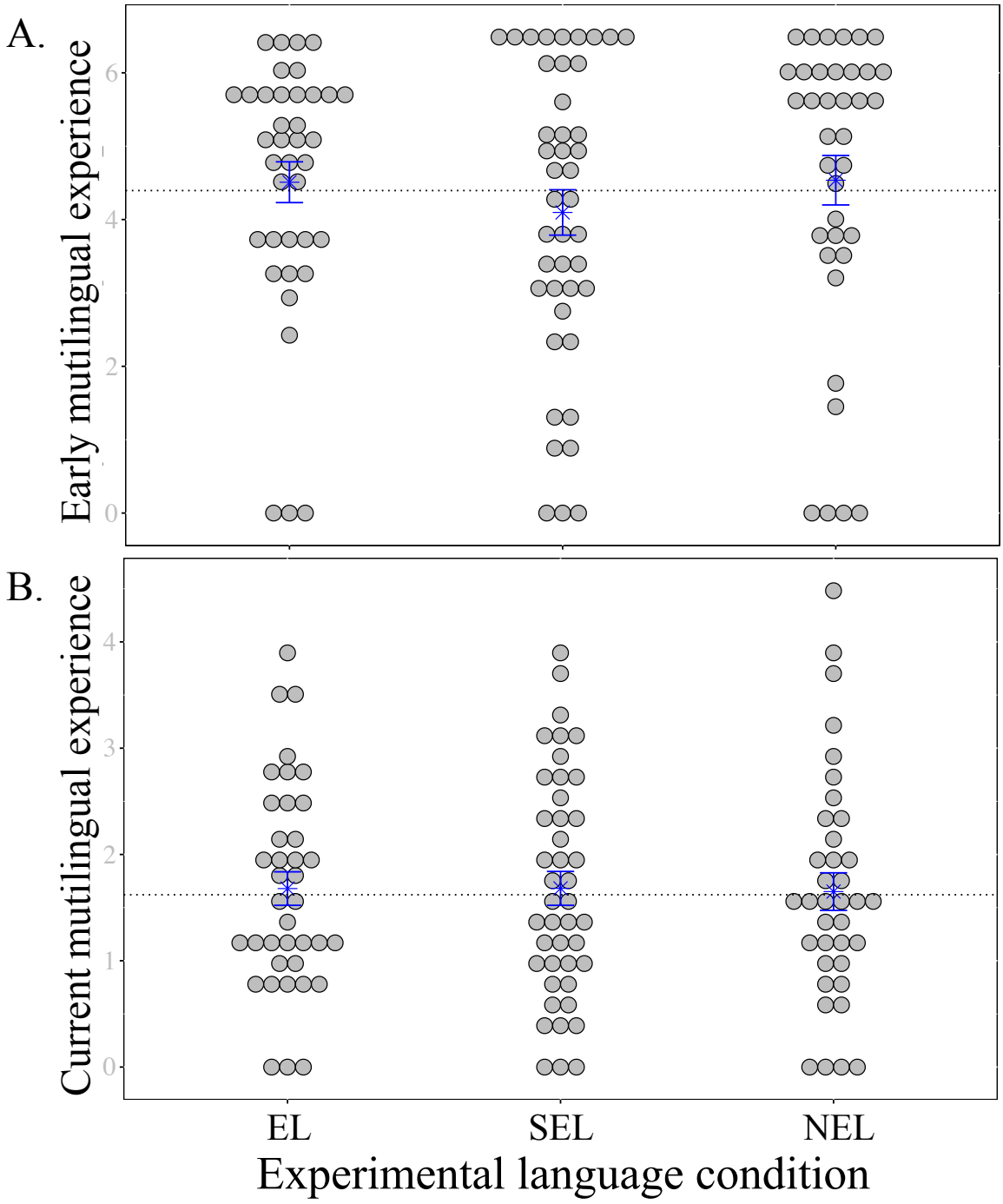

Figure 3.1 Individual participants' ratings for early multilingual experience (Panel A) and current multilingual experience (Panel B). Higher numbers reflect more multilingual experience; lower numbers reflect less multilingual experience (with 0 = monolingual). The dashed line represents the group-wide mean. Blue stars and brackets are the mean +/- one standard error by experimental language condition.

3.2.4.2 Specific language experience

Specific language experience was operationalized in two ways. First, the language conditions (English, semi-English, and non-English) are the primary manipulation of specific language experience (i.e., I predict differences between these three conditions based on their decreasing degree of familiarity to native speakers of English).

Second, it was hypothesized that bi- and multilingual individuals would have experience with non-English sounds that might facilitate their parsing of the non-English languages. Specific language experience was therefore calculated as the degree of phonemic inventory overlap between a participant's second language and the language condition he/she was exposed to. [19] A second language was defined as the language (other than English) that a participant ranked themselves as having the highest proficiency with (i.e., the average score across Speaking and Listening). This resulted in a total of 19 languages (listed in Appendix B.3). Phonemic inventory lists were found for each language (sources listed in Appendix B.3), with three exceptions. These were as follows: one participant listed the language name as Dene Tza, which might refer to one of several Dene languages; one participant recorded their relative proficiency, but did not note the name of any language; two participants listed proficiency in reading and writing Latin, but as these ratings did not include speaking and listening, no inventory overlap was calculated. It should be noted that the sources consulted for deriving phonemic inventories included a wide range of phonetic specificity, making an accurate assessment of the presence/absence of a particular segment impossible; moreover, these inventories cannot speak

---

[19]Only the language with highest proficiency was selected for the sake of simplicity. Participants' proficiencies in additional listed languages were, by definition, lower; to include these values might therefore have required a way to account for the asymmetry in proficiency.

to the particular dialect spoken by the participant, nor whether the productions in the familiarization languages were produced with the characteristics typically found in that language/dialect reported in the sources. Nevertheless, a phonemic overlap score was computed by dividing the number of segments uniquely shared between the 2nd language and experimental language by the total number of segments of the experimental language. This was done separately for vowels and consonants, as the relative difficulty of encoding non-native phonemes has been shown to differ for the two types of sounds (Cutler, Weber, Smits, & Cooper, 2004), and SL has been shown to rely on consonants more than vowels (Bonatti, Peña, Nespor, & Mehler, 2005). Voiced plosives and the two back vowels of the SEL overlap with English phonemes; they were thus removed from the calculation of the language overlap scores, since all speakers are known to share this overlap in phoneme space.

To demonstrate how this calculation was done, I will walk through an example language. If a learner in the SEL condition spoke French as a second language, their French inventory was determined to not overlap in consonant space, and to entirely overlap with the vowel space that remains after the English-like sounds are accounted for. In other words, French does not have an apical rhotic trill, palatal lateral, or ejective obstruents (the remaining consonant contrasts, once the voiced plosives are removed due to overlap with English), therefore the learner received a 0 for consonant specific language experience.[20] French does share the rounded high front and mid front vowels with the semi-English language; therefore, the learner would receive a 1.0 for vowel

[20]Certain Canadian dialects of French do have an apical rhotic trill (Pulleyblank, personal correspondence). If a participant noted a specific dialect, I sought an inventory for that dialect; however, if no dialect was noted, inventory lists were compiled with respect to the most common variant/accessible inventory list.

specific language experience. If the same learner were in the non-English language condition, on the other hand, he/she would receive a score of .125 for consonant specific language experience (i.e., one out of 8 possible segments exists in both languages: the uvular trill), and a .25 for vowel specific language experience (i.e., one of 4 possible segments is shared: the mid-front rounded vowel). It should be noted that because the inventories of the artificial languages are small, the possible unique values for consonant or vowel specific language experience are very narrow (6-8 values for consonants; 3-5 for vowels).

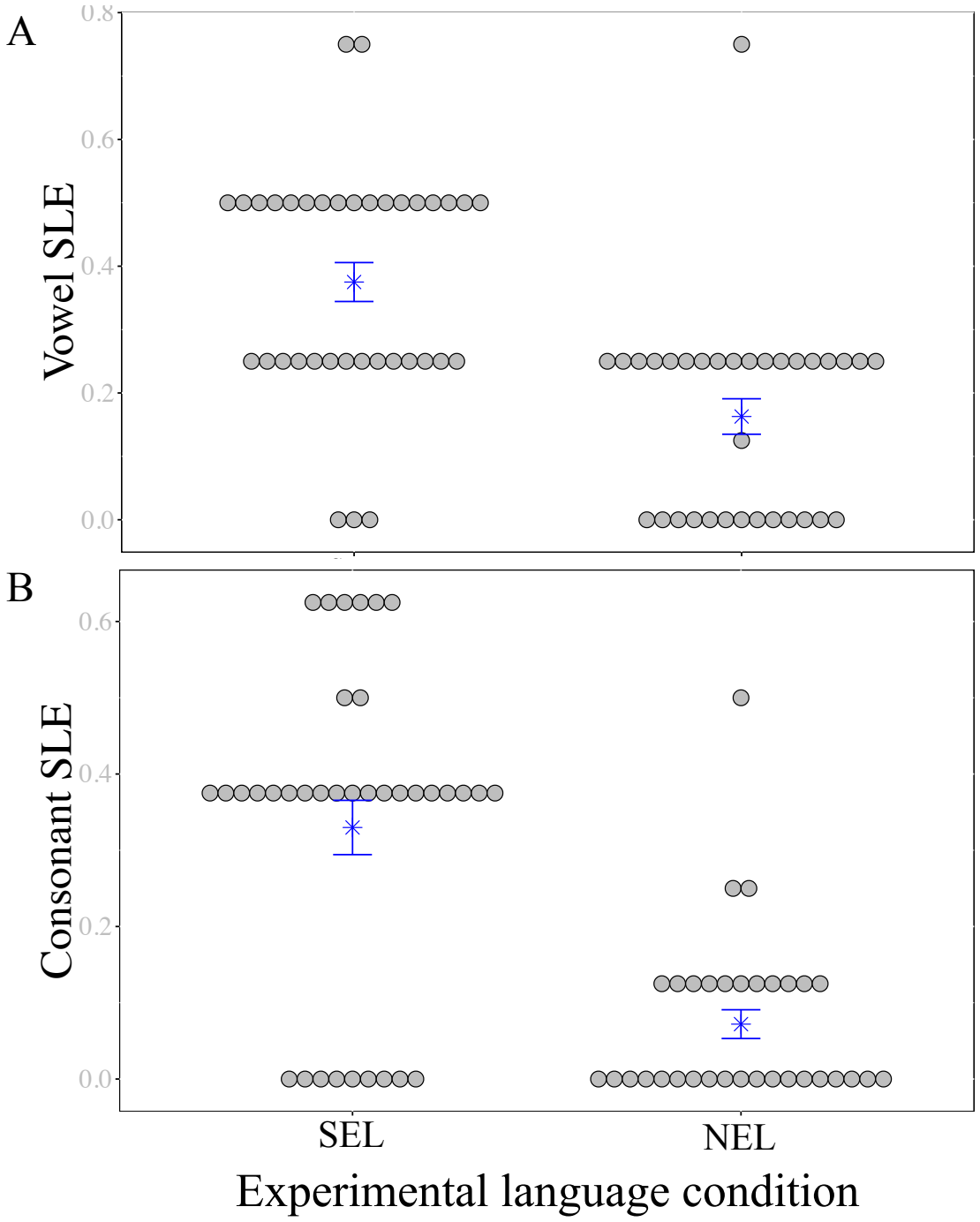

Figure 3.2 Individual participants' Specific Language Experience (SLE) scores for vowels (Panel A) and consonants (Panel B). SEL stands for semi-English language, NEL stands for non-English language. Higher numbers reflect more overlap between the speaker's 2nd language and the experimental language condition phoneme inventories; lower numbers reflect less overlap (with 0 = no overlap). Blue stars and brackets are the mean plus/minus one standard error by experimental language condition.

3.2.4.3 Music

Participants rated their music proficiency on a scale of 0 – 4. As I predicted that proficiency (as opposed to number of instruments) would correlate with auditory tuning (e.g., Tierney, Krizman, & Kraus, 2015), I selected the highest self-rating, and entered these scores as a continuous variable.

3.2.4.4 Age

Age is coded as a continuous variable. Due to the non-normal distribution in the sample, it has been scaled and transformed by the natural logarithm.

3.3 Results

Mean performance by experimental language condition is plotted in Figure 3.3. Performance in all three conditions was significantly above chance (note: these means collapse performance across all word versus any non-word foil trial types, which was not reported on in Chapter 2 as the questions of interest there concerned differences between trial types) (English-language: M = 64.0%, SD = 10.8, 95% CI = [61.0, 67.8], $t(34) = 7.46$, $p < .0001$, $d = 1.30$; Semi-English: M = 61.8%, SD = 11.5, 95% CI = [58.1, 65.6], $t(36) = 6.37$, $p < .0001$, $d = 1.03$; Non-English: M = 56.9%, SD = 9.6, 95% CI = [53.7, 60.1], $t(34) = 4.36$, $p = .0001$, $d = 0.72$), and significantly differs by conditions ($F(2, 104) = 4.15$, $p = .02$). Post-hoc Tukey tests reveal that the non-English performance was significantly worse in comparison to the English-language (mean difference = -7.14, adjusted $p = .02$), but that the semi-English does not differ from either the English (mean difference = -2.19, adjusted $p = .66$) or non-English (mean difference = 4.94, adjusted $p = .12$).

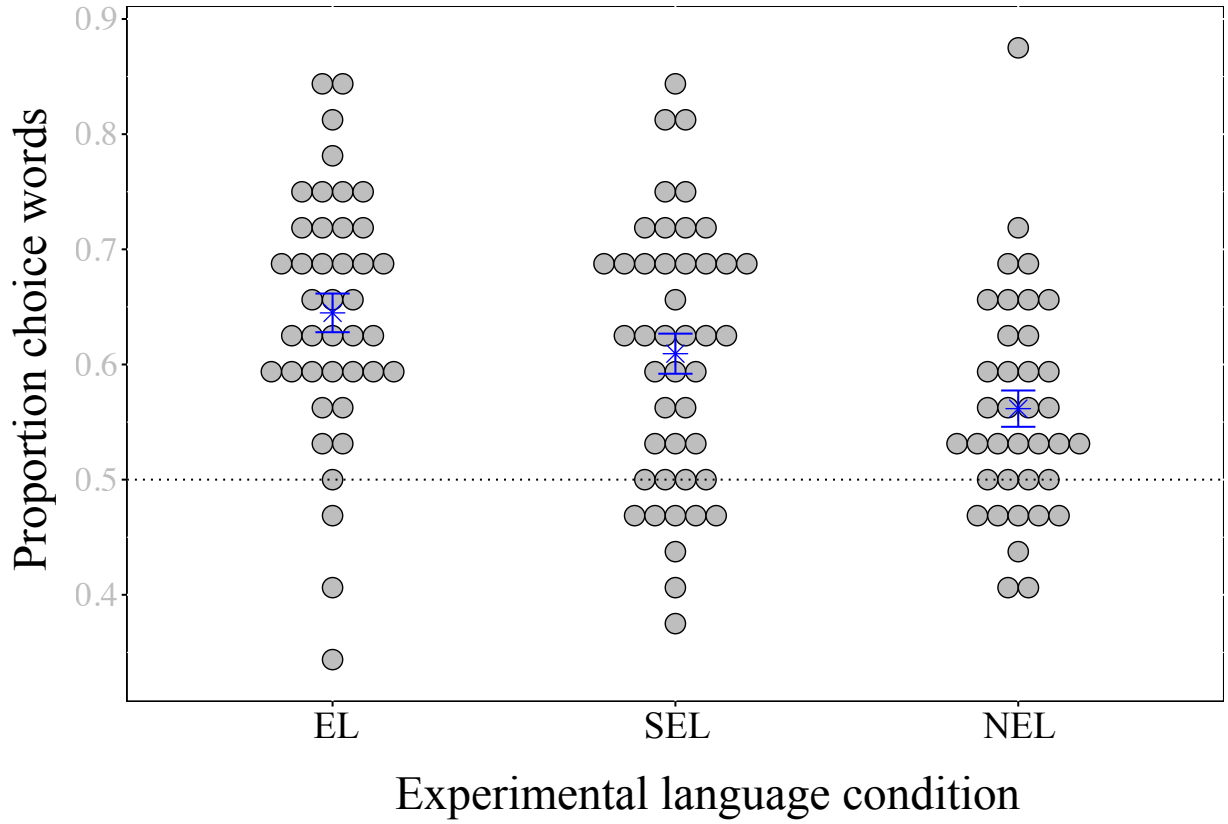

Figure 3.3 Mean performance by language condition. EL stands for native English language; SEL is for semi-English language; NEL is for non-English language. Each dot represents an individual participant's mean performance on all trials pitting a word against a non-word foil. The dotted line represents chance performance. Stars reflect group means plus/minus one standard error.

3.3.1 Correlations

The correlations between each predictor (specific language experience, multilingualism, music, and age) and SL performance can be found in Table 3.3. None of the predictors significantly correlated with performance across the entire sample, though both early and current multilingual experience reflected a very small positive association (Early lingual: $r(105) = 0.18$, $p = .06$; Current lingual: $r(105) = 0.12$, $p = .22$). Specific language experience can only be examined in the semi-English and non-English conditions. Spearman correlations (corrected for

ties) reveal small, positive associations between overlapping consonant and vowel inventories and performance; however, none of these patterns reach significance.

I also examined the correlations between each predictor and performance within each language condition. Neither music (Spearman correlations, corrected for ties) nor current multilingualism bore any relationship to performance in any language condition. Early multilingual experience, however, facilitated performance in the semi-English (SEL) condition ($r(35) = .34$, $p = .04$). Correlations between the log-transformed normalized scores for age and performance revealed a non-significant, negative relationship in the English language (EL) condition ($r(33) = -.29$, $p = .09$), but positive association in the SEL ($r(35) = .31$, $p = .06$).

	EL	SEL	NEL	Combined
Early Lingual Experience	.16	**.34***	.08	.18
	[-.18, .47]	[.01, .59]	[-.26, .40]	[-.01, .36]
Current Lingual Proficiency	.23	.14	.04	.12
	[-.12, .52]	[-.19, .45]	[-.30, .36]	[-.07, .30]
Music	-.02	.00	-.09	.02
	[-.35, .31]	[-.32, .32]	[-.41, .25]	[-.17, .21]
Age	-.29	.31	-.04	-.02
	[-.57, .05]	[-.01, .58]	[-.37, .29]	[-.20, .17]
Specific language *Consonant*		.12	.12	.24
		[-.21, .43]	[-.22, .44]	[.05, .41]
Specific language *Vowel*		.15	.29	.27
		[-.18, .45]	[-.05, .57]	[.08, .44]

Table 3.3 Correlations between the predictors and proportion choice words over non-words by experimental language condition and combined. EL stands for English language (Exp. 1); SEL is Semi-English language (Exp. 2); NEL is non-English language (Experiment 3). Values in square brackets indicate the 95% confidence interval for each correlation; * indicates $p < .05$. Spearman's rho statistics are provided for music, specific language experience Consonant and Vowel predictors. 95% CI are estimated by the formula tanh(arctanh(r +/- 1.96/sqrt(n – 3))).

3.3.2 Mixed Effects modeling

I explored the interaction between familiarization condition (EL, SEL, and NEL) and current multilingual proficiency (Current Lingualism), early multilingual experience (Early

Lingualism), musical skill, and age on Word-choice proportion. Current Lingualism and Early Lingualism were found to be highly correlated ($r(106) = 0.702$, $p < .0001$); these factors were therefore run in separate models, and models were compared by examining the Akaike information criterion (AIC) values.[21] As in previous analyses, models are run in sets so as to alternate the reference level of any non-binary categorical variable (in this case, Language condition). Mixed effects logistic regression model sets were specified for all 2-way interactions between subject-specific factors and language conditions and with random intercepts for subjects. The model structure is defined in standard R notation below.

Choice ~ Language condition * Lingualism (Current or Early) +
Language condition * Age +
Language condition * Music +
(1 | Subject)

The set of models that include the factor Early Lingual performed slightly better than those with Current Lingual (a difference of 2.4 points in AIC value); the pattern of results, however, is nearly identical across the two sets.[22] The results of the model set with the factor Early Lingual (3 models to alternate each language condition as reference) are shown in Table 3.4. These models reveal that early multilingual experience facilitated performance in the semi-English language (OR = 1.10, $p = .008$); the direction of the estimated effect is the same, though

[21]Note: these models cannot be compared using Likelihood ratio test statistics, as the number of parameters is identical across models. I therefore compare the numerical values and pattern of results.

[22]The one exception is that current lingualism has a significant, facilitative effect on the Native language condition (OR = 1.20, CI = [1.01 – 1.43], $p = .04$); the estimate for this parameter is in the same direction, but is of lower magnitude and not significantly different from chance for early lingualism in the parallel model (OR = 1.07, $p = .12$; see Table 3.4).

lower in magnitude, in the English and non-English languages, and does not reach significance. Age, on the other hand, facilitated performance in the SEL (OR = 2.82, p = .012), but was associated with poorer performance in the English language (OR = 0.47, p = 0.03) and non-English language (OR = 0.87, p = .7). Finally, musical skill had no relationship to performance in any condition. Mean performance differed between the native- and non-native experimental language conditions (OR = 1.40, p = .041), but there was no evidence in the model for a significant difference between the semi-native and either the native- or non-native language conditions. Mean performance by condition is plotted in Figure 3.3, followed by plots of the relationship between segmentation performance and early multilingual experience (Figure 3.4), current multilingual proficiency (Figure 3.5), music (Figure 3.6) and age (Figure 3.7).

	EL = reference			SEL = reference			NEL = reference		
	Odds Ratio	*CI*	*p*	*Odds Ratio*	*CI*	*p*	*Odds Ratio*	*CI*	*p*
Fixed Effects									
(Intercept)	1.89	1.48 – 2.42	**<.001**	1.78	1.38 – 2.30	**<.001**	1.36	1.10 – 1.67	**.004**
EL				1.06	0.75 – 1.51	.738	1.40	1.01 – 1.92	**.041**
SEL	0.94	0.66 – 1.34	.738				1.31	0.95 – 1.82	.101
NEL	0.72	0.52 – 0.99	**.041**	0.76	0.55 – 1.05	.101			
Early Lingual	1.07	0.98 – 1.17	.120	1.10	1.03 – 1.18	**.008**	1.02	0.95 – 1.09	.623
Music	0.96	0.87 – 1.07	.489	1.02	0.91 – 1.14	.746	0.98	0.88 – 1.09	.696
Age	0.47	0.24 – 0.93	**.030**	2.82	1.25 – 6.34	**.012**	0.87	0.38 – 1.99	.743
EL:Early Lingual				0.97	0.87 – 1.09	.638	1.05	0.94 – 1.18	.349
SEL:Early Lingual	1.03	0.92 – 1.15	.638				1.08	0.98 – 1.19	.111
NEL:Early Lingual	0.95	0.85 – 1.06	.349	0.92	0.84 – 1.02	.111			
EL:Music				0.95	0.81 – 1.10	.478	0.98	0.85 – 1.14	.827
SEL:Music	1.06	0.91 – 1.23	.478				1.04	0.89 – 1.21	.614
NEL:Music	1.02	0.88 – 1.18	.827	0.96	0.83 – 1.12	.614			
EL:Age				0.17	0.06 – 0.48	**.001**	0.55	0.19 – 1.59	.266
SEL:Age	5.94	2.07 – 17.05	**.001**				3.24	1.02 – 10.31	**.047**
NEL:Age	1.83	0.63 – 5.34	.266	0.31	0.10 – 0.98	**.047**			
Random Effects									
$\tau_{00,\ Subject}$					0.041				
$\tau_{00,\ Item}$					0.054				
$N_{Subject}$					107				
$ICC_{Subject}$					0.011				
Observations					3424				
Deviance					4491.852				

Table 3.4 Generalized linear model results predicting proportion choice words over non-words by early lingual experience, music, and age by language conditions

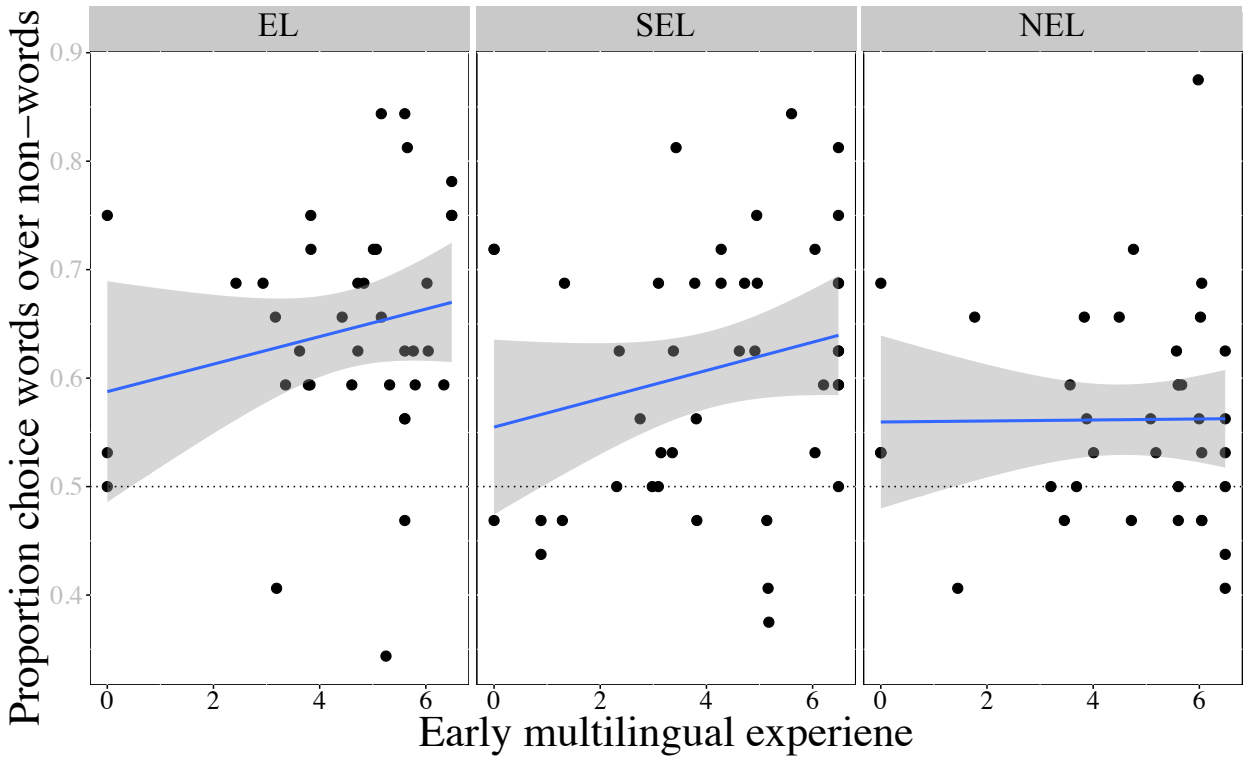

Figure 3.4 The effect of early multilingual experience on SL across experimental language conditions. EL stands for native English language; SEL is for semi-English language; NEL is for non-English language. Each dot represents an individual participant's mean performance on all word versus non-word trials. The dotted line reflects chance performance. The blue line and surrounding grey bars reflect the line of best fit and 95% confidence interval.

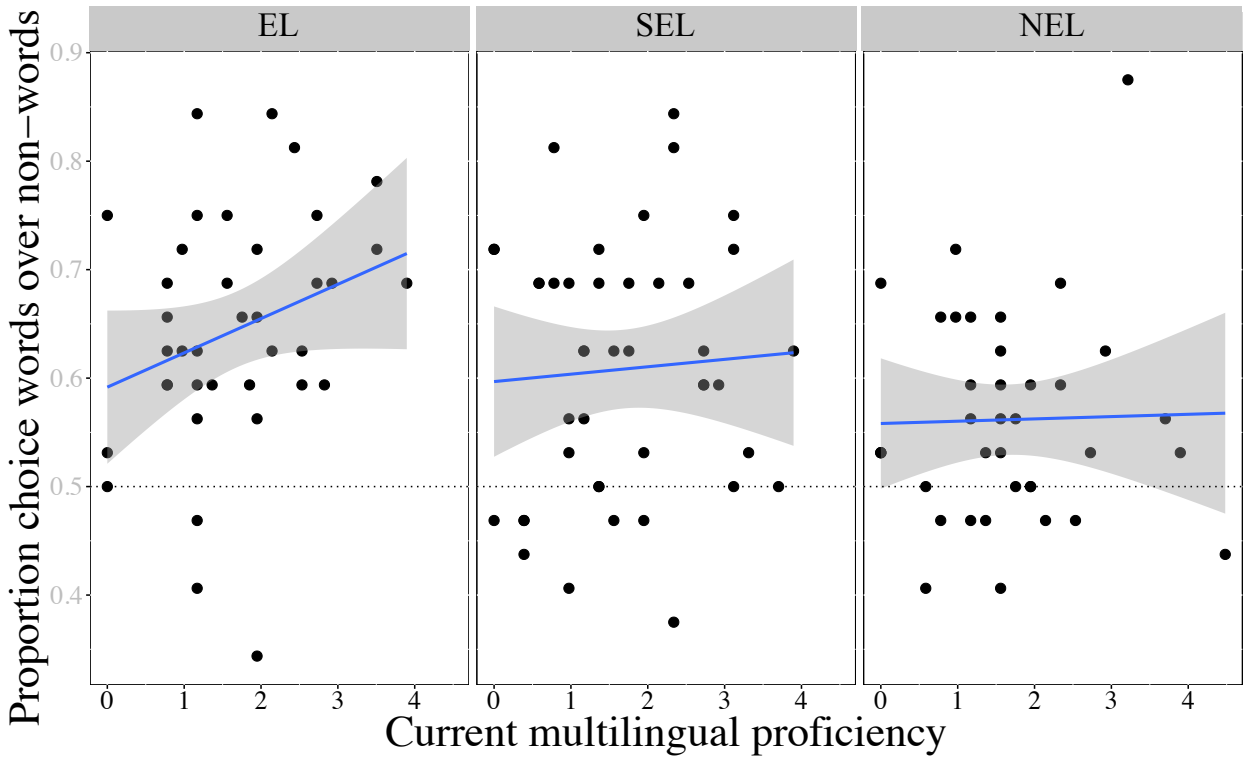

Figure 3.5 The effect of current multilingual proficiency on SL across experimental language conditions. EL stands for native English language; SEL is for semi-English language; NEL is for non-English language. Each dot represents an individual participant's mean performance on all word versus non-word trials. The dotted line reflects chance performance. The blue line and surrounding grey bars reflect the line of best fit and 95% confidence interval.

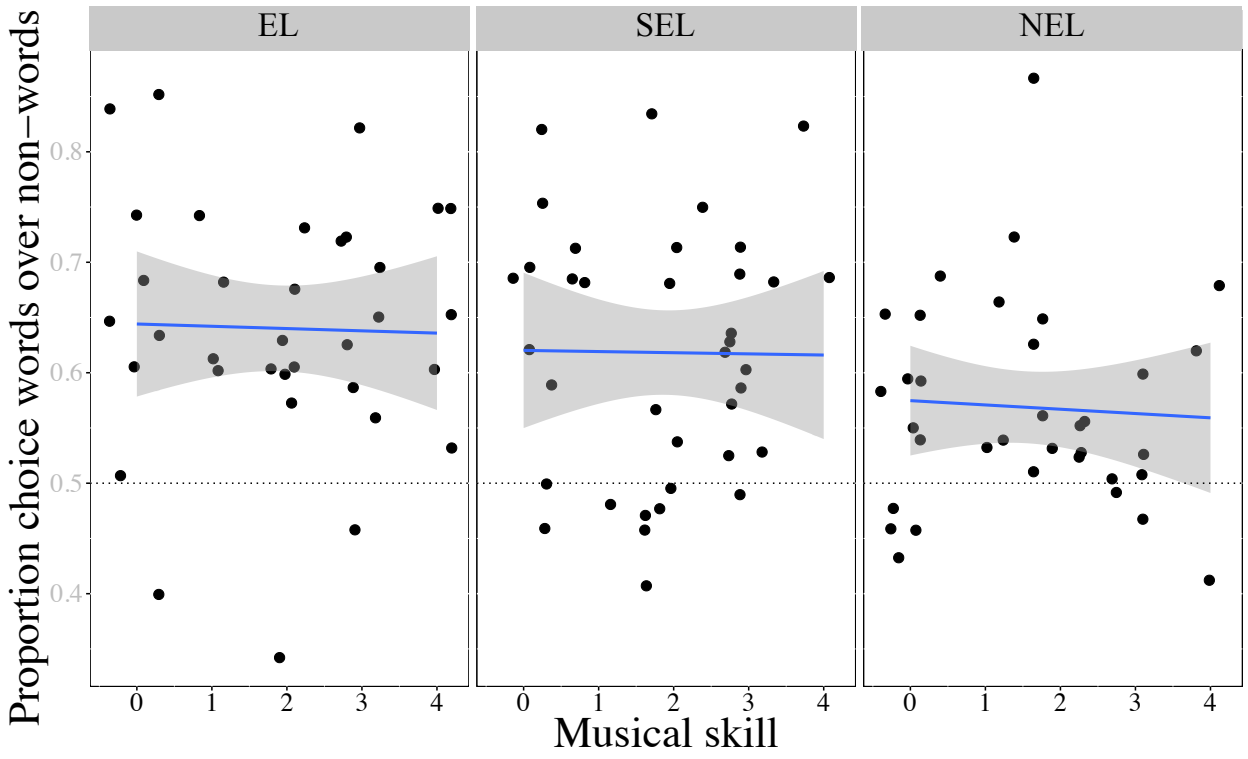

Figure 3.6 The effect of musical skill on SL across experimental language conditions. EL stands for native English language; SEL is for semi-English language; NEL is for non-English language. Each dot represents an individual participant's mean performance on all word versus non-word trials. The dotted line reflects chance performance. The blue line and surrounding grey bars reflect the line of best fit and 95% confidence interval.

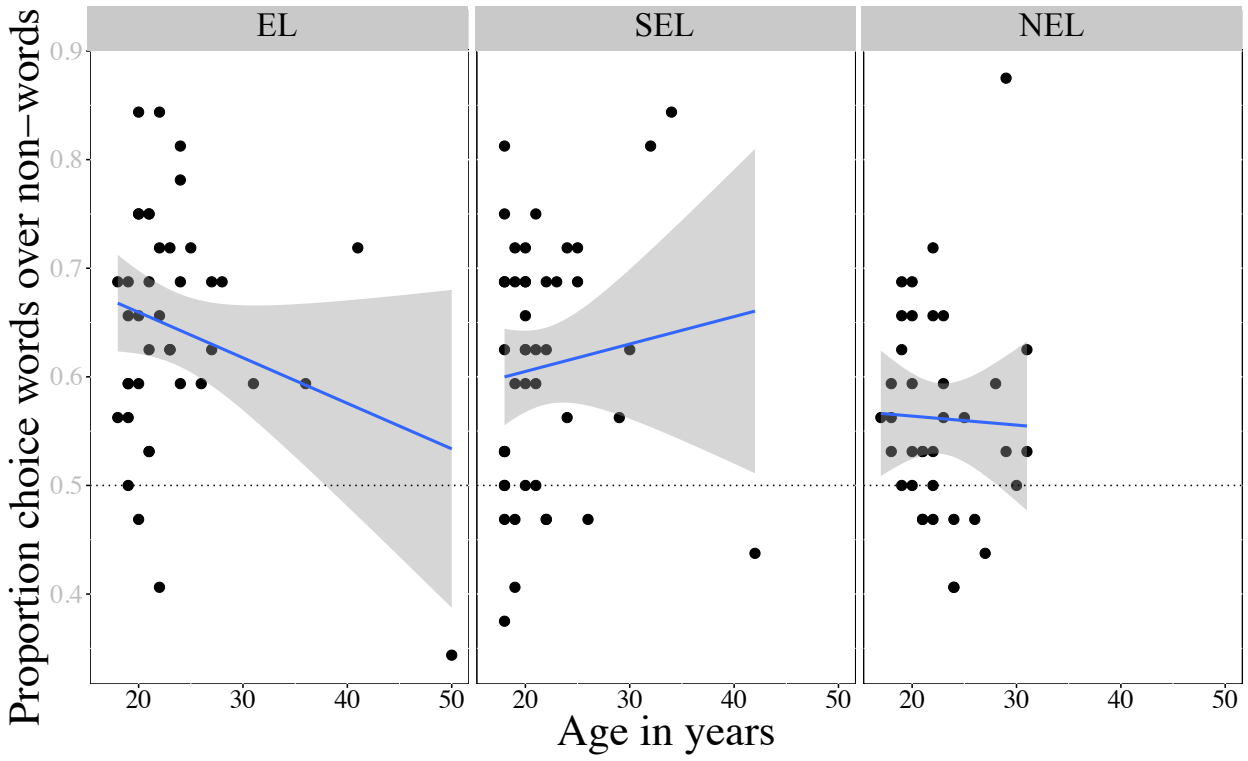

Figure 3.7 The effect of age on SL across experimental language conditions. EL stands for native English language; SEL is for semi-English language; NEL is for non-English language. Each dot represents an individual participant's mean performance on all word versus non-word trials. The dotted line reflects chance performance. The blue line and surrounding grey bars reflect the line of best fit and 95% confidence interval.

As can be seen in Figure 3.7, the distribution of participant age was heavily right-skewed ($\gamma_1 = 2.41$): though the median age of the sample was 21, the range extended to 50 years. Reducing the range to under 30 years of age reduces the sample by 9 participants and creates a more normal distribution ($\gamma_1 = 0.88$). Running the same generalized models on this restricted dataset eliminates the age effects; I will return to this in the discussion below.

I next ran the same models, but restricted to the semi-English and non-English conditions in order to examine the potential contribution of specific language experience. These models reinforced the correlational results: early lingual experience facilitated performance in the SEL (see Table 3.5 below for model summary).

Model Structure:
Choice ~ Language condition * Early lingual +
Language condition * Music +
Language condition * Age +
Language condition * Specific language consonant +
Language condition * Specific language vowel +
(1|Subject)

	SEL = reference level			NEL = reference level		
	OR	*CI*	*p*	*OR*	*CI*	*p*
Fixed effects						
(Intercept)	1.37	0.95 – 1.96	.089	1.08	0.83 – 1.41	.563
SEL				1.26	0.81 – 1.98	.304
NEL	0.79	0.51 – 1.24	.304			
Early lingual	1.15	1.05 – 1.26	**.003**	1.05	0.94 – 1.17	.368
Music	1.02	0.91 – 1.13	.756	0.99	0.90 – 1.09	.795
Age	2.17	0.82 – 5.71	.118	1.24	0.56 – 2.75	.602
Specific lang Vowel	1.37	0.95 – 1.97	.094	2.33	0.84 – 6.49	.106
Specific lang Consonant	1.73	0.67 – 4.48	.258	2.78	0.36 – 21.81	.330
SEL : Early Lingual				1.26	0.81 – 1.98	.304
NEL : Early Lingual	0.92	0.80 – 1.06	.226			
SEL : Music				1.03	0.89 – 1.19	.686
NEL : Music	0.97	0.84 – 1.12	.686			
SEL : Age				1.75	0.50 – 6.15	.382
NEL : Age	0.57	0.16 – 2.01	.383			
SEL : Spec lang Vowel				0.59	0.20 – 1.74	.336
NEL : Spec lang Vowel	1.70	0.57 – 5.06	.336			
SEL : Spec lang Cons				0.62	0.06 – 6.00	.681
NEL : Spec lang Cons	1.61	0.17 – 15.59	.682			
Random effects						
$\tau_{00,\ Subject}$			0.004			
$N_{Subject}$			63			
$ICC_{Subject}$			0.001			
Observations			2016			
Deviance			2699.25			

Table 3.5 Generalized linear model results predicting proportion choice words over non-words by early lingual experience, music, and age by language conditions.

3.4 Discussion

In this study, I predicted that specific language experience, bi/multilingualism, and musical skill would positively impact learners' abilities to detect statistical patterns in continuous

streams of speech, while age would negatively impact performance. I framed these predictions in what I termed the *encoding* hypothesis – that is, that a perceiver's ability to encode perceptual stimuli would fundamentally impact statistical learning. The results reveal partial support for this claim. On the one hand, I found that learners performed worse overall on language conditions with increasingly unfamiliar (to a native English speaker's ear) sounds. There was no clear impact of specific language experience, although a larger degree of overlap consistently patterned with better performance in both the semi-English and non-English language conditions. Moreover, early multilingual experience facilitated performance in the semi-native language condition, providing support for the claim that bilingualism – either through a global cognitive advantage, or by having established a more flexible or efficient auditory system – facilitates statistical learning. On the other hand, contrary to my predictions, musical experience did not correlate with performance in any condition, and age facilitated, rather than impaired, performance in the semi-native language condition.

A number of caveats must be noted for each of these effects. I will address them in reverse order. As mentioned in the analysis, age was heavily right-skewed, making the age-related estimates susceptible to extreme outliers. Indeed, when removing these outliers, the age-related effects no longer surface in the regression model. The original estimates, then, may be spurious. It is also possible, however, that reducing the sample to individuals under 30 actually removes an effect that is underlyingly there; in other words, performance at these younger ages may not yet show the improvement/decline that increasing age actually incurs. Distinguishing between these possibilities, unfortunately, cannot be determined by this sample.

Musical skill did not correlate with performance in any condition. This finding accords with similar results in Wang and Saffran (2014): while bilingualism facilitated segmentation of a

tonal language, musical skill was unrelated in either monolingual or bilingual participants. It may be worth noting, however, that the metric reported here (similarly constructed as in Wang & Saffran, 2014) differs from measures that have been used more broadly in explorations of the impact of musical training on cognitive skills. In the current study, participants were asked to rate themselves on a scale of 0 – 4, but were not asked how often they practiced or whether they were still routinely training on that instrument. In the literature discussed above, relationships between musical skill and cognitive measures are typically found in groups of highly trained musicians as opposed to non-musician control groups. Most of our participants would likely fall into the "non-musician" category of those samples. There were 10 individuals who rated themselves as professional-level musicians. Their mean performance was 65%, compared to the mean performance of 61% in the remainder of the sample ($t(10.5) = -1.03$, $p = .33$). As with the predictor age, the current sample is simply not diverse enough to either confirm or reject the encoding hypothesis as it pertains to the effect of musical training.

Finally, the sample does appear to support some role of bi-/multilingual experience on segmentation performance, but in somewhat unexpected ways. That is, multilingualism positively impacted performance in the native (current bilingual proficiency) and semi-native (early bilingual experience) language conditions, but not in the non-native language condition. This might be interpreted as evidence that – as appeared to be the case in Chapter 2 – there was insufficient learning in the non-native condition across the board, thereby impeding our ability to detect any subtle effects (see Siegelman and Frost, 2015 for a discussion of the importance of variability for individual difference predictions). This failure to learn as efficiently in the non-English language condition accords with *encoding* hypothesis: i.e., lack of familiarity with the stimuli impeded learning of the embedded statistical structure. I also examined evidence for a

direct link between relative familiarity and learning outcomes by correlating a talker's specific language experience with their performance. While prior experience with both consonants and vowels was positively associated with performance in both language conditions, these correlations were negligible to small, and non-significant. This may be a meaningful null effect; however, it is important to note that the phonetic overlap values are derived from sparse and likely rather inaccurate data. They are also only calculated for each learner's second language, and do not take into account any of the other languages spoken by the learner.

Taken together, these results offer some support for the claim that early bilingualism tunes auditory capacities, which in turn impacts statistical learning. It would be premature, however, to conclude – as proposed in the introduction of this chapter – that reported differences in performance across modalities (e.g., Conway & Christiansen, 2005; Emberson et al., 2011) or different SL tasks (e.g., Siegelman & Frost, 2015) are a function of differences in prior knowledge states. Future studies that endeavor to push the encoding hypothesis further will need to use better measures of the predicted individual differences (e.g., a direct test of perceptual familiarity of the acoustic stimuli), and improved measures of the SL capacity itself (e.g., a more implicit measure of learning, such as implemented in serial-reaction-time tasks or through neuroimaging of entrainment to the underlying structure during familiarization).

Chapter 4: Developmental change in Statistical Learning

Chapters 2 and 3 have revealed that adults' representations reflect both the adjacent syllable relationships, as well as something about the position of syllables within high TP-defined chunks. I hypothesized that stronger evidence for these positional learning effects might emerge under conditions of increased perceptual load – either through a reduced ability to encode the sounds (Perruchet & Poulin-Charronnat, 2012), or through a division of attentional resources (Finn et al., 2014). I found that reduced ability to encode sounds as a function of familiarity did not increase positional learning effects; rather, there was evidence that a severe reduction in phonetic accessibility (the non-English condition) limited learning overall, and may have restricted the level of analysis to immediately adjacent syllables. Altering attentional resources, on the other hand, led to relatively high levels of learning overall, and greater evidence for a role of positional information in learners' extracted word representations. The role of attention as a functional contributor to SL success was further reflected in individual difference predictors: multilingualism facilitated performance on learning from streams composed of *both* native language and semi-familiar sounds.

In the present chapter, I re-examine these questions through a different lens. From infancy through to adolescence, learners differ along the two dimensions that were manipulated in the adult studies – that is, in the quality and stability of their phonological representations (e.g., Werker & Tees, 1984; Hazan & Barrett, 2000; Houston & Jusczyk, 2000; Zamuner, Moore, & Desmeules-Trudel, 2016; Rigler, Farris-Trimble, Greiner, Walker, Tomblin & McMurray, 2015) and the maturity of their executive function skills (e.g., Welsh, Pennington, & Groisser, 1991; Huizinga, Dolan, & van der Molen, 2006; see Blakemore & Choudhury, 2006 for review

on executive function development). We might therefore expect differences in child statistical learning outcomes in comparison to adults'. Much of the research to date, however, has suggested that auditory SL is isomorphic across development (e.g., Raviv & Arnon, 2016; though note that this is in contrast to visual SL, e.g., Arciuli & Simpson, 2011). As with the adult studies, I proposed that a closer examination of children's extracted representations might provide a more powerful means of revealing potential developmental differences, and thus further elucidate the mechanisms involved in SL.

I present the data from an experiment with 7- to 13-year-olds in which they listened to a stream of English sounds, and then answered 56 2AFC questions that pitted the high TP items from the stream (words) against lower TP items from the stream (part-words), and words or part-words against novel combinations that manipulated position-based information (fake-words).

4.1 Background

Research has demonstrated successful segmentation of continuous streams of sounds via statistical learning in newborns (Teinonen et al., 2009; Kudo et al., 2011), infants (e.g., 6-month-olds: Hay & Saffran, 2012; 8-month-olds: Saffran, Aslin, & Newport, 1996; 11-month-olds: Graf Estes & Lew-Williams, 2015; 14-month-olds: Graf Estes, Gluck, & Bastos, 2015), children (e.g., 6- to 7-year-olds: Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; 6.5- to 14-year-olds: Evans, Saffran, & Robe-Torres, 2009; 5- to 12-year-olds: Raviv & Arnon, 2017), and adults (e.g., 17- to 50-years-old: Black & Hudson Kam, *submitted*[23]; 60- to 84-years-old: Neger, Rietveld, & Janse, 2014) – with little evidence to suggest any difference in learning outcomes

[23]This manuscript is based on the data presented in Chapter 2 of this thesis.

between infancy and adulthood (see Saffran et al., 1997 and Raviv & Arnon, 2017, for direct comparisons). As learners' phonetic sensitivities are known to shift (dramatically) across this timespan (e.g., Werker & Tees, 1984; Rigler et al, 2016), it is reasonable to hypothesize that – if auditory SL fails to shift across development – it is because SL operates at the level of age-invariant perceptual primitives. In the current study, then, we might expect child learning patterns to parallel those of the adults in the familiar sounds learning condition (Section 2.2). There are reasons, however, to question this hypothesis.

Statistical learning studies across multiple domains have demonstrated that it is an iterative process (see Saffran & Kirkham, 2018 for discussion). Learners are capable of extracting nested structures using the same mechanism (Thompson & Newport, 2007) – and these representations undergo a transformation, such that certain dimensions become more salient/definitive of the object's identity (Fiser & Aslin, 2005). Moreover, the material that is attended to carries with it previously learned associations that are not always relevant to the stream itself (Zhao & Yu, 2016). These facts suggest that SL can take place over any number of levels of representation, and that the process itself creates new levels of representation, that are then available for future SL. Indeed – this is the reason that SL is an appealing potential mechanism for language acquisition.

How, then, are we to understand the lack of developmental differences in the auditory word-segmentation SL paradigms? There are two possibilities: (1) there are no differences between infancy and adulthood in our capacity to use TPs to segment streams of syllables; (2) there are differences, but the extant paradigms have been insufficient to uncover them. I address both of these possibilities in turn. For the first possibility to hold, we would propose a more nuanced version of the perceptual primitive SL hypothesis outlined above. That – absent

available higher-order representations – SL takes place over perceptual primitives that are developmentally invariant. The success of so many age groups at learning the same syllable-level structure then may have more to do with the syllable itself being such a perceptual primitive, and thus available for parsing to even the youngest infant (e.g., Bertoncini & Mehler, 1981; Jusczyk & Derrah, 1987; Eimas, 1999).

On the other hand, however, this would mean that any additional knowledge of syllables (and the segments they are composed of) is irrelevant to the adult learner. This possibility seems less plausible: we know that learners bring their existing knowledge to the process of statistical learning, and that it can impact their learning. For example, when learners are exposed to a stream that violates their native language phonotactic expectations, statistical learning performance is impaired (Finn & Hudson Kam, 2008; Mersad & Nazzi, 2011). Studies have also shown that learners easily attend to and extract statistical relationships between segments (Newport & Aslin, 2004; Finn & Hudson Kam, 2008) and that such learning looks remarkably similar to SL learning over syllables, despite the fact that such representations (phonemic ones) show a great deal of change with development. In addition, we know that when perceptual units are highly unfamiliar, learning is slowed down (Gebhart et al., 2009; Graf Estes et al., 2015), or limited to less complex associations/networks (Thiessen, 2010).

Finally, evidence from visual SL studies suggest improvement in SL capacities from early infancy through adolescence (Bulf et al., 2011; Arciuli & Simpson, 2011; Schlichting, Guarino, Schapiro, Turk-Browne, & Preston, 2017), and a subsequent decline (Janacsek, Fiser, & Nemeth, 2012). Although it is possible that visual and auditory SL are supported by entirely different mechanisms and neural systems (Frost, Armstrong, Siegelman & Christiansen, 2015; Li, Zhao, Shi, & Conway, 2018), work also suggests the mechanisms operate in similar (if not

identical) (e.g., Kirkham, Slemmer, & Johnson, 2002, Saffran & Kirkham, 2018), and integrated ways (Mitchel, Christiansen, & Weiss, 2014). It is possible, then, that auditory SL similarly shifts across development, but that the nature of these differences when using such familiar and simple structures as CV syllables has gone undetected. Indeed, this may be the case, as there are surprisingly few direct comparisons between adult and child performance to even evaluate. There are only two studies that directly compare children's and/or adults' performance on a purely linguistic SL task across a wide age range (Saffran et al.; Raviv & Arnon, 2017). Neither find evidence for change between early childhood and adulthood; however, participants were tested on TP-defined words versus zero-TP non-word foils. Given our results in the non-native learning condition – where participants were more successful at distinguishing high from zero-TP contrasts than from lower TP contrasts – I suggest that this may represent too blunt a tool to detect developmental change.

It should be noted that Saffran, Johnson, Aslin, and Newport (1999) conducted a developmental comparison of SL of non-linguistic tone sequences that did contrast word and part-word test items. Both infants and adults showed significant learning; this performance was compared to previously run linguistic tasks, with no differences noted between conditions. This comparison, however, was within a given age group – the infant and adult paradigms are sufficiently different (explicit 2AFC versus the more implicit measure of looking-time preference) that it is difficult to compare relative magnitudes of learning in the different age groups. Thus, while this data suggests that neither infants nor adults found tones more difficult to learn from than language sounds, we do not know whether or how their respective courses of learning might have differed. The present study, therefore, aims to clarify this question.

Finally, research has suggested that attention plays an important role in SL. For example, asking learners to track unrelated auditory or visual signals while simultaneously attending to a continuous artificial language significantly impairs learning of the embedded trisyllabic structures (Toro, Sinnett, & Soto-Faraco, 2005), as does having learners draw pictures during the familiarization exposure (Ludden & Gupta, 2000, cf., Saffran et al., 1997, and Evans et al., 2009, for different results). There is a potentially more nuanced view of the impact that attention has on SL, however: Finn et al. (2015) discovered that directing participants to concentrate their attention on the stimuli (i.e., whether they were told to look for words, categories, or word-order) resulted in more veridical tracking of adjacent TPs, while passive listening led to the extraction of more abstract categories. This finding is echoed in my own work (Experiment 4, Chapter 2): adults faced with two unrelated perceptual streams engage in more position-based (i.e., abstract) encoding than do learners faced with a (familiar) auditory stream alone. Executive function – including the ability to sustain and direct attention – continues to develop past adolescence (Enns & Girgus, 1985; Davidson, Amso, Cruess Anderson, & Diamond, 2006; McKay, Halperin, Schwartz & Sharma, 2009). We might therefore expect more position-based (or abstract-like) encoding from children than what emerges from the same auditory-only paradigm with adults.

I will examine the children's performance using the same logic laid out in Chapter 2 – that is, I will compare the children's performance against the ordering relationships predicted by the TP- and position-encoding hypotheses. These predictions are depicted in Figure 4.1 (repeated from Figure 2.3). In addition, I posit the following predictions, given the preceding literature and findings of Chapter 2:

(1) Children will successfully segment the stream (at some level) at all ages

(2) Younger children will show more evidence for position-based encoding (see Figure 4.1).

(3) Older children will converge to the adult pattern in the native-English language condition from Chapter 2 (Experiment 1).

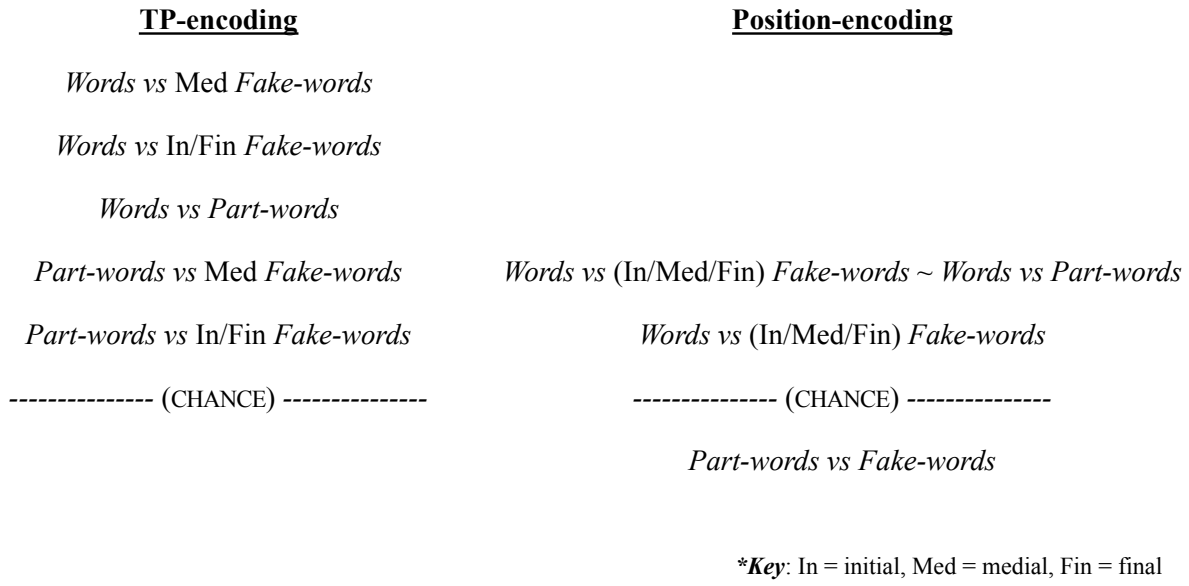

Figure 4.1 Predictions according to the TP- and position-encoding hypotheses (repeated from Figure 2.3). Trial types are plotted according to the ordering relationship of relative proportion choice, but not absolute differences from chance (the dotted line). Performance above chance means higher proportion choice of the first sequence type listed (e.g., *words* in the trial type "Words vs. PW"). Performance below chance means higher proportion choice of the second sequence type listed (e.g., *fake-words* in the trial type "Part-words vs Fake-words").

4.1.1 Methods

The experiment paradigm parallels that used with the adult participants. Any areas of difference are noted in detail below.

4.1.1.1 Participants

Seventy-seven children between the ages of 7 and 13 were recruited through the Living Lab at Science World, the local science museum. Of these, 8 were excluded due to: failure to follow instructions or complete the task (n = 3), parental report of a language-related disorder (n

= 4), and lack of signature on the parent consent form (n = 1).[24] The final sample thus consisted of 69 participants (37 female; see Table 4.1 for gender breakdown by age). Children came from a wide range of language and cultural backgrounds, and not all children were native speakers of English (defined as living in an English environment before age 3; non-native English speakers: n = 14). All children had consent to participate given by a legal guardian, and had provided their individual assent to participate. Children received a sticker for their participation.

Age in years	*Female*	*Male*	*Age range*
7	5	5	7;0, – 7;11
8	6	6	8;0 – 8;11
9	5	4	9;0 – 9;9
10	4	3	10;0 – 10;8
11	7	6	11;0 – 11;11
12	6	4	12;0 – 12;11
13	4	4	13;0 – 13;9

Table 4.1 Participants by gender and age.

4.1.1.2 Materials.

The materials are identical to those used in Experiment 1 (Language A) of Chapter 2. The inventory of sounds and words can be found in Table 2.1(page 35).

4.1.1.3 Procedure

The Living Lab consists of two testing rooms and a central waiting room in an area of the science museum that is separated from the remainder of the museum by glass walls. Parents with

[24]The parent read the consent form, filled out the language background questionnaire and verbally consented to the child's participation; however, after the parent and child had completed the study and left the lab, we discovered the parent had not signed the form.

children who appeared to be in the appropriate age range were approached in the public areas of the museum by the author or a trained research assistant. The author/research assistant would give a brief explanation of the project, and ask if the parent thought the child might be interested in participating. If the parent/s and child agreed, they were then brought back to the Living Lab waiting room for the experiment.

In the lab, parents were given a consent form and language background questionnaire (see Appendix B.1). Once the parent had read and signed the consent form, the child and parent were asked if the child was still interested in participating, and if they were comfortable sitting in a testing room with the door closed for the duration of the study. If the child and parent agreed, the child was seated at a desk in the study room. He/she was first presented with an assent form; this was summarized auditorily for children age 5-8; 9-13 year olds were given the choice to read the form on their own, or have it presented to them by the researcher.

If the child provided their assent to participate, they were instructed that they would be listening for the next few minutes to a made-up language called Vesutian, and that they would then be asked some questions about that language. The participants were prompted to put on headphones, and use the keyboard to answer questions. The researcher told the participant they would first be given some test trials to familiarize them with the question-answer procedure; the researcher stayed close by as the child went through these four test trials to ensure proper fitting of the headphones, understanding of the keyboard, and understanding of the procedure. When these test trials ended, the researcher checked that the participant understood, and that he/she was ready to listen to the familiarization stream. The researcher then prompted the participant to continue the rest of the procedure by following the instructions presented on the computer screen. The researcher remained in the room, but at a distance, for the remainder of the study.

The study itself was identical in form to the native language condition discussed in Chapter 2 Section 2.2.1.3 (pg. 58). I repeat the basic procedure here for clarity: participants first heard four training trials in which they were asked to indicate which of two sound files sounded more like the word "say". They were then asked to listen quietly to a made-up language, which was presented for two minutes. Finally, they were presented with 56 2AFC trials in which words were pitted against part-words or fake-words, and part-words and fake-words were pitted against each other. The experiment was presented through E-prime 2.0 Experimental Software (Psychology Software Tools, Pittsburgh, PA). While the adults were tested on desktop computers in sound-attenuated rooms, the children were tested via a Panasonic CF-F9 laptop computer in a quiet testing room inside the science museum (described above).

4.1.1.4 Analysis Plan

As in the previously reported adult studies, the data are analyzed by main trial type, syllable position manipulation, and trial. In addition to the effects of main trial type, syllable position, and trial, I also examine the effect of age as a continuous linear predictor, in order to probe for change in learning patterns across development. I predicted that learning would vary as a function of age and trial type – for instance, that younger children would perform relatively poorly on words versus part-words, but well on words versus medial fake-words, whereas the oldest children might perform equally well across both (as in the adult sample). For syllable manipulation trial types, I predicted that differences in performance should be most apparent in the younger children, and become less pronounced as children age. As such, I look for interactions between age, main trial type, and syllable position. Age is coded as a continuous predictor, but is graphically presented in bins by year in Figure 4.4 for visualization purposes.

In addition, I hypothesized that children may learn more or less from the testing conditions as they age; for example, it is theoretically possible that younger children might appear to perform similarly to older children, but due to rapid learning across the trials rather than entering the testing phase with a similar degree of knowledge. I therefore also look for interactions between trial, age, and trial types (both main and syllable position) to account for this kind of possibility.

Mixed effects models are constructed as follows: I first attempt a fully specified model, which includes all fixed effects and interactions, and in which the random effects structure consists of interactions, slopes, and intercepts for all within-subject variables grouped by subject intercepts, and intercepts for test items (included as a control variable).[25] Models are run with higher optimizer iterations (up to 200,000,000), and then successively pruned (beginning with the covariance of the random effects structure) until model convergence is reached. When multiple models are run on the same analysis (i.e., in order to rotate the reference level of a categorical variable), the simplest model structure required for convergence is applied across each model run. All model results are reported in terms of odds ratios, their 95% confidence intervals (derived via *Wald* tests), and associated *p*-values. Statistical analysis was done in R (Version 3.3.3), using the packages lme4 and sjPlot.

[25]As the children were all run on the same language and testing items, there may be random variation associated with the items that can be captured by the generalized model. Item was not included in the model structure of the adult sample, though, as half of the participants in the English language condition (Experiment 1) were exposed to one language, and the other half were exposed to a different language and set of testing items. For the sake of comparison across the four adult experiments, item was withheld from all models.

Finally, I also examine the raw correlations between individual children's performance on the different trial types and syllable manipulations, to determine patterns of learning (e.g., to see whether good learners, as indexed by performance on word versus part-word trials, are also better at certain syllable position trial types). As there is insufficient power to detect patterns by year of age, these results are reported collapsed across the entire sample.[26] Unlike in the adult sample, children were responding on a laptop keyboard (Panasonic CF-F9) and not via button box; there is therefore too much noise in the timing accuracy to justify an analysis of reaction time data.

4.1.2 Results

Proportion choice aggregated by individual across main trial types and syllable manipulations is presented in Figure 4.1.

[26]The adult correlations averaged around 0.3; assuming a similar effect size in the children would require a sample of 85 children per age group.

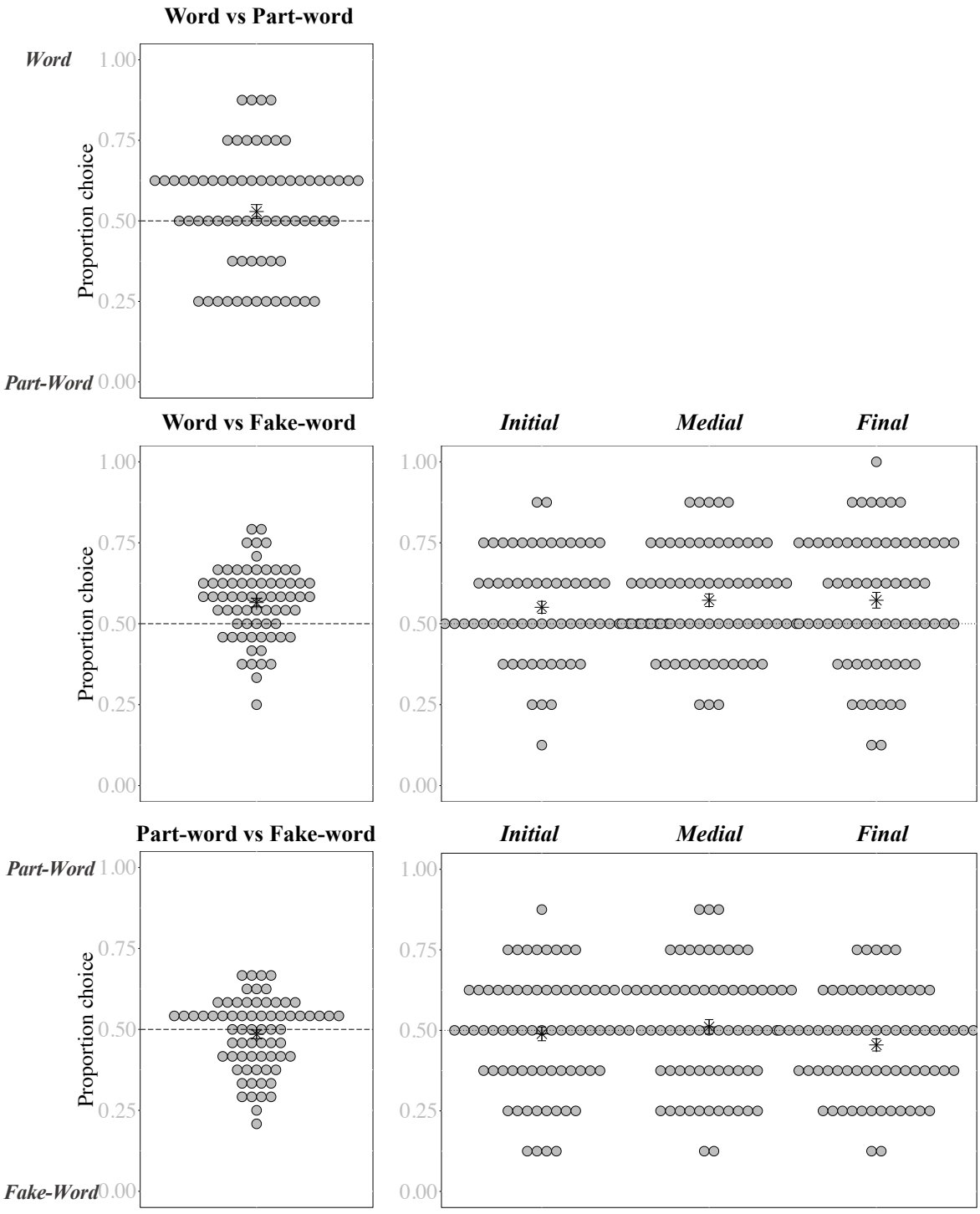

Figure 4.2 Proportion choice by trial type and syllable position manipulation. Dots reflect individual participant mean scores. Stars reflect mean accuracy scores; error bars are plus/minus 1 standard error. Chance is 0.5 (the dotted line).

4.1.2.1 Words versus Part-Words.

I first examined children's performance on trials that pitted words against part-words. Given the poor performance of adults in the non-native sounds condition on this contrast, I predicted that younger children would similarly struggle, given their relatively poor phonological representations. T-test comparisons of proportion choice to change performance (50%) reveal that the children did not distinguish words from part-words (M = 52.9%, SD = 18.0%, 95% CI = [48.6%, 57.2%], $t(68) = 1.34$, $p = .18$, $d = 0.16$). To look for change across the age-span, and to determine whether there was learning over the course of the experiment, I fitted the data to a generalized mixed effects model specified for the two-way fixed effects interaction between age and trial, and random slopes for trial by subject intercepts. This model revealed no change over the tested age range (OR = 1.14, $p = .18$). Trial was not significant (OR = 0.99, $p = .21$), nor was the interaction between trial and age (OR = 0.99, $p = .36$).

4.1.2.2 Words versus Fake-Words.

Results are first reported for all word versus fake-word trials as a whole, and then broken down by syllable manipulation type.

4.1.2.2.1 Combined

T-test comparisons of performance against chance (50%) reveal that children successfully distinguished words from fake-words overall (M = 56.5%, SD = 11.0%, 95% CI = [53.9%, 59.2%], $t(68) = 4.92$, $p < .0001$, $d = 0.59$). A mixed effects model was run to determine whether

performance varied by age, over the course of the experiment, or both.[27] This model revealed a small but significant decrease in performance over the course of the experiment (OR = 0.99, p = .003), and improved performance (though non-significant) with increasing age (OR = 1.11, p = .070), but no interaction between the two (OR = 1.00, p = .96). Performance by age and trial can be seen in Figures 4.2 and 4.3, respectively.

4.1.2.2.2 Syllable Manipulations

Independent *t*-tests comparing group-level performance to chance (50%) reveal that participants chose words significantly more often than fake-words across all syllable positions: Initial (M = 55.0%, SD = 15.4%, 95% CI = [51.4%, 58.8%], $t(68) = 2.74$, $p = .008$, $d = 0.32$), Medial (M = 57.2%, SD = 16.2%, 95% CI = [53.3%, 61.0%], $t(68) = 3.71$, $p = .0004$, $d = 0.44$), and Final (M = 57.2%, SD = 20.0%, 95% CI = [52.4%, 62.1%], $t(68) = 2.99$, $p = .004$, $d = 0.36$). We are also interested in whether there are differences in performance across the different syllable positions, whether there is change in performance on the syllable types as children develop, and whether learning (as indexed by change over trial) differs by syllable position, age, or both.

Full model results for a generalized mixed effects model fit to this data can be found in Table 4.2.[28] There were no significant differences across syllable position, and no interactions.

---

[27]Choice ~ Age * Trial + (Trial | Subject) + (1 | Item)

[28]The first model attempted had the following structure:

Choice ~ Age * Syllable Position * Trial + (Syllable Position * Trial | Subject) + (1 | Item)

This model failed to converge; model convergence could not be reached for each syllable position reference level until the 3-way interaction was removed from the fixed effects structure. The final model structure that converged for every syllable reference is listed at the top of Table 4.2.

Children's performance decreased over the course of the experiment when initial- (OR = 0.99, p = .034) or medial- (OR = 0.99, p = .014) syllable manipulations served as reference levels, but not when final-syllable manipulations was the reference (OR = 1.00, p = .805). When final-syllable manipulation was the reference level, there was a significant positive effect of age (OR = 1.24, p = .018). In other words, there is some evidence that children become more confused by initial- and medial-syllable fake-words over the course of the experiment, and that children get better at rejecting final-syllable fake-words as they mature. Figure 4.3 shows the patterns by age; figure 4.4 shows the patterns by trial.

Model structure:

Choice ~ Age * Syllable Position + Trial * Syllable Position + Trial * Age + (1 | Subject) + (1 | Item)

	Model 1 Reference level = Initial			**Model 2** Reference level = Medial			**Model 3** Reference level = Final		
	OR	***CI***	***p***	***OR***	***CI***	***p***	***OR***	***CI***	***p***
Fixed Effects									
(Intercept)	1.25	0.87 – 1.80	.227	1.33	0.93 – 1.91	.120	1.36	0.95 – 1.96	.093
Syllable: Initial				0.94	0.56 – 1.56	.804	0.92	0.55 – 1.52	.739
Syllable: Medial	1.07	0.64 – 1.77	.804				0.98	0.59 – 1.63	.933
Syllable: Final	1.09	0.66 – 1.81	.739	1.02	0.62 – 1.70	.933			
Trial	0.99	0.98 – 1.00	**.035**	0.99	0.98 – 1.00	**.010**	1.00	0.99 – 1.01	.626
Age	1.08	0.90 – 1.30	.386	1.01	0.84 – 1.21	.927	1.25	1.04 – 1.50	**.017**
Syll Initial * Trial				1.00	0.99 – 1.02	.799	0.99	0.98 – 1.01	.225
Syll Medial * Trial	1.00	0.98 – 1.01	.799				0.99	0.97 – 1.00	.133
Syll. Final * Trial	1.01	0.99 – 1.03	.225	1.01	1.00 – 1.03	.133			
Syll Initial * Age				1.08	0.84 – 1.38	.572	0.87	0.67 – 1.12	.273
Syll Medial * Age	0.93	0.72 – 1.20	.572				0.81	0.63 – 1.04	.096
Syll Final * Age	1.15	0.89 – 1.48	.273	1.24	0.96 – 1.59	.096			
Age * Trial	1.00	0.99 – 1.01	.869	1.00	0.99 – 1.01	.869	1.00	0.99 – 1.01	.869
Random Effects									
$\tau_{00,\ Subject}$		0.032							
$\tau_{00,\ Item}$		0.205							
$N_{Subject}$		69							
N_{Item}		24							
$ICC_{Subject}$		0.009							
ICC_{Item}		0.058							
Observations		1656							
Deviance		2133.080							

Table 4.2 Model results proportion choice by syllable position, age, and trial in Word versus Fake-word trial types.

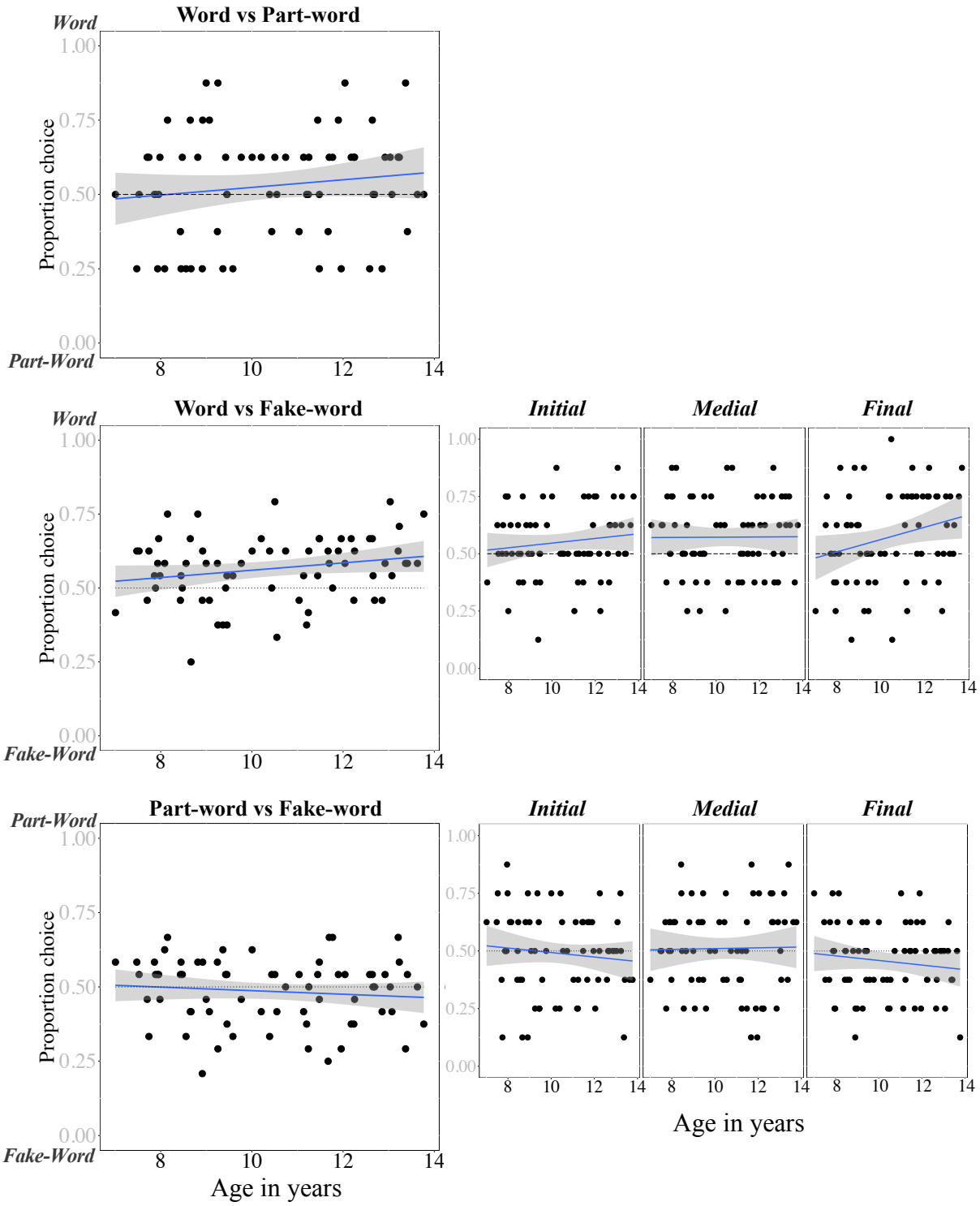

Figure 4.3 Proportion choice by trial type, syllable manipulations, and age Dots represent individual subject means. Chance performance is represented by the dotted line at 0.5. Blue lines reflect best linear fit, grey proportions reflect the 95% CI.

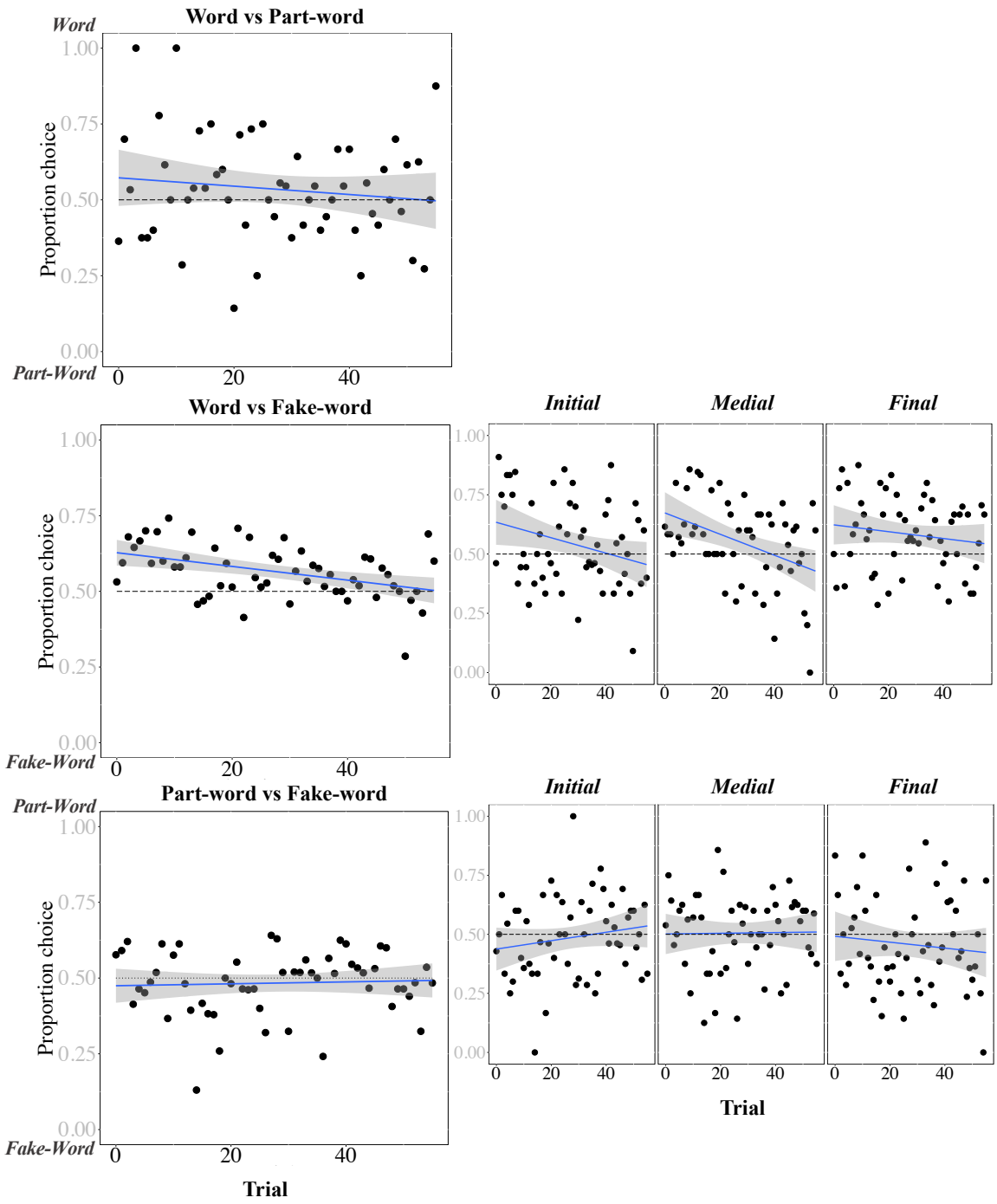

Figure 4.4 Proportion choice by trial type and trial. Dots represent trial means across subjects. Chance performance is represented by the dotted line at 0.5. Blue lines reflect best linear fit, grey proportions reflect the 95% CI.

4.1.2.3 Word versus Part-Word compared to Word versus Fake-Word.

The position-encoding hypothesis predicts that children should find (at least some) fake-words more confusing than part-words in contrast to words, but that this confusion will diminish as they get older. The TP-encoding hypothesis, on the other hand, predicts that children will find fake-words (which always involve 0.0 TPs) easier to reject in comparison to part-words (which involve non-zero TPs across both syllable transitions). Although the specific models run on each set of data separately showed that this was the case – children distinguished words from fake-words but not words from part-words, the difference in performance between word vs. fake-word and word vs. part-word test trials is not significant ($t(68) = -1.58$, $p = .12$, $d = 0.19$). To examine, however, whether the ability to distinguish words from fake-words and part-words changed across age, whether any syllable position manipulations differed from word versus part-word trials, and whether these factors interacted with learning across the experiment, a mixed effects model with contrast type (i.e., word versus part-word, word versus initial fake-word, word versus medial fake-word, word versus final fake-word), trial, age, and their interaction as fixed effects, with random intercepts for subjects was fitted to the data. None of these factors significantly contributed to performance (Table 4.3).

Model structure:
Choice ~ Contrast type * Age * Trial + (1 | Subject)

	OR	*CI*	*p*
Fixed Effects			
(Intercept)	1.13	0.95 – 1.35	.155
Initial Syll	1.08	0.85 – 1.37	.523
Medial Syll	1.16	0.91 – 1.47	.237
Final Syll	1.19	0.94 – 1.51	.155
Trial	0.99	0.98 – 1.00	.196
Age	1.12	0.94 – 1.34	.220
Initial Syll * Trial	1.00	0.98 – 1.01	.532
Medial Syll * Trial	0.99	0.98 – 1.01	.391
Final Syll * Trial	1.01	0.99 – 1.02	.449
Initial Syll * Age	0.97	0.76 – 1.23	.776
Medial Syll * Age	0.90	0.71 – 1.15	.413
Final Syll * Age	1.11	0.87 – 1.42	.416
Trial * Age	1.00	0.98 – 1.01	.425
Initial * Age * Trial	1.00	0.99 – 1.02	.726
Medial * Age * Trial	1.00	0.99 – 1.02	.548
Final * Age * Trial	1.01	0.99 – 1.02	.426
Random Effects			
$\tau_{00,\ Subject}$		0.033	
$N_{Subject}$		69	
$ICC_{Subject}$		0.010	
Observations		2208	
Deviance		2976.618	

Table 4.3 Model results for generalized linear model predicting choice by contrast type, age, and trial for Word versus PW and Word versus FW trial types

4.1.2.4 Part-Words versus Fake-Words.

The position-encoding hypothesis predicts that children will choose fake-words over part-words, at least in some syllable manipulations, and that younger children will be more likely to do so than older children. The TP-encoding hypothesis predicts that children should choose part-words over fake-words, and that this might shift from at-chance performance earlier in life as compared to later. Results are reported below, first as main effects, and then broken down by syllable positions.

4.1.2.4.1 Combined.

Participants did not prefer either fake-words or part-words (M = 48.5%, SD = 10.9%, 95% CI = [45.9%, 51.1%], $t(69) = -1.15$, $p = .26$, $d = 0.14$). This global pattern did not change with age (OR = 0.95, CI = [0.85, 1.05], $p = .32$) or trial (OR = 1.00, CI = [1.00, 1.01], $p = .55$), or their interaction (OR = 1.01, CI = [1.00, 1.01], $p = .068$).[29]

4.1.2.4.2 Syllable Manipulations

Mean performance across age can be seen in Figure 4.2. As a group, performance differed from chance in the final syllable position manipulation, but not the initial or medial positions: Initial (M = 48.9%, SD = 17.8%, 95% CI = [44.6%, 53.2%], $t(68) = -0.51$, $p = .61$, $d = .06$), Medial (M = 51.1%, SD = 18.8%, 95% CI = [46.6%, 55.6%], $t(68) = 0.48$, $p = .63$, $d = .06$), and Final (M = 45.5%, SD = 15.6%, 95% CI = [41.7%, 49.2%], $t(68) = -2.41$, $p = .02$, $d = .29$). To see if these means differed from each other, changed over development, the course of the experiment, or in interaction, a logistic mixed effects model was fitted to the data. These models are reported in Table 4.4. There is no significant effect of age or trial, nor any interactions between the three factors.

[29]Final pruned model structure: Choice ~ Age * Trial + (0 + Age * Trial | Subject) + (1 | Item)

Model Structure: Choice ~ Syllable position * Age * Trial + (1 | Subject) + (0 + Trial | Subject) + (1 | Item)

	Model 1 Reference level: Initial			**Model 2** Reference level: Medial			**Model 3** Reference level: Final		
	OR	*CI*	*p*	*OR*	*CI*	*p*	*OR*	*CI*	*p*
Fixed Effects									
(Intercept)	0.95	0.76 – 1.20	.692	1.05	0.83 – 1.32	.698	0.83	0.65 – 1.04	.110
Syll Pos: Initial				0.91	0.66 – 1.27	.580	1.16	0.83 – 1.60	.388
Syll Pos: Medial	1.10	0.79 – 1.52	.574				1.27	0.91 – 1.76	.154
Syll Pos: Final	0.87	0.62 – 1.20	.388	0.79	0.57 – 1.09	.154			
Trial	1.01	1.00 – 1.02	.227	1.00	0.99 – 1.01	.962	1.00	0.99 – 1.01	.583
Age	0.92	0.77 – 1.10	.385	1.02	0.85 – 1.21	.867	0.92	0.77 – 1.10	.373
Initial * Trial				1.01	0.99 – 1.02	.414	1.00	0.99 – 1.02	.217
Medial * Trial	0.99	0.98 – 1.01	.414				1.00	0.99 – 1.02	.671
Final * Trial	0.99	0.98 – 1.01	.217	1.00	0.98 – 1.01	.671			
Initial * Age				0.91	0.71 – 1.16	.453	1.00	0.78 – 1.28	.983
Medial * Age	1.10	0.86 – 1.40	.454				1.10	0.86 – 1.41	.443
Final * Age	1.00	0.78 – 1.28	.983	0.91	0.71 – 1.16	.443			
Trial * Age	1.00	0.99 – 1.01	.486	1.00	0.99 – 1.02	.397	1.01	1.00 – 1.02	.085
Initial * Trial * Age				1.00	0.98 – 1.01	.915	0.99	0.98 – 1.01	.464
Medial * Trial * Age	1.00	0.99 – 1.02	.915				1.00	0.98 – 1.01	.532
Final * Trial * Age	1.01	0.99 – 1.02	.464	1.00	0.99 – 1.02	.532			
Random Effects									
$\tau_{00,\ Subject}$					0.023				
$\tau_{00,\ List2}$					0.000				
$N_{Subject}$					69				
N_{List2}					24				
$ICC_{Subject}$					0.007				
ICC_{List2}					0.000				
Observations					1656				
Deviance					2233				

Table 4.4 Generalized linear model results of the effect of age, trial, and syllable position on proportion choice part-words versus fake-words

4.1.2.5 Correlations.

If learning is driven primarily by TPs, performance should be positively correlated across the three main trial types – that is, in each trial type, choice of the higher TP item will lead to performance above chance. A negative correlation between word versus non-word (i.e., part-word or fake-word) and part-word versus fake-word trials would suggest position-based encoding. Correlations are first presented across the main trial types, and then by syllable position.

4.1.2.5.1 Combined.

Though non-significant, there are positive correlations between performance on the word versus part-word trials and word versus fake-word trials ($r(68) = 0.20$, $p = .10$), and word versus fake-words and part-words versus fake-word trials: $r(68) = 0.21$, $p = .08$. Unlike in the adult sample, there was no relationship between word versus part-word and part-word versus fake-word trials ($r(68) = 0.07$, $p = .59$). The positive correlation between word versus fake-word and part-word versus fake-word trials is consistent with the interpretation that child learners were driven primarily by TP strength.

4.1.2.5.2 Syllable manipulations.

Correlations across the syllable manipulation and word versus part-word conditions are very low (absolute value average $r = 0.1$; see Table 4.5). There is a significant positive correlation between performance on the word versus part-word trials and word versus medial fake-word trials ($r(68) = .32$, $p = .007$), and between word versus final fake-word trials and part-word versus initial fake-word trials ($r(68) = .29$, $p = .02$). Word versus medial fake-word and

word versus final fake-word trials were also positively correlated, though not significantly ($r(68) = .22,\ p = .07$).

Variable		1	2	3	4	5	6
1. Word vs PW							
2. Word vs FW	*Initial*	.00 [-.23, .24]					
3.	*Medial*	.32** [.09, .52]	.14 [-.10, .36]				
4.	*Final*	.07 [-.17, .30]	-.04 [-.27, .20]	.22 [-.02, .43]			
5. PW vs FW	*Initial*	.07 [-.16, .31]	-.01 [-.25, .22]	.16 [-.08, .38]	.29* [.05, .49]		
6.	*Medial*	.06 [-.18, .29]	.17 [-.07, .39]	.09 [-.15, .32]	.05 [-.19, .28]	.18 [-.06, .40]	
7.	*Final*	-.02 [-.25, .22]	.00 [-.24, .24]	.05 [-.19, .28]	-.07 [-.30, .17]	-.02 [-.25, .22]	.10 [-.14, .32]

Table 4.5 Correlations by trial type and syllable manipulation

4.1.3 Discussion

In this study, I looked for evidence that SL performance varies as a function of age, and hypothesized that previous studies have failed to find such evidence due to a too-coarse examination of the nature of the extracted representations. I hypothesized that less stable phonological representations in early childhood and less mature attentional resources would impact SL, and lead to greater evidence for position-based encoding (based on the results of Experiments 2 and 4 in Chapter 2). I will discuss each of the specific predictions I made in turn.

The position-encoding and TP-encoding hypotheses, as described in Chapter 2, predict a different order of relative performance on the different trial types. The group-level performance on trial types aligns more closely with the TP-encoding hypothesis than the position-encoding hypothesis. Children performed best on word versus medial fake-words, followed by word versus initial- and final-fake-words. They were at chance on word versus part-word contrasts, as well as part-word versus initial- and medial-fake-word contrasts. The one contrast that did not align with the TP-encoding predictions is the part-word versus final fake-word contrasts: here children chose fake-words more frequently than part-words (d = -.29). These patterns are presented graphically in Figures 4.5 (children's performance in comparison to the TP-encoding and Position-encoding predictions) and 4.6 (children's performance in comparison to the adult data). As can be seen from Figure 4.6, the strength of children's performance (i.e., how successful they were at segmentation) most closely resembles that of the adults who were exposed to non-English sounds (Chapter 2, Experiment 3). Aside from their poor performance on word versus part-word trials, the ordered relationship between test trials most closely resembles that of the adults exposed to a video simultaneously with the audio stream (Chapter 2, Experiment 4).

TP-encoding	**Position-encoding**	**CHILDREN'S PERFORMANCE**
Words vs Med *Fake-words*		*Words vs* Med *Fake-words*
Words vs In/Fin *Fake-words*		*Words vs* In *Fake-words*
Words vs Part-words		*Words vs* Fin *Fake-words*
Part-words vs Med *Fake-words*	*Words vs* (In/Med/Fin) *Fake-words ~ Words vs Part-words*	*Words vs Part-words*
Part-words vs In/Fin *Fake-words*	*Words vs* (In/Med/Fin) *Fake-words*	*Part-words vs* Med *Fake-words*
---------- (CHANCE) ----------	---------- (CHANCE) ----------	---------- (CHANCE) ----------
	Part-words vs Fake-words	*Part-words vs* In *Fake-words*
		Part-words vs Fin *Fake-words*

Key: In = initial, Med = medial, Fin = final

Figure 4.5 Predicted performance compared to actual performance. Item types plotted above the 50% chance performance line indicate greater proportion choice of the first listed item (e.g., greater proportion choice *Words* over *Part-words*). Items plotted below the 50% chance performance line indicate greater proportion choice of the second listed item (e.g., greater proportion choice *Fake-words* over *Part-words).* Distance from chance or from other item types does not reflect absolute differences in performance, rather the predicted/actual relative order of performance.

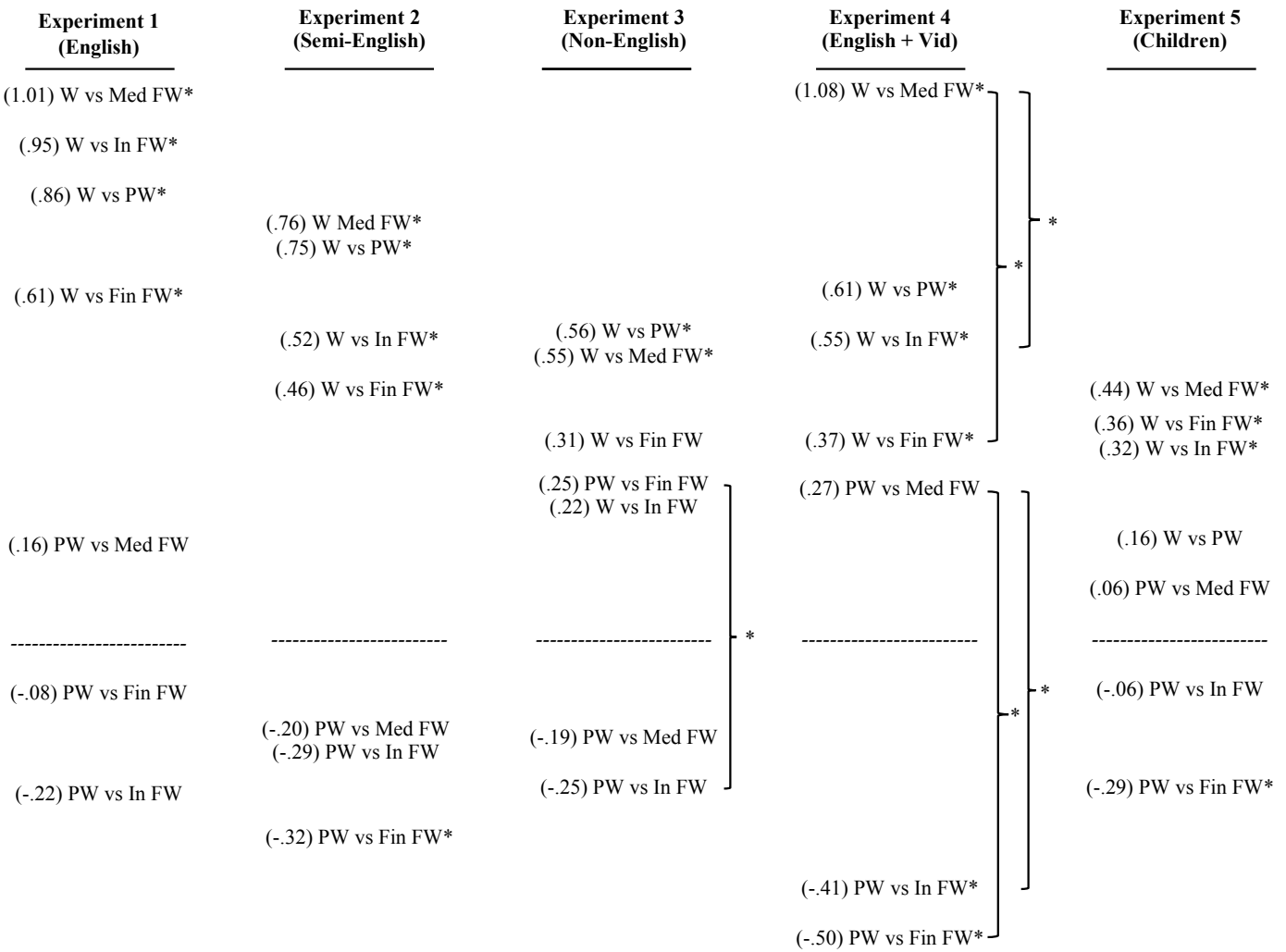

Figure 4.6 Ordered relationship of performance on all trial types and syllable manipulations. W stands for *words;* FW stands for *fake-words*; PW stands for *part-words*. In stands for *initial;* Med stands for *medial*; Fin stands for *final*. * indicate trial types that are significantly different from chance; brackets and *s indicate trial types that are significantly different from one another. Performance on PW vs FW trials was never contrasted with performance on W vs PW or W vs FW trials.

There was little clear evidence of change across development. It is unexpected that the children were unable to distinguish words from part-words at any point across the age span tested. This contrast has become a standard test of statistical learning success since the original Saffran, Aslin, and Newport (1996) paper, and is frequently used successfully with very young infants. In the adult studies presented in this dissertation, learners succeeded in distinguishing words from part-words in all language and attention conditions, even when evidence from the rest of the paradigm indicated low levels of learning overall (i.e., the non-English condition, which had smaller effect sizes and no internal cohesion to participant strategies/learning across trial types). There was, however, a significant effect of age for word versus final-syllable fake-words, such that they became easier to reject as children grew older (OR = 1.25, p = .017); this patterned in the same direction, though non-significantly, for all of the word versus non-word contrasts (see Tables 4.2 and 4.3). This finding is compatible with either the TP-encoding or position-encoding hypotheses: both predict that children will improve at distinguishing words from non-word foils.

In contrast, there are asymmetrical effects of learning across the experiment. Medial- and initial-syllable fake-words became more difficult for all-age children to reject further into the experiment, a pattern that was mirrored (non-significantly) for word versus part-word trials – but was not for word versus final-syllable fake-words. To better understand these modelled effects, I present the model predicted fits in Figure 4.7 below. The model suggests that older children perform similarly on all four contrast types at the beginning of the experiment, but performance over trials suffers in all conditions except for words versus final-syllable manipulated fake-words. Younger children, on the other hand, succeed on the medial-syllable manipulated fake-words at the beginning of the task, but are at chance for final-syllable fake-word and part-word

contrasts. This simplified picture of (very messy) data presents a plausible, possible interpretation: namely, it appears that while the younger children do not distinguish words from part-words, older children do. They perform increasingly worse on this contrast over the course of the experiment, however – with the net result being performance only slightly above chance. This result must be interpreted with caution, of course – the interaction fails to reach significance based on the actual data at hand. It is, however, an interpretation that aligns with known facts from the literature: children as young as 7 years old succeed at SL tasks (e.g., Saffran et al., 2008), but children across this span are susceptible to task demands that may artificially impact performance (e.g., Schiff & Knopf, 1985; Burkart & Rueth, 2013).

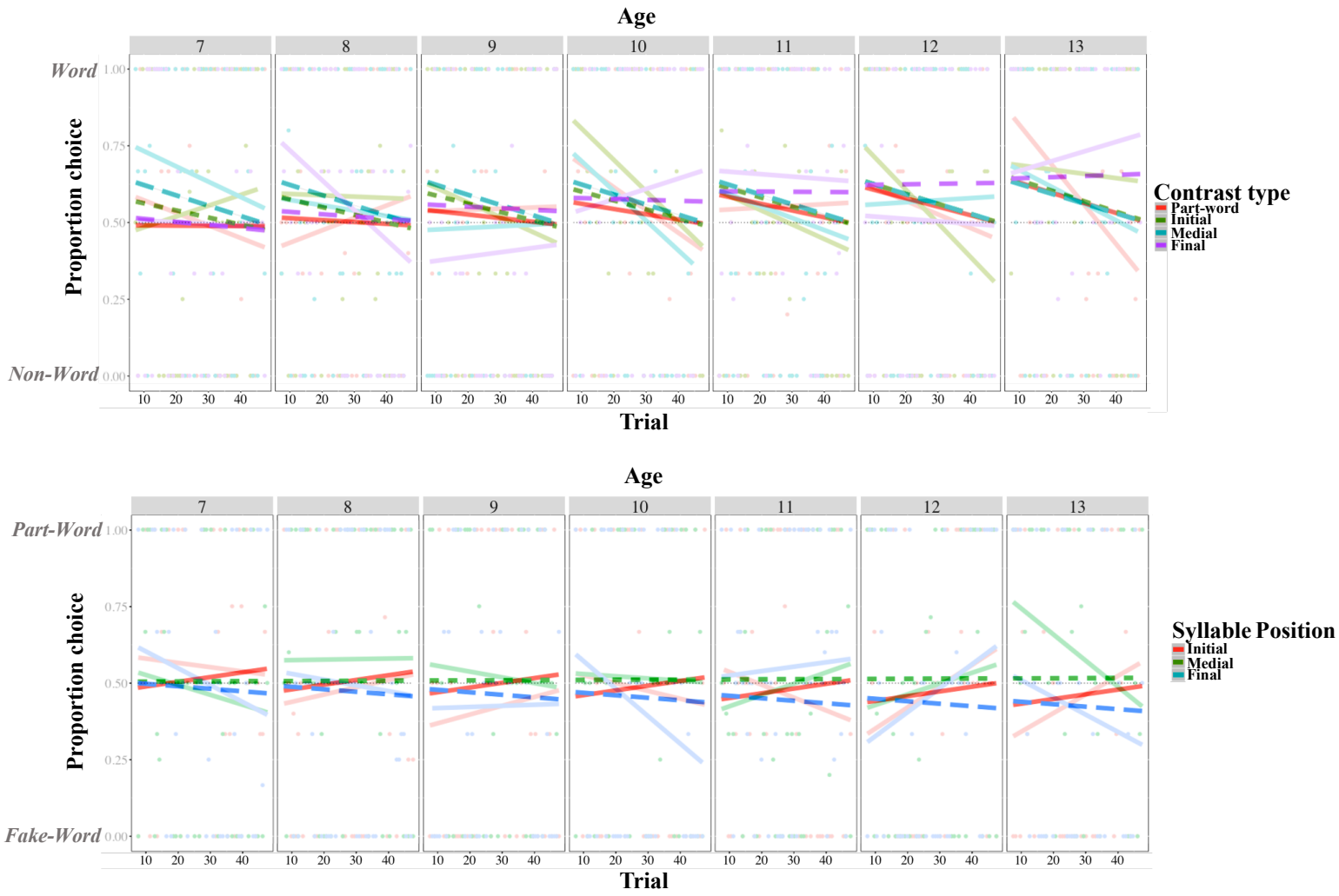

Figure 4.7 Predicted fits by contrast type, age, and trial In the top panel, words versus part-word and word versus fake-word (initial, medial, and final) are plotted; the bottom panel shows part-word versus fake-word (initial, medial, and final) trials. Transparent dots reflect the actual mean proportion choice scores by trial; transparent lines reflect the linear best fit of the data by age, contrast type, and trial. The bold lines reflect the model predicted fit lines by age, contrast type, and trial.

When forced to choose between part-words and fake-words, if children are associating syllables with particular positions within a word, they would be more likely to choose fake-words (or fake-words of some syllable position manipulations) over part-words. I also predicted that this pattern of performance would decrease (i.e., leading to chance performance) as children age and develop greater attentional control. On the other hand, if children are using TPs to make their 2AFC decisions, they should opt for part-words – an effect that would get stronger as the children age. As a group, children were at chance across initial and medial-syllable manipulation trials. There was a small but significant pattern (d = -.29), however, for the children to choose final-syllable manipulated fake-words over their part-word counterparts. There was no evidence from the regression models that this pattern changed as a result of learning from trials, or across development. As can be seen in the bottom panel of Figure 4.5, however, there is some suggestion from the model-predicted fit lines that, while all ages consistently fail to choose between part-words and medial-syllable fake-words, older children initially choose both initial and final syllable fake-words over part-words, but over the course of the experiment increasingly choose part-words over initial syllable fake-words, and final-syllable fake-words over part-words.

Finally, the correlations between conditions can also speak to the two hypotheses. Unlike in the adult experiments (except for the non-English language condition), the children's performance on word versus part-word and part-word versus fake-word tasks was uncorrelated. We might expect that this is due to their relatively poor performance on the word versus part-word trials – i.e., perhaps the children simply failed to learn enough about the statistical structure of the familiarization stream across the board. This is countered, however, by the fact that the children were able to distinguish words from novel combinations (fake-words) – and that

learners who scored higher on these trials were also better able to choose words over part-words. Finally – children's performance on word versus fake-word trials was positively correlated with performance on part-word versus fake-word trials. This was not found in any of the adult experiments, and suggests that children are relying on adjacent TPs.

It is worth noting that across 4 of 5 experiments (the current study and all adult studies except in the non-native condition), participants were more likely to choose fake-words of initial and final-syllable manipulations over part-words in comparison to the medial-syllable manipulated fake-words. This effect is small (and hence non-significant in nearly each study individually), however, we might interpret this pattern as follows: learners (both adults and children) rely on both TP and position-based information to make their decisions about word identity, and both sources of information are weighted in memory. When both syllable transitions are 0, the word candidate incurs too many violations, and is more likely to be rejected. With a single 0 syllable transition, and correct syllable positions, however, the relatively higher TP – but positionally illicit – candidate incurs more violations.

Though these results seem to suggest developmental differences in statistical learning (i.e., that children are worse statistical learners than adults), there are unrelated sample differences that may have impacted performance. Unlike the adult sample, the children were not all native speakers, nor did they all live in Canada. While the adult population was widely diverse in language backgrounds and degree of multilingualism, they were screened to only include learners who had begun learning English before the age of three. In addition, the entire sample currently resided in Vancouver, though their permanent homes might have been elsewhere. Their experience of the sounds in the "native" language condition was therefore one of native speakers who are surrounded in their daily lives by the sounds encountered in the

familiarization stream. The children's sample, which represents much more diversity in native language background, may thus be confounded by a relative lack of familiarity with the speech sounds that is independent of development. In the following section, I examine the impact of these factors on child performance, in addition to the individual difference predictors of bilingualism and musical ability, for further comparison with the adult data.

4.2 Secondary Analysis: Individual differences

As mentioned above, this sample differs from the adult sample in a few, potentially significant, ways. The adult sample was coded for current versus early bilingualism, using both age of acquisition and current self-rated proficiency. In the child sample, however, current proficiency (described below) is highly correlated with the age of acquisition ($r(68) = .51$, $p < .0001$). I have therefore combined the two items into a composite bilingualism measure as follows:

(1) **Bilingualism** = (the negative of) z-score age of acquisition of his/her second language + z-score proportion of time the child uses the second language.[30]

A similar metric was designed for musical skill:

(2) **Music**: (the negative of) z-scored age of onset of musical training + z-scored parent-reported proficiency on a musical instrument.[31]

---

[30]Some parents wrote that their child spoke Lang 1, e.g., 100% of the time (at home) and Lang 2 100% of the time (at school). In these cases, the percentages have been divided by the number of languages reported. So, for instance, in the given example, the data have been recoded as 50% usage of Lang 1 and 50% usage of Lang 2.

[31]When multiple instruments and proficiencies were reported, the highest score was used.

As outlined in Chapter 3, bilingualism is associated with enhanced executive function skill (e.g., Bialystok & Martin, 2004; Prior & MacWhinney, 2010); this led me to predict that children who score higher on the bilingualism scale should outperform those who are lower on the scale in all trials pitting words against non-words of any type. They should also show a reduced position-based effects. Though musical skill did not appear to impact adult SL, I have maintained the same factor structure here for the sake of comparison across the two samples.

Finally, I examined two additional factors that were not necessary in the adult sample. First, this sample was not limited to native speakers of English. I therefore created a binomial categorical variable to reflect whether the child had native-speaker knowledge. Secondly, while some of the adult learners permanently live in other locations, they were all residing in Canada at the time they participated. This is not the case for the children; rather, 18 (of 69 total) were based in Europe or Asia. These two factors were thus operationalized in the following way:

(3) **Native**: coded as '1' if the child began learning English before the age of 3

(4) **Live**: coded as '1' if the child currently lived in Canada.

If native language background is the factor driving the lower performance in the child sample, we should see better performance by native English speakers. If home country/language environment impacts SL, I predict that participants who live in Canada will outperform their international peers.

	1: Native/Canada	0: Non-native/elsewhere	Missing data
Native	56	12	1
Live	47	18	4

Table 4.6 Demographics. Number of children who were identified as being Native (or non-Native) speakers of English and who, at the time of testing, lived in North America (or elsewhere), and numbers with missing data.

The analysis has been separated in to two sections: (1) correlations between all individual difference predictors and performance, and (2) modeling of the effects of individual difference predictors and age on performance.

4.2.1 Correlations

There does not appear to be a relationship between overall performance on word versus non-word trials and whether or not the child spoke English as a native language ($t(15.5) = 0.24$, $p = .8$), nor whether he/she lives in Canada ($t(34.5) = 0.14$, $p = .9$). There are small, non-significant correlations with the bilingualism measure ($r(67) = .20$, $p = .09$) and musical proficiency ($r(59) = .18$, $p = .18$). These latter two patterns are reflected in Panels A and B of Figure 4.8. Proficiency scores are plotted as z-score values based on the calculation that was outlined in Section 4.2. Negative numbers indicate very low proficiency (i.e., a -3 for bilingualism indicates that the child was monolingual; -3 means that the child had no musical training), while higher scores reflect greater proficiency (i.e., for Bilingualism, children with higher scores had been exposed to languages other than English from earlier ages, and spoke those languages a higher percentage of the time).

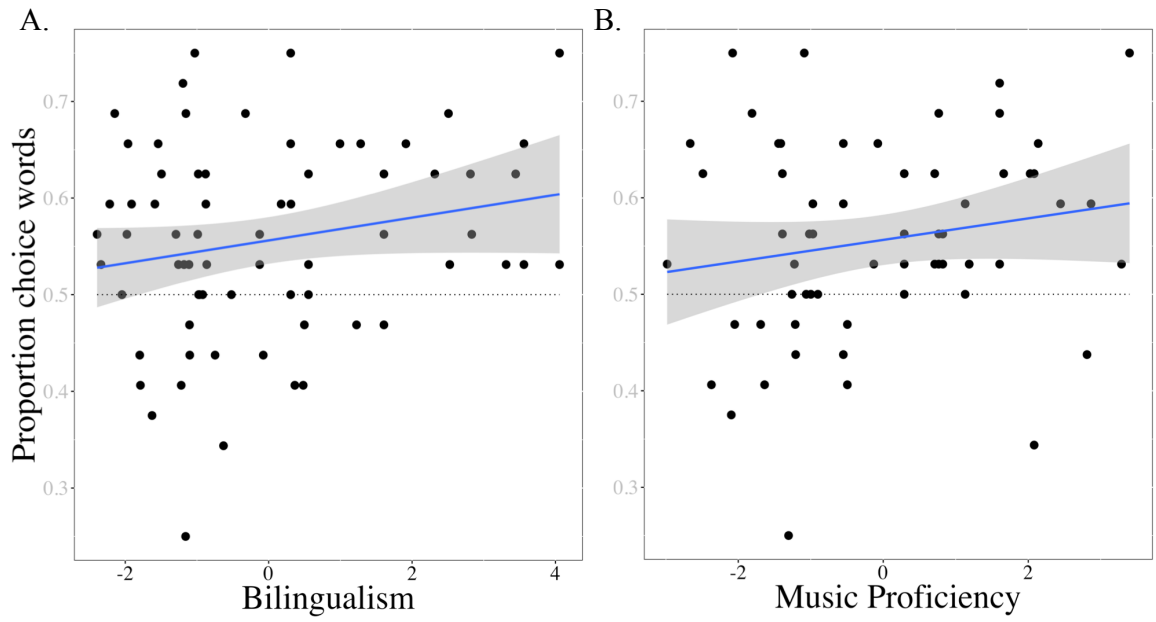

Figure 4.8 Relationship between proportion choice words and bilingualism (Panel A) and musical proficiency (Panel B) In both metrics, lower scores reflect less proficiency (i.e., monolingual or very little second language proficiency; or no or very little musical training/skill). Zero reflects the group mean.

4.2.1.1 Mixed effects modeling

Logistic mixed effects regression models, run separately by trial type (i.e., word versus part-word, word versus fake-word, part-word versus fake-word), were fitted to the data to predict performance by syllable position, bilingualism, musical proficiency, and age, controlling for trial.

4.2.1.1.1 Words versus PW

There is no effect of any factor on word versus part-word trials (see Table 4.7).

Model structure:
Choice ~ Bilingualism * Age + Music proficiency *
Age + Syllable position + Trial +
(Trial | Subject)

	OR	*CI*	*p*
Fixed Effects			
(Intercept)	1.13	0.92 – 1.39	.235
Lingualism	1.09	0.97 – 1.23	.127
Age	1.03	0.83 – 1.27	.817
Music	1.07	0.87 – 1.32	.496
Trial	0.99	0.98 – 1.00	.181
Lingualism : Age	1.00	0.89 – 1.11	.934
Age : Music	0.99	0.80 – 1.22	.903
Random Effects			
$\tau_{00,\ Subject}$	0.050		
ρ_{01}	-1.000		
$N_{Subject}$	65		
$ICC_{Subject}$	0.015		
Observations	520		
Deviance	683.460		

Table 4.7 Generalized model predicting choice words over part-words by Lingualism, Age, Music, and Trial.

4.2.1.1.2 Words versus FW

Trial is significant across all three reference levels for syllable position (OR = .99, *p* = .004). There is no other significant effect (see Table 4.8).

Model structure:
Choice ~ Bilingualism * Age + Music proficiency * Age + Syllable position + Trial + (Trial | Subject)

	Model 1 Reference level: Initial			**Model 2** Reference level: Medial			**Model 3** Reference level: Final		
	OR	***CI***	***p***	***OR***	***CI***	***p***	***OR***	***CI***	***p***
Fixed Effects									
(Intercept)	1.22	1.01 – 1.47	**.034**	1.30	1.08 – 1.57	**.006**	1.33	1.11 – 1.61	**.002**
Initial				0.94	0.73 – 1.20	.626	0.92	0.71 – 1.17	.483
Medial	1.06	0.83 – 1.36	.626				0.97	0.76 – 1.25	.831
Final	1.09	0.85 – 1.40	.483	1.03	0.80 – 1.32	.831			
Bilingualism	1.03	0.96 – 1.10	.441	1.03	0.96 – 1.10	.441	1.03	0.96 – 1.10	.441
Age	1.05	0.93 – 1.19	.456	1.05	0.93 – 1.19	.456	1.05	0.93 – 1.19	.456
Music	1.08	0.95 – 1.22	.224	1.08	0.95 – 1.22	.224	1.08	0.95 – 1.22	.224
Trial	0.99	0.98 – 1.00	**.004**	0.99	0.98 – 1.00	**.004**	0.99	0.98 – 1.00	**.004**
Bilingualism : Age	1.02	0.95 – 1.09	.654	1.02	0.95 – 1.09	.654	1.02	0.95 – 1.09	.654
Age : Music	1.01	0.89 – 1.14	.911	1.01	0.89 – 1.14	.911	1.01	0.89 – 1.14	.912
Random Effects									
$\tau_{00,\ Subject}$					0.025				
ρ_{01}					1.000				
$N_{Subject}$					65				
$ICC_{Subject}$					0.008				
Observations					1560				
Deviance					2101.510				

Table 4.8 Generalized model predicting choice words over fake-words by Bilingualism, Age, Music, and Trial.

4.2.1.1.3 PW versus FW

Age (OR = 0.89, p = .05) is the only significant effect – suggesting that as children age, they are more likely to choose fake-words over part-words. Interestingly, however, the interaction between age and bilingualism is in the predicted direction, though it fails to reach the .05 significance cut-off (OR = 1.06, p = .07). In other words, children who are higher on the bilingualism scale are increasingly likely to choose part-words as they get older (Table 4.9).

Model structure:
Choice ~ Bilingualism * Age + Music proficiency * Age + Syllable position + Trial + (Trial | Subject)

	Model 1 Reference level: Initial			**Model 2** Reference level: Medial			**Model 3** Reference level: Final		
	OR	*CI*	*p*	*OR*	*CI*	*p*	*OR*	*CI*	*p*
Fixed Effects									
(Intercept)	0.96	0.81 – 1.15	.692	1.01	0.85 – 1.21	.898	0.82	0.69 – 0.99	**.036**
Initial				0.95	0.75 – 1.22	.701	1.17	0.92 – 1.49	.211
Medial	1.05	0.82 – 1.34	.700				1.23	0.96 – 1.57	.103
Final	0.86	0.67 – 1.09	.211	0.82	0.64 – 1.04	.103			
Bilingualism	1.04	0.97 – 1.10	.250	1.04	0.97 – 1.10	.250	1.04	0.97 – 1.10	.250
Age	0.89	0.79 – 1.00	**.052**	0.89	0.79 – 1.00	**.052**	0.89	0.79 – 1.00	**.052**
Music	1.04	0.93 – 1.16	.530	1.04	0.93 – 1.16	.530	1.04	0.93 – 1.16	.530
Trial	1.00	0.99 – 1.01	.973	1.00	0.99 – 1.01	.973	1.00	0.99 – 1.01	.973
Bilingualism:Age	1.06	1.00 – 1.13	.071	1.06	1.00 – 1.13	.071	1.06	1.00 – 1.13	.071
Age : Music	0.97	0.86 – 1.09	.612	0.97	0.86 – 1.09	.613	0.97	0.86 – 1.09	.612
Random Effects									
$\tau_{00,\ Subject}$					0.002				
ρ_{01}					-1.000				
$N_{Subject}$					65				
$ICC_{Subject}$					0.001				
Observations					1560				
Deviance					2141				

Table 4.9 Generalized model predicting choice part-words over fake-words by Bilingualism, Age, Music, and Trial.

4.2.2 Individual differences: Discussion

Younger learners are characterized both by less stable phonological representations and less mature executive function skills. Their statistical learning performance seems to reflect the impact of both these dissociable skillsets: they learn less about the language, similar to the adult learners of non-native speech sounds, but they also show position-based effects (which were clearest in the attentional-resource taxing adult condition) that interact with correlates of executive function. Specifically, children who are higher on the bilingualism scale should be higher on executive function measures; therefore, I predicted that they would learn better, and show *less* knowledge of the syllable positions. The evidence does not support the hypothesis that bilingualism facilitated learning from the stream, as I found no effect for bilingualism (or age or music) across word versus non-word trial types. The analysis did reveal, however, limited support for the hypothesis that bilingualism would decrease position-based effects: bilingual children were increasingly likely to choose part-words over fake-words across age. At the same time, age was predicted to lead to increasing proportion choice part-words over fake-words; on the contrary, it was associated with increasing proportion choice fake-words.

This analysis also sought to determine whether the differences in performance between children and adults (e.g., the children's failure to discriminate words from part-words versus adult success on the same trials) was driven by sample differences such as native language and community backgrounds. This does not appear to be the case – there is no evidence for difference in performance across the native and non-native learners in the child data, nor those who live locally versus internationally.

4.2.3 Conclusion

The experiment described in this chapter used developmental change as a lens on the impact of prior knowledge and attentional resources on statistical learning. Contrary to previous results in the literature, I found evidence for developmental differences in the outcome of SL. The major, unexpected point of difference was the children's failure to discriminate words from part-words at any stage in development. This finding, in tandem with the ordered relationship of performance on word versus non-word trials (i.e., 0.0 TPs were easier to reject than 0.33 TPs), and the positive correlation between word versus fake-word and part-word versus fake-word trial types, suggests that younger children are using primarily adjacent TPs in segmenting the speech stream.

Given the results of the adult studies, where semi-familiar sounds and increased attentional demands caused greater reliance on position-based information in the stream, this is a somewhat surprising finding. It is also, however, not the only effect: children also showed a propensity to choose final-syllable fake-words over part-words (a position-based decision), and were simultaneously more resistant to interference of final-syllable fake-words over the course of the experiment. Why would this happen? One possible explanation might simply be fatigue; the experiment was rather dull (as several children made quite clear). The fact that the pattern is asymmetrical, however (i.e., discrimination of final-syllable fake-words from words remains stable) makes this explanation less plausible. Instead, the (older) learners' consistent success on final-syllable fake-words may suggests that they have established as a chunk in memory the transition between the medial and final syllables of words. When they hear a trisyllabic item that breaks that TP, they recognize that it is wrong (or vice versa – they recognize the transition that is right). If this is the correct interpretation, it would also predict that children should perform

similarly well on the medial syllable manipulated fake-words, which also break the transition between medial and final syllables. They do not, however – like initial syllable fake-words, children successfully reject them at the beginning of the task, but become increasingly confused by them as the task goes on.

To account for this discrepancy, I consider two proposals. First: it may be that when learners conclude the familiarization phase, they have learned sufficiently about the stream to reject any zero-TP sequence. As they hear multiple repetitions of test items, some of which create novel associations between syllables (i.e., fake-words), children begin to learn these new items. This learning is not evenly distributed across syllable positions, however. Rather, children learn the novel relationships from the left edge before the right (i.e., the novel transitions between initial and medial syllables). They do not (for some reason) learn as quickly about novel associations between medial and final syllables. The second option is that children become increasingly aware of the non-adjacent TP – i.e., the relationship between the first and final syllables. In medial syllable manipulations, this relationship is maintained; therefore, despite the fact that the adjacent TPs are zero (and so should be the easiest to reject from a TP perspective), the non-adjacent relationship signals word-hood.

The data presented here can speak to these two accounts. If the children are learning the novel initial- and medial-syllable fake-word combinations, and not learning the final-syllable fake-word combinations, we should expect to see greater relative proportion choice of initial- and medial-syllable fake-words over trial, and comparatively lower proportion choice final-syllable fake-words. As already discussed, however, this is not the pattern observed; rather, the children consistently fail to choose between medial-syllable fake-words and part-words, increasingly choose part-words over initial-syllable fake-words, and increasingly opt for final-

syllable fake-words over part-words as the experiment goes on. The non-adjacency account makes a similarly unmet prediction: if learners are extracting the non-adjacent dependency, and using that to make word judgments (making them increasingly confused during word versus fake-word trials), they should similarly choose medial-syllable fake-words in the fake-word versus part-word contrast. But, mean performance on these items is in the direction of a part-word preference, and does not change across age or trial. The source of the asymmetry in encoding across syllable positions thus remains a puzzle.

To sum: in comparison to the adult studies, the children appear to rely more heavily on adjacent TP strength in making decisions on a 2AFC task. At the same time, they evidence some patterns that are inconsistent with TP-tracking alone, and there is weak evidence to support the hypothesis that bilingual children – who are argued to have more advanced executive function skills (Carlson & Meltzoff, 2008; Bialystok & Viswanathan, 2009; Bialystok, 2011) – are less susceptible to the position-based effects.

Chapter 5: General discussion

Over the last two decades, the field of language acquisition has been transformed by evidence that infants, even within hours of birth, are capable of detecting statistically defined patterns in the sensory information that surrounds them, and of storing some representation of these patterns (at least temporarily) in memory (Teinonen, Feldman, Näätänen, Alku, & Huotilainen, 2009; Bulf, Johnson, & Valenza, 2011; Kudo, Nonaka, Mizuno, Mizuno, & Okanoya, 2011). At the same time, an extensive body of work has revealed functional constraints on infant perception, constraints that are thought to fundamentally shape and streamline the information infants and children attend to at different points in development (e.g., Cooper & Aslin, 1990; Hudson Kam & Newport, 2005, 2009; Kuhl, 2007; Yoshida & Smith, 2008; Kidd, Piantadosi, & Aslin, 2012, 2014). These two streams of research have led to an increasing number of proposals that language, an aspect of human cognition previously thought to depend on specific innate knowledge, may in fact emerge from the interplay of low-level learning mechanisms and perceptual or cognitive constraints (Chater & Christiansen, 2010; Newport, 2016; Aslin, 2017).

The search for evidence that statistical learning (SL) is one such low-level learning mechanism in language acquisition has indeed met with great success. For example, learners across the developmental span have been shown to be able to track distributions of sounds, and to impute categories that relate to these distributions –at the level of allophones (Noguchi & Hudson Kam, 2018), phonemes (Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008; Yoshida, Pons, Maye, & Werker, 2010; Olejarczuk & Kapatsinski, 2016), word classes (Endress & Mehler, 2009b), and phrasal units (Thompson & Newport, 2005). In a related (possibly

distinct; see Thiessen, Kronstein, & Hufnagle, 2013, for discussion) form of SL, learners have been shown to extract independent chunks that are based on statistical relationships between sub-units across development (e.g., Saffran, Aslin, & Newport, 1996, and many more), as well as the ordering relation between chunks (Saffran & Wilson, 2003; Finn, Lee, Kraus, & Hudson Kam, 2014). Despite the apparent ubiquity and power of this learning mechanism, however, there is much that remains a mystery. For instance – are the fields of implicit learning and SL researching the same phenomenon, overlapping phenomena, or distinct and unrelated learning processes (Perruchet & Pacton, 2006; Christiansen, in press)? What are the psychobiologically plausible models of SL? Should we conceive of the process as one of tracking transitional probabilities (Aslin, Saffran, & Newport, 1998), or one of encoding semi-random chunks that are reinforced (or not) in memory (Perruchet & Vinter, 1998; Thiessen, 2017)? Is SL a single, domain-general mechanism, or are there independent, domain-specific mechanisms that operate along similar, but distinct principles (Siegelman, Bogaerts, Christiansen, & Frost, 2017)? These mysteries (among others) constrain our ability to determine the extent to which SL contributes to language acquisition, or to evaluate the broader question regarding the nature of the relationship of language acquisition to general perceptual and cognitive constraints.

In this dissertation, I have sought to better understand the mechanism(s) underlying SL by examining the nature of the representations extracted from an SL task. Specifically, I asked two questions: (1) what do learners *learn* from a SL task? And (2) does a change in the available, underlying representations lead to different (e.g., more/less abstract) learning outcomes? In the following paragraphs, I will outline the particular form of SL I investigated, and review the experimental results. I will then discuss what I believe we can conclude from this work, and where future work should focus to bring more light to these questions.

5.1 SL: segmenting words from continuous speech

The word-segmentation SL literature emerged as an answer to the question: how do learners extract 'word' candidates from continuous streams of speech? As such, the literature largely presumes that the outcome of SL involves some type of coherent chunk from the auditory stream. There is research to support this hypothesis; for example, Graf Estes, Evans, Alibali and Saffran (2007) found that 17-month-old infants were better able to learn the association between high TP units they had been exposed to and novel objects than they were low TP units (see also Hay, Pelucchi, Graf Estes, & Saffran, 2011). This kind of result might be interpreted in a number of ways, however. It is possible that infants have stored all TPs, and that chunks associated by higher TPs are more readily available for association with semantics. Under this account, the infants have not 'extracted' independent chunks in advance of the subsequent exposure to semantics. This interpretation would match with studies that find learners continue to entrain to adjacent TPs (including low TPs), rather than extract higher-order structures from the stream (Peña Bonatti, Nespor, & Mehler, 2002; Endress & Mehler, 2009).

On the other hand, however, there is evidence that learners do impute *boundaries* between elements according to their TP structure. For example, recent work has used the oscillatory electrical signals produced by neuronal activity to indicate successful statistical learning. In these studies, learners begin to show a spike in oscillatory activity at exactly the frequency with which the higher order structure occurs. In other words, if a learner is exposed to syllables every 300 milliseconds, and these syllables are arranged in consistent trisyllabic chunks, the learner will initially exhibit a peak in neural activity at 3.3 Hz (the rate of syllable presentation), that is quickly followed by a peak in activity at 1.1 Hz (the rate of trisyllabic 'word' presentation) (Batterink & Paller, 2017; see also Kabdebon, Peña, Buiatti, & Dehaene-

Lambertz, 2015; Ding, Melloni, Zhang, Tian, & Poeppel, 2016). It would appear then, that some aspect of the neural system has detected the most predictable structure – which in turn suggests the imputation of some kind of *boundary* between those chunks.

Work that has examined the event-related-potentials that develop in response to particular syllables within continuous streams of sound offers additional support for the idea that learners impute bounded, independent chunks. Specifically, these studies find that after brief exposure to the structured stream, learners begin to exhibit larger magnitude N100 (Sanders, Newport, & Neville, 2002; Sanders & Neville, 2003; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009) and N400 (Abla, Katahira, & Okanoya, 2008; Mandikal Vasuki, Sharma, Ibrahim, & Arciuli, 2017) responses to the syllables that belong at the left-edge of a high-TP sequence. The N100 may simply reflect the lower predictability of the first syllable of a trisyllabic sequence in comparison to the second or third syllables, which is consistent either with learning that yields knowledge of TPs or learning that yields independent chunks. The N400, however, is generally thought to reflect the process of lexical retrieval. In the context of speech segmentation, then, it may indicate that sequences with lower transitional probability or co-occurrence frequency are treated fundamentally differently than those with higher TP or co-occurrence – in other words, providing a mechanism for hypothesizing a boundary. Finally, as was discussed in Chapter 1, work has shown that learners under some circumstances show reduced/inhibited memory for the components *within* a high-TP defined chunk (Giroux & Rey, 2009; Fiser & Aslin 2005). It is difficult to imagine a scenario in which this would occur without some representation of the larger chunk itself.

In this dissertation, I probed this question further. I asked (1) what is the nature of the output of SL? Do learners extract word-like chunks from a SL experience, or do they acquire

relative TP strength between syllables? I also asked (2) whether input representations would promote different trajectories of learning, which might in turn elucidate (1).

5.2 Summary of findings

To answer these questions, I proposed three specific hypotheses: (i) learners' prior knowledge would impact the accessibility of units to SL, and thereby modify the process of learning; (ii) SL involves more than veridical TP-tracking; and (iii) the interaction of prior knowledge and the underlying mechanisms of SL would relate to differences in learning outcomes across development. In Chapters 2 - 4 I tested these hypotheses by exposing adult and child learners to a continuous stream of sounds and examining the outcome of their learning, and the individual difference predictors that might contribute to learning. In particular, I asked whether the outcome representations were based solely on the strength of the underlying TP structure (termed the *TP-encoding* hypothesis), or if there was evidence for chunk-like knowledge of the trisyllabic structure – the relative position of syllables (termed the *Position-encoding* hypothesis). I probed whether prior knowledge impacts SL by (1) manipulating the degree of familiarity adult learners had with the auditory stimuli, and (2) testing children, whose phonological representations are still developing (e.g., Rigler et al., 2015).

5.2.1 Chapter 2 summary

In Chapter 2, four experiments saw adult learners exposed to a continuous stream of linguistic sounds that had been arranged in a TP-defined structure. At test, participants were asked to choose between items in a 2-alternative forced choice (2AFC) task. These items were designed to probe learners' emergent representations for evidence of greater reliance on the

embedded TP-structure versus the positional nature of syllables within a high-TP chunk. The first experiment sought to establish native-English speaking learners' baseline performance on learning from a continuous, nonsense stream of English sounds (Experiment 1). In Experiments 2 and 3, native-English listeners were exposed to a range of unfamiliar sounds that I termed semi-English (Experiment 2) and non-English (Experiment 3). Finally, in Experiment 4 I shifted participants' ability to perceive the stream by dividing attention between the stream of familiar English sounds and an unrelated, silent video cartoon (Experiment 4).

In all four experiments, learners successfully discriminated high-TP sequences (words) from low, but non-zero, TP sequences (part-words) – one of the standard tests that has been historically used to determine successful segmentation of a speech stream by SL (e.g., Saffran et al., 1996). In Experiment 1 (English-language), learners successfully discriminated words from part-words, and words from fake-words (trisyllabic sequences with at least one 0.0 TP, but positional fidelity of syllables), but were unable to consistently choose either part-words or fake-words when these items were pitted against each other. The latter outcome was not predicted: I anticipated that learners would choose part-words if learning involves veridical tracking of TPs, or fake-words if learning involves chunking of word-like units. There was correlational evidence to support the position-encoding hypothesis: participants who had higher scores on the word versus part-word trials were more likely to choose fake-words over part-words ($r(40) = -0.36$, $p = .02$), a relationship that was particularly strong for final-syllable fake-words ($r(40) = -0.40$, $p < .001$). In other words – better segmentation was associated with more chunk-like behavior at test.

The learning outcomes of Experiment 2 (Semi-English Language) closely paralleled the outcomes of Experiment 1 (English Language), though there were some small differences that lent additional support to the position-encoding hypothesis. Namely, participants chose final-

syllable fake-words over part-words significantly above chance – suggesting that, despite the 0.0 TP between the medial and final syllable, participants preferred items with positional fidelity over the 0.33 TP sequence that crossed a word boundary (d = -.32). Though not significantly different from chance, performance on initial and medial-syllable fake-words patterned in the same direction.

Learners exposed to non-English sounds (Experiment 3) showed significantly worse learning overall. Comparison across the three different language conditions (see Chapter 3) revealed that performance decreased as sounds became less familiar: participants in the non-English condition were significantly worse at choosing words over all non-word types in comparison to participants in the English-language condition ($F(2, 104) = 4.15$, $p = .02$; Tukey's HSD mean difference = -7.14, adjusted $p = .02$). Performance in the semi-English condition was intermediate – it did not statistically differ from either the English-language (mean difference = -2.19, adjusted $p = .66$) or non-English conditions (mean difference = 4.94, adjusted $p = .12$).

Taken together, the results of the three experiments provided evidence that was consistent with both TP-tracking and position encoding. Learners in all three language conditions performed numerically better on word versus zero-TP fake-words (i.e., medial-syllable fake-words) than they did fake-words with one 1.0 TP (i.e., initial- and final-syllable fake-words). High performance on medial-syllable fake words is predicted by the TP-encoding hypothesis (as these sequences involve 0.0 TPs only, and so should be the easiest to detect and reject). The TP-encoding hypothesis further predicts (1) that learners' performance on all of the word versus fake-word contrasts should exceed that of word versus part-word, and (2) learners should choose part-words over fake-words; neither of these predictions, however, were borne out by the data.

The range of evidence to support position-based learning in Experiments 1-3, however, was not particularly conclusive. The patterns across the three sets diverged in unexpected ways (e.g., different syllable position correlations), and the effect sizes of contrasts designed to pit position-based versus TP-encoding against each other were very small (Cohen's d = .20 to .32). Moreover, the condition designed to create the greatest perceptual difficulty (the most acoustically/phonetically distant from English-language phonology, i.e., the non-English language) appeared to present too high a burden – learners simply did not extract enough reliable information from the brief auditory presentation. I therefore extended the hypothesis and proposed that a different means of increasing perceptual load - taxing the attentional resources available to the learners - might yield the predicted position-based effects. In the fourth experiment, I therefore exposed learners to an unrelated, silent cartoon at the same time that they attended to the auditory stream. At test, learners showed more distinct patterns of position-based learning. Learners' performance on word versus fake-word trials varied by syllable manipulation: while medial syllable fake-words were easy to reject (d = 1.08), initial (d = .55) and final (d = .37) syllable fake-words were significantly more confusing. Moreover, participants were more likely to choose initial- and final-syllable fake-words than the trisyllabic sequences they had actually encountered in the stream.

Why might this be the case? It could be that learners simply didn't encode the stream with as high fidelity as their counterparts in Experiment 1 due to distraction. If this is so, we might find that more acoustically similar fake-words are more confusable than less acoustically similar fake-words. Though the evidence is not determinative, the data pattern in the right direction. Thus, we are left with two possibilities: (1) it may be that learners' representations are defined by TPs, but that sequences that are highly similar to a high-TP chunk resonate with the

memory trace and so are better options than (veridical) low-TP chunks; (2) it may be that syllables are encoded with particular positional information in mind, in addition to the adjacent TPs/relationships between syllables. I argue in the conclusion of Chapter 2 that since learners' preferences for fake-words over part-words is asymmetric (i.e., they prefer initial- and final-syllable fake-words over part-words, but prefer part-words over middle-syllable fake-words), the evidence is more consistent with a mechanism of learning that encodes both position of and statistical relationships between syllables.

5.2.2 Chapter 3 summary

In Chapter 3, I examined the same data for evidence that different underlying representations and/or different learning capacities would impact SL performance as a whole. I proposed that specific underlying representations (via language experience), auditory skill (indexed through musical training or bi/multilingualism), or the enhanced cognitive skill related to multilingualism would improve upon SL performance. I also proposed that age would negatively impact performance. I found multilingualism did facilitate SL, and argued that this effect is due to a global cognitive advantage in addition to any benefit derived from specific language experience. Specific language experience was also found to play a role, as evidenced by the fact that participants performed worse on the non-native sounds as compared to native and semi-English sounds. There was no effect of musical experience, and – while age significantly impacted performance (in both positive and negative ways, depending on condition) – I argue that the sample is too unbalanced to make strong conclusions regarding its potential impact (or lack thereof).

5.2.3 Chapter 4 summary

Finally, in Chapter 4 I merged the approaches of Chapters 2 and 3 to examine SL across development. I discuss the results of a study in which I familiarized 7- to 13-year-old children to a 2-minute stream of English syllables, arranged in such a way that four TP-defined trisyllabic words continuously repeated, with no prosodic or acoustic cue to the word boundaries. The children then completed an identical 2AFC task as the adults in the English-language experiment (Experiment 1, Chapter 2), in which the learners' representations are probed for evidence of TP-strength versus positional knowledge. I hypothesized that younger learners would demonstrate less proficient learning overall but that they would improve with age. I also hypothesized that children would show a higher propensity for positional knowledge (i.e., greater proportion choice fake-words) at younger ages, as a result of their lower levels of attentional control (a hypothesis that developed out of the experiments presented in Chapter 2).

The results indicate that children's emergent representations from the familiarization period were less stable than those of adults, but that they did improve on the task as they aged. Children failed to discriminate words from part-words – one of the standard tests that has been used to evaluate SL performance in infants, children, and adults – but did succeed on trials that pitted non-word foils with at least one 0.0 TP against a trisyllabic word. The overall order of their performance scores largely accords with the TP-encoding hypothesis, with the exception of one contrast: children (as a group) chose final-syllable fake-words over part-words.

The two-way interaction of trial and age was not significant in the mixed effects models fit to this data; however, the pattern of results across all word versus non-word trials, and the predicted fits created by the models suggest that this failure on the word versus part-word contrast may derive from older children's learning (or un-learning) over the course of the

experiment. In other words, older children may successfully discriminate words from part-words at the beginning of the experiment, but gradually lose this ability as they complete additional test trials. I argue in the chapter that this pattern may not be simply due to decreased attention or fatigue, but rather is evidence of learning from the test trials themselves. This is because change in performance across trials is asymmetric: children have an increasingly difficult time distinguishing words from initial- and medial-syllable fake-words, but are consistently capable of distinguishing words from final-syllable fake-words. Though not significant, there is an echo of this pattern in the part-word versus fake-word trials: older children appear more likely to choose initial- and final-syllable fake-words over part-words, but are driven towards part-words over initial-syllable fake-words, and final-syllable fake-words over part-words as the trials go on.

Finally, I additionally hypothesized that age and bilingualism – a purported contributor to superior executive-function skills – would lead the children to perform more similarly to adults (i.e., a greater proportion choice of higher-TP items over lower-TP items with minimal interference from position-based encoding), whereas younger and/or monolingual children would perform more similarly to adults whose attention was divided by an unrelated visual stream (i.e., taxing their executive-function skills; stronger preferences for low TP but positionally licit foils). This predicts that age and bilingualism should correlate with greater proportion choice part-words in the part-word versus fake-word trials. This met with mixed results: bilingualism was associated numerically (but non-significantly) with greater proportion choice part-words (in line with the prediction), but age was associated with fake-word choice (contra the prediction). It is unclear how to interpret these results at this stage. Operationalizing factors such as degree of bilingualism or musical proficiency is not straightforward (Byers-Heinlein, 2015); furthermore,

performance was quite low overall, which may mean that there is insufficient variability to yield reliable evidence for individual difference factors.

5.3 Discussion

Over the preceding three chapters, I have tested whether: (i) learners' prior knowledge impacts SL, (ii) SL involves more than veridical TP-tracking, and (iii) the interaction of prior knowledge and the underlying mechanisms of SL would relate to differences in learning outcomes across development. I determined that (i) differences in underlying representations can impact SL, as indicated by the relatively poorer performance on tasks that involved semi-native and non-native sounds in comparison to native sounds. Across both children and adults, learners appear to attend to syllable positions within a trisyllabic sequence in addition to TP structure, which I interpret as evidence in support of hypothesis (ii) – that SL is characterized by a mechanism beyond veridical TP-tracking. Contrary to my prediction, however, smaller/less stable representations did not *enhance* these positional effects; rather, the performance of adults exposed to unfamiliar sounds and children was primarily characterized by much lower levels of learning overall. On the other hand, executive function (a factor known to change across development) was found to play the predicted moderating role.

Do the results that suggest position-encoding align with previous reports on asymmetrical encoding across syllable positions (Saffran, Newport, & Aslin, 1996; Saffran, Johnson, Newport, & Aslin, 1999)? Saffran and colleagues found that learners were better able to reject non-words of the structure ABX than the structure XBC, where ABC represents the high-TP, trisyllabic word, and X represents a random syllable. The trial types that most closely parallel these structures in this dissertation are the final-syllable (ABX) and initial syllable (XBC) fake words.

There was no evidence in Chapter 2 that for the adult learners performance was better on trials pitting final-syllable fake-words (i.e., $A_1B_1C_x$) against words, as compared to trials pitting initial-syllable fake-words ($A_xB_1C_1$) against words, with one small exception. In the non-English language condition, participants were more likely to choose initial-syllable fake-words over part-words than they were final-syllable fake-words over part-words; they did not, however, significantly prefer *either* part-words or final-syllable fake-words. Implicit measures of increased processing demands (i.e., RT) likewise did not support a special role for final-syllable fake-words, though there was some evidence that initial-syllable, or both initial- and final-syllable fake-words, incurred greater effort. In the children, however, performance on word versus final-syllable fake-words is more robust to interference across trials than on other contrasts. On the other hand, we also determined across the set of experiments that learners prefer initial and final-syllable fake-words over part-words, and that better discrimination of words from part-words was correlated with participants' selection of edge-manipulated fake-words. The data thus appear to suggest that edges are processed and remembered differently than material embedded in the middle of a chunk.

The position-encoding hypothesis, as set out in this dissertation, details properties of the *output* of SL. While this hypothesis does not necessitate a specific underlying mechanism, I have argued that position-based encoding would not emerge from an exclusively TP-tracking mechanism. What are the alternatives? A full review of the current computational literature as it relates to SL is beyond the scope of this chapter; however, it is worth noting a few points. First – many 'chunking' accounts of SL, such as instantiated in PARSER (Perruchet & Vinter, 1998) are not designed such that the position of syllables is explicitly encoded by the system, nor would the machinery ever yield units such as the *fake-words* described in this dissertation. For example,

in PARSER, the possible outputs are tied to the sequences that have actually been encountered in the familiarization stream. This means that part-words – sequences that are encountered, though less frequently, in the input stream – are logical possible outcomes of a PARSER learning simulation. Fake-words, however, will never emerge, as they involve novel syllable combinations. It is therefore unlikely that a chunking model such as PARSER would ever predict fake-word preference over part-words – despite the fact that it uses a non TP-tracking computational mechanism to account for SL.

It is possible, however, to look for position-based effects in learning, even with the output of a model that does not explicitly encode position. For example, the computational model PARSER yields a potential 'lexicon' at each step of learning when trained on a given corpus. This lexicon will include a variety of singleton syllables and syllable combinations as candidate percepts.[32] These can be broken into the following categories: singleton syllables, two-syllable combinations, trisyllabic combinations (i.e., 'words' and 'part-words'), and larger sequences. I ran 25 learning iterations of PARSER using Language A from Experiment 1 as a test case. Combining the candidate percepts from the final step of learning across these 25 runs yields a total of 506 possible percepts, 277 of which have a 'strength' above 1. What are the characteristics of these percepts?

Singleton syllables that correspond to the initial syllables of words emerged 38 times (32 values greater than 1, the standard 'threshold' for learning), medial syllables emerged 14 times (10 greater than 1), and final syllables emerged 40 times (34 greater than 1). This distribution

[32]Simulations were run using the model's default settings for decay rate (0.05) and interference rate (0.005)

does not appear to be random – out of all singleton syllables (n = 92), edge-based syllables occur approximately 40% of the time each, whereas medial syllables occur 15% of the time. If singleton syllables emerged at random, we would expect a roughly equivalent distribution of 33% for each syllable position type. Bisyllabic combinations occurred 9.5% of the time overall (i.e., 48 of 506 total units) and were roughly split between those that began with an initial syllable and those that began with a medial syllable from a word. None, however, began with a final syllable – that is, none involve two syllables from two different words. The model is of course very good at finding words. Of the 96 trisyllabic percepts, 92 were words (95.8%). Part-word segmentations emerged extremely rarely, either as trisyllabic units (4 times, 1 value above 1) or in combination with additional syllables (36 times out of 506 total, 12 instances out of the 277 units above strength 1). Finally, combinations of more than three syllables that involved an initial word followed by additional syllables were quite common (234 times out of 506; 62 above 1). Of these, the vast majority (86.8%) ended with a final syllable (i.e., they represent multi-word chunks, and consist of exactly 6, 9, 12, or 15 syllables).

In other words, learning via a process of chunking (as instantiated via PARSER) appears to establish and highlight the edges of chunks. These results are not entirely consistent with the behavioural results presented in this dissertation. For example, the preponderance of multi-syllabic chunks that begin with word-initial syllables in the PARSER data suggests a stronger effect for initial edges; the behavioural results, on the other hand, are more equivalent between the two edges, or may in fact be more consistent with a final syllable-based effect in learning. However, the results of this brief analysis suggest that one way forward in determining the underlying mechanism(s) of SL is through a more fine-grained evaluation of existing computational models of SL phenomena.

There are a number of important limitations that must be acknowledged. First, the 2AFC paradigm involves presenting isolated chunks to learners after exposure to a continuous stream. The learning experience of a participant, however, does not stop when the experimenter switches from 'training' to 'test' in a particular paradigm. Therefore, it is undoubtedly the case that participants continued to learn – and specifically, potentially learned something about either the position and/or TP-based structures they are being tested on. Learners were exposed to an equal number of instances of part-words and words (n = 32); however, they heard a higher number of instances of fake-words overall (n = 48). Trial was included as a factor in all analyses to model the potential effect of this confound. There was little evidence of change in performance over the course of the experiment, except in two cases: Experiment 4 (video + NL) of Chapter 2, and the developmental sample (Chapter 4).

It is of note that Experiment 4 is precisely the experiment in which we see the greatest degree of position-based knowledge over TP-based knowledge. If the higher proportion of fake-word test items leads to learning of those items, we would expect the choice of fake-words to increase across trial. While there was no evidence for change over the course of the experiment in choice of words over fake-words, this prediction was upheld for trials pitting part-words against initial- and medial-syllable fake-words. However, participants also chose final-syllable fake-words over part-words (OR = 0.71, p = .007) – an effect that did not change over the course of the experiment (OR = 1.00, p = .903). The developmental sample was similarly mixed: under certain conditions participants were increasingly likely to choose part-words (in word versus part-word trials, and part-word versus initial-syllable fake-word trials), in others, increasingly likely to choose fake-words (words versus initial- and medial-syllable fake-words; part-words versus final-syllable fake-words). Taken together, I believe these results suggest that effects

related to initial- and medial-syllable fake-words may be susceptible to learning during the testing phase, but that effects related to final-syllable fake-words may be more reliable.

5.4 Future directions

The results summarized above leave open an array of questions. For example, are the reported developmental differences primarily related to underlying differences in executive function skills, and do these differences lead to actual shifts in learning outcome – or merely differences in *testing* outcome? When learners exposed to unfamiliar sounds are given more time to learn, does the process of learning follow the same trajectory on a longer timeline? Or are there qualitative differences in learning outcomes? To answer these questions, I have run a similar set of studies with 8-month old infants and 3- to 6-year olds, as well as adults exposed to longer durations of the non-English sounds; hopefully, the data from these projects will help shed additional light on the underlying mechanism(s) of SL. The most critical questions that remain, however, are (1) whether the position-based differences in the adult sample reflect the actual learning process, or whether they have somehow been derived from the testing protocol itself, and (2) what the psychobiological mechanism/s is/are that underpin these statistical learning outcomes.

The answers to both of these questions will likely require a different kind of paradigm, one that is able to tap into the learning process as it is happening as opposed to after the fact, which requires explicit decision-making processes. There have been a number of recent innovations designed to implicitly track learning in the segmentation SL paradigm through behavioural means (e.g., Siegelman et al., 2017). At the same time, combining behavioural and neuroimaging techniques has the potential to bring greater clarity to the phenomenon. For

example, one way to determine differential encoding by syllable position would be to harness the now well-documented oscillatory response to multi-syllabic structures (e.g., Batterink & Paller, 2017; Kabdebon et al., 2015). In other words, participants could be entrained to a consistent, structured stream for several minutes, at which point, unbeknownst to the participant, the stream is subtly altered such that certain syllables are unexpected. These unexpected syllables/sounds could be parametrically varied across different syllable locations within high-TP chunks, to observe whether certain syllable manipulations cause greater impedance to the consistent neuronal activity. In addition/alternatively, the use of novel statistical techniques, such as neural decoding through multi-variate pattern analysis (see Haxby, Connolly, & Guntupalli, 2014 for overview), may prove a fruitful means of detecting how the brain encodes structure during the course of learning.

I believe that understanding the neurobiological mechanisms and cognitive processes that underlie statistical learning is important for understanding how (and whether) statistical learning is involved in language acquisition. I further propose that a deeper understanding of these processes may elucidate apparent differences in child and adult language learning trajectories (e.g., critical period effects; see Thiessen, Girard, & Erickson, 2016). The work presented in this dissertation is a small step down this road; using behavioural analysis techniques, I have found that attentional skills and prior knowledge impact the course of auditory statistical learning, and that both of these skills may underpin differences between children's and adults' learning in the same task. At the same time, however, the effects presented herein are small, in places contradictory, and inevitably confounded by the experimental protocol. Future work that combines on-line behavioural measures with neuroimaging and computational modelling will be

necessary to extend these findings and deepen our understanding of the nature of statistical learning and its relationship to language acquisition.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952-964.

Agrawal, Y., Platz, E. A., & Niparko, J. K. (2008). Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. *Archives of Internal Medicine*, *168*(14), 1522-1530.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Ninth European Conference on Speech Communication and Technology*.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439.

Altmann, G., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 899.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, *32*(41), 14156-14164.

Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science, 21*(2).

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464-473.

Arciuli, J., & Simpson, I. C. (2012a). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, *517*(2), 133-135.

Arciuli, J., & Simpson, I. C. (2012b). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286-304.

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711).

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2).

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170-176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Barbosa, P. A. &Albano, E. C. (2004). Brazilian Portuguese, *Journal of the International Phonetic Association*, *34*(2), 227–232.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, *2*, 324.

Batterink, L. J. & Paller, K. A. (2017). Online neural monitory of statistical learning. *Cortex, 90,* 31-45.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv* preprint arXiv:1506.04967.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253-3258.

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916-12921.

Bertinetto, M. & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151.

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant behavior and development*, *4*, 247-260.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In de Boysson-Bardies B., de Schonen S., Jusczyk P., McNeilage P., Morton J. (Eds.) *Developmental neurocognition: Speech and face processing in the first year of life* (289-304). Springer, Dordrecht.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *110*(3), 461-468.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325-339.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103-119.

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*(3), 494-500.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, *29*(4), 711.

Black A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A Meta-Analysis**.** In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 124-129.

Black, A. & Hudson Kam, C.L. (2013, November). The role of memory and representations in statistical learning. Poster at the Acoustical Society of America Conference; San Francisco.

Black, A. & Hudson Kam, C. L. (2015, June). Representations extracted during statistical learning: developmental differences in ages 5-10 and adults. Interdisciplinary Advances in Statistical Learning Conference; San Sebastian, Spain.

Black, A. & Hudson Kam, C. L. (2016a, November). The impact of phonological knowledge of statistical learning. Boston University Conference on Language Development, 41, Boston, USA.

Black, A. & Hudson Kam, C. L. (2016b, June). Prior knowledge and individual differences impact statistical learning. Fifth Implicit Learning Seminar, Lancaster, UK.

Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, 47(3-4), 296-312.

Boersma, P. & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3, retrieved 2012 from http://www.praat.org/.

Bogaerts, L., Siegelman, N., & Frost, R. (2016). Splitting the variance of statistical learning performance: A parametric investigation of exposure duration and transitional probabilities. *Psychonomic Bulletin & Review*, *23*(4), 1250-1256.

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451-459.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298-304.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Bulgarelli, F., Benitez, V., Saffran, J., Byers-Heinlein, K., & Weiss, D. J. (2017). Statistical Learning of Multiple Structures by 8-Month-Old Infants. *Proceedings of the Annual Boston University Conference on Language Development*. *Boston University Conference on Language Development*, *41*, 128–139.

Burkart, J. M., & Rueth, K. (2013). Preschool children fail primate prosocial game because of attentional task demands. *PLoS One*, *8*(7), e68440.

Byers-Heinlein, K. (2015). Methods for studying infant bilingualism. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing*. Cambridge, UK: Cambridge University Press.

Cardona, G. & Suthar, B. (2007). Gujarati. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan Languages* (660 – 698), New York, NY: Routledge.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental science*, *11*(2), 282-298.

Cashon, C. H., Ha, O. R., Estes, K. G., Saffran, J. R., & Mervis, C. B. (2016). Infants with Williams syndrome detect statistical regularities in continuous speech. Cognition, 154, 165-168.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131-1157.

Christiansen, M. H. (in press). Implicit-statistical learning: A tale of two literatures. *Topics in Cognitive Science.*

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: effects of music background. *Frontiers in Neuroscience*, 10, 229.

Clynes, A. & Deterding, D. (2011). Standard Malay (Brunei). *Journal of the International Phonetic Association*, *41*, 259–268.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In Cole, R. A. (Ed.) *Perception and production of fluent speech* (133-163). Lawrence Erlbaum Associates, Inc.

Conway, C. M. & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(1), 24-39.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584-1595.

Corretge, R. (2012). Praat vocal toolkit: a free plugin for Praat with automated scripts for voice processing. http://www.praatvocaltoolkit.com/.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *30*(5), 1119.

Cunillera, T., Gomila, A. & Rodriguez-Fornels, A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neuroscience, 9*(1), 23.

Currie Hall, K. Allen, B., Fry, M. Mackie, S. & McAuliffe, M. (2015). Phonological CorpusTools [Computer program]. https://phonologicalcorpustools.github.io/CorpusTools/.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3-4), 133-142.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668-3678.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: an artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General*, *128*(2), 165.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158.

Dixon, R. M. W. & Aikenvald, A. Y. (2002). Word: a typological framework. In Dixon, R. M. W. & Aikhenvald, A. Y. (Eds.) *Word: a cross-linguistic typology* (1-41). New York: Cambridge University Press.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dumay, N., Content, A., & Frauenfelder, U. H. (1999). Acoustic-phonetic cues to word boundary location: Evidence from word spotting. *Proceedings of the 14th International Congress of Phonetic Sciences*, University of California, Linguistics Department, Berkeley.

Dum-Tragut, J. (2009). *Armenian: Modern Eastern Armenian (Vol. 14)*. Amsterdam/Philadelphia: John Benjamins Publishing.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322-1331.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition, 2*(3), 189-220.

Eimas, P. D. (1999). Segmental and syllabic representations in the perception of speech by young infants. *The Journal of the Acoustical Society of America*, *105*(3), 1901-1911.

Elordieta, G. (2014). The word in phonology. In Ibarretxe-Antuñano, I. & Mendívil-Giró J. L. (Eds.) *To be or not to be a word: new reflections on the definition of word,* (6-65). Newcastle upon Tyne: Cambridge Scholars Publishing.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, *64*(5), 1021-1040.

Endress, A. D. (2010). Learning melodies from non-adjacent tones. *Acta Psychologica*, *135*(2), 182-190.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247-299.

Endress, A. D., & Mehler, J. (2009a). The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*(3), 351-367.

Endress, A. D., & Mehler, J. (2009b). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, *62*(11), 2187-2209.

Enns, J. T., & Girgus, J. S. (1985). Developmental changes in selective and integrative visual attention. *Journal of Experimental Child Psychology*, *40*(2), 319-337.

Erickson, L. C., Thiessen, E. D., & Graf Estes, K. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language*, *72*, 49-58.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206-EL212.

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, *133*(2), 408-413.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-335.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477-499.

Finn, A. S., & Hudson Kam, C. L. (2015). Why segmentation matters: Experience-driven segmentation errors impair "morpheme" learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1560.

Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When it hurts (and helps) to try: the role of effort in language learning. *PloS one*, *9*(7), e101806.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119-130.

Forest, T. A. (2017). The impact of attention on people's ability to learn two statistical patterns simultaneously. Masters' Thesis, retrieved from TSpace Repository, University of Toronto.

Francois, C., & Schön, D. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, *2*, 167.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107-125.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic bulletin & review*, *16*(3), 486-490.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, *33*(2), 260-272.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*(2), 229-252.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178-186.

Goyet, L., Nishibayashi, L. L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PloS One*, *8*(11), e79646.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254-260.

Graf Estes, K., Gluck, S. C. W., & Bastos, C. (2015). Flexibility in statistical word segmentation: finding words in foreign speech. *Language Learning and Development, 11*(3), 252-269.

Graf Estes, K., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. Developmental psychology, 51(11), 1517.

Greenberg, S. (1999). Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*(2-4), 159-176.

Hanulíková, A. & Hamann, S. (2010). Slovak, *Journal of the International Phonetic Association*, *40*(3), 373–378.

Harris, Z. (1955). From phoneme to morpheme. *Language*, *31*(2), 190-222.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*(2001), B53-B64.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435-456.

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, *17*(6), 610-641.

Hay, J. F., Pelucchi, B. Graf Estes, K. & Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology, 63*(2), 93-106.

Hayes, B., & Abad, M. (1989). Reduplication and syllabification in Ilokano. *Lingua*, *77*(3-4), 331-374.

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In Hayes, J.R. (Ed.), *Cognition and the Development of Language* (221-234). New York, NY: Wiley.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377-396.

Healy, A. F., & Cutting, J. E. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 73-83.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570.

Hualde, J. I. (2005). *The Sounds of Spanish with Audio CD*. New York: Cambridge University Press.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151-195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30-66.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017-2036.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658.

Hutson, J., Palmer, S. & Mattys, S. (2016). Speech segmentation by statistical learning in young, middle-aged, and older adults. Paper presented at *The Fifth Implicit Learning Seminar*, Lancaster, UK.

Itō, J. & Mester, R. A. (1995). Japanese phonology. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (817-838), Blackwell Handbooks in Linguistics, Blackwell Publishers.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, *15*(4), 496-505.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association, 33*(1), 103-107.

Johnson, E. K. (2012). Bootstrapping language: are infant statisticians up to the job? In Rebuschat, P. & Williams, J. N. (Eds.) *Statistical learning and language acquisition,* 55-89, Walter de Gruyter, Inc., Boston/Berlin.

Johnson, E. K., & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, *13*(2), 339-345.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*(9), 323-328.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, *23*(5), 648.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465-1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3-4), 159-207.

Kabdebon, C., Peña, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25-36.

Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*(1), 111-116.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795-1804.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795-1804.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*(1), 171.

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184-193.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145-149.

Kirby, J. P. (2011). Vietnamese (Hanoi Vietnamese), *Journal of the International Phonetic Association*, *41*(3), 381–392.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In Cole, R. J. (Ed.) *Perception and Production of Fluent speech*, 243-288.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, *128*(1), 34-40.

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., & Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Developmental Science*, *14*(5), 1100-1106.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). P*hilosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Ladefoged, P & Maddieson, I. (1996). *The Sounds of the World's Languages*, Oxford: Blackwell.

Lee, H., & Noppeney, U. (2014). Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. *Frontiers in Psychology*, *5*, 868.

Leslau, W. (1997). Chapter 21: Amharic phonology. In Alan S. Kaye & Peter T. Daniels (Eds.) *Phonologies of Asia and Africa (including the Caucasus)*, *Volume 1* (399-430). Winona Lake, Indiana: Eisenbraus.

Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*(2), 241-246.

Li, X., Zhao, X., Shi, W., & Conway, C. (2018). Lack of Cross-modal Effects in Dual-modality Implicit Statistical Learning. *Frontiers in Psychology*, *9*, 146.

Llamzon, T. A. (1966). Tagaolog Phonology, *Anthropological Linguistics, 8*(1), 30-39.

Ludden, D., & Gupta, P. (2000, January). Zen in the art of language acquisition: Statistical learning and the less is more hypothesis. In Proceedings of the Annual Meeting of the *Cognitive Science Society, 22*(22).

Lüdecke, D. (2017). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.4.0, https://CRAN.R-project.org/package=sjPlot.

MacKay, D. G. (1970). Spoonerisms: the structure of errors in the serial order of speech. *Neuropsychologia, 8*(3), 323-350.

Mandikal Vasuki, P. R., Sharma, M., Ibrahim, R. K., & Arciuli, J. (2017). Musicians' online performance during auditory and visual statistical learning tasks. *Frontiers in Human Neuroscience, 11*, 114.

Mareschal, D., & French, R. M. (2017). TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B*, *372*(1711), 20160057.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576-585.

Massaro, D. W. (1972). Perceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*(2), 124.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.

McKay, K. E., Halperin, J. M., Schwartz, S. T., & Sharma, V. (1994). Developmental analysis of three aspects of information processing: Sustained attention, selective attention, and response organization. *Developmental Neuropsychology*, *10*(2), 121-132.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 22-33.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-B42.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 298-305.

Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory & Cognition*, *39*(6), 1085-1093.

Mersad, K., & Nazzi, T. (2012). When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language learning and Development*, *8*(3), 303-315.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 31.

Mitchel, A. D., Christiansen, M. H., & Weiss, D. J. (2014). Multimodal integration in statistical learning: evidence from the McGurk illusion. *Frontiers in Psychology*, *5*, 407.

Mohammed, M. A. (2001). Modern Swahili Grammar. Nairobi: East African Educational Publishers Ltd.

Morén, B. (2006). Consonant-vowel interactions in Serbian: features, representations, and constraint interactions. *Lingua*, *116*(2006), 1198-1244.

Moreno, S., Lee, Y., Janus, M., & Bialystok, E. (2015). Short-Term Second Language and Music Training Induces Lasting Functional Brain Changes in Early Childhood. *Child Development*, *86*(2), 394-406.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex*, *19*(3), 712-723.

Morrison, J. M., & Kam, C. H. (2018). Phonological form influences memory for form-meaning mappings in adult second-language learners. *PsyArXiv,* preprint: 10.17605/OSF.IO/SP3CX

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894-15898.

Neger, T., Janse, E., & Rietveld, T. (2015). Correlates of older adults' discrimination of acoustic properties in speech. *Speech, Language and Hearing*, *18*(2), 102-115.

Neger, T. M., Rietveld, T., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in Human Neuroscience*, *8*, 628.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Language and Cognition*, *8*(3), 447-461.

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance 1. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127-162.

Noguchi, M. & Hudson Kam, C. L. (2018). The emergence of allophonic perception of unfamiliar speech sounds: the effects of contextual distribution and phonetic naturalness. *Language Learning.* DOI: 10.1111/lang.12267.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355-376.

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, *7*, 1245.

Olejarczuk, P., & Kapatsinski, V. (2016). Attention allocation in phonetic category learning. *Proceedings of Cognitive Modeling in Linguistics*, *14*, 148-156.

Omigie, D., & Stewart, L. (2011). Preserved statistical learning of tonal and linguistic material in congenital amusia. *Frontiers in Psychology*, *2*, 109.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225-237.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745-2750.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232-258.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244-247.

Peña, M. Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.

Penha, B. R. G. (2014). Aging effects in speech statistical learning: a behavioral and computational study. Doctoral dissertation.

Perruchet, P. (2005). Statistical approaches to language acquisition and the self-organizing consciousness: A reversal of perspective. *Psychological Research*, *69*(5-6), 316-329.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*(7), 1299-1305.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233-238.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, *17*(2-3), 97-119.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language*, *66*(4), 807-818.

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica*, *149*, 1-8.

Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*, *34*(2), 255-285.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246-263.

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *The Quarterly Journal of Experimental Psychology: Section A*, *55*(2), 485-503.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115-154.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human perception and performance*, *8*(2), 297.

Poljac, E., de-Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, *123*(2), 308-312.

Potter, C. E., Wang, T. and Saffran, J. R. (2017), Second Language Experience Facilitates Statistical Learning of Novel Linguistic Materials. *Cognitive Science*, *41*, 913–927.

Price, A., & Shin, J. C. (2009). The impact of Parkinson's disease on sequence learning: perceptual pattern learning and executive function. *Brain and Cognition*, *69*(2), 252-261.

Prinzmetal, W., McCool, C. & Park, S. (2005). Attention: reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General, 134*(1), 73.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. Bilingualism: *Language and Cognition*, *13*(2), 253-262.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2016). Incremental implicit learning of bundles of statistical patterns. *Cognition*, *157*, 156-173.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, *46*(4), 377-396.

Räsänen, O., Doyle, G., & Frank, M. C. (2018). Pre-linguistic segmentation of speech into syllable-like units. *Cognition*, *171*, 130-150.

Raviv, L., & Arnon, I. (2017). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. Developmental Science, e12593.

Riad, T. (2014). *The Phonology of Swedish*. Oxford, UK: Oxford University Press.

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. Wiley Interdisciplinary Reviews: *Cognitive Science*, *1*(6), 906-914.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608-635.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition, 81*(2), 149-169.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110-114.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181-203.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, *8*(2), 101-105.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273-284.

Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, *105*(2), 466-476.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700-703.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, *15*(3), 228-240.

Santolin, C., & Saffran, J. R. (2017). Constraints on statistical learning across species. *Trends in Cognitive Sciences*, *22*(1), 52-63.

Schiff, A. R., & Knopf, I. J. (1985). The effect of task demands on attention allocation in children of different ages. *Child Development*, *56*(3), 621-630.

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, *29*(1), 37-51.

Schwab, J. F., Schuler, K. D., Stillman, C. M., Newport, E. L., Howard Jr, J. H., & Howard, D. V. (2016). Aging and the statistical learning of grammatical form classes. *Psychology and Aging*, *31*(5), 481.

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, *23*(9), 994-999.

Shapiro, Michael C. (2003). Hindi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan Languages* (250 – 285), New York, NY: Routledge.

Shin, J. (2015). Vowels and Consonants. In L. Brown & J. Yeon (Eds.) *The Handbook of Korean Linguistics* (3-21); West Sussex, UK: John Wiley & Sons, Ltd.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical experience influences statistical learning of a novel language. *The American Journal of Psychology*, *126*(1), 95.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Phil. Trans. R. Soc. B*, *372*(1711), 20160059.

Siegelman, N., Bogaerts, L. & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behaviour Research Methods, 49*(2), 418-432.

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2017). Redefining "learning" in statistical learning: what does an online measure reveal about the assimilation of visual regularities? *Cognitive Science,* DOI: 10.1111/cogs.12556.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105-120.

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, *291*, 244-252.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In Ferguson, C. A., & Slobin, D. I. (Eds.) *Studies of Child Language Development,* (75-208), New York: Holt, Rinehart, & Winston.

Smith, E. R., Branscombe, N. R., & Bormann, C. (1988). Generality of the effects of practice on social judgment tasks. *Journal of Personality and Social Psychology*, *54*(3), 385.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Soyuzmultfilm (Producer), & Khitruk, F. (Director). (1969). *Vinni Pukh* [Motion picture]. The Soviet Union: Soyuzmultfilm.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73-108.

Takahasi, M., Yamada, H., & Okanoya, K. (2010). Statistical and prosodic cues for song segmentation learning by Bengalese finches (Lonchura striata var. domestica). *Ethology*, *116*(6), 481-489.

Tamminen, J., & Gaskell, M. G. (2008). Short Article: Newly Learned Spoken Words Show Long-Term Lexical Competition Effects. *Quarterly Journal of Experimental Psychology*, *61*(3), 361-371.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*(1), 21.

Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cognitive Science, 34*(6), 1093-1106.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B*, *372*(1711).

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, *139*(4), 792.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, *39*(4), 706.

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, *112*(32), 10062-10067.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, *3*(1), 1-42.

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, *112*(32), 10062-10067.

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432-444.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867-875.

Tummeltshammer, K., Amso, D., French, R. M., & Kirkham, N. Z. (2017). Across space and time: Infants learn from backward and forward visual statistics. *Developmental Science*, *20*(5).

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934-1945.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367-376.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62*(6), 1297-1311.

Van Gijn, R., & Zúñiga, F. (2014). Word and the Americanist perspective. *Morphology*, *24*(3), 135-160.

Vouloumanos, A., Brosseau-Liard, P. E., Balaban, E., & Hager, A. D. (2012). Are the products of statistical learning abstract or stimulus-specific? *Frontiers in Psychology*, *3*.

Vuong, L., Meyer, A., & Christiansen, M. (2011, January). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33).

Wade, T. (2011). *A Comprehensive Russian Grammar*. Oxford, UK: John Wiley & Sons, Ltd.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953.

Wanrooij, K., Escudero, P., & Raijmakers, M. E. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*(5), 307-319.

Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech segmentation in a simulated bilingual environment: A challenge for statistical learning?. *Language Learning and Development*, *5*(1), 30-49.

Weise, R. (2000). *The phonology of German*, Oxford: University Press.

Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge University Press.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131-149.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Wickham, H. (2017). https://cran.r-project.org/package=tidyverse.

Wittke, K., Spaulding, T. J., & Schechtman, C. J. (2013). Specific language impairment and executive functioning: Parent and teacher ratings of behavior. *American Journal of Speech-Language Pathology*, *22*(2), 161-172.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565-585.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420-433.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229-248.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891-921.

Zamuner, T. S., Moore, C., & Desmeules-Trudel, F. (2016). Toddlers' sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology*, *152*, 136-148.

Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association, 21*(1), 46-48.

Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, 146, 217-222.

Zimmer, K. & Orgun, O. (1999). Turkish, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (154-158), Cambridge: Cambridge University Press.

Appendices

Appendix A

This appendix contains materials relevant to the studies discussed in Chapters 2, 3, and 4. The tables below present the full trial list for the two statistically defined languages (see Chapter 2, Experiment 1). Lists are represented with English letters; the physical realization of the semi-English (Chapter 2, Experiment 2) and non-English (Chapter 2, Experiment 3) trial lists employed the structurally parallel sounds that were encountered in familiarization (see Tables 2.8 and 2.15, Chapter 2, or Table 3.1, Chapter 3). Sequences written in all caps represent the trisyllabic nonce words from the familiarization stimuli. Sequences that are underlined represent fake-words.

A.1 Language A (Chapter 2 – Experiment 1, EL; Experiment 2, SEL)

Words vs Part-words

Sound1	***Sound2***
GOLABU	rogola
BIDAKU	bubida
TUPIRO	titupi
PADOTI	kupado
labubi	GOLABU
dakugo	BIDAKU
piropa	TUPIRO
dotitu	PADOTI

Words vs Fake-words

Initial		*Medial*		*Final*	
GOLABU	bilabu	GOLABU	godabu	GOLABU	golaku
BIDAKU	tudaku	BIDAKU	bipiku	BIDAKU	bidabu
TUPIRO	papiro	TUPIRO	tudoro	TUPIRO	tupiku
PADOTI	tudoti	PADOTI	padati	PADOTI	padoro
tulabu	GOLABU	gopibu	GOLABU	golati	GOLABU
godaku	BIDAKU	bilaku	BIDAKU	bidaro	BIDAKU
gopiro	TUPIRO	tularo	TUPIRO	tupiti	TUPIRO
godoti	PADOTI	papiti	PADOTI	padobu	PADOTI

Part-words vs Fake-words

Initial		*Medial*		*Final*	
labubi	tulabu	labupa	gopibu	labupa	golati
dakugo	godaku	dakutu	bilaku	dakutu	bidaro
piropa	gopiro	pirogo	tularo	pirogo	tupiti
dotitu	godoti	dotigo	papiti	dotigo	padobu
bilabu	rogola	godabu	kugola	golaku	kugola
tudaku	bubida	bipiku	robida	bidabu	robida
papiro	titupi	tudoro	butupi	tupiku	butupi
tudoti	kupado	padati	ropado	padoro	ropado

A.2 Language B (Chapter 2 – Experiment 1, EL)

Words vs Part-words

Sound1	*Sound2*
DATUBI	tubipi
GOTIBU	tiburo
PIDOPA	dopada
ROKULA	buroku
padatu	DATUBI
bigoti	GOTIBU
lapido	PIDOPA
kulago	ROKULA

Words vs Fake-words

Initial		*Medial*		*Final*	
DATUBI	gotubi	DATUBI	dakubi	DATUBI	datubu
GOTIBU	rotibu	GOTIBU	godobu	GOTIBU	gotibi
PIDOPA	rodopa	PIDOPA	pitipa	PIDOPA	pidobi
ROKULA	gokula	ROKULA	rotila	ROKULA	rokubu
pitubi	DATUBI	dadobi	DATUBI	datupa	DATUBI
datibu	GOTIBU	gotubu	GOTIBU	gotila	GOTIBU
dadopa	PIDOPA	pitupa	PIDOPA	pidola	PIDOPA
pikula	ROKULA	rotula	ROKULA	rokupa	ROKULA

Part-words vs Fake-words

Initial		*Medial*		*Final*	
gotubi	padatu	dakubi	ladatu	datubu	ladatu
rotibu	bigoti	godobu	pagoti	gotibi	pagoti
rodopa	lapido	pitipa	bupido	rokubu	biroku
gokula	kulago	rotila	biroku	pidobi	bupido
tubipi	pitubi	tubiro	dadobi	tubiro	datupa
tiburo	datibu	tibuda	gotubu	tibuda	gotila
dopada	dadopa	dopago	pitupa	kulapi	rokupa
buroku	pikula	kulapi	rotula	dopago	pidola

A.3 Language A (Chapter 2 – Experiment 3, NEL)

Words vs Part-words

Sound1	*Sound2*
GOLABU	rogola
BIDAKU	dakugo
TUPIRO	titupi
PADOTI	dotitu
labubi	GOLABU
bubida	BIDAKU
piropa	TUPIRO
kupado	PADOTI

Words vs Fake-words

Initial		*Medial*		*Final*	
GOLABU	bilabu	GOLABU	godabu	GOLABU	golaku
BIDAKU	tudaku	BIDAKU	bipiku	BIDAKU	bidabu
TUPIRO	papiro	TUPIRO	tudoro	TUPIRO	tupiku
PADOTI	tudoti	PADOTI	padati	PADOTI	padoro
bilabu	GOLABU	gopibu	GOLABU	golati	GOLABU
godaku	BIDAKU	bilaku	BIDAKU	bidaro	BIDAKU
gopiro	TUPIRO	tularo	TUPIRO	tupiti	TUPIRO
godoti	PADOTI	papiti	PADOTI	padobu	PADOTI

Part-words vs Fake-words

Initial		*Medial*		*Final*	
rogola	bilabu	labupa	gopibu	kugola	golaku
dakugo	godaku	robida	bipiku	dakutu	bidaro
titupi	papiro	pirogo	tularo	butupi	tupiku
dotitu	godoti	ropado	padati	dotigo	padobu
tulabu	labubi	godabu	kugola	golati	labupa
tudaku	bubida	bilaku	dakutu	bidabu	robida
gopiro	piropa	tudoro	butupi	tupiti	pirogo
tudoti	kupado	papiti	dotigo	padoro	ropado

A.4 Language A (Chapter 4)

Words vs Part-words

Sound1	*Sound2*
GOLABU	rogola
BIDAKU	dakugo
TUPIRO	titupi
PADOTI	dotitu
labubi	GOLABU
bubida	BIDAKU
piropa	TUPIRO
kupado	PADOTI

Words vs Fake-words

Initial		*Medial*		*Final*	
GOLABU	bilabu	GOLABU	godabu	GOLABU	golaku
BIDAKU	tudaku	BIDAKU	bipiku	BIDAKU	bidabu
TUPIRO	papiro	TUPIRO	tudoro	TUPIRO	tupiku
PADOTI	tudoti	PADOTI	padati	PADOTI	padoro
bilabu	GOLABU	godabu	GOLABU	golati	GOLABU
tudaku	BIDAKU	bipiku	BIDAKU	bidabu	BIDAKU
papiro	TUPIRO	tudoro	TUPIRO	tupiku	TUPIRO
tudoti	PADOTI	padati	PADOTI	padoro	PADOTI

Part-words vs Fake-words

Initial		*Medial*		*Final*	
rogola	bilabu	godabu	rogola	rogola	golaku
dakugo	tudaku	bipiku	dakugo	dakugo	bidabu
titupi	papiro	tudoro	titupi	titupi	tupiku
dotitu	tudoti	padati	dotitu	dotitu	padoro
bilabu	labubi	labubi	godabu	golaku	labubi
tudaku	bubida	bubida	bipiku	bidabu	bubida
papiro	piropa	piropa	tudoro	tupiku	piropa
tudoti	kupado	kupado	padati	padoro	kupado

Appendix B

This appendix contains materials relevant to the studies discussed in Chapter 3. The language background questionnaire is presented in B.1, the exit interview in B.2, and a table of all the second languages reported, as well as their phonetic/phonemic inventory overlap score with the SEL and NEL, and the sources from which those inventories were derived can be found in B.3.

B.1 Language Background Questionnaire

Language Background Questionnaire Subject Number _____

• What cities or towns have you lived in? List first the place where you were born, and list each town or city you have lived in.

birth until age ____ in town/city _____
age ___ until age ____ in town/city _____
age ___ until age ____ in town/city _____
age ___ until age ____ in town/city _____
age ___ until age ____ in town/city _____

• What languages do you speak (include your native language(s))? When did you start learning this language? How would you rate your proficiency in reading, writing, speaking, and understanding it? (1) not at all, (2) poorly, (3) fairly well, (4) fluently.

language	age	reading	writing	speaking	understanding
_____	___	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	___	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	___	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	___	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

What languages does your family speak (include native language(s))? How would you rate their proficiency in reading, writing, speaking, and understanding it? (1) not at all, (2) poorly, (3) fairly well, (4) fluently.

language	family member	reading	writing	speaking	understanding
_____	_____	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	_____	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	_____	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
_____	_____	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4

• How much do you enjoy learning new languages (please circle one)? (1) not at all, (2) a little, (3) a lot, (4) it's one of my favorite activities.

• Do you play any instruments (include voice, if you sing)? When did you start learning, and how long did/have you played? How would you rate your skill level? (1) beginner, (2) intermediate, (3) advanced, (4) professional.

• Do you have any speech or hearing disorders? No Yes
If "yes", please specify:

B.2 Exit Interview

How was that?
Were there any sounds that caught your attention more than others?
List any syllables/sounds/sequences that caught your attention/stood out
What were you thinking about while you were listening?
By the time you got to the end of the 2 min of speech (before the question-answer part), did you feel like you knew the "words" of this language?
How confident did you feel in your answers?
Did you feel that this changed over the course of the task (that it got easier or harder as time went on)?
Did you feel like you made your choices more based on what was wrong or what was right?
Were there any sounds that were uncomfortable to listen to or aversive?
Did you think that this might be a real language?
Did you recognize the voice?
Have you taken a language acquisition class before?
Have you heard of statistical learning?
Was this language familiar to you?

B.3 Table of participants' 2nd languages, specific language scores, and inventory sources

Language	SNL		NNL		Inventory Source
	Cons.	*Vowel*	*Cons.*	*Vowel*	
Amharic	0.6	0	0.125	0.25	Leslau (1997); https://en.wikipedia.org/wiki/Amharic
Arabic	0.4	0	0.25	0	https://en.wikipedia.org/wiki/Arabic_phonology
Armenian	0.2	0	0.125	0	Dum-Tragut (2009); https://en.wikipedia.org/wiki/Armenian_language
Belizean	0	0	0	0	https://en.wikipedia.org/wiki/Belizean_Spanish
Cantonese	0	1	0	0.25	Zee (1991); https://en.wikipedia.org/wiki/Cantonese_phonology
Chinese	0	0.4375	0	0.125	see Cantonese, Mandarin
French	0	1	0.125	0.25	Ladefoged & Maddieson (1996); https://en.wikipedia.org/wiki/French_phonology
German	0.2	1	0.125	0.5	Wiese (2000); https://en.wikipedia.org/wiki/Standard_German_phonology
Gujarati	0	0	0	0	Cardona & Suthar (2003); https://en.wikipedia.org/wiki/Gujarati_phonology
Hindi	0.20	0	0	0	Shapiro (2003); https://en.wikipedia.org/wiki/Hindustani_phonology
Ilocano	0.20	0	0	0.25	Hayes & Abad (1989); https://en.wikipedia.org/wiki/Ilocano_language#Phonology
Italian	0.4	0	0.125	0	Bertinetto & Loporcaro (2005); https://en.wikipedia.org/wiki/Italian_phonology
Japanese	0	0	0	0.25	Ito & Mester (1995); https://en.wikipedia.org/wiki/Japanese_phonology
Kiswahili	0	0	0.25	0	Mohammed (2001); https://en.wikipedia.org/wiki/Swahili_language
Korean	0	0	0	0.25	Shin (2015); https://en.wikipedia.org/wiki/Korean_phonology
Malay	0.20	0	0	0	Clynes & Deterding (2011); https://en.wikipedia.org/wiki/Malay_phonology
Mandarin	0	0.5	0	0	Duanmu (2007); https://en.wikipedia.org/wiki/Standard_Chinese_phonology
Polish	0.2	0	0	0.25	Jassem (2003); https://en.wikipedia.org/wiki/Polish_phonology
Portugese	0.2	0	0.25	0.25	Barbosa & Albano (2004); https://en.wikipedia.org/wiki/Portuguese_phonology
Punjabi	0	0	0	0	https://en.wikipedia.org/wiki/Punjabi_language
Russian	0.2	0	0	0.25	Wade (2011); https://en.wikipedia.org/wiki/Russian_phonology
Sanskrit	0	0	0	0	https://en.wikipedia.org/wiki/Sanskrit
Serbian	0.4	0	0.125	0	Moren (2006); https://en.wikipedia.org/wiki/Serbo-Croatian_phonology

Language	SNL		NNL		Inventory Source
	Cons.	*Vowel*	*Cons.*	*Vowel*	
Slovak	0.4	0	0.125	0	Hanulíková & Hamann (2010); https://en.wikipedia.org/wiki/Slovak_phonology
Spanish	0.4	0	0.125	0	Hualde (2005); https://en.wikipedia.org/wiki/Spanish_phonology
Swedish	0.2	1.00	0	0.75	Riad (2014); https://en.wikipedia.org/wiki/Swedish_phonology
Tagalog	0	0	0	0	Llamzon (1966); https://en.wikipedia.org/wiki/Tagalog_phonology
Telugu	0	0	0	0	https://en.wikipedia.org/wiki/Telugu_language
Turkish	0	1	0	0.5	Zimmer & Orgun (1999); https://en.wikipedia.org/wiki/Turkish_phonology
Urdu	0	0	0	0	https://en.wikipedia.org/wiki/Urdu
Vietnamese	0	0	0	0.25	Kirby (2011); https://en.wikipedia.org/wiki/Vietnamese_phonology