Spring 2018

Incremental Processing in a Non-Speech Domain

Kathleen Peters
University of Iowa

INCREMENTAL PROCESSING IN A NON-SPEECH DOMAIN

by

Kathleen Peters

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Psychology

Bob McMurray
Thesis Mentor

Spring 2018

All requirements for graduation with Honors in the
Psychology have been completed.

Toby Mordkoff
Psychology Honors Advisor

INCREMENTAL PROCESSING IN A NON-SPEECH DOMAIN

by

Kathleen Peters

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in Psychology

Bob McMurray
Thesis Advisor

Spring, 2018

All requirements for graduation with Honors in Psychology have been completed.

J. Toby Mordkoff
Psychology Honors Advisor

Abstract

Incremental processing is the process of using auditory information as it unfolds over time. It has been well established within known and novel words but has not been examined in non-speech domains. Tone sequences were used to examine incremental processing in a non-speech domain. Novel pseudo-word served as a control. Participants were trained to map tone sequences or pseudo-words to novel objects, and then they underwent testing trials using eye-tracking in the visual world paradigm to evaluate the degree to which they processed incrementally. Results showed increased fixation proportions to the referents of words sharing onset phonemes compared to the unrelated items indicating people were accessing potential interpretations from the very beginning of the word. This was not observed with the tone sequences. This suggests incremental processing in the pseudo-words and no incremental processing in the tone sequences.

In language, words unfold over time. As a result, a critical problem in speech perception is the fact that not all the information for a word is available at once. It is only at the end of a word that the entirely of that word can be known. This contrasts, with written word recognition (e.g. reading) when the entire word is available at any time. This temporal unfolding creates several challenges.

First, each word does not begin with a unique sound. For example, the sound /bi/ begins the words *bed*, *bell*, *belt*, and many others. Thus, during the early moments of word recognition when the listener has only heard the first couple phonemes (e.g., /bi/), it is impossible to know what word is being heard (though many can be ruled out). The uncertainty can only be resolved upon hearing more phonemes in the word. This problem is known as *temporary ambiguity* (Marslen-Wilson, 1987).

Second, the end of a word (and hence the start of the next word) cannot be known until the word is over. If the sound /bi/ is heard, the word could be *bell*, *bedroom*, or *benefactor*, all of which end at different times. This *segmentation ambiguity* adds additional uncertainty and difficulty to understanding spoken words.

To help solve these problems, listeners process speech incrementally (Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986). That is, listeners use information to immediately access the lexicon as the signal unfolds over time. For example, upon hearing the sound /bi/, the words *bed*, *bell*, *bedroom*, etc. are possibilities of what that word being heard could be and are active in such a way that they are being considered as the word being heard. As more of the word is heard, the set possibilities get winnowed down until a single option remains, the word that is ultimately recognized.

This property of incremental processing has been demonstrated in numerous studies of lexical processing (Allopenna, Magnuson, & Tanenhaus, 1998; Marslen-Wilson, 1987). Incremental processing has not been examined in non-speech domains. It could be a property of processing that is unique to speech. Another possibility is incremental processing can be used in a variety of domains, including speech. The current experiment examines incremental processing in a non-speech domain.

Incremental Processing in Language Comprehension

The strongest evidence for incremental processing in language comes from work with the Visual World Paradigm (VWP). Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivey (1995) gave participants objects preceded by descriptive words. For example, participants could be told to “touch the starred yellow square” (Tanenhaus et al., 1995). They found eye-movements were initiated after the distinguishing point in the phrase. For example, if only one object was starred, the distinguishing point would be the word *starred*. If no ambiguity about the target object was present, looks to the target would be initiated before the end of the word. In addition, when ambiguity was present (e.g., there, were two starred objects), a slight delay in initiating looks to the target was observed (Tanenhaus et al., 1995). These findings suggest participants are using information as it become available to them, a finding that is predicted if people are immediately and simultaneously activating multiple interpretations.

Within speech, competing words are activated immediately and simultaneously after the onset of the word (Allopenna et al., 1998; Marslen-Wilson, 1987; Spivey, Grosjean, & Knoblich, 2005; Tanenhaus et al., 1995). In addition, a given word does not have to reach its point of uniqueness in order for words to become active (Marslen-Wilson, 1987). For example, when trying to distinguish between the words *bed* and *bell*, the uniqueness point is at the third phoneme of each word. Upon hearing /bi/, both *bed* and *bell* will become active even though which word is being said is not yet known.

Evidence of immediate activation of words was initially found by examining cohort competitors. An example of a cohort competitor could be *bed*, if the target word is *bell*. Both *bed* and *bell* begin with the same sounds, /bi/, and are thus cohorts of each other. Zwiterlood (1989) was among the first to find activation in cohort competitors before the uniqueness point of the target word had been heard. Allopenna et al. (1998) used eye-tracking in the VWP. Participants heard a spoken word (e.g. *beaker*) and saw four objects on the screen, a target (e.g. *beaker*), a cohort (e.g. *beetle*), a rhyme (e.g. *speaker*), and an unrelated item (e.g. *carriage*). They found increased looks to cohort competitors compared to unrelated items. Increased looks to cohorts suggests they were being considered as an option for what the auditory word could be. If participants were waiting for all necessary information to then determine the target word, no increased looks to cohorts would be found. In contrast, if participants are using auditory information as it arrives (e.g. processing the information incrementally), then increased looks to the cohort competitors would be found because cohorts share beginning phonemes with the target word.

Allopenna et al. (1998) also found increased looks to rhymes compared to unrelated items. However, cohort fixations peaked earlier and higher than rhyme fixations. Increased looks to the rhymes would suggest a flexibility within the speech recognition system. That is to say, activation of words is not an all or none phenomenon using only the first few phonemes to provide clues as to what words to activate. Flexibility is necessary to compensate for situations where speech input is not ideal, such as when trying to converse with someone in a crowded, noisy room (Brouwer & Bradlow, 2014). As a result, increased looks to words that rhyme with the target can be seen. This suggests the rhyme is being considered briefly, despite a difference in the initial phoneme(s).

One limitation of the VWP is that participants can only fixate one thing at a time, making it unclear whether both words are truly active in parallel. Spivey et al. (2005) overcame this with mouse tracking. Participants heard a spoken word (e.g. *candle*) that matched one of two pictures presented on the screen. Meanwhile, their mouse movements were tracked. They found greater curvature when the competing picture was a cohort competitor than an unrelated word. This suggests initiation of movement of the mouse while the target of the trial was still ambiguous. Importantly, these findings also suggest that multiple words can be activated simultaneously. If the target of a given trial is ambiguous when movement is initiated, the ambiguity present at that time would imply different options are being considered simultaneously.

In contrast to these studies, a recent study found less incremental processing in prelingually deafened cochlear implants users or normal hearing individuals given severely degraded auditory information (McMurray, Farris-Trimble, & Rigler, 2017). This suggests there are circumstances where incremental processing is not used, a result which differs from all other findings in word recognition. Nonetheless, the vast majority of studies support incremental processing for words (Allopenna et al, 1998; Marslen-Wilson, 1987; Spivey et al., 2005).

Incremental Processing in Recently Learned Words

The prevalence of incremental processing raises the question: can it be observed in non-speech domains? If incremental processing can be observed in a non-speech domain, that would suggest it is a property of processing that can be utilized by different domains, including speech.

Alternatively, incremental processing could be exclusive to speech. One way to test for incremental processing in a non-speech domain would be to teach people non-speech sequences and then test for evidence of incremental processing. The process of teaching people non-speech sequences would be necessary because no non-speech sequences exist that are comparable to speech. That is to say, people have considerable practice in their everyday lives matching know objects to known words. If somewhere were to hear the word *bell*, they would be able to identify what object the speaker was talking about. The same cannot be said for any non-speech sequences, and thus, training is necessary for people to learn any non-speech sequences. However, to observe incremental processing in non-speech sequences, it must be able to be seen in newly learned linguistic items (e.g. pseudo words). Several studies with artificial lexica confirm that incremental processing can be seen in recently learned pseudo-words.

A number of such studies have used the VWP to examine incremental processing in artificial lexica (Apfelbaum & McMurray, 2017; Creel, Aslin, & Tanenhaus, 2006; Farris-Trimble & McMurray, 2018; Magnuson, Tanenhaus, Aslin, & Dahan, 2003). In these studies, participants learn to map pseudo-words in an artificial lexicon to novel objects (Apfelbaum & McMurray, 2017; Creel, Aslin, & Tanenhaus, 2006; Farris-Trimble & McMurray, 2018; Magnuson et al., 2003). Using this paradigm, Creel, Aslin, and Tanenhaus (2006) found greater phonemic overlap in words within the artificial lexicon (e.g. *bamo* and *bami* have more overlap than *bamo* and *kanu*) resulted in higher error rates. In addition, when participants made an error on a trial, they were more likely to select the cohort of the target object than an unrelated item. However, a rhyme did not have as great of an effect as cohorts on error rate. This suggests people can be influenced by phonemic overlap, such as the overlap between a cohort and a target, in a relatively short period of time. Also, the beginning phonemes of a word seems to have more of an impact on lexical competition than the ending phonemes.

Magnuson et al. (2003) used the VWP to more precisely characterize the time course of processing. They created sets of four novel words with one target word, a cohort competitor, a rhyme competitor, and an unrelated word. Participants were trained to map pseudo-words to novel objects and completed 14 blocks of training trials and two blocks of testing trials over two days. Each day consisted of seven training blocks and one testing block using eye-tracking. At the end of the testing block on day one, eye-tracking results showed increased fixations to cohort and rhymes over the unrelated item, a pattern which is highly similar to patterns of looking found in real words in Allopenna et al. (1998). However, increased fixation proportions to the rhyme appeared slightly later in the time course of processing the pseudo-word than those to the cohort (Magnuson et al. 2003). If the novel words Magnuson et al. (2003) used were not processed incrementally, increased looks to the cohort or rhyme would not have been found. Thus, these findings demonstrate incremental processing of newly learned words occurs relatively quickly, and that seven training blocks was sufficient time for participants to treat the novel words similarly to real words.

Apfelbaum and McMurray (2017) examined when, in the time course of the word, participants begin to learn pseudo-word to novel object pairings. During training, one group, the simultaneous group, was presented with the novel objects simultaneously with the pseudo word for that trial. The other, group, the delay group, heard the pseudo word 1000 msec before they saw the objects appear on their screen. During testing, a VWP was used and eye-movements were tracked. They found greater fixation proportions to competitors such as cohorts and rhymes in the simultaneous group when compared to the delay group. The difference between the two groups suggests participants were not waiting until the end of the word to start learning (Apfelbaum & McMurray, 2017). These results demonstrate competition can still occur even when there is still uncertainty in what the word will be. In addition, these finding provide further support for the processing of newly learned words incrementally.

Together, these studies using artificial lexica demonstrate incremental processing can be seen in newly learned linguistic items. Therefore, significant experience with a new item is not necessary to begin to process that item incrementally. However, these studies do not support the broader hypothesis that incremental processing is seen in other domains of cognition for several reasons. First, participants were all adults (Apfelbaum & McMurray, 2017; Creel, Aslin, & Tanenhaus, 2006; Farris-Trimble & McMurray, 2018; Magnuson et al., 2003) who would have significant experience with processing language. Many of the novel words used were constructed using phonemes present in English (Apfelbaum & McMurray, 2017; Farris-Trimble & McMurray, 2018; Magnuson et al., 2003). As a result, participants' experience with processing words incrementally in English could bias them towards processing these English-sounding novel words incrementally. If this is the case, participants may not extend this processing strategy towards less speech-like sequences.

Competition in Non-Linguistic Domains

The process of correctly identifying a word (either a real word or pseudo-word in an artificial lexicon) is often described as a process of resolving competition (Marslen-Wilson & Tyler, 1980; McClelland and Elman, 1986). However, the need to resolve competition is not just limited to speech perception. Some evidence of competition has been found with color (Farmer, Anderson, & Spivey, 2007; Huette & McMurray, 2010). Most work with color has focused on categorical perception, the idea of classifying things such as color as belonging to one of two distinct categories. Color exists on a continuum with different colors being defined as different ranges in wavelength. As a result, some color categories will be closer together in wavelength than others. For example, green and blue are closer in wavelength than green and red. This continuous property creates an opportunity for parallel activation of different color categories. For example, a greenish blue hue could briefly activate both green and blue. This then creates the need to resolve the competition between activated categories.

In a control condition for an experiment on language, Farmer, Anderson, and Spivey (2007) presented three colored squares (red, green, and greenish-blue) and instructed participants to click on the green square. They used mouse-tracking to examine the path the mouse took to the green square. They found greater curvature toward the greenish-blue box than the red box. This suggests the color categories of green and blue were activated together with the greenish-blue box being considered as a possible correct answer for that trial. Further evidence of competition between different colors was found by Huette and McMurray (2010). Participants classified colors along a green-blue spectrum as either green or blue. Mouse tracking data revealed that the closer a given hue was to the boundary between green and blue, the greater the curvature toward the competitor color. This provides evidence for a dynamic competition process that is sensitive to changes within the color category.

Evidence of competition between color suggests the resolution of competition is not limited to just speech. It also supports the notion of simultaneous activation of multiple candidates in a non-speech domain. However, it does not provide evidence for incremental processing, since both candidates are activated simultaneously by the static visual cue. In fact, several key differences exist between speech and color that may make it difficult to generalize real-time processing. First and most importantly, speech stimuli unfold over time, and the color swatches used in these studies do not.

The Present Experiment

The current study examines whether incremental processing, as seen in speech perception, can be observed in a non-speech domain. In this experiment, participants were trained to match novel tone sequences to novel objects, and subsequently tested for incremental processing using the VWP. The tone sequences were built to have overlap in the first three tones, to create cohort competitors. Like speech, these tone sequences unfold overtime. As a result, the identity of a particular sequence can only be known at the end of the sequence. Like color, the tone sequences are not speech-like. The evidence of competition between different color categories (Farmer, Anderson, & Spivey, 2007; Huette & McMurray, 2010) would suggest competition could be seen between the different tone sequences. However, competition in color categories cannot address the question as to the existence (or not) of incremental processing in non-speech domains.

In a second (control) condition, participants matched pseudo-words to the same objects. Pseudo-words had a similar structure to the tones (replacing tones with phonemes). The tone sequences do not sound like speech, and the pseudo-words serve as a speech-like comparison to the tone sequences. We expected to observe evidence for incremental processing for these items.

Participants were first trained to map the 16 tone sequences or pseudo-words to objects. For each trial, they heard a tone sequence or pseudo-word and were then asked to click on the corresponding object. Participants completed training blocks of two-alternative forced choice (2AFC) and four-alternative forced choice (4AFC) to help them learn the tone sequence-object or pseudo-word-object pairing. They then underwent testing trials using the VWP. If incremental processing is present in both conditions, eye-tracking results should show increased looks to cohort competitors when compared to unrelated items. If incremental processing is not present, eye-tracking results should not show any difference between cohort and unrelated competitors.

Methods

Participants

Participants consisted of 40 University of Iowa students who received course credit for compensation. They were native, monolingual English speakers with normal hearing and normal or corrected-to-normal vision (by self-report). Participants underwent informed consent according to an IRB approved protocol.

Design

There were two conditions: a melody condition and a pseudo-word condition. Participants were randomly assigned to one of the two conditions in a between-subject design. Participants first completed a familiarization task to acquaint them with the tone sequences or pseudo-words. This consisted of 32 trials where each of the 16 tone sequences or pseudo-words were presented twice. Participants then underwent training to help them learn the tone sequence-object or pseudo-word-object pairings. Training consisted of 928 trials with 208 trials of two-alternative forced choice (2AFC) and 720 trials of four-alternative forced choice (4AFC). Each tone sequence or pseudo-word was the target 13 times for the 2AFC portion and 45 times for the 4AFC portion. The final task participants completed was testing to examine the level of incremental processing. Testing consisted of 128 VWP trials with each tone sequence or pseudo-word appearing as the target eight times.

Stimuli

Tone Sequences. Sixteen tone sequences formed the auditory stimuli for the melody condition (Table 1). Each tone sequence contained six tones, two of which came from the C Major chord (C, E, and G). The other four tones came from notes outside of the C Major chord (e.g., A, A#, B, etc.). One of the tones from the C major chord appeared within the first three tones. Having tones appear from the C major chord gave all the tone sequences a common feel and harmonicity. This allowed for the slight feel of the tone sequences having a "key." No tone appeared more than once in a given tone sequences, so that each of the six tones were unique within the sequence. Each tone sequences contained at least one ascending and one descending interval (providing a contour cue that listeners could use instead of attending to the tones). This was done to prevent tone sequences from being entirely increasing or decreasing. In addition, each tone sequences had to contain an interval between consecutive tones that was greater than or equal to five half-steps, so that a given tone sequences did not have tones that only came from a limited range. Also, each transition between tones consisted of at least one whole step (two half-steps).

Each tone sequences had a counterpart, a cohort, for which the first three tones were identical but the last three tones were different (Table 1). With 16 tone sequences, there were eight cohort pairs and thus eight unique patterns for the first three tones. To ensure that the point of uniqueness could be easily perceived as such, the fourth tone of each cohort pair differed by at least six half-steps from fourth tone of its cohort pair. In addition, key intervals that are easily recognizable, such as an octave, or fifth, were avoided for the fourth tones of cohort pairs.

Sixteen offsets, the last three tones, were created and assigned to non-cohort pairs (Table 1). For offsets, we wanted some overlap (so that participants could not ignore the onsets), thus, we constructed eight pairs, which used the same three tones, but the order of the tones did not match exactly. For example, melody one contains the offset E4, D3, and F3, in that order. Those same tones appear in melody sixteen, but in the new order, D3, F3, and E4. Offset pairs were created so that no tone would be unique to a single melody. Creating offset pairs with the same tones but differing orders prevented unique tones while still allowing for unique endings.

Table 1: Structure of the tone sequences used in the experiment.

Melody	Cohort Pair	Cohort Tones			Offset Pair	Offset Tones		
1	1	A5	G4	C#5	1	E4	D3	F3
2	1	A5	G4	C#5	2	A#5	C4	G#4
3	2	E3	D5	F#3	3	D#4	G4	A3
4	2	E3	D5	F#3	4	C5	F5	B5
5	3	F4	A#3	G3	5	G#5	E5	D5
6	3	F4	A#3	G3	6	F#4	B3	C3
7	4	G#3	C4	A4	7	G5	D#3	A#4
8	4	G#3	C4	A4	8	C#3	D4	G3
9	5	B4	F#5	E5	7	A#4	D#3	G5
10	5	B4	F#5	E5	2	C4	A#5	G#4
11	6	C3	B3	D#3	5	E5	D5	G#5
12	6	C3	B3	D#3	8	G3	D4	C#3
13	7	G5	D#5	A#4	4	F5	B5	C5
14	7	G5	D#5	A#4	6	C3	B3	F#4
15	8	C#4	G#4	C5	3	G4	A3	D#4
16	8	C#4	G#4	C5	1	D3	F3	E4

Tones were selected from a three-octave range from C3 to B5. Each of the notes from the three octaves were balanced across all tone sequences so that each octave had the same number of notes present in the tone sequences. In addition, the first half (cohort portion) and second half (offset portion) of the tone sequences also contained even numbers of notes from the three octaves. Each individual note was present a minimum of two times and a maximum of four times across all tone sequences. In addition, an individual note was present a maximum of two times in each half of the tone sequences. Each tone of a tone sequences played for 115 msec. In addition, approximately 11.6 msec of silence was added between each tone and at the beginning and end of the melody. In total, each tone sequences lasted 771 msec, 690 msec of tones sounding and 81 msec of silence. Each tone sequence had exactly one exemplar.

Pseudo-words. Sixteen pseudo-words formed the auditory stimuli for the word condition (Table 2). All pseudo-words were recorded by a male, native English speaker. One exemplar of each pseudo-word was used. This was done so that both the melody and word conditions contained the same number of exemplars.

Table 2: Pseudo-words used. The International Phonetic Alphabet (IPA) notations for each syllable appears in the last two columns.

Word	Orthographic Form	Cohort IPA	Offset IPA
1	Vamgeer	/væm/	/gir/
2	Vamsot	/væm/	/sɑːt/
3	Koshveed	/kɑːʃ/	/vid/
4	Koshpith	/kɑːʃ/	/pIθ/
5	Weermaf	/wir/	/mæf/
6	Weerbik	/wir/	/bik/
7	Gislaz	/gIs/	/læz/
8	Gisnauf	/gIs/	/nɑːf/
9	Pauntos	/pɑːn/	/tɑːs/
10	Paunkib	/pɑːn/	/kip/
11	Bizdeev	/bIz/	/div/
12	Bizfam	/bIz/	/fæm/
13	Deengthip	/diŋ/	/θIp/
14	Deengzal	/diŋ/	/zæl/
15	Talreeg	/tæl/	/rig/

For the pseudo-words, phonemes replaced the tones from the tone sequences. Each pseudo-word had two syllables and contained six phonemes in a CVCCVC format. No phoneme was repeated within a pseudo-word. While not English words, all pseudo-words obeyed phonotactic rules of the English language. Consonants used were present a minimum of two times and a maximum of four times across all pseudo-words. Vowels used were present a minimum of four times and a maximum of eight times across all pseudo-words. 50 msec of silence was added to the beginning and end of each pseudo-word. The average length of each pseudo-word was 794 msec, including the silence, and 694 msec without the silence.

Like the tone sequences, each pseudo-word had a counterpart (cohort) that began with the same three phonemes and differed in the last three (Table 2). With the 16 pseudo-words, there were eight cohort pairs and thus eight unique patterns for the first three phonemes. Each pseudo-word began with a stop constant. In addition, the third phoneme (second consonant) of each pseudo-word was a fricative. Having a fricative as the third phoneme allowed the transition between the first and second syllables be more consistent with typical patterns seen in English.

Eight offsets (last three phonemes) were created and assigned to non-cohort pairs (Table 2). There was no overlap in phonemes of the offsets in cohort pairs. Like the tone sequences, the order of the phonemes in offsets pairs did not match exactly. For example, in word one the offset was *geer* (/gir/). In word 15, the same offset appears again but in the reverse order, *reeg* (/rig/). To maintain the offset format of CVC, all offsets appear in the reverse order in the offset pair. As with the melodies, offset pairs were created so that no phoneme would be unique to a single pseudo-word.

Visual Stimuli. The visual stimuli consisted of 16 novel objects on a white background that were unfamiliar to participants and could not be easily named (Figure 1). Objects were randomly assigned to tone sequences or pseudo-words such that each participant had different tone sequence-object or pseudo-word-object pairings.

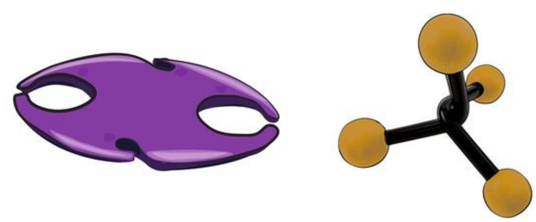

Figure 1. Examples of the novel objects used

Procedure

Familiarization. Participants first heard the tone sequences or pseudo-words being used in the experiment. This was done so that participants could be familiarized with the tone sequences or pseudo-words prior to having to learn to map them to referents. Participants were instructed to press the spacebar to initiate each tone sequence or pseudo-word. While the tone sequence or pseudo-word played, participants saw a blank screen for 1000 msec per tone sequence or pseudo-word. During the 1000 msec, pressing the spacebar, or any other keyboard keys, would not end the trial. This was done to ensure that participants listened to the entire tone sequence or pseudo-word. All tone sequences or pseudo-words played twice, in a random order for a total of 32 trials.

Training. The training paradigm was loosely based on that of Magnuson et. al. (2003) which included both two-alternative forced choice (2AFC) and four-alternative forced choice (4AFC) training. Thus, training was broken up into two portions: 208 trials of 2AFC followed by 720 trials of 4AFC for a total of 928 training trials. Participants were told they would learn tone sequence-object or pseudo-word-object pairs and that their job would be to match the tone sequence or pseudo-word to the correct novel object. They were also told they would need to guess at first, but that their response should come more informed over time.

For 2AFC trials, participants saw two novel objects along with a blue circle in the center of the screen. Both objects were 300 by 300 pixels and equidistant from the center. They appeared 50 pixels from the outside edge of the screen and were vertically aligned in the center of the screen. After 1000 msec, the circle became red. Once the circle became red, participants needed to click on it. Upon clicking on the red circle, a tone sequence or pseudo-word played, and the circle disappeared. The participant then clicked on one of the objects, ending the trial. Fifty msec after clicking on the object, participants received feedback on the correctness of the trial. If the object they clicked on was correct for the tone sequence or pseudo-word they heard, “Correct” would appear in the middle of the screen. Conversely, “Incorrect” would appear on the screen if the object clicked on was incorrect. All feedback screens appeared for 350 msec.

The 208 2AFC trials were divided into two blocks of 104 trials. On a given trial, the novel objects seen consisted of the target objects and a random foil object. For each of the 16 tone sequence/object or pseudo-word/object pairs, 13, out of 15 possible foils, were randomly selected to appear as a competitor object. For each tone sequence or pseudo-word, the novel object paired with the cohort tone sequence or pseudo-word had to be one of the 13 visual competitors. Each tone sequence or pseudo-word was a target a total of 13 time appearing with each of the 13 novel objects only once. The order of the trials was random.

4AFC trials were identical to 2AFC trials with the exception that there were four novel objects on the screen instead of two. One object appeared in each of the four corners and were equidistant from the circle in the center. The 720 4AFC trials were divided into six blocks of 120 trials each. Each tone sequences or pseudo-words was the target 45 times. Competitors for each trial were selected randomly, with replacement, with all novel objects appearing as competitors a roughly equal number of times.

Testing. After training, participants were tested and were moved to a different room for eye-tracking. They then through a standard 9-point calibration.

Next, participants completed 128 trials of testing. Testing trials were identical to 4AFC training trials with a few notable differences. First, participants received no feedback after object selection. Second, the same four objects always appeared together on a given trial. This was done so that the frequency at which objects did or did not occur together was the same. If the target object on each trial always appeared with the cohort object and different unrelated objects, this could indicate to participants that the two objects had a relationship and could affect their response. By having the same objects always appear together, participants would be unable to infer the objects that were cohort pairs based only on sight. With 16 objects, there were four different sets of four objects that participants saw. Within the sets of four objects, there were two object pairs. For example, melodies 1 and 2 as well as 3 and 4 were cohort pairs. The objects associated with melodies 1, 2, 3, and 4 would appear together whenever melody 1, 2, 3, or 4 was played.

Eye-Tracking recording and analysis. Throughout testing, eye movements were tracked using a head mounted SR Research Eyelink II eye-tracker. When possible, corneal reflection and pupil were used to determine where participants were looking. With some participants, only pupil could be obtained. Both eyes were tracked when possible, but only the eye with the better calibration was used. For analysis, we used Eyelink Anal version 3.31.a (McMurray, 2018). We collapsed saccades and fixations into a single look such that the beginning of one saccade until the end of the next fixation formed a single look. When examining where a look was located, the boundaries of the ports were extended 100 pixels past the edge of the object. This did not result in any overlap between the regions of interest.

Results

Out of the 40 participants, the data from 37 were included in the final analysis. One was excluded due to not complying with the research task. An additional two were excluded due to accuracy scores of less than 50% during testing. All three excluded were in the melody condition.

Training

Figure 2 shows the accuracy scores for both conditions from the first block of four-alternative forced choice (4AFC) training (Block 1) to testing (Block 7). Prior to block 1 of the 4AFC training, participants completed two blocks of two-alternative forced choice (2AFC) training. During the 2AFC trials, participants displayed rapid learning that allowed participants in the word condition to approach ceiling and to be just above 60% in the melody condition. Accuracy scores between the melody and word conditions show several differences. First, the overall accuracy scores across all blocks are lower in the melody condition compared to the word condition. Second, the melody condition shows a gradual increase in accuracy throughout the blocks. In particular, the melody condition shows an increase of almost 10% between the last 4AFC training block (block 6) and testing (block 7). The word condition shows little increase in

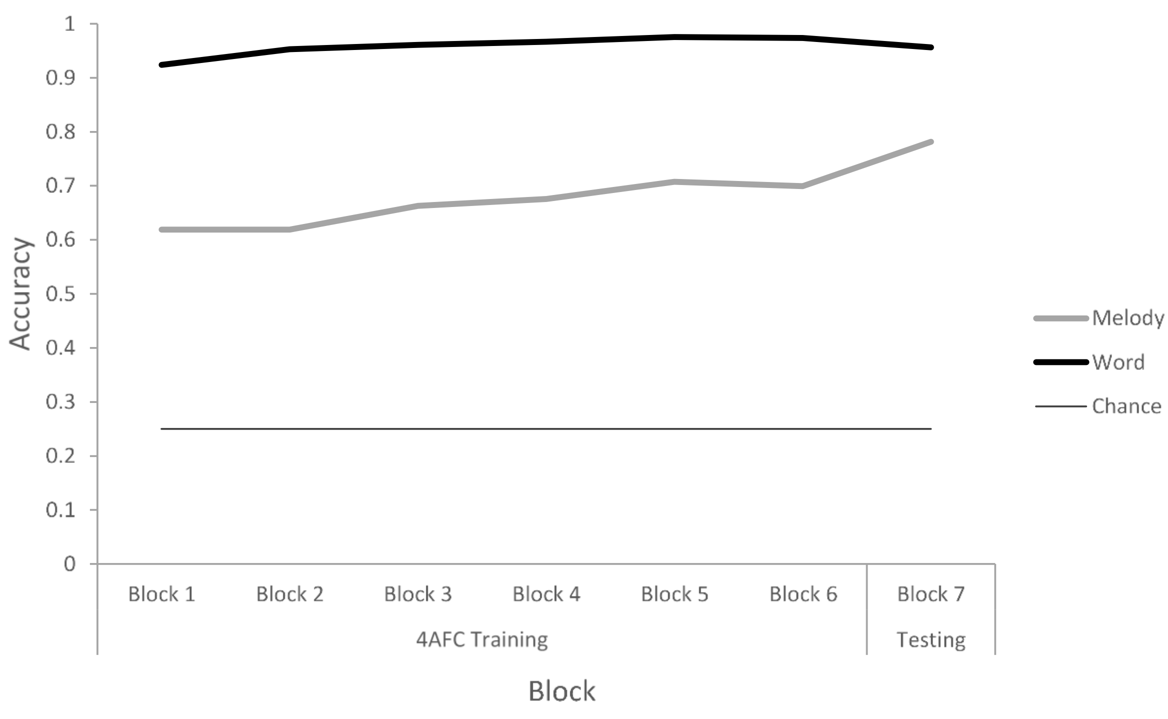

Figure 2: Accuracy across different blocks. Block 1 is the first block in 4AFC training.

accuracy levels during the blocks. This can probably be attributed to the accuracy scores being fairly close to ceiling levels by the first 4AFC training block of the word condition.

A repeated measures ANOVA was conducted on the accuracy data with block as a within-subjects factor and condition as a between-subjects factor. The results showed a significant effect of block ($F(6,210) = 16.908$, $p<.001$). This was due to the fact accuracy scores increased across the blocks. There was also a main effect of condition ($F(1, 35) = 66.426$, $p<.001$) with the word condition having higher accuracy scores than the melody condition. Finally, a significant interaction between block and condition was found ($F(6) = 9.376$, $p<.001$).

To understand this interaction, a repeated measures ANOVA was conducted examining the effect of block separately for the two conditions. A significant effect of block ($F(6,96) = 13.004$, $p < .001$, $\eta_p^2 = .448$) was found in the melody condition. For the word condition, a significant effect of block ($F(6,114) = 5.653$, $p < .001$, $\eta_p^2 = .229$) was also found. This suggests that accuracy scores in the melody and word conditions significantly increased over the different blocks.

Testing

For the eye-tracking data during testing, only correct trials were examined. Fixation proportions for target, cohort, and unrelated items were calculated at each four-millisecond slice separately in each condition (Figure 3). In the melody condition, results show fixation proportions for the target begin to show a difference around 600 msec after the onset of tone sequence (Figure 3A). In addition, cohort and unrelated items do not seem to have much of a difference throughout the time course in the melody condition and reach around .125 fixation proportions at its peak. In contrast, the word condition shows the fixation proportions for the unrelated item beginning to differ from the target around 500 msec after the onset of the pseudo-word and the cohort beginning to differ from the target around 700 msec. In the word condition, the cohort reaches around .2 fixation proportions at its peak, and the unrelated item reaches around .125 fixation proportions at its peak.

When comparing fixation proportions of the target in the melody condition and the target in the cohort condition, several differences emerge (Figure 3C). First, the target for the word condition receives higher fixation proportions earlier on then the target for the melody condition. For example, fixation proportions for the target in the word condition reach a level of .2 around 600 msec, and the target in the melody condition does not reach a level of .2 until around 900 msec. Second, the target in the word condition receives higher fixation proportions at the end of the time course than the target in the melody condition. At the end of the time course, the target in the word condition reaches a fixation proportion level of around .9. At the same time, the target in the melody condition has reached a level of around .65.

To assess these findings statistically, we computed the Area Under the Curve (AUC) from 400 to 2400 msec and analyzed this with ANOVA.

We started by examining the fixations to the cohort and unrelated items. Here, incremental processing would predict greater fixation to the cohort than unrelated items; if listeners are not processing incrementally, we predict no difference between cohort and unrelated items will be found. A 2 (condition) x 2 (object type [cohort and unrelated]) ANOVA was run to compare AUC between cohort and unrelated items across the two conditions (Figure 3A,B, 4). Results showed a significant main effect of object type ($F(1) = 30.838$, $p<.001$) with cohort competitors receiving more fixations than unrelated items. Condition was not significant ($F(1) = .837$, $p = .367$). This says that number of fixations did not differ between the melody and word conditions. In addition, a significant interaction was found between object type and condition (F

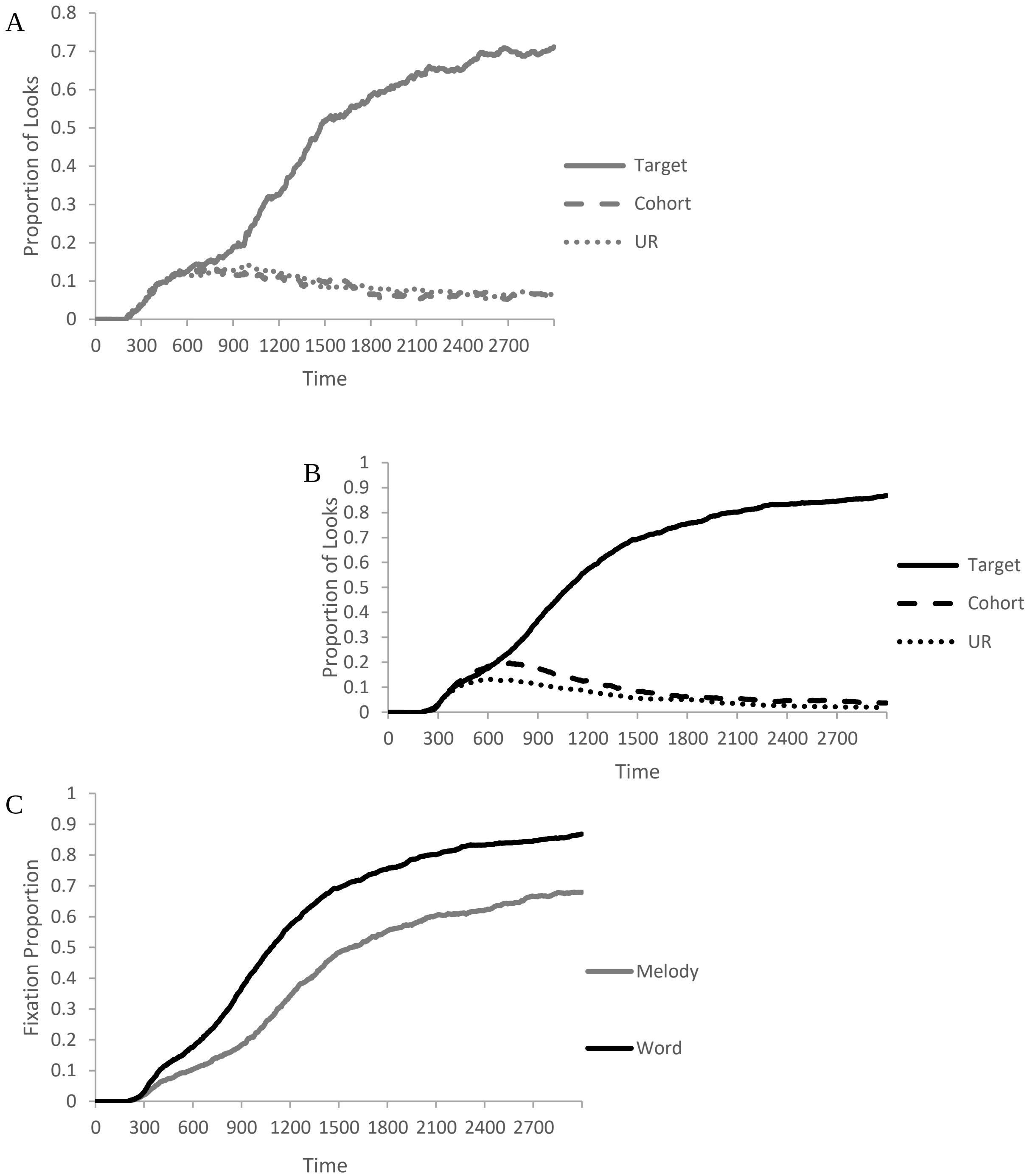

Figure 3: The Panel A shows fixation proportions for target, cohort, and unrelated items in the melody condition across time. The Panel B shows fixation proportions for target, cohort, and unrelated items in the word condition across time. The Panel C shows fixation proportions for the target across both conditions.

(1) = 30.668, p<.001). This was because the distribution of fixations for cohort and unrelated items differed between the melody and word conditions.

To further examine this interaction, a repeated measures ANOVA was conducted comparing the cohort and unrelated items for the melody and word conditions separately. For the melody condition, no significant difference was found between cohort and unrelated items (F (1,16) < .001, p = .991, $\eta_p^2 < .001$). For the word condition, a significant difference was found between the cohort and unrelated items (F (1,19) = 56.690, p < .001, $\eta_p^2 = .749$). This reveals that there was not a detectable difference in fixation proportions between the cohort and unrelated items in the melody condition, but there was a difference between the two items in the cohort condition.

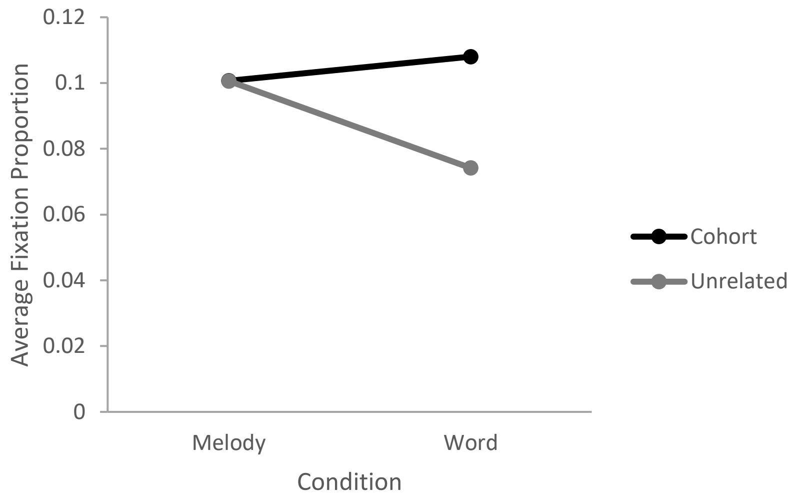

Figure 4: Average fixation proportion of Cohort and Unrelated items across conditions.

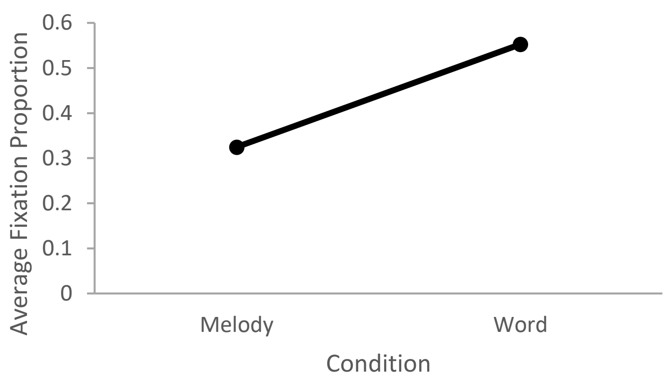

Figure 5: Average fixation proportion of target items across condition.

Next, we examined the target to see if there were any differences in fixation proportions between the conditions. A univariate ANOVA was conduction on the effect of target across conditions (Figure 5). A significant effect between the melody and word conditions was found ($F(1) = 56.079$, $p<.001$). This says that fixation proportions to the target differed between the melody and word conditions.

Discussion

This study examined incremental processing in a non-speech domain. Participants were trained to match tone sequences or pseudo-words to novel objects and then tested for incremental processing in the VWP. We start by discussing the accuracy results and then turn to the eye-tracking results.

Accuracy

Between the melody and word conditions, an interaction between block and condition was found with accuracy scores. This would suggest a learning difference between the two conditions with the words being far easier to learn. In addition, accuracy scores at the end of training were significantly higher in the word condition than in the melody condition. This suggests that participants in the word condition ultimately learned the pseudo-word-object pairings better than the participants in the melody condition learned the tone sequence-object pairings.

There were also clear differences early in training. Prior to the first block of four-alternative forced choice (4AFC) training blocks, participants completed two blocks of two-alternative force choice (2AFC) blocks where rapid learning was found. The first training block of 4AFC trials showed accuracy scores of 62% for the melody condition and 92% for the word condition. While participants did seem to continue to learn the tone-sequences during the 4AFC training trials, as shown by a significant effect of block for both conditions, the bulk of learning the tone sequence-object or pseudo-word-objects pairs seemed to occur during the 2AFC trials.

Within the accuracy scores for the melody condition, an almost 10% increase in accuracy was found between the last block of 4AFC training (block 6) and the testing block (block 7). One explanation for this is the elimination of offset pairs in testing. During 4AFC training, all objects appeared together at some point, including the object associated with offset pair for the target of a given trial. During testing, offset pairs never appeared together, and thus, the opportunity to mistake the two for each other, never occurred. The target tone sequence was mistaken as the offset object about 3% of the time across all 4AFC training trials. If similar errors were made between training trials and testing trials, this mistake does not account for all of the difference between block 6 (training) and block 7 (testing). Another possible explanation lies with another difference between training trials and testing trials. During testing, the same four objects always appeared together and could be broken down into two cohort pairs (e.g. melody 1 and 2 as well as 3 and 4 were cohort pairs and the objects that were associated with those tone sequences always appeared together during testing). During 4AFC training, there were no restrictions in what objects appeared together. As a result, sequences that contained tones closer together as well as similar contours could appear together. During testing, pairs were chosen in such a way that minimized both overlap and contour. By eliminated the objects associated with similar tones and/or contour during test, possible incorrect tone sequence-object mappings were also eliminated.

Accuracy scores for the melody condition never approached ceiling levels. In contrast, accuracy scores for the word condition were fairly close to ceiling levels by the first 4AFC training block. This difference between the two conditions could potentially be attributed to difference in experience levels. While participants had never heard the pseudo-words used in the word condition, phonemes for the pseudo-words came from the English language and resembled English words. The adult participants would have a significant amount of experience learning and processing English words. As a result, previous experience learning word-object pairings could have helped boost learning in the word condition. In the melody condition, participants had little to no experience with anything similar to the tone sequences heard. For those in the melody condition, previous experience learning word-object pairs would not probably help with learning the tone sequence-object pairs because the tone sequences do not resemble speech.

Eye-tracking results

In addition to these differences in accuracy, large differences were also observed in the eye-tracking results. If both condition were processed incrementally, we would have expected to find increased fixation proportions to the cohort compared to the unrelated items. If no incremental processing was found, we would have expected to find no difference in fixation proportions in the cohort compared to the unrelated items. In the word condition, increased fixation proportions to the cohort was found over unrelated items. This indicates incremental processing and replicates previous eye-tracking findings with other artificial lexica (Magnuson et al. 2003) as well as findings with known words (Alloppena et al., 1998). In the melody condition, no difference in fixation proportions was found between the cohort and unrelated items. This suggests the tone sequences were not processed incrementally and is more consistent with the results found in McMurray et al. (2017) for prelingually deaf cochlear implant users who show less incremental processing.

Could the differences in accuracy have led to the differences in incrementality? A better understanding of the pairings between the objects and the auditory information might lead to better representations and links between the objects and the auditory information. In turn, this might help process information more incrementally because there is a greater understanding and knowledge about the sounds. However, visual examination of eye-tracking data from participants who had accuracy score during testing of greater than or equal to 90%, in the melody condition, showed no visual difference between cohort and unrelated items. However, only four people fit this criterion. As a result, there is not enough statistical power to be able to rule out differences in learning as an explanation for the results found.

Beyond the competitor fixations, a crucial difference between the melody and word conditions was the time course of fixations to targets. In the word condition, fixation proportions increased much earlier than in the melody condition. In addition, in the word condition, participants reached much higher levels of fixation proportions at the end of the time window than in the tone-sequence. This could be a result of differences in accuracy scores that affect overall levels of confidence. If a participant is unsure of which object is paired with the tone sequence they are hearing, they overall look less at any of the displayed objects and wait until they have heard more of the tone sequence (and thus have more information about the sequence) to begin looking at the displayed objects. Participants in the melody condition could be waiting until the uniqueness point (located at the fourth tone, approximately 386 msec into the tone sequence) to begin to initiate fixations. Future research could explore reasons behind the observed differences between fixation proportions. Perhaps one way this could be done would be by bolstering accuracy within tone sequences (and thus eliminating the learning difference between the melody and word conditions in the present experiment).

Overall, the results do not support incremental processing within the non-speech domain of the tone sequences. While some of the findings could be influenced by learning differences between the two conditions, this is not the only possible explanation. Incremental processing could be unique to speech. If this were the case, no evidence of incremental processing would be seen outside of speech, or speech-like, domains. Evidence of parallel activation and competition does exist within color, a domain outside of speech (Farmer, Anderson, & Spivey, 2007; Huette & McMurray, 2010). However, differences between color and speech prevent the competition that is found in color to be called incremental processing. Further evidence is needed within domains similar to the tone sequences to conclude incremental processing is not present.

References

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.

Apfelbaum, K. S., & McMurray, B. (2017). Learning During Processing: Word Learning Doesn't Wait for Word Recognition to Finish. *Cognitive Science, 41*(S4), 706-747.

Brouwer, S. & Bradlow, A. R. (2014). Context Variability During Speech-in-Speech Recognition. *The Journal of the Acoustical Society of America*, *136*(1), 26-32.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language, 54*(1), 1-19. doi:10.1016/j.jml.2005.09.003

Farmer, T. A., Anderson, S. E., & Spivey, M. J. (2007). Gradiency and Visual Context in Syntatic Garden-Paths. *Journal of Memory and Language, 57*(4), 570-595.

Farris-Trimble, A., & McMurray, B. (2018). Morpho-phonolgical regularities Influence the Dynamics of Real-Time Word Recognition: Evidence from Artificial Language Learning. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 9*(1), 2.

Huette, S. & McMurray, B. (2010). Continuous Dynamics of Color Categorization. *Psychonomic Bulletin & Review,* 17(3), 348-354.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The Time Course of Spoken Word Learning and Recognition: Studies with Artificial Lexicons. *Journal of Experimental Psychology: General, 132*(2), 202-227.

Marslen-Wilson, W. D. (1987). Functional Parallelism in Spoken Word Recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. & Tyler, L. K. (1980). The Temporal Structure of Spoken Language Understanding. *Cognition*, *8*(1), 1-71.

McClelland, J. L. & Elman, J. L. (1986) The TRACE Model of Speech Perception. *Cognitive Psychology, 18*(1), 1-86.

McMurray, B. (2018) *EyelinkAnal*. Available from: http://osf.io/c35tg.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for Lexical Access: Cochlear Implants or Sevverely Degraded Input Lead Listeners to Process Speech Less Incrementally. *Cognition, 169*, 147-164

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous Attraction Toward Phonological Competitors. *Proceedings of the National Academy of Sciences of the United States, 102*(29), 10393-10398.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, Julie, C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science, 268*(5217), 1632-1634

Zwitserlood, P. (1989). The Locus of the Effects of Sentential-Semantic Context in Spoken-Word Processing. *Cognition 32*(1), 25-64