Kicking or Being Kicked? Using Acoustics as Unconscious Cues to the Passive

Melinh K. Lai

Interdisciplinary Honors Thesis

A thesis in partial fulfillment of the requirements

of the

Interdisciplinary Honors Thesis

Written under the direction of

Karin Stromswold, PhD, MD

Department of Psychology, Rutgers Center for Cognitive Science

and

Paul de Lacy, PhD

Department of Linguistics

School of Arts and Sciences, Rutgers University

2014-2015
**Introduction**

In English, even well-formed sentences can have multiple or ambiguous meanings. Consider the sentence *The boy saw the man with the binoculars*. This sentence has two potential interpretations. In one interpretation of the sentence, the boy uses the binoculars to see the man; in the other, the man has the binoculars and the boy simply happens to see him. Therefore, depending on how the sentence is parsed, the owner of the binoculars can be either the boy or the man. The sentence is an example of *permanent ambiguity* because even after hearing the entire sentence, it is not clear whether the speaker intended the boy or the man to be the owner of the binoculars.

Well-formed sentences can also be *temporarily ambiguous*. In a temporarily ambiguous sentence, the intended meaning may not initially be clear. However, as more of the sentence is heard, it becomes apparent that only one parse can accurately describe the full sentence, and then the sentence’s meaning becomes unambiguous. Some notorious examples of temporarily ambiguous sentences are called *garden-path sentences* because the syntactic structure of the sentence can initially mislead a listener to the wrong interpretation before “redirecting” the listener to another. In garden-path sentences, the initial parse of a sentence will generally be the more intuitive one;
it will also be wrong. An example of a garden-path sentence is *The government plans to raise taxes were defeated*. A person hearing this sentence might initially believe that the speaker is referring to a current plot by the government to raise taxes (i.e., that *government* is being used as a noun and *plans* as a verb). Upon encountering the word *were*, however, the listener would have to reevaluate their initial parse of the sentence before eventually realizing that *government* is actually being used as an adjective and *plans* as a noun. In this way, the listener is forced to backtrack in order to reevaluate the sentence’s syntax, as if retracing his or her steps on a winding and unfamiliar garden path.

![Figure 2. The two parses that are formed for the sentence *The government plans to raise taxes*. The initial parse (A) is found to be ungrammatical upon encountering the word *were*, and the parser then forms a new parse (B).](image)

In some sense, *all* sentences are temporarily ambiguous, since a well-formed sentence’s meaning will not be clear to the listener until enough information has been processed to arrive at the correct parse. Sentence processing could be off-line or on-line. In *off-line processing*, listeners wait to hear the entirety of a sentence before beginning to parse it. However, converging evidence, including the existence of garden-path sentences, indicates that people do not process sentences off-line. In contrast, the *on-line processing* model posits that listeners actively process linguistic information as they receive it. Put another way, they are actively piecing together a potential interpretation of the sentence as they hear each word. In on-line sentence processing,
listeners are sometimes able to arrive at the correct parse of a sentence before they even hear the sentence in its entirety.

What, then, are the mechanisms by which people are able to correctly predict the correct parse for a temporarily ambiguous sentence? There are a number of sentence processing models that attempt to provide an answer to this question. One type of model posits that syntactic parsing occurs in two stages, and is therefore appropriately known as a two-stage model. In the first stage, a parser sorts the incoming sequence of words into phrase categories and then builds a plausible syntactic structure. This process is completely insensitive to any contextual information and relies solely on the categorization of the words. Meaning is computed in the second stage, when semantic rules are applied to the proposed syntactic structure (Ferreira & Clifton, 1986; Ferreira & Henderson 1990). The garden-path model is an example of a two-stage sentence processing model (Frazier, 1979; 1987). It works under the assumption that a parser begins to build a syntactic structure as he or she receives the stream of words; when the parser encounters a word in the stream that does not “fit” into the syntactic structure that was initially formed, a new syntactic structure is formed. Two stage models like the garden-path model use many various structural heuristics, such as the late closure, minimal attachment, and main assertion heuristics to parse sentences. The late closure heuristic tells a parser to continue a syntactic structure from the phrase or clause that is currently being processed. The minimal attachment heuristic dictates that a parser should build a syntactic structure with the fewest nodes possible. The main assertion heuristic says that when two structures are possible, the parser should build the one where new elements attach to the main assertion of the sentence, as opposed to, say, a relative clause. Experimental evidence suggests that these heuristics are deployed in flexible ways and are sometimes dependent on the thematic assignments of the second stage to correct potential errors (Traxler & Frazier, 2008). Because these heuristics apply
only to the structural components of sentence processing, such as syntax or morphology, non-structural information such as word meaning can be processed more modularly, and the second stage allows for more flexibility in the final parse.

The most prominent alternative to two-stage models are constraint satisfaction models, where syntactic parsing and semantic interpretation occur simultaneously (MacDonald et al., 1994; Spivey-Knowlton & Sedivy, 1995, Tanenhaus et al., 1995; Trueswell et al., 1993). In constraint satisfaction models, information from multiples sources comes together to constrain the parser’s interpretation. These sources of information include lexical biases, such as word frequency in a given context, frequency of a word class in a context, and (in some models) the argument structure of the verb. Referential and visual context can also be a source of information that a parser uses to constrain the number of potential interpretations of a sentence. Yet another source of information is the prosody with which a sentence is uttered. Many studies have shown that the intonation—the rhythm, stress, or pitch of sentences—of an ambiguous sentence can change depending on a speaker’s intended meaning. In addition, listeners tend to correctly interpret a sentence by picking up on the prosodic cues of a person’s speech (see, for example, Barrow et al., 2005; Beach, 1991; Beach et al., 1996; Birch and Clifton, 2002; Carlson et al., 2001; Dahan et al., 2002; Katz et al., 1996; Kielgaard and Speer, 1999; Kraljic and Brennan, 2005; Lehiste, 1972, 1973; Lehiste et al., 1976; Schaefer et al., 2000; Snedeker and Trueswell, 2003; Speer et al., 1996).

However, a review of relevant databases reveals little research has investigated the role of more local acoustic cues such as pitch, amplitude, or morpheme duration in disambiguating sentences. A study conducted by Stromswold et al. (under review) suggests that adult listeners may attend to acoustic cues to help predict syntactic structure. The stimuli in Stromswold et al.’s study were active sentences like The girl was push-ing the boy, and passive sentences like The girl
was push-ed by the boy. Notice the two sentences are morphosyntactically identical up until the suffix of the verb (the –ed of pushed and the –ing of pushing). In this study, adult participants were asked to listen to the active and passive sentences before choosing one of two pictures on a screen that best matched the meaning of the sentence. Eye-tracking equipment captured the participants’ eye movements as they completed the task. Analyses of eye-gaze patterns revealed that adults began to look at the correct picture prior to hearing the verb ending, indicating that they “knew” whether a sentence was active or passive even though the part of the sentence they were hearing was still syntactically ambiguous.

One possible explanation as to how the participants were able to consistently look at the correct picture before hearing the verb participle is that the audio recordings used by Stromswold et al. contained intonational cues that were either consciously or unconsciously provided by the original speaker. These cues would then have allowed the participants in the eye-tracking study to predict the syntactic structure of the sentence (i.e., whether it was active or passive). However, analysis of the intonational contours of the active and passive sentences used by Stromswold et al. via standard ToBI transcription conventions showed no consistent patterns or differences between the sentences. In addition, there were no systematic intonational differences in the first noun phrase (NP) or the second NP in both the active and passive sentences. Put another way, there were no distinct intonational patterns that differentiated active sentences from passive sentences. However, examination of the duration of individual morphemes revealed that there were no significant differences in the length of duration for the first the, the first noun, or the auxiliary was (in both active and passive sentences). More importantly, the active verb stems (e.g., the push in the girl was push-ing the boy) were about 35 milliseconds shorter than the passive verb stems (e.g., the push in the girl was push-ed by the boy).
Why, then, were verb stems of the active sentences used by Stromswold et al. significantly and consistently shorter than the verb stems of the passive sentences? In their discussion of Stromswold et al.’s stimuli, Gahl and Garnsey (2004) suggest that the shorter duration of active verb stems conforms to a phenomenon where words shorten when they are used highly frequently (Hooper, 1976) or when they have a high probability of occurring in a given context (Jurafsky et al., 2001). Thus the shorter verb stems of active sentences reflect the tendency to shorten words that are highly frequent or highly probable. However, Stromwold et al. provide several reasons to believe that neither explanation can adequately describe the cause of the durational differences. First, because the progressive auxiliary was (as in the active verbal phrase was pushing) is more frequent and more probable than the passive auxiliary was (as in was pushed), the explanation posed by Gahl and Garnsey predicts that the progressive auxiliary should be shorter than the passive auxiliary; as previously stated, Stromswold et al. found no significant differences in length between the two auxiliaries. Second, Stromwold et al. also found that the second the (as in the determiner used in the second NP) of their active sentences was significantly shorter than in their passive sentences. This occurrence cannot be explained by the frequency or probability of the word, since it is the same word being used in the same way in both sets of sentences.

Stromwold et al. propose three phonetic explanations for why passive verb stems are longer than active verb stems. The first possibility is that the lengthening reflects the fact that speakers tend to lengthen the final syllable before the start of a new prosodic phrase (Scott 1982, Turk & Shattuck-Hufnagel 2007). Transitive verbs, such as kiss, can be the final word of a passive phrase (e.g., The boy was kissed). However transitive verbs are not felicitous as the final word of an active verb phrase (e.g., *The boy was kissing). Given that prosodic phrase boundaries tend to align with syntactic phrase boundaries (Kreiman 1982; Lehiste & Wang 1977), it is not surprising
that in Stromswold et al.’s study, passive verb stems were prosodic phrase-final in over 70% of the passive stimuli, whereas active verb stems were phrase-final in only 50% of the active stimuli. This could have caused passive verb stems to sound longer than active verb stems. The second phonetic explanation relates to an observation that the stems of monosyllabic words are longer than the stems of polysyllabic words (Beckman and Edwards 1990, Lehiste 1972). Therefore, the stems of active verbs are shorter because the progressive inflection -ing forms its own syllable, whereas the passive inflection can often be realized as an alveolar stop (/t/ or /d/).

Another possible reason that active stems were shorter than passive stems is that speakers have a tendency to lengthen or shorten utterance in order to maintain a constant average predictability (Aylett and Turk, 2004; Levy and Jaeger, 2007; Jaeger, 2010). The combination of a longer duration of the following suffix (-ing vs. -ed) and the fact that actives are more frequent and probable than passives (COCA, Davies 2008), might result in speakers choosing to reduce the length of the active verb stems. Regardless of the reason behind the passive verb stem lengthening, the fact remains that passive verb stems in Stromswold et al.’s stimuli were consistently longer than active verb stems and adults’ eye gaze during a comprehension study showed that they looked at images depicting the correct interpretation of active and passive sentences before they even heard the verb ending. This suggests that adults could predict whether a spoken sentence was active or passive while it was still syntactically ambiguous. To determine the robustness of this finding, a follow-up gating study was performed in which participants listened to the beginning fragments of sentences (e.g., the girl was push—) and then had to guess whether the fragment was part of an active or passive sentence (e.g., choosing between the girl was push-ing the boy or the girl was push-ed by the boy). Although participants said that they were guessing randomly, the researchers found that they correctly predicted
whether a sentence would be active or passive about 85% of the time (Stromswold et al., under review).

Two sets of questions remain regarding the relationship between acoustics and syntax. The first concerns how robust these acoustic findings are. For example, do all speakers provide acoustic cues for syntactic structure and do all speakers provide the same cues? In another production study conducted by Rehrig et al. (2015), seven monolingual American English speakers were recorded saying a variety of active and passive sentences. These target sentences had similar syntactic structures as the sentences from Stromswold et al.’s first study in that corresponding active and passive sentences were temporarily ambiguous up until the verb participle –ed/-ing. Using Praat software (Boersma & Weenink 2015), Rehrig et al. marked boundaries between individual morphemes in each sentence and then performed analyses of F0, intensity, and duration. Consistent with Stromswold et al.’s findings, they found that, for all speakers and for all 16 verbs, passive verb stems were longer than active verb stems, indicating that passive verb stem lengthening is a robust phenomenon.

The second set of questions concerns the usefulness of these cues. Are all listeners equally sensitive to acoustic cues when they are present? How salient must they be for people to use them? What happens when such cues are not present? To answer these questions, using the sentences said by one of Rehrig et al.’s participants, I conducted a gating comprehension experiment to determine whether monolingual English speakers were able to guess whether a sentence was active or passive when the sentence was truncated at the verb stem.
Methods

Participants

A total of forty Rutgers University students participated in the experiment. Two were eliminated because they performed at chance level on “catch trials” (see below). The remaining thirty-eight participants were monolingual American English speakers with normal hearing and normal or corrected-to-normal vision.

Stimuli

The experiment consisted of 100 trials in which participants listened to truncated recordings of active and passive sentences before choosing one of two sentences on a monitor that they believed to be the best completion of the truncated audio. The 100 trials consisted of 32 target active sentences, 32 target passive sentences, 24 unambiguous catch trials, and 12 ambiguous fillers. The speaker in the audio recordings was a participant in a production study conducted by Rehrig et al. during the summer of 2014, who is a native female speaker of English with a mid-Atlantic American accent and was naïve to the purposes and findings of Rehrig et al.’s study. Sentences were recorded digitally using a high-quality head-mounted AKG C420 close talk microphone and a Marantz PMD 670 solid state receiver in a sound-attenuated booth at Rutgers University. Recordings were saved as WAV files at 44.1 kHz 16-bit mono.

The verbs and nouns in the sentences were chosen for their ability to be used in semantically reversible sentences so that the same noun and verb pair could be used in both an active and a passive sentence. The sentences used by Rehrig et al. (and subsequently in this project) were based off of a series of pre-drawn images depicting anthropomorphized animals performing a variety of actions. The sentences were purposefully structured so that semantic cues could not be used to disambiguate between active and passive sentences. A total of 16 different
action verb stems were used for the active and passive target sentences. For the target sentences, the verbs *chase*, *comb*, *kick*, *kiss*, *lick*, *pat*, *pinch*, *poke*, *punch*, *push*, *scrub*, *shove*, *tickle*, *touch*, *trap*, and *wash* were chosen based on a series of phonological constraints, all of which were violated by at least one verb. The constraints are explained here in the order of most to least importance.

First, verbs were chosen if they did not begin with /s/ or /z/; this constraint was violated by the verb *scrub*. Second, verbs should manifest the –ed passive participle ending as [t] or [d]; this was violated by *pat*. Third, verbs should have been monosyllabic; this was violated by *tickle*. Fourth, verb onsets should have been stops; this was violated by *scrub*, *like*, and *wash*. Note that *chase* was not considered to violate this constraint because the onset was treated as a cluster starting with [t]. Finally, verbs were chosen if they had a CVC order (i.e., they avoided consonant clusters); this was violated by *scrub*, *trap*, and *chase* in the onset position and by *punch* and *pinch* in the coda position. Furthermore, these verbs were chosen because they are actional verbs that can also be used as passives. When used as passives, they can have an animate patient and overt animate agents (Levin, 1993; Levin and Rappaport). 4 verbs (*comb*, *scrub*, *shove*, *tickle*) had voiced stem codas while the remaining 13 (*chase*, *kick*, *kiss*, *lick*, *pat*, *pinch*, *poke*, *punch*, *push*, *touch*, *trap*, *wash*) had unvoiced stem codas. 11 verbs (*scrub*, *pat*, *lick*, *kick*, *poe*, *tickle*, *comb*, *punch*, *pinch*, *touch*) had stop stem codas while the remaining 5 (*shove*, *chase*, *wash*, *kiss*, *push*) had non-stop stem codas.

In sentences requiring agent and patient thematic roles, the animate NPs were both animals. Pairing of the agent and patient animal NPs for each sentence was decided by what was depicted in the pre-drawn images used by Rehrig et al. The animate NPs used in the study were *bear*, *cat*, *cow*, *dog*, *duck*, *elephant*, *fox*, *frog*, *hippo*, *kangaroo*, *lion*, *monkey*, *mouse*, *pig*, *rabbit*, *rhino*, *sheep*, *turtle*, and *zebra*. 
**Target sentences.** Participants listened to truncations of active sentences with object NPs (e.g., *The pig was kissing the sheep*) and passive sentences with agentive by phrases (e.g., *The pig was kissed by the sheep*) and chose from two complete sentences that would appear on a monitor, one of which was the active completion with the other being the passive completion. For both the active and the passive sentences, the audio was truncated immediately before the onset of the verbal suffix (-ing/-ed). Because both active and passive sentence contained the auxiliary *was*, all target sentences were morphosyntactically ambiguous until the onset of the verbal inflection. Participants therefore heard *The pig was kiss-* for both the active and passive forms of the sentence.

The 64 experimental sentences were divided into 2 equally-sized lists. In each of the two lists, each verb appeared 4 times in a pseudo-randomized order (one time each in an active and passive sentence with animal pair NPs). The sentences on the 2 lists differed only in the order in which they were presented to the participants (e.g., List 2 presented the sentences in the reverse order of List 1). There were a total of 16 scenarios, with each scenario consisting of a pair of animals and one verb. Each scenario occurred in the active and passive forms, with each animal in a pair appearing as the subject and the object for both the active and passive forms (16 verbs x 2 active/passive x 2 subject/object = 64 sentences).

**Ambiguous fillers.** In addition to the 64 experimental sentences, the experiment also included 12 filler sentences. All of these sentences were active and were truncated before the onset of the verbal suffix -ing so as to be morphosyntactically ambiguous, similarly to the active target sentences. The two sentences that would appear on the screen would both be active and differed only in the identity of the patient NP. For example, sentence options given for the truncated audio *The bear was hold-* might be *The bear was holding the book* or *The bear was holding the cookie*. The verbs *wear* and *hold* were used in these sentences because they are non-
actional and can have inanimate patients; they also do not take the –ed passive participle. This makes them crucially different from active target sentences. Because these verbs take inanimate objects, the inanimate NPs shirt, ball, medal, book, and cookie were used in these sentences.

Unambiguous catch trials. To ensure that participants remained on task, 24 unambiguous catch trials were included in the experiment. These sentences used the prepositions behind and near as well as the comparative adjectives cleaner and dirtier. Sentences containing the prepositions behind and near were truncated immediately before the onset of the second determiner the of the second NP in the sentence. Sentences containing the comparative adjectives cleaner and dirtier were truncated immediately before the onset of the comparative suffix –er. All of these sentences were truncated so as to make to the identity of the second noun phrase ambiguous. The two sentences that would subsequently appear on the monitor (see Procedure) differed by the identity of the first NP, instead of the second. For instance, during a catch trial a participant might hear The hippo was behind—; the two sentences that would then appear on the monitor might be The hippo was behind the rhino or The rhino was behind the hippo. Since participants would have been told the identity of the first noun phrase in the sentence audio, they would have had enough information to choose the correct sentence on the monitor.

Procedure

Each trial began with a cross appearing at the center of a computer monitor for 500 milliseconds, with participants instructed to visually fixate on the cross. The fixation point was followed by the presentation of the truncated sentence recording, which was presented binaurally via high quality Sennheiser HD202 headphones. After a 300 millisecond interval following the offset of the auditory stimulus, two written sentences appeared simultaneously on
the left and right hand sides of the monitor. For targets, one sentence would be active and the other would be passive. Catch trials would differ in the identity of the first noun phrase of either sentence, while non-catch filler trials would differ in the identity of the second noun phrase. Participants were instructed to press a key on the left side of the keyboard (the q key) if they thought the left sentence was the complete sentence of the auditory truncation and instructed to press a key on the right side of the keyboard (the p key) if they thought the right sentence was the complete sentence. The intertribal interval was 500 milliseconds. Participants completed 4 practice trials before beginning the experiment. Participants were told to make their best guess for each trial, and were not given feedback on their performance. At the end of the experiment, the majority of participants reported that they felt like they were guessing what the origin of the sentence fragment was.

Results

Because there are two options for each trial, by chance alone, the probability of guessing correctly on any one trial is .50. Participants who were correct for 18 or more of 24 catch trials performed at above chance levels ($p < .05$ cumulative binomial probability), participants who were correct on between 8 and 17, inclusive, performed at chance levels, and participants who were correct on fewer than 8 trials performed at below chance level ($p < .05$ cumulative binomial probability). Of the 40 participants, 38 performed at above chance levels, and 2 performed at chance levels (see Figure 3). All subsequent results are for the 38 participants who performed at above chance levels on the catch trials.
Figure 3. Participants’ scores for catch trials. 38 participants scored above chance levels (at least 18 correct). Two scored at chance levels and were not included in the remaining analyses

**Accuracy results**

Participants who were correct on at least 40 out of the 64 target sentences were performing at above chance levels \( (p < .05 \text{ cumulative binomial probability}) \). For the target active and passive sentences, 23 participants (61%) performed above chance levels, with the remaining 15 participants (39%) performing at chance levels (between 26 and 39, inclusive, correct). None of the participants performed below chance levels on the targets.
The 64 target sentences consisted of 32 active sentences and 32 passive sentences.

Participants who were correct for 21 sentences were performing above chance levels for that category \(p < 0.05\) cumulative binomial probability). For the passive target sentences, 34 of the 38 participants (89%) performed above chance levels. The remaining 4 (11%) performed at chance levels (between 13 and 20, inclusive, correct). None performed below chance levels. For the active target sentences, 8 participants (21%) performed above chance levels. 14 participants (37%) performed at chance levels. The remaining 16 participants (42%) performed below chance levels.

**Figure 4.** Participants’ scores for target sentences. 23 participants scored above chance levels (at least 40 correct). 15 scored at chance levels (between 26 and 39, inclusive, correct).
A one way ANOVA with subject as a random variable revealed that, overall, participants guessed the correct completion of passive sentences more often than active sentences (84% vs. 47%, $F(1,37) = 59.42, p < .001$). Individual subject analyses revealed that 6 participants performed above chance on both passives and actives, 12 performed above chance on passives and at chance on the actives, 16 performed above chance on passives and below chance on actives, 2 performed
at chance on passives and above chance on actives, and 2 performed at chance on both passives and actives.

Because all of the filler sentences were active sentences, there were in total more active sentences than passive sentences in the experiment (32 active targets + 36 active fillers = 68 active sentences total vs. 32 passive targets). Therefore there is reason to believe that encountering over twice as many active sentences as passive sentences may have created an expectation bias within the participants to complete more truncations as passive sentences. Comparison of accuracy data from the first 15 trials versus the last 15 trials indicates that participants generally correctly completed target sentences more often earlier in the experiment than later on. A 2 (active/passive) x 2 (first 15/last 15) ANOVA with subject as a random variable revealed that participants tended to be more accurate during the first 15 trials than the last 15 trials (65.33% and 60.53%, respectively $F(1, 37) = 2.76, p = .105$). The marginal interaction between the two factors was caused by participants performing more accurately on active sentences on the first 15 trials than on the last 15 trials (50.00% and 41.58%, respectively $F(1, 37) = 4.22, p = .047$).

Depending on what version of the experiment a participant received, the first and last 15 trials contained either 9 or 10 target sentences. Participants would have had to correctly complete at
least 7 or 8 sentences, respectively, in order to perform above chance levels ($p < .05$ cumulative binomial probability). Individual analyses revealed that for the first 15 trials, 11 participants (29%) performed above chance levels. 26 participants (68%) performed at chance levels and 1 (3%) performed below chance levels. For the last 15 trials, only four participants (11%) performed above chance levels. The remaining thirty-four (89%) performed at chance levels. None of the participants performed below chance levels on the last 15 trials. Perhaps in an experiment that contains only the 32 active targets and the 32 passive targets, where there would be less influence of an expectation bias, more participants would perform at higher than chance levels, and accuracy data for active vs. passive sentences would be closer to equal levels.

![Figure 8](image.png)

**Figure 8.** Effect of earliness or lateness on accuracy. Error bars denote standard errors.
Effects of coda voicing on accuracy. The voicing of a syllable's coda can affect the length of the nucleus, with vowels generally being longer when followed by a voiced consonant (Sharf, 1962; Lehiste and Peterson, 1961; House 1961). Therefore, accuracy data were also analyzed using a 2 (active/passive) x 2 (voiced/unvoiced coda) ANOVA with subject as a random variable. Participants correctly guessed the completion of passive sentences more often than they did for active sentences (86% vs. 38% respective, $F(1, 37) = 92.30, p < .001$). There was also a main effect of stem coda voicing, with participants being more accurate when the coda of the verb stem was unvoiced than when the coda was voiced (68% vs. 56% respective, $F(1, 37) = 26.99, p < .001$). There was a significant interaction between the two factors ($F(1, 37) = 36.13, p < .001$) for active sentences. Post hoc analyses revealed that participants were significantly better on passives with voiced codas than unvoiced codas ($F(1, 37) = 8.68, p = .006$), whereas they were significantly more accurate on actives with unvoiced codas than voiced codas ($F(1, 37) = 50.78, p < .0005$). The lengthened vowel in verbs with voiced codas may have caused participants to mistakenly believe they were hearing a passive sentence when in fact they were hearing an active sentence.

**Figure 9.** Interaction of earliness/lateness of experiment trial and syntax on accuracy. Error bars denote standard errors. Star represents significance level $p < .05$. 

Effects of coda voicing on accuracy. The voicing of a syllable's coda can affect the length of the nucleus, with vowels generally being longer when followed by a voiced consonant (Sharf, 1962; Lehiste and Peterson, 1961; House 1961). Therefore, accuracy data were also analyzed using a 2 (active/passive) x 2 (voiced/unvoiced coda) ANOVA with subject as a random variable. Participants correctly guessed the completion of passive sentences more often than they did for active sentences (86% vs. 38% respective, $F(1, 37) = 92.30, p < .001$). There was also a main effect of stem coda voicing, with participants being more accurate when the coda of the verb stem was unvoiced than when the coda was voiced (68% vs. 56% respective, $F(1, 37) = 26.99, p < .001$). There was a significant interaction between the two factors ($F(1, 37) = 36.13, p < .001$) for active sentences. Post hoc analyses revealed that participants were significantly better on passives with voiced codas than unvoiced codas ($F(1, 37) = 8.68, p = .006$), whereas they were significantly more accurate on actives with unvoiced codas than voiced codas ($F(1, 37) = 50.78, p < .0005$). The lengthened vowel in verbs with voiced codas may have caused participants to mistakenly believe they were hearing a passive sentence when in fact they were hearing an active sentence.
Collapsed across the 38 participants, there were a total of 228 active voiced trials and 228 passive voiced trials. In total, the 38 participants correctly completed 54 active voiced sentences correctly. A cumulative binomial distribution shows that participants would have needed to be correct on at least 126 sentences to be at above chance levels and between 103 and 125 to be at chance levels. Therefore, participants performed at below chance levels for active sentences with voiced stem codas \( p < .05 \). Of the total 228 passive voiced trials, participants correctly completed 203. Therefore participants as a whole performed above chance levels for passive sentences with voiced stem codas \( p < .05 \). Collapsed across the 38 participants, there were a total of 988 active

![Figure 10. Effect of stem coda voicing on accuracy. Error bars denote standard errors. Stars represent significance level \( p < .001 \).](image1)

![Figure 11. Interaction of stem coda voicing and syntax on accuracy. Error bars denote standard errors. Stars represents significance level \( p < .001 \).](image2)
unvoiced trials and 988 passive unvoiced trials. In total, participants completed 515 active unvoiced sentences correctly. A cumulative binomial distribution indicates that participants would need to have been correct on 520 trials to be performing above chance levels; therefore participants as a whole were performing at chance levels for active sentences with unvoiced stem codas \((p < .05)\). Of the 988 passive unvoiced sentences, participants as whole correctly completed 818 sentences and were therefore performing above chance levels for the passive sentences with unvoiced stem codas \((p < .05)\).

**Effects of coda articulation on accuracy.** The phonological manner of articulation of the coda has also been shown to affect the overall length of the syllable (Van Santen, 1992). Syllables ending in stops are generally shorter than syllables ending in non-stops. Accuracy data were also analyzed using another 2 (active/passive) x 2 (stop coda/non-stop coda) ANOVA with subject as a random variable. This analysis once again showed that participants correctly guessed the completion of passive sentences more frequently than active sentences (84\% vs 49\% respective, \(F(1, 37) = 50.29, p < .001\)). There was a main effect of verb stem coda type, with participants correctly completing sentences using verbs whose stems end with a non-stop coda than with verb stems that end with a stop coda (69\% vs. 64\% respective, \(F(1, 37) = 7.82, p = .008\)). There was a significant interaction between the two factors \((F(1, 37) = 10.58, p = .002)\). Post hoc analyses indicate that this interaction is only significant for active sentences, where participants performed better on active sentences with non-stop codas \((F(1, 37) = 14.48, p = .001)\). Perhaps this occurs because verb stems that end in stops make it more apparent to participants that any verb stem lengthening they hear is due to the passive voice, and not by other factors. Participants may otherwise be attributing an increased verb stem length to the non-stop in the coda position of the verb stem.
Collapsed across the 38 participants, there were a total of 836 active stop trials and 836 passive stop trials. In total, the 38 participants correctly completed 366 active stop sentences correctly. A cumulative binomial distribution shows that participants would have had to be correct on 441 sentences to perform above chance levels and between 398 and 440 to be at chance levels; therefore participants as a whole performed at chance levels for active sentences with stop stem codas ($p < .05$ cumulative binomial probability). Of the total 836 passive stop trials, participants correctly completed 703. Again, participants would have had to be correct on 441 to perform at above chance levels. Therefore participants as a whole performed at above chance levels for
passive sentences with stop stem codas (\(p < .05\)). Collapsed across the 38 participants, there were a total of 380 active non-stop trials and 380 passive non-stop trials. In total, participants completed 203 active non-stop sentences correctly. A cumulative binomial distribution indicates that participants would have needed to be correct on 206 trials to be performing at above chance levels and between 174 and 205 trials to be at chance levels; therefore participants as a whole were performing at chance levels for active sentences with non-stop stem codas (\(p < .05\)). Of the 380 passive non-stop sentences, participants as whole correctly completed 318 sentences and were therefore performing above chance levels for the passive sentences with non-stop stem codas (\(p < .05\) cumulative binomial distribution probability).

Reaction time results

A one way ANOVA with subject as a random variable was used to analyze reaction time data. Participants responded to truncated recordings of passive sentences faster than they did for truncated active sentences (1532 ms vs. 1581 ms respective, \(F(1, 37) = 8.86, p = .005\). Because participants also guessed the correct completion of passive sentences more often than with passive sentences, this indicates that no speed-accuracy trade-off occurred.

Figure 14. Effect of syntax (active/passive) on reaction time. Error bars denote standard errors. Stars represents significance level \(p < .01\).
**Effects of coda voicing on reaction time.** As with the accuracy data, data for reaction times were next analyzed in a 2 (active/passive) x 2 (voiced/unvoiced coda) ANOVA with subject as a random variable. Participants completed truncations of passive sentences more quickly than for active sentences (1515 ms vs. 1568 ms respective, $F(1, 37) = 5.90, p = .020$). There was a significant effect from the verb stem coda voicing (1517 ms for voiced codas vs. 1565 ms for unvoiced codas, $F(1, 37) = 5.39, p = .026$). There was no significant interaction between the two factors. However, planned comparisons revealed that participants were marginally faster on passive sentences with voiced codas than unvoiced codas ($F(1, 37) = 3.60, p = 0.066$).

![Figure 15. Effect of stem coda voicing on reaction time. Error bars denote standard errors. Star represents significance level $p < .05$.](image1)

![Figure 16. Interaction of stem coda voicing and syntax on reaction time. Error bars denote standard errors](image2)
Effects of coda articulation on reaction time. Reaction time data were also analyzed using a 2 (active/passive) x 2 (stop/non-stop) ANOVA with subject as a random variable. Once again, participants were faster on passive sentences than active sentences (1530 ms vs. 1570 ms respective, $F(1, 37) = 6.09, p = .018$). There was no significant effect of stem coda type (1568 ms for stops vs 1531 ms for non-stops, $F(1, 37) = 3.15, p = .084$). The interaction of syntax and the stem coda type was not significant. However, planned comparisons revealed that participants were marginally faster on active sentences with non-stop codas than stop codas (1539 ms vs 1600 ms respective, $F(1, 37) = 4.82, p = .034$).

**Figure 17.** Effect of manner of articulation on reaction time. Error bars denote standard errors.

**Figure 18.** Interaction of manner of articulation and syntax on reaction time. Error bars denote standard errors. Star represents significance level $p < .05$. 

---

25
**Verb by verb analysis**

To determine whether the effect of syntax was valid for all verbs, a separate one way ANOVA with verb stem as a random variable was used to analyze the accuracy data of each verb. Results of this analysis show that the main effect of syntax on accuracy was significant for all verbs (see Table 1).

<table>
<thead>
<tr>
<th>Verb</th>
<th>Active</th>
<th>Passive</th>
<th>F-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>chase</td>
<td>0.539</td>
<td>0.803</td>
<td>10.98</td>
<td>0.002</td>
</tr>
<tr>
<td>comb</td>
<td>0.566</td>
<td>0.75</td>
<td>4.67</td>
<td>0.037</td>
</tr>
<tr>
<td>kick</td>
<td>0.539</td>
<td>0.895</td>
<td>20.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>kiss</td>
<td>0.553</td>
<td>0.816</td>
<td>9.39</td>
<td>0.004</td>
</tr>
<tr>
<td>lick</td>
<td>0.487</td>
<td>0.829</td>
<td>14.89</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>pat</td>
<td>0.342</td>
<td>0.539</td>
<td>5.61</td>
<td>0.023</td>
</tr>
<tr>
<td>pinch</td>
<td>0.553</td>
<td>0.908</td>
<td>25.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>poke</td>
<td>0.539</td>
<td>0.921</td>
<td>23.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>punch</td>
<td>0.605</td>
<td>0.855</td>
<td>13.78</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>push</td>
<td>0.697</td>
<td>0.895</td>
<td>6.25</td>
<td>0.017</td>
</tr>
<tr>
<td>scrub</td>
<td>0.211</td>
<td>0.921</td>
<td>97.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>shove</td>
<td>0.316</td>
<td>0.882</td>
<td>40.60</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>tickle</td>
<td>0.184</td>
<td>0.868</td>
<td>91.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>touch</td>
<td>0.645</td>
<td>0.855</td>
<td>6.35</td>
<td>0.016</td>
</tr>
<tr>
<td>trap</td>
<td>0.145</td>
<td>0.908</td>
<td>87.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>wash</td>
<td>0.566</td>
<td>0.789</td>
<td>7.53</td>
<td>0.009</td>
</tr>
</tbody>
</table>
Finally, to determine whether there was a relationship between comprehension and production, the sum of the average scores for active and passive sentences for each verb was correlated to the normalized difference in the duration of the active and passive forms for each verb. There was no significant correlation between comprehension and production ($r = 0.196, p = 0.23$).

Table 1. Mean accuracy scores, F-stats and p-values by verb. Within-groups df = 1. Between-groups df = 37.

Figure 19. Effect of syntax on accuracy for each verb. Error bars denote standard errors.

Finally, to determine whether there was a relationship between comprehension and production, the sum of the average scores for active and passive sentences for each verb was correlated to the normalized difference in the duration of the active and passive forms for each verb. There was no significant correlation between comprehension and production ($r = 0.196, p = 0.23$).

Figure 20. Correlation of comprehension results to production difference in active and passive stems.
Discussion

This project was conducted as part of a larger follow up study to Stromswold et al.’s original investigation, which was one of the first to focus on the comprehension of spoken actives and passives by adult speakers. Taken as a whole, these results indicate that participants were faster and more accurate at identifying passive sentences than active sentences. In fact, participants could identify passive sentences at above chance levels whereas the only performed at chance levels when predicting active sentences. This suggests that any potential cue that they pick up only signals upcoming passive sentence structures, and that there is no complementary active cue, such as a verb stem shortening. Effects of coda voicing and articulation on the stem vowel indicate that listeners’ comprehension of a verb stem lengthening cue may be distorted by phonetic factors. The complex interaction of phonetic features with prediction of syntactic structure may indicate that the mechanism behind predicting sentence structure relies on constraints, and that listeners use verb stem lengthening to constrain the number of potential parses of a sentence to passive structures only. Given that active sentences are more common than passive sentences in English (COCA, Davies 2008), it would be useful to have a cue that allowed listeners to quickly predict a syntactic structure that they may not otherwise have expected. It is important to note, however, the inconsistency of the results with original expectations makes it unclear whether participants’ ability to predict passive sentence structure is necessarily due to their unconscious processing of a passive verb stem lengthening effect.

Vowels tend to lengthen when followed by a voiced consonant (Sharf, 1962; Lehiste and Peterson, 1961; House 1961). If verb stem lengthening is in fact a cue to the passive voice, then vowel lengthening due to a voiced coda could potentially distort a listener’s perception of a stem lengthening caused by the passive voice. This appears to be the case, as the ANOVA for accuracy of the voiced vs unvoiced codas shows that participants were more accurate for sentences with
unvoiced verb stem codas. However, the ANOVA for reaction times of voiced vs unvoiced codas shows that participants were slower for unvoiced codas; therefore there was a speed/accuracy trade-off for sentences with unvoiced verb stem codas. Perhaps the longer vowel of verbs with voiced stem codas was enough to disrupt participants’ detection of longer verb stems caused by the passive voice. In support of this, the interaction of syntax (i.e., active vs. passive) and stem coda voicing shows that participants were more accurate on sentences with unvoiced codas for the active sentences and on sentences with voiced codas for the passive sentence. Here it would seem that participants may have mistaken the lengthened vowel of verbs with voiced codas as the verb stem lengthening cue to the passive voice, even when the sentence was actually active. In a future study, it would be prudent to return to the original recordings from Rehrig et al (2015) to see if the vowel is in fact the phoneme that is most lengthened in passive verb stems.

A coda’s manner of articulation can also have an effect on the overall length of the syllable, with syllables ending with stops generally being shorter than those ending in non-stops (Van Santen, 1992). One might guess that a longer verb stem caused by a non-stop coda might cause participants to misattribute the lengthening to an anticipation of a passive sentence structure. However, the ANOVAs on the effect of the coda’s manner of articulation on accuracy and reaction time indicate otherwise, with participants being faster and more accurate for the non-stop codas than for the stop codas. Moreover, the interaction of the manner of articulation and syntax was caused only by participants performing faster and more accurately on active sentences with non-stop codas than on active sentences with stop codas. This is the first indication that a cue to the passive voice may not be as straightforward as a lengthening of the verb stem.

It is possible, however, that participants were actually responding to an entirely different cue that was unintentionally coded within the stimuli. Upon revisiting Rehrig et al.’s recordings,
which were sliced between the morphemes of each sentence using Praat software by researchers who did not have prior phonetic backgrounds (myself included), it was found that Rehrig et al.’s protocol included the closure of the passive verb participle ending, [t] and [d], in the verb stem slices. This was found to be true for both stop and non-stop verb stem codas. Therefore, the truncated audio recordings used in this study may have unintentionally contained an additional cue to the passive voice. This event further necessitates a return to the original stimuli so that they can be examined on a deeper phonological level to determine if there are any other inconsistencies between the active and passive forms.

![Image](https://example.com/image.png)

**Figure 21.** Example slices of (A) *lick*, which ends with a stop coda and (B) *push*, which ends with a non-stop coda, in passive sentences. Notice that both stems include the closure of the passive verb participle, [t].

The most damning piece of evidence against the use of verb stem lengthening as a cue to the passive voice is the lack of a correlation between participants’ accuracy performance on the comprehension gating experiment and the normalized degree of verb stem lengthening for each passive verb from the audio recordings. Considering how robust Rehrig et al.’s findings were, one would expect a high correlation between the two measures. Despite this glaring anomaly, it is possible that these results do not accurately represent the relationship between the degree of
passive verb stem lengthening by a speaker and the comprehension of this lengthening by a listener as a cue to the passive voice due to the high potential for participants to form an expectation bias during the course of the experiment.

Because there were so many more active sentences than passive sentences throughout the course of the experiment (due to all of the fillers being active), there was a high potential for participants to form an expectation bias. Similar to how people may try answer true/false questions with relatively equal amounts of “true” and “false” responses, participants in this experiment may have been trying to provide an equal number of active and passive responses. There is some evidence to support the formation of the expectation bias in the analysis of participants’ performances on the first 15 trials of the experiment in comparison to their performances on the last 15 trials. Initial analyses revealed that performance on the targets within the first 15 trials was no different from performance on targets within the last 15 trials. However, post-hoc analyses showed that participants were marginally better on the active targets at the beginning of the experiment than at the end of the experiment.

**Follow-up Experiment**

To determine whether a potential expectation bias had an effect on the outcomes of the original experiment, a follow-up experiment that consisted only of the target active and passive sentences was run on a new group of participants. Preliminary data and results are described below.

**Participants**

Twelve Rutgers University students participated in the experiment. All 8 participants were monolingual English speakers with normal hearing and normal or corrected-to-normal vision.
Stimuli

The stimuli in the follow-up experiment were the same as the active and passive target sentences in the original comprehension gating experiment (16 verbs x 2 active/passive x 2 subject/object = 64 sentences).

Procedure

The procedure for the follow-up experiment was identical to the procedure of the original comprehension gating experiment.

Preliminary Results

Accuracy

Participants who were correct on at least 40 out of the 64 sentences were performing at above chance levels ($p < .05$ cumulative binomial probability). For the target active and passive sentences, 2 participants (16.7%) performed at above chance levels for both actives and passives, 5 participants (41.7%) performed at above chance levels for passives and at chance levels for actives (between 26 and 39, inclusive, correct), 1 participant performed at above chance levels for actives and at below chance levels for actives, and 4 were at chance levels for both actives and passives. These results seem to suggest that participants were still better at passives overall.

A one way ANOVA with subject as a random variable revealed a marginally significant difference in accuracy for active and passive sentences (70% passives vs 54% actives, $F(1, 11) = 5.15, p = .044$).

A one way ANOVA with subject as a random variable revealed that the effect of stem coda voicing on accuracy that was found in the original comprehension experiment disappeared in the follow-up experiment (55% voiced vs 64% unvoiced, $F(1, 11) = 4.15, p = .066$). Likewise, there was
no significant interaction between stem coda voicing and syntax. However, planned comparisons revealed that participants were significantly better at identifying active sentences with unvoiced stem codas than with voiced stem codas, which is what was found in the original experiment (57% vs 40%, respective, $F(1, 11) = 7.08, p = .022$).

Yet another one way ANOVA with subject as a random variable revealed that the effect of manner of articulation of the coda also disappeared in the follow-up experiment (61% stops vs 65% non-stops, $F(1, 11) = 2.43, p = .148$). There was no significant interaction between manner of articulation and syntax. Although planned comparisons did not reveal any significant differences between stops and non-stops in both active and passive sentences, it appears that participants may prefer stops to non-stops in active sentences and non-stops to stops in passive sentences. These findings are more in-line with what was originally expected than what was found in the first experiment.

Reaction time

A one way ANOVA with subject as a random variable showed that participants were not significantly faster at either passives or actives (1909 ms passives vs 2135 ms actives, $F(1, 11) = 4.37, p = .061$).

A one way ANOVA with subject as a random variable revealed no effect of stem coda voicing on reaction time (1935 ms voiced vs 2042 ms unvoiced, $F(1, 11) = 1.91, p = .194$). Likewise, there was no significant interaction between stem coda voicing and syntax. Another one way ANOVA with subject as a random variable revealed no effect of manner of articulation of the coda on reaction time (2012 ms stop vs 2043 ms non-stop, $F(1, 11) = .327, p = .579$). Likewise, there was no significant interaction between stem coda voicing and syntax.
Discussion of follow-up experiment results

Preliminary results from the 12 participants of the follow-up experiment show that participants correctly completed passive sentences at above chance levels more often than they did for active sentences. Because there were no fillers in this experiment, the number of active sentences was balanced by the number of passive sentences, and there was no potential for the formation of an expectation bias over the course of the experiment. These results therefore suggest that the tendency to perform more accurately on passive sentences than on active sentences is fueled by some other phenomenon. However, a one way ANOVA revealed no significant effect of syntax on participants’ accuracy or reaction time.

In addition to the established preference for passives over actives, the results of the analyses for the effects of coda voicing are more in-line with what was originally expected of the first experiment. Participants were expected to be better at identifying passive sentences when the verb had a voiced coda and active sentences when the verb had an unvoiced coda. Although the follow-up experiment did not have a significant interaction between syntax and coda voicing, it was found that participants were indeed significantly better at identifying active sentences when the verb had an unvoiced coda than when it had a voiced coda. Based on these preliminary findings, it is quite possible that with more participants, the overall interaction would eventually become significant and follow the same trend that was found in the original experiment.

Furthermore, the findings from the analyses for the effects of coda articulation are also more in-line with original expectations. In the original experiment, participants were unexpectedly better at identifying active sentences if the verb stems ended with a non-stop coda. Although there was no significant interaction between syntax and coda articulation, results are nonetheless indicating that participants are better at identifying active sentences when the verb
ends with a stop coda and passive sentences with the verb ends with a non-stop coda, as originally expected. Therefore, it is possible that with more participants, this interaction will eventually become significant and follow a trend more in-line with original expectations.

Because this follow-up experiment balanced the number of active and passive trials to avoid the possibility of an expectation bias forming, these preliminary results suggest that an expectation bias did in fact have an effect on the results of the first experiment. However, it is important to note that without any catch trials or fillers, it was impossible to tell whether participants in the follow-up experiment were staying on task for the entirety of the experiment. Therefore, it would be prudent to confirm these findings by running an additional experiment that includes active and passive catch trials and fillers in order to ensure that data comes from participants who are on task and that the number of active and passive trials remains balanced throughout the experiment.

**Concluding remarks**

The present study used a comprehension gating experiment to investigate whether listeners could predict upcoming syntactic structure using an acoustic verb stem lengthening cue. Results revealed that participants were able to predict passive sentences with greater accuracy and speed than active sentences, which could suggest that the verb stem lengthening cue could constrain potential parses to passive structures. Additionally, results also showed that the passive verb stem lengthening cue is more easily transmitted in certain phonetic environments (i.e., voiced codas and possibly non-stop codas). Preliminary results of a follow-up experiment that removed the potential for the formation of an expectation bias support these findings. While further studies are needed to explore how parsing mechanisms use acoustic information to
predict sentence structure, it is apparent in this research that adults can use this information to accurately and efficiently process sentences.

Acknowledgements

I am incredibly grateful to my two advisers, Karin Stromswold and Paul de Lacy for their guidance and insight throughout the entirety of this thesis. I would also like to thank Joe Munson for his help with running the follow-up experiment, as well as Gwen Rehrig, Nora Beierg, Liz Chalmers, Nic Schrum, Aldo Mayro, and Rachel Shu for their help with Praat analyses and audio slicing. This work was generously supported by the Aresty Foundation and the National Science Foundation (REU CNS-1062735; IGERT DGE 0549115).
References


Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. Probabilistic relations between words: Evidence from reduction in lexical production.


