Listening to Speech and Non-speech Sounds Activates Phonological and Semantic Knowledge Differently

James Bartolotti[a,1], **Scott R. Schroeder**[b], **Sayuri Hayakawa**[a], **Sirada Rochanavibhata**[a], **Peiyao Chen**[a], **Viorica Marian**[a]

[a]Northwestern University, Department of Communication Sciences and Disorders, 2240 Campus Drive, Evanston, IL, United States 60208

[b]Hofstra University, Department of Speech-Language-Hearing Sciences, Hempstead, NY, United States 11549

Abstract

How does the mind process linguistic and nonlinguistic sounds? The current study assessed the different ways that spoken words (e.g., "dog") and characteristic sounds (e.g., <barking>) provide access to phonological information (e.g., word-form of "dog") and semantic information (e.g., knowledge that a dog is associated with a leash). Using an eye-tracking paradigm, we found that listening to words prompted rapid phonological activation, which was then followed by semantic access. The opposite pattern emerged for sounds, with early semantic access followed by later retrieval of phonological information. Despite differences in the time courses of conceptual access, both words and sounds elicited robust activation of phonological and semantic knowledge. These findings inform models of auditory processing by revealing the pathways between speech and non-speech input and their corresponding word forms and concepts, which influence the speed, magnitude, and duration of linguistic and nonlinguistic activation.

Keywords

Speech comprehension; Sound processing; Phonology; Semantic competition; Eye-tracking; Psycholinguistics

---

The human auditory system provides fast, precise feedback for interpreting and interacting with the environment. Sounds that are characteristic of an entity carry nuanced information; a dog's barking sound, for example, typically reveals that specific dog's size and location. Spoken language, in contrast, is comprised of words that refer to, but are not characteristic of, a concept. For example, the spoken word "dog" is a poor location cue, as the speaker that produced it may be referring to a dog that is far away, or not even present. Models of

auditory processing take into account words' and sounds' unique features, and propose that the two types of input access conceptual knowledge via different routes (Chen & Spence, 2011). In the current study, we compare the time courses of semantic and phonological activation by spoken words and characteristic sounds.

During speech processing, individual words' phonological form and semantic meaning are rapidly accessed (Connolly & Phillips, 1994; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). That is, hearing "dog" cues access to the representation of the word DOG in the mental lexicon, as well as semantic features that constitute the concept of dog, such as "has fur," "barks," and "has four legs." Evidence of phonological and semantic activation can also be observed in the form of spreading activation to related words or concepts. Eye-tracking studies have shown that when people hear a spoken word, such as "dog," and are asked to find its corresponding picture in a display, they often briefly look at pictures representing words that are phonologically related (e.g., doctor) or semantically related (e.g., leash) (Allopenna, Magnuson, & Tanenhaus, 1998; Huettig & McQueen, 2007; Yee & Sedivy, 2006). Current models of spoken word processing, such as TRACE (McClelland & Elman, 1986) predict how spoken words cause activation to spread throughout the lexico-semantic system over time. These models can be adapted to incorporate non-speech sounds, but there is currently insufficient data comparing word and sound processing needed to train such a model.

Characteristic sounds, similar to spoken words, trigger access to semantic information (Chen & Spence, 2011, 2013; Edmiston & Lupyan, 2015), as evidenced by semantic priming effects in multiple paradigms (e.g., Dudschig et al., 2018; Orgs et al., 2008). Sounds can also activate corresponding phonological and lexical information; when asked to identify a barking sound, one can easily label it as a "dog." However, the rate and extent of a sound's spreading activation to phonological and semantic information has not been well-defined.

The aim of the current study is to compare the manner and rate with which both spoken words and characteristic sounds provide access to information associated with a concept. We designed a visual world eye-tracking experiment that assessed spreading activation from auditorily-presented targets to their phonological and semantic competitors. Participants heard an auditory cue (a spoken word, e.g., "dog" or characteristic sound, e.g., <arf-arf>[1]) and selected the matching picture in an array of four images while their eye movements were tracked. In phonological activation trials, a picture of a phonological onset competitor was present on the screen (e.g., a picture of a cloud when the target was the word "clock" or a <tick-tock> sound). In semantic activation trials, a picture of a thematic semantic competitor was present on the screen (e.g., a picture of a bone when the target was the word "dog" or an <arf-arf> sound). Activation of phonological and semantic competitors was assessed on trials where the target picture was not present in order to isolate the effect of the auditory cue.

Our investigation is motivated by the multisensory framework proposed by Chen and Spence (2011). Chen and Spence propose that spoken words and characteristic sounds cue access to

_____

We use double quotes " " to denote spoken words, and chevrons < > to denote non-speech sounds.

concepts via different pathways: Words have a direct connection to phono-lexical representations, whereas characteristic sounds connect directly to semantic representations. The different pathways that each kind of auditory stimulus takes through the lexico-semantic system leads to a number of possible predictions. If spoken words (e.g., "dog") provide more direct access to phono-lexical representations than characteristic sounds (e.g., <arf-arf>), activation of phonologically-related information (e.g., the word "doctor") should be faster in response to words. Using our paradigm, this pattern should manifest as earlier visual fixations to the phonological competitor (e.g., a picture of a doctor) when cued by words compared to sounds. Relatedly, if sounds provide more direct access to semantic representations relative to words, we may expect to see faster activation of semantically related information (e.g., the concept of a leash), manifesting as earlier fixations to the semantic competitor (e.g., a picture of a leash) when cued by sounds compared to words. In addition to timing, the magnitude of activation can shed light on the processes underlying conceptual access. For instance, given that matching a picture to a word, but not a sound, necessarily requires access to lexical information, we may expect to see a larger magnitude of phonological activation in response to words than to sounds. If instead we observe comparable phonological access, this would suggest that sounds also strongly activate phono-lexical information regardless of whether doing so is necessary for the task at hand.

Method

Design

The study followed a 2 (auditory input: characteristic sounds and spoken words) × 2 (competitor type: phonological and semantic) design, with auditory input manipulated between participants and competitor type repeated within participants.

Participants

Thirty monolingual English speakers participated in the study. Fifteen completed the characteristic sound input condition (14 females, M_{age} = 20.33 years, SD_{age}=2.94 years) and 15 completed the spoken word input condition (13 females, M_{age}=21.93 years, SD_{age}=2.84 years). Eye-tracking data for one participant in the characteristic sound condition was lost due to equipment error. Participants in the sound and word conditions did not differ in non-verbal IQ scores (*Wechsler Abbreviated Scale of Intelligence*; WASI, PsychCorp, 1999), phonological memory scores (digit span and nonword repetition subtests of the *Comprehensive Test of Phonological Processing*; CTOPP, Wagner, Torgesen, & Rashotte, 1999), or English receptive vocabulary scores (*Peabody Picture Vocabulary Test;* PPVT, Dunn, 1997). See Table 1 for participant characteristics.

Materials

Fifteen sets of stimuli were created for each competitor type, phonological and semantic (See Appendix A for the full stimuli list). The 15 phonological sets included three critical items: A target (e.g., clock), a phonological onset competitor (e.g., cloud) whose name overlapped with the target, and a control (e.g., lightbulb) that did not overlap. The 15 semantic sets also included three critical items: a target (e.g., rain), a thematic semantic competitor (e.g., umbrella), and a control (e.g., lungs).

Trial types.—Each of the 15 phonological and semantic sets was used to create 8 types of trials (for a total of 240 trials). Three of the 8 trial types were included in the analyses[2]: (1) Target-Only: one target (e.g., "cat" or "dog") with three filler objects, (2) Competitor-Only: one phonological ("cast") or semantic ("bone") competitor with three filler objects, and (3) Control: four filler objects (see Figure 1 for example stimuli). Control trials were created by replacing the competitor with a fourth filler object, thereby ensuring that competitor pictures and filler pictures appeared in the same locations and contexts, and appeared the same number of times. The full set of 8 trial types included displays in which both the target and competitor were present; however, the choice to focus the analyses on Target-Only, Competitor-Only, and Control trials ensured that fixations to each focal and filler object represented an independent measure of attention. In other words, the proportion of target, competitor, or location-matched filler fixations on a given trial did not directly impact the likelihood of looking at other objects relevant to the analysis.

Stimuli.—In order to ensure that the target and competitor stimuli only overlapped in phonology or semantic association (and that the filler items did not share characteristics with either), each target, competitor, and filler item was carefully selected for use in either the phonological or semantic condition. The two conditions therefore had distinct stimulus sets. Potential confounds due to variation in stimuli were minimized by examining responses to targets and competitors relative to control items within the same set, and by controlling for visual, phono-lexical, and auditory stimulus characteristics.

Visual.—Pictures were black and white line drawings primarily obtained from the International Picture Naming Database (Bates et al., 2000). For objects whose images were not in the International Picture Naming Database, pictures were chosen from Google Images and were independently normed by 20 English monolinguals using Amazon Mechanical Turk (http://www.mturk.com). Pictures that appeared in the same display were selected to account for unintentional overlap, including shape, saturation (i.e., none of the pictures were darker than the others), line thickness, primary function, phonological similarity, or semantic similarity (excepting the experimental manipulations of phonological and semantic similarity between competitors and targets).

Phono-lexical.—Words representing the target, competitor, and control did not differ from each other in word frequency (SUBTLEXUS; Brysbaert & New, 2009), phonological neighborhood size (CLEARPOND; Marian, Bartolotti, Chabal, & Shook, 2012), orthographic neighborhood size (CLEARPOND; Marian et al., 2012), familiarity, concreteness, or imageability (MRC Psycholinguistic Database; Coltheart, 1981).

Auditory.—The words representing the 30 target items were recorded at 44.1 Hz by a Midwestern female speaker of Standard American English, and the spoken word and characteristic sound stimuli were amplitude normalized. Due to the fact that many

Note that each phonological and semantic set had a secondary competitor that was not included in the present analysis: Spanish phonological competitors and categorical semantic competitors. Spanish competitors were included in the paradigm for consistency with a future Spanish-English bilingual study, but were not analyzed for the English monolingual population used in the present study. Categorical competitors (e.g., cow-sheep) were excluded in favor of thematic competitors, as items that are categorically related often look and/or sound alike and would be a weaker test of semantic relatedness due to these other sources of overlap.

continuous sounds do not have a fixed ending point, as words do, the duration of characteristic sounds (M = 1319.22, SE = 201.27, $Range$ = [329, 3868]) was significantly longer than that of spoken words (M = 654.78ms, SE = 24.46, $Range$ = [502, 864]; $t(17)$=3.30, p = .004). Note that duration was not significantly correlated with response times for either words (r = .12, n.s.) or sounds (r = .32, n.s.).

An independent norming study was used to calculate word and sound identification times using an auditory gating paradigm. These measures were then incorporated into the model in order to control for input- and item-related variability. Each auditory stimulus was used to create up to ten sound clips that started at word/sound onset and varied in length from 100ms to 1000ms, in 100ms intervals. Twenty participants, recruited on MTurk, listened to each sound clip and typed the word they heard or the name of the object that produced the sound. An item's identification point was calculated as the first sound clip to exceed 80% accuracy across participants. A number of items failed to reach this criterion even after the maximum duration of 1000ms that was used in the gating task. As only identifiable items could be expected to activate semantic and phonological representations that share features with competitors, unidentified items were excluded from analyses. Furthermore, in order to ensure that the word and sound conditions varied only in presentation and not in referents, items were only included if they were reliably identified in both conditions (that is, if they met the predetermined 80% identification threshold during the gating task in both the word and the sound conditions). As a result, final analyses included 18 of the original 30 items (identified in Tables A1 and A2 of the Appendix). Identification times of included items did not differ between words (M=377.78ms; SD=87.82) and sounds (M=333.33ms; SD=190.97; $t(17)$=0.846, p=.409). The difference in identification times between words and sounds did not differ between phonological (M=62.5ms; SD=199.55) and semantic trials (M=30ms; SD=249.67; $t(15.99)$=0.307, p=.762).

Procedure

During the visual search task, participants heard the auditory input (words or sounds) through closed-back headphones and their eye-movements were tracked using an Eyelink 1000 eye-tracking system recording at 250 Hz or higher (4-ms sampling resolution). On each trial, a fixation cross was shown on the screen for 1500ms before the four objects appeared on the display (see Appendix, Tables A3 and A4 for full list of stimuli on each trial). The objects were shown for 500ms, before the participants heard either a characteristic sound or a spoken word that corresponded with the target object (e.g., <tick-tock> or "clock"). Participants were instructed to mouse-click on the target picture as quickly as possible if the target was present, and to click on the fixation cross in the center of the screen if the target picture was absent. To indicate that the participant's response was recorded, the picture (or the fixation cross) that was selected was bordered with a green box as soon as the mouse was clicked. Regardless of when the participant responded, the visual display remained on the screen for 4500 ms after the onset of the auditory input. Before the experiment began, participants completed a set of practice trials meant to familiarize them with the task, followed by 240 test trials. In post-experiment debriefings, all participants reported noticing semantic relationships present on some trials. The phonological relationships were noticed by some participants in the spoken word condition, but by no

participants in the characteristic sound condition (see Vandeberg, Guadalupe, & Zwaan, 2011).

Data Analysis

Model syntax for all analyses is described in Appendix B. Variables in all models were coded using weighted Helmert effect coding.

Accuracy.—Two linear mixed effects models were created with by-subjects averaged data and by-items averaged data. Models included fixed effects of Auditory-input (word, sound), Condition (phonological, semantic), and Focal Object (target, competitor, control) and their interactions, as well as a random intercept of either subject or item[3]. For both accuracy and RT, Focal Object was contrast coded to first compare the target-present condition against the combined target-absent competitor and control conditions, followed by a comparison of competitor and control conditions. Significance levels for fixed effect estimates were obtained using *t*-tests and the Satterthwaite approximation for degrees of freedom. Follow-up pairwise comparisons on model estimates used the Tukey correction for multiple comparisons.

Response time.—Analysis of RT was restricted to correct trials only. RTs were measured in milliseconds starting from the onset of the auditory target, and times that were longer than an upper threshold of mean plus two standard deviations (calculated individually for each condition across all subjects and items) were replaced with the threshold value (4.37% of trials). The RT model included fixed effects of Auditory-input, Condition, and Focal Object plus their interactions, as well as random intercepts of both subject and item.

Eye movements.—The time course of visual fixations to semantic and phonological competitors was analyzed using growth curve analysis (GCA) (Mirman, Dixon, & Magnuson, 2008; Mirman, Magnuson, Graf Estes, & Dixon, 2008), a form of multilevel regression optimized to assess change over time. GCA was used over a generalized additive mixed model approach (Baayen, Vasishth, Kliegl, & Bates, 2017) due to the interpretability of GCA orthogonal polynomials which describe independent components of a fixation time course (e.g., slope, central peak shape). The tradeoff is that GCA is less robust at managing autocorrelation within a timeseries; we mitigate this by downsampling and aggregating the fixation time courses[4].

Accuracy was also analyzed with mixed effects logistic regression including both subject and item as random effects. However, two conditions (Phonological Word Control and Semantic Word Control) had 100% accuracy and zero variance across all subjects and items, and as a result, the logistic regression was unable to fit the data.

Within this window, Level-1 models captured the effect of time on changes in visual fixations using fourth-order orthogonal polynomials. Level-2 submodels captured the effects of experimental manipulations and individual differences on each of the polynomial time terms present in the Level-1 models through a combination of population means, fixed effects, and random effects. In these models, the intercept captures a constant effect across the analysis window, the linear term captures changes in overall growth or decay, the quadratic term captures symmetric effects primarily in the center of the window, the cubic term captures asymmetric effects, and the quartic term captures symmetric effects that are largely localized to window onsets and offsets. The full Level-2 model included all time terms and random effects of item on all time terms, plus additional fixed effects of each variable of interest (auditory-input, condition, and focal object) and of each item's auditory identification time (scaled score) on all time terms. Interactions between the temporal parameters and auditory-input or condition primarily relate to phonological/semantic competitor activation in the following ways: A difference in the speed of activation (i.e., latency effect) where one condition peaks in the middle of the window and one condition peaks earlier or later will be captured by a combination of quadratic and cubic effects, where the central peak has a stronger quadratic component and the non-central peak has a larger cubic component. A difference in the strength of activation (i.e., a

Visual fixations to targets, competitors, and controls were averaged into 25ms bins for each trial with a correct response. These values were then aggregated to calculate item averages for each condition in each time bin (item averages were used instead of subject averages so that auditory identification times of sounds and words could be included as an additional predictor in the model). Fixations to location-matched control items were subtracted from fixations to the focal object (target or competitor) to obtain measures of relative fixations (positive values indicate increased looks to the target/competitor relative to its matched control baseline). Relative fixations were analyzed from 200ms post-word onset (the time required to plan and execute an eye movement, Viviani, 1990) up to the time bin closest to the average RT across items, calculated individually for each competitor/control condition (Phonological-Word 1275ms; Semantic-Word 1275ms; Phonological-Sound 1200ms; Semantic-Sound 1275ms) and each target/control condition (Phonological-Word 1275ms; Semantic-Word 1250ms; Phonological-Sound 1125ms; Semantic-Sound 1225ms).

A linear mixed effects regression model was designed including fixed effects of Auditory-input (word, sound), Condition (phonological, semantic), Focal Object (target, competitor), and auditory identification time plus their interactions on all time terms, as well as a random effect of item for each polynomial time term (Table 5, and see Appendix Table B1 for full model specification and all effects). The inclusion of auditory identification time significantly improved the model's fit ($\chi^2(40)=632.92$, $p<.001$). Parameter-specific p-values were obtained using the Satterthwaite approximation for degrees of freedom, and a Holm-Bonferroni correction was applied to control for multiple comparisons within each model.

Results

Accuracy

We found a significant three-way interaction with the first Focal Object contrast (target vs. competitor/control) in the by-subject model (by-subjects, $t(135)=2.28$, $p=.025$), as well as with the second Focal Object contrast (competitor vs. control) in both models (by-subjects, $t(135)=-3.38$, $p<.001$; by-items, $t(96)=-2.60$, $p=.011$; Figure 2, Table 2). Follow-up pairwise comparisons indicated that the Semantic-Sound Competitor had lower accuracy than all other conditions (all ps<.001, by-subjects and by-items); no other comparisons were significant. The majority of errors (76.5%) in the Semantic-Sound condition were caused by clicks on the semantic competitor (see Table 3 for by-item accuracy scores). Accuracy was not correlated with nonverbal IQ, vocabulary size, or phonological memory, in either the full sample or within the sound or word conditions (all ps>.05). All incorrect trials were excluded from further analyses.

Response Time

The RT analysis revealed a significant main effect of the second Focal Object contrast (competitor vs. control; *Estimate*=−95.57, *SE*=86.97, $t(1497.3)=-4.13$, $p<.001$) indicating that participants were 95.57ms slower to affirm the absence of the target when a competitor

---

peak height effect) where two peaks of different heights are centered in the window but with similar onset and offset activations will be captured by quadratic and quartic effects with opposite signs; the quadratic component captures central peak height, and the quartic balances out onset and offset activation levels.

was present (Figure 3, Table 4). Response time was not correlated with nonverbal IQ, vocabulary size, or phonological memory, in either the full sample or within the sound or word conditions (all ps>.05).

Eye Movements

Target and competitor fixations.—Focal Object had a significant effect on the intercept, with relatively more fixations to the target than to the competitor overall, as well as on the linear, quadratic, and cubic terms, reflecting differences in the proportion of fixations over time. In addition, there was a Focal Object by Condition interaction on the quadratic term, as well as a three-way interaction between Auditory-input, Condition, and Focal Object on the intercept (see Figure 4, Table 5)[5]. The remainder of the manuscript will focus on competitor fixations in target-absent trials, which address our theoretical question of conceptual access during auditory processing. Further analyses of target fixations in no-competitor trials can be found in Appendix C.

Competitor fixations.—We began with a linear mixed effects regression on relative looks to the competitor (competitor minus control) with fixed effects of Auditory-input (word, sound), Condition (phonological, semantic), and identification time, plus their interactions on all time terms, as well as a random effect of item for each polynomial time term (see Appendix, Table B2 for all effects and model specifications). Because there were significant interactions between Auditory-input and Condition on the quadratic ($p<.001$) and quartic ($p=.001$) time terms, we followed-up with separate analyses for phonological and semantic trials with Auditory-input, identification time, and interactions with time terms as fixed effects.

Phonological competitors.—There were significant effects of Auditory-input on the quadratic and cubic time terms (Table 6; Appendix Table B3). The combination of effects captures the earlier peak of phonological activation for words (Figure 5, left, red) compared to sounds (Figure 5, left, orange), which peaked in the middle of the window.

Semantic competitors.—For semantic competitors, there was a significant effect of Auditory-input on the intercept, such that there were more relative fixations to the competitor after hearing words compared to sounds (Ms=0.065 and 0.048, respectively; Table 6; Appendix B4). There were also significant effects of Auditory-input on the quadratic and quartic time terms, reflecting a single large peak in the middle of the window for words (Figure 5, right, blue) compared to two smaller peaks at the beginning and end of the window for sounds (Figure 5, right, green). This latter pattern is notably distinct from other competitor effects (that is, the effects of phonological competition from both words and sounds, and of semantic competition from words). In most cases, we observe a clear rise and fall in competitor fixations around a single central inflection point, likely reflecting the

Note that effects of Condition (phonological, semantic) were only expected to emerge on competitor trials, as items on target and control trials did not share either phonological or semantic features with each other. However, given that the target, competitor, and control items differed between the phonological and semantic sets (e.g., clock, cloud, lightbulb for phonological and rain, umbrella, lungs for semantic), Condition was included in all models to account for any confounding stimulus characteristics. We additionally took care to minimize the influence of potential confounds by examining the effects of Auditory Input and Condition on competitor and target activation relative to controls within the same set.

consideration and subsequent discounting of the competitor as a potential candidate for selection (i.e., transient activation; Mirman et al., 2008). In contrast, semantic competition elicited by sounds emerges early and only temporarily declines before reemerging in the latter half of the window, suggesting that it takes longer for participants to discount semantic competitors activated by sounds relative to words.

In sum, we observed distinct effects of auditory input for phonological and semantic competition. The magnitude and pattern of phonological activation is comparable for words and sounds, but the time courses differ, with words eliciting and resolving phonological competition earlier than sounds. Semantic activation, in contrast, is characterized by qualitatively different patterns in response to words and sounds: while the overall magnitude of semantic activation is greater for words than sounds, sounds elicit recurring activation of semantic competitors even in cases in which the correct response is ultimately made.

Presentation order.—As an exploratory analysis, we examined whether competitor activation was moderated by the order in which the stimuli were presented. For instance, it may be the case that individuals would attend to a competitor image of a bone more or less depending on whether they had previously seen the associated target image of a dog. If competitor fixations increased following exposure to related stimuli (i.e., targets within the same set), looks to the cohort may reflect implicit or explicit associations learned over the course of the task, rather than naturally emerging coactivation patterns. While no effects of order were observed for phonological-word competitors (p=.104) or semantic-sound competitors (p=.245), later presentations resulted in significantly fewer fixations to phonological-sound competitors (*Estimate*=−0.05, *SE*=0.01, t(95.64)=−4.47, p<.0001), but not controls (*Estimate*=−0.01, *SE*=0.001, t(99.99)=−1.39, p=.330). Similarly, later presentations resulted in significantly fewer fixations to semantic-word competitors (*Estimate*=−0.04, *SE*=0.01, t(111.86)=−4.43 p<.0001), but not controls (*Estimate*=0.001, *SE*=0.007, t(104.3)=0.18, p=.858). This pattern indicates that the observed competitor effects are not driven by associations between stimuli acquired throughout the course of the task, and provides further evidence that words and sounds activate phonological and semantic information in distinct ways. Specifically, we observe that competition arising from purportedly direct routes of processing (i.e., words to phonological information and sounds to semantic information) remain stable over repeated presentations, while competitor activation arising from potentially more indirect pathways (words to semantic information and sounds to phonological information) is most robust at first exposure and declines over time (see Appendix C for more details).

Stimulus Manipulation Check Study: Effects of Auditory Input on Phonological, Semantic, and Visual Similarity

A follow-up study was conducted to confirm the effectiveness of our manipulations; that is, that targets and competitors were perceived as more phonologically/semantically similar to each other than targets and controls. The follow-up study also enabled us to examine additional mechanisms that may underlie differences observed between words and sounds. Specifically, we explored the possibility that words and sounds may influence not only how and when phonological and semantic information are activated, but also the representations

themselves. Past research has shown that words, particularly concrete nouns as used in the current study, activate prototypical semantic concepts (Hampton, 2016; Lupyan & Thompson-Schill, 2012), while characteristic sounds may be more likely to activate specific referents linked to their original source (Edmiston & Lupyan, 2013, 2015; Lupyan & Thompson-Schill, 2012). In the context of the current study, it is possible that the conceptual representations elicited by words (e.g., a general concept of a dog) were perceived as more closely related to the competitors compared to potentially more specific representations activated by sounds (e.g., a particular type of dog performing a specific activity). Such a result could potentially account for the relatively greater semantic activation observed in the word condition. On the other hand, although identified sounds may evoke more specific referents relative to words, there may be greater ambiguity regarding the boundaries of what a given sound represents. For instance, the word "dog" clearly refers to a dog and not a wolf, but a howling sound may strongly activate both even if the listener ultimately identifies the actual source. Indeed, a number of studies utilizing electrophysiological (Hendrickson et al., 2015) as well as behavioral (Chen & Spence, 2018; Saygin, Dick, & Bates, 2005) measures have found evidence that sounds can elicit more diffuse or less fine-grained representations relative to words. If so, sounds may trigger stronger activation of semantic associates, which could have contributed to the relatively lower accuracy rate observed in the semantic-sound condition, as well as the recurring fixations towards semantic competitors throughout the analysis window.

Method

Twenty-eight English monolinguals on MTurk judged the similarity between the target and the competitor and control stimuli used in the primary experiment. As in the original study, participants were randomly assigned to either the word or sound condition and made similarity judgments for each of the 18 sets of stimuli included in the primary analysis. On each trial, participants were presented with a single drawing of either a competitor or control item paired with the auditory presentation of a target sound or word. Participants were asked to consider the referents associated with the auditory and visual stimuli and rate their similarity in terms of their labels (phonological) and association to each other (semantic), as well as their functional and visual similarity, each on a scale between 0 (not at all similar) to 100 (very similar). The latter two measures were included to identify potentially confounding stimulus characteristics that could elicit preferential fixations towards the competitor relative to the control item. We first examined the fixed effects of Auditory input (effect coded as sound: −.53; word: +.47), Condition (phonological: −.57; semantic: +.43), and Comparison Object (control: −.49; competitor: +.51) on each measure of similarity to the targets (phonological, semantic/association, semantic/functional, visual) using linear mixed effects models with random effects for item and subject. We then reran the primary eye-tracking analyses including each similarity rating as a covariate. In this case, each condition (phonological, semantic) was examined separately using linear mixed effects models with the relative proportion of fixations to the competitor object (competitor minus control) as the outcome variable, with fixed effects of Auditory input (words, sounds), relative similarity ratings (competitor minus control), identification time, and interactions

with each of the time terms, and random effects of item for each time term. *P*-values were corrected for multiple comparisons using the Holm-Bonferroni method.

Results

Phonological (Label) Similarity

Phonological (*Estimate* = 39.73, *SE*=4.44, *t*(21.65)=8.95, *p*<.0001), but not semantic (*Estimate*=−0.01, *SE*=4.03, *t*(22.62)=−0.003, *p*=.998), competitors were perceived as more phonologically similar to the target relative to control items, resulting in a significant interaction between Condition (phonological, semantic) and Comparison Object (competitor, control) (*Estimate*=−39.63, *SE*=5.48, *t*(15.96)=−7.24, *p*<.0001). There were no effects of or interactions with auditory input (all *p*s>.05), suggesting that the labels associated with both word and sound targets were perceived as more similar to the phonological competitors than controls.

Semantic (Association) Similarity

Semantic (*Estimate*=76.86, *SE*=4.07, *t*(20.33)=18.91, *p*<.0001), but not phonological (*Estimate*=−1.17, *SE*=4.51, *t*(19.88)=−0.26, *p*=.798), competitors were perceived as more semantically associated with the target relative to their control items, resulting in a significant Condition × Comparison Object interaction (*Estimate*=77.72, *SE*=5.73, *t*(16.26)=13.56, *p*<.0001). While this was the case in both the sound (*Estimate*=81.39, *SE*=4.56, *t*(29.38)=17.87, *p*<.0001) and word conditions (*Estimate*=72.34, *SE*=4.47, *t*(27.86)=16.20, *p*<.0001), there was a significant three-way interaction between Auditory Input, Condition, and Comparison Object (*Estimate* = −10.59, *SE*=4.10, *t*(763.38)=−2.59, *p*<.01), such that participants listening to sounds perceived relatively greater semantic similarity between the semantic targets and competitors compared to participants listening to words.

Semantic (Functional) Similarity

Perhaps unsurprisingly, semantic (*Estimate*=21.30, *SE*=5.34, *t*(18.17)=3.99, *p*<.001), but not phonological (*Estimate*=1.03, *SE*=5.94, *t*(17.89)=0.17, *p*=.863), competitors were perceived as more functionally similar to the target relative to controls, resulting in a significant interaction between Condition and Comparison Object (*Estimate*=20.25, *SE*=7.74, *t*(16.01)=2.62, *p*=.019). There were no effects of or interactions with Auditory Input (all *p*s>.05).

Visual Similarity

Items in the semantic condition were rated as more visually similar to the targets than those in the phonological condition (i.e., a main effect of Condition; *Estimate*=4.67, *SE*=1.96, *t*(20.20)=2.38, *p*=.027). However, there was no significant difference between competitors and controls, no interaction between Condition and Comparison Object, and no effects of or interactions with Auditory Input (all *p*s>.05).

Reanalysis of Eye-Tracking Data Controlling for Phonological, Semantic, and Visual Similarity

In addition to confirming that our manipulation of stimuli characteristics was successful, we assessed the robustness of our primary findings by rerunning the eye-tracking analysis while including each of the similarity ratings as a covariate. The inclusion of similarity ratings did not notably change the pattern of eye-tracking results, with two exceptions: (1) for phonological activation, including ratings of semantic similarity (either association or function), made the previously significant effect of auditory input on the cubic time term marginal; for semantic activation, including ratings of semantic similarity (either association or function) or visual similarity made the previously significant effect of auditory input on the intercept no longer significant (see Appendix Table B5 for comparison of p-values across all models). Recall that this latter effect reflected the increase in semantic activation by words compared to sounds.

In sum, the follow-up study confirmed the successful manipulation of stimuli characteristics, but also revealed differences in functional similarity between semantic competitors and controls (i.e., semantic targets were perceived as more functionally similar to their competitors than controls), as well as in visual similarity between items in the semantic and phonological conditions (semantic targets were perceived as more visually similar to both competitors and controls relative to phonological targets). If semantic targets were in fact more functionally similar to their competitors than control items, our interpretation of semantic competitor effects would remain largely the same, as increased activation of competitors due to functional overlap also implicates semantic access. Greater visual similarity between targets and semantic items compared to phonological items may have led to greater overall fixations to competitors and controls in the semantic condition compared to the phonological condition. However, our interpretations of competitor type and auditory input effects should remain unaffected, as analyses were always centered on the fixations to competitors relative to controls (which did not significantly differ from each other in visual similarity). Perhaps most intriguing is the finding that semantic targets were perceived as more semantically similar to their competitors in the sound condition compared to the word condition. Though the inclusion of semantic similarity in the eye-tracking models did not notably alter the previously observed patterns of semantic activation, this finding suggests that auditory input may alter not only how concepts are accessed, but also the conceptual representations themselves.

General Discussion

The current study examined relative activation of word-form (i.e., phonological) knowledge and meaning (i.e., semantic) knowledge while listening to spoken words or characteristic sounds using eye-tracking in a visual world paradigm. By examining the shape and time course of visual fixations to phonological and semantic competitors, we found that spoken words triggered behavior consistent with a mode of processing that began with rapid phonological activation, followed by strong, but relatively slower semantic access. Characteristic sounds, on the other hand, prompted early and recurring semantic processing that preceded and extended past the activation of phonological competitors.

The earlier activation of phonological information for words compared to sounds is consistent with Chen and Spence's (2011) model of auditory processing, as is the relative timing of phonological and semantic activation within each of the auditory input conditions (phonological → semantic for words, and semantic → phonological for sounds). Evidence of earlier semantic activation for sounds than words, however, is less robust – while there was slightly greater semantic activation for sounds compared to words at the beginning of the time window, we did not observe a clear temporal shift of otherwise similar fixation curves, as we saw for phonological activation. Instead, our analyses revealed that words and sounds were associated with qualitatively different patterns of activation. Words prompted a single large peak in the center of the window, which likely reflects the transient activation and subsequent discounting of semantic competitors. Sounds elicited two smaller peaks at the beginning and end of the window, potentially reflecting recurring activation and a later stage at which semantic competitors are disregarded as potential candidates for selection.

One possibility is that the different patterns of semantic activation for words and sounds result from differences in how likely it is that a given auditory input will activate not only the direct referent, but also semantically related concepts. In contrast to words, environmental sounds, such as a dog barking, are more likely to be encoded and experienced in the immediate presence of its source (a dog), as well as any objects or contexts that are associated with it (e.g., a leash). The causal relationship between sounds and their referents also increases the likelihood that semantically similar concepts (e.g., a dog and a wolf) will become associated with similar sounds (e.g., howling). As a result, even if the direct referent of an auditory input receives comparable activation by words and sounds, the conceptual boundaries between the referent and both thematically- and categorically-related concepts may be less distinct when listening to sounds compared to words (Chen & Spence, 2018; Hendrickson, et al., 2015; Saygin, Dick, & Bates, 2005). Indeed, our behavioral data are in line with this hypothesis, as accuracy on semantic-sound trials was significantly lower than all other conditions, almost exclusively as a result of participants incorrectly selecting the competitor. Most tellingly, results from our follow-up study revealed that semantic targets and competitors were rated as more closely associated when elicited by sounds compared to words. Although the processes underlying the differential effects of auditory input for phonological and semantic trials are still speculative, our findings suggest that auditory input can influence both conceptual activation as well as representation. Notably, while the timing and patterns of phonological and semantic activation differed between inputs, we eventually observed robust activation of phonological and semantic information in both word and sound conditions. Particularly surprising was the magnitude of phonological activation in the sound condition, which was comparable in magnitude to that of words.

Lastly, a key finding in support of Chen and Spence's (2011) auditory processing model was that participants fixated the phonological competitor several hundred milliseconds earlier when cued by a word compared to a sound. One important distinction between conditions, however, was the direct overlap between the phonological targets and their competitors. While the phonological sound competitor did not share any features with the target stimulus *itself* (but rather the label associated with the sound stimulus), the onsets of the word condition's target and competitor objects were highly similar by design. The earlier fixations towards the competitor in response to words may therefore reflect access to phonological

information as well as more overlap between the competitor label and the auditory stimulus. Direct comparisons between conditions should therefore be interpreted with these potential differences in mind, and considered in conjunction with the relative time courses of phonological and semantic activation within input conditions. Moving forward, it will also be important to determine how differences in speech and non-speech sound activation extend to other phonological relationships (e.g., offset overlap or cross-linguistic competitors), as well as to other semantic relationships (e.g., taxonomic or functional associations).

In conclusion, we have identified critical similarities and differences in how humans process two types of auditory input – linguistic spoken words and non-linguistic sounds. Spoken words facilitated rapid access to phonological information, but activated semantic information more gradually. The relatively delayed semantic activation (compared to phonological activation) by words may reflect bottom-up processes, such that spoken words activate phonological and lexical forms, which then activate the corresponding semantic concepts. Non-linguistic sounds prompted early and prolonged semantic processing, which preceded robust phonological activation. The later onset of phonological activation by sounds may be a function of top-down processes, such that sounds activate associated semantic concepts, which then feed down to activate corresponding lexical and phonological information. Differences in the time course, magnitude, and persistence of phonological and semantic activation reveals features of the cognitive architecture used to process auditory input. Together, our findings suggest that spoken words and characteristic sounds elicit distinct, but highly dynamic patterns of activation, with interactions observed among linguistic and non-linguistic processing, auditory and visual processing, and phonological and semantic processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Allopenna PD, Magnuson JS, & Tanenhaus MK (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. 10.1006/jmla.1997.2558

Baayen H, Vasishth S, Kliegl R, & Bates D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. Journal of Memory and Language, 94, 206–234. 10.1016/J.JML.2016.11.006

Bates E, Andonova E, D'Amico S, Jacobsen T, Kohnert K, Lu C, … Pleh C. (2000). Introducing the CRL international picture naming project (CRL-IPNP). Center for Research in Language Newsletter, 12(1).

Brysbaert M, & New B. (2009). Moving beyond Ku era and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for

American English. Behavior Research Methods, 41(4), 977–990. 10.3758/BRM.41.4.977 [PubMed: 19897807]

Chabal S, & Marian V. (2015). Speakers of different languages process the visual world differently. Journal of Experimental Psychology: General, 144(3), 539–550. 10.1037/xge0000075 [PubMed: 26030171]

Chen Y-C, & Spence C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. Journal of Experimental Psychology: Human Perception and Performance, 37(5), 1554–1568. 10.1037/a0024329 [PubMed: 21688942]

Chen Y-C, & Spence C. (2013). The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. Multisensory Research, 26(4), 371–386. 10.1163/22134808-00002420 [PubMed: 24319929]

Chen YC, & Spence C. (2018). Audiovisual semantic interactions between linguistic and nonlinguistic stimuli: The time-courses and categorical specificity. Journal of Experimental Psychology: Human Perception and Performance, 44(10), 1488–1507. 10.1037/xhp0000545 [PubMed: 29708379]

Coltheart M. (1981). The MRC psycholinguistic database. Quartely Journal of Experimental Psychology, 33(4), 497–505. 10.1080/14640748108400805

Connolly JF, & Phillips NA (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. Journal of Cognitive Neuroscience, 6(3), 256–266. 10.1162/jocn.1994.6.3.256 [PubMed: 23964975]

Dunn LM (1997). Examiner's Manual for the PPVT-III: Peabody Picture Vocabulary Test-Third Edition. Circle Pines, MN: American Guidance Service.

Edmiston P, & Lupyan G. (2013). Verbal and nonverbal cues activate concepts differently, at different times. Proceedings of the 35th Annual Conference of the Cognitive Science Society, 2243–2248.

Edmiston P, & Lupyan G. (2015). What makes words special? Words as unmotivated cues. Cognition, 143, 93–100. 10.1016/j.cognition.2015.06.008 [PubMed: 26117488]

Glaser WR, & Glaser MO (1989). Context effects in stroop-like word and picture processing. Journal of Experimental Psychology. General, 118(1), 13–42. 10.1037/0096-3445.118.1.13 [PubMed: 2522504]

Hampton JA (2016). Categories, prototypes, and exemplars. In Reimer N (Ed.), Routledge Handbook of Semantics (pp. 125–141). New York: Routledge.

Hendrickson K, Walenski M, Friend M, & Love T. (2015). The organization of words and environmental sounds in memory. Neuropsychologia, 69, 67–76. 10.1016/j.neuropsychologia.2015.01.035 [PubMed: 25624059]

Huettig F, & McQueen JM (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. Journal of Memory and Language, 57(4), 460–482. 10.1016/j.jml.2007.02.001

Iordanescu L, Grabowecky M, Franconeri S, Theeuwes J, & Suzuki S. (2010). Characteristic sounds make you look at target objects more quickly. Attention, Perception & Psychophysics, 72(7), 1736–1741. 10.3758/APP.72.7.1736

Marian V, Bartolotti J, Chabal S, & Shook A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. PLoS ONE, 7(8), e43230. 10.1371/journal.pone.0043230

McClelland JL, & Elman JL (1986). The TRACE Model of Speech Perception. Cognitive Psychology, 18, 1–86. [PubMed: 3753912]

Mirman D, Dixon JA, & Magnuson JS (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of Memory and Language, 59(4), 475–494. 10.1016/j.jml.2007.11.006 [PubMed: 19060958]

Mirman D, Magnuson JS, Graf Estes K, & Dixon JA (2008). The link between statistical segmentation and word learning in adults. Cognition, 108(1), 271–280. 10.1016/j.cognition.2008.02.003 [PubMed: 18355803]

PsychCorp. (1999). Wechsler abbreviated scale of intelligence (WASI). San Antonio, TX: Harcourt Assessment.

Saygin AP, Dick F, & Bates E. (2005). An on-line task for contrasting auditory processing in the verbal and nonverbal domains and norms for younger and older adults. Behavior Research Methods, 37(1), 99–110. 10.3758/BF03206403 [PubMed: 16097349]

Shook A, & Marian V. (2013). The bilingual language interaction network for comprehension of speech. Bilingualism, 16(2), 304–324. 10.1017/S1366728912000466

Torgesen J, Wagner RK, & Rashotte CA (n.d.). Comprehensive test of phonological processing. Austin, TX: Pro-Ed.

Van Petten C, Coulson S, Rubin S, Plante E, & Parks M. (1999). Time course of word identification and semantic integration in spoken language. Journal of Experimental Psychology. Learning, Memory, and Cognition. 10.1037//0278-7393.25.2.394

Vandeberg L, Guadalupe T, & Zwaan RA (2011). How verbs can activate things: Cross-language activation across word classes. Acta Psychologica, 138(1), 68–73. 10.1016/j.actpsy.2011.05.007 [PubMed: 21696694]

Viviani P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. Reviews of Oculomotor Research, 4, 353–393. [PubMed: 7492533]

Yee E, & Sedivy JC (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. J Exp Psychol Learn Mem Cogn, 32(1), 1–14. 10.1037/0278-7393.32.1.1 [PubMed: 16478336]
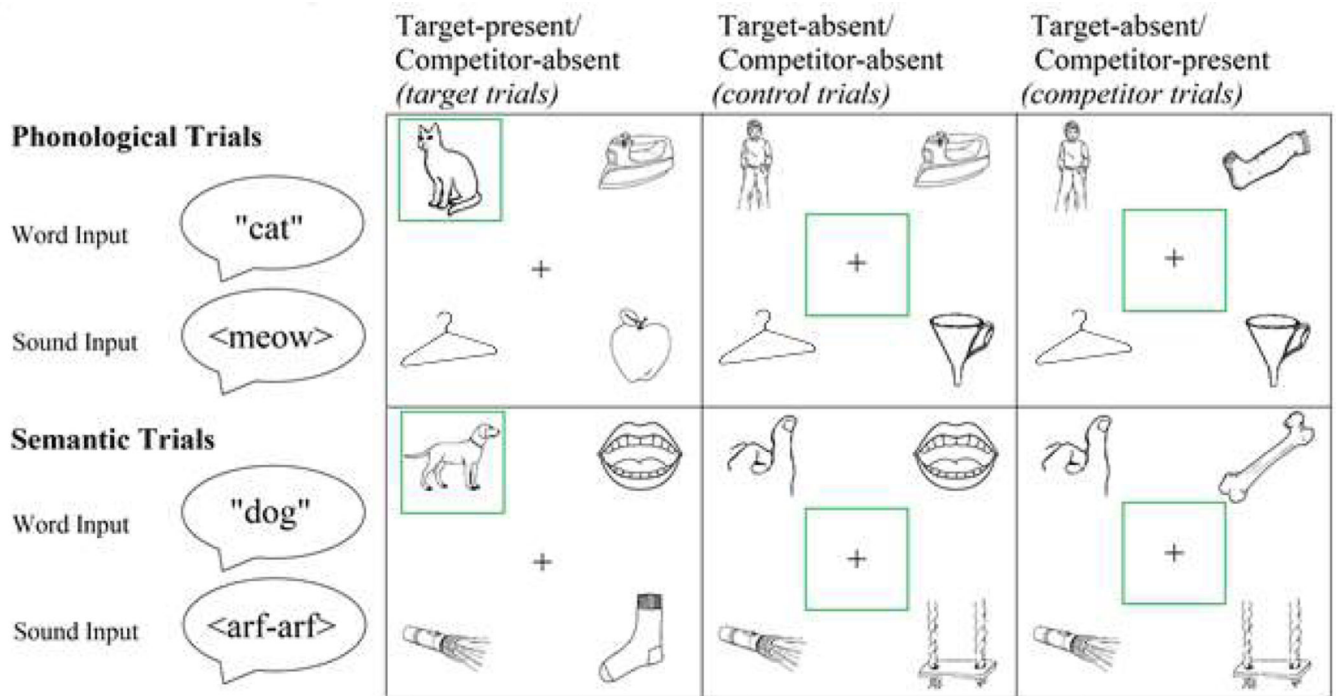
Figure 1.
Examples of stimulus displays. Green boxes indicate the correct response. Target displays (left column) included three filler items and the target (e.g., cat; dog). Control displays (middle column) included four filler items. Competitor displays (right column) included three filler items and a phonological (e.g., cast; top) or semantic (e.g., bone; bottom) competitor.
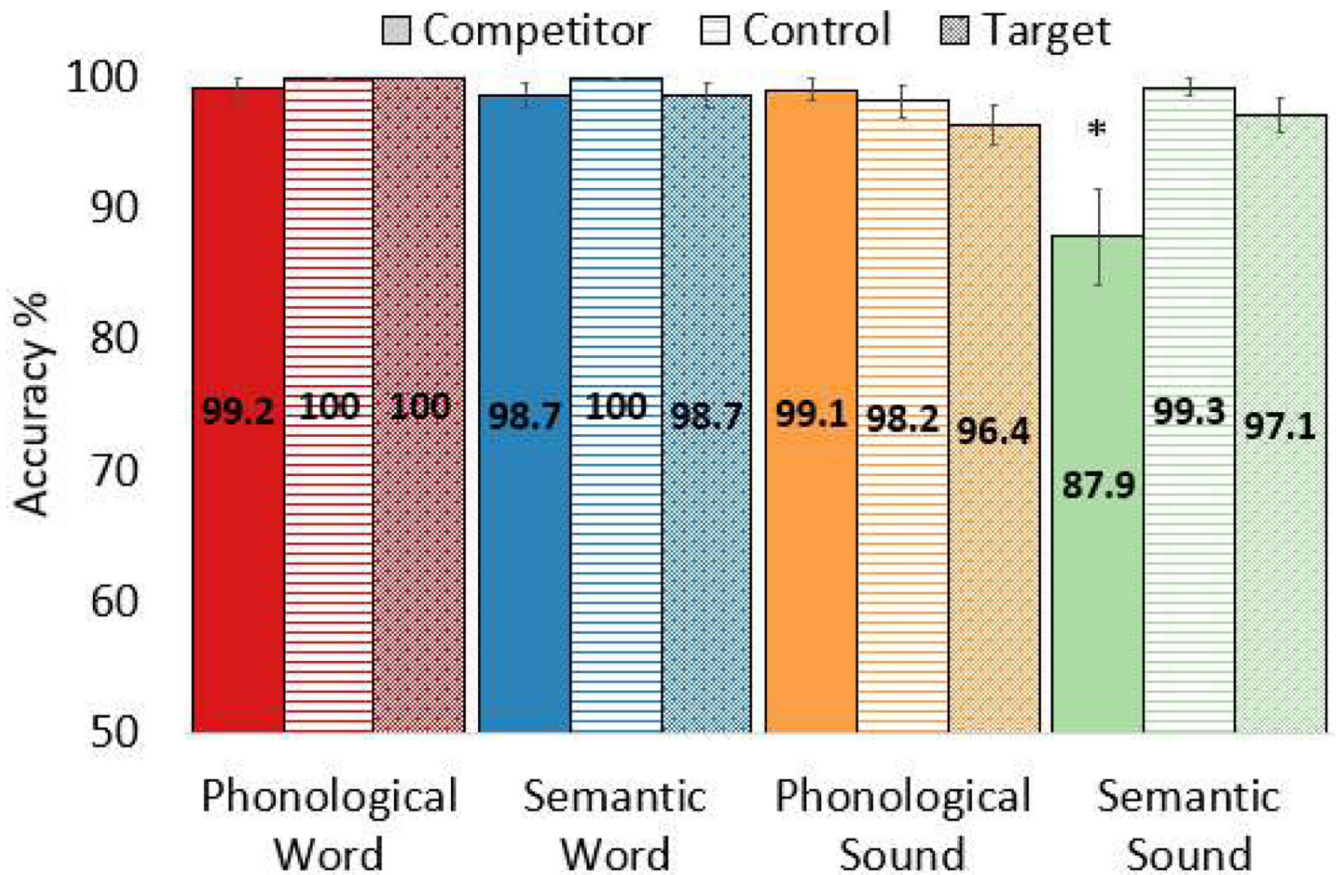
Figure 2.
Accuracy by auditory input, condition, and focal object. In trials with a characteristic sound and a thematic semantic competitor (light-green bar, solid fill), accuracy was lower than each of the other eleven conditions (ps<.001 by-subjects and by-items). Error bars represent standard error.
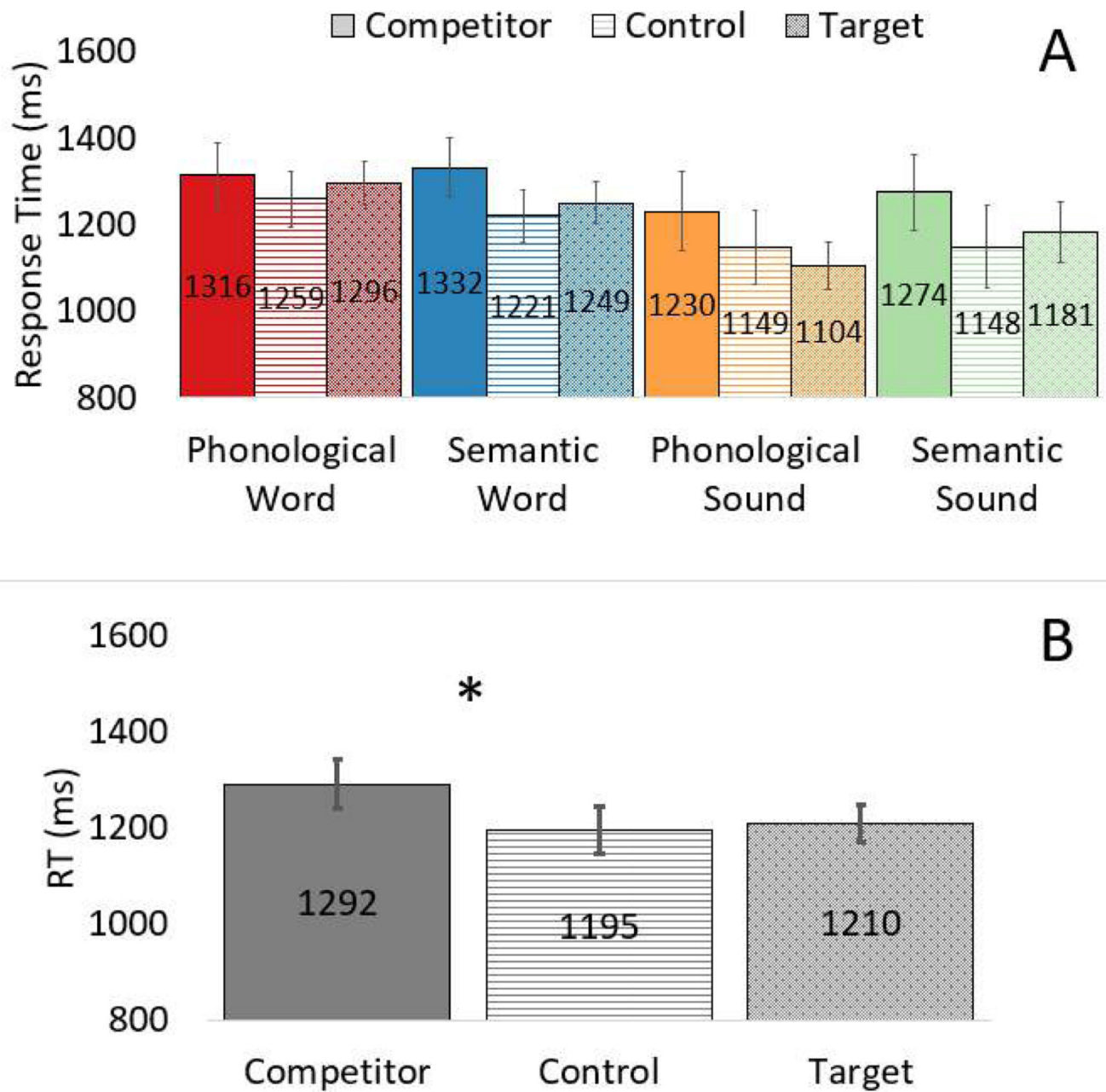
Figure 3.
(A) Response time by auditory input, condition, and focal object. (B) A main effect of focal object indicated that competitor trials (solid) were responded to slower than control trials (dashed). Error bars represent standard error.
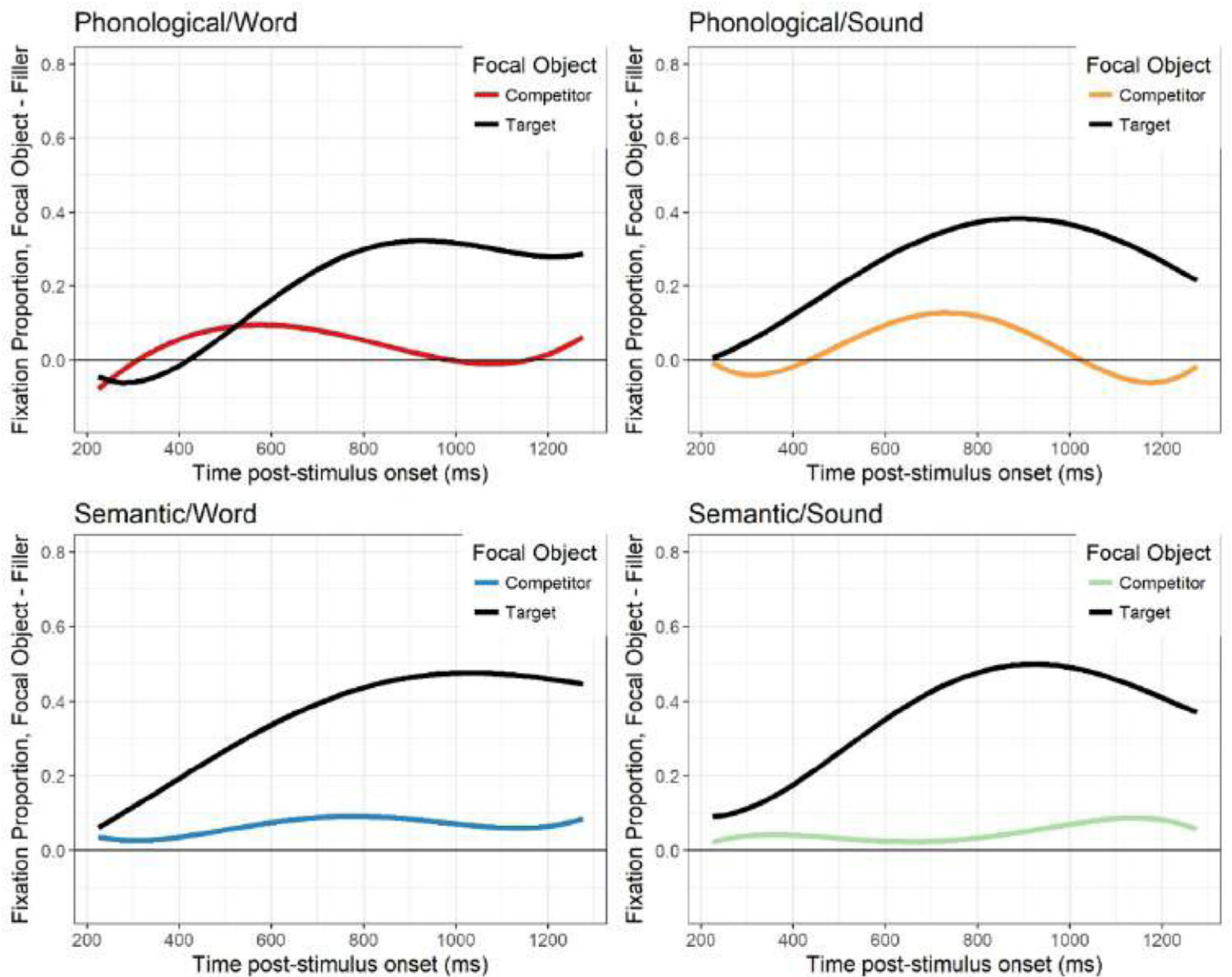
Figure 4.
Activation of targets (black) and phonological (red/orange) and semantic (blue/green) competitors in response to spoken words (red/blue) and characteristic sounds (orange/green). Curves represent growth curve model fits for difference waves comparing target/competitor fixations to control baselines when identification time was held constant (set to mean values). Positive values indicate more looks to the target/competitor than the control.
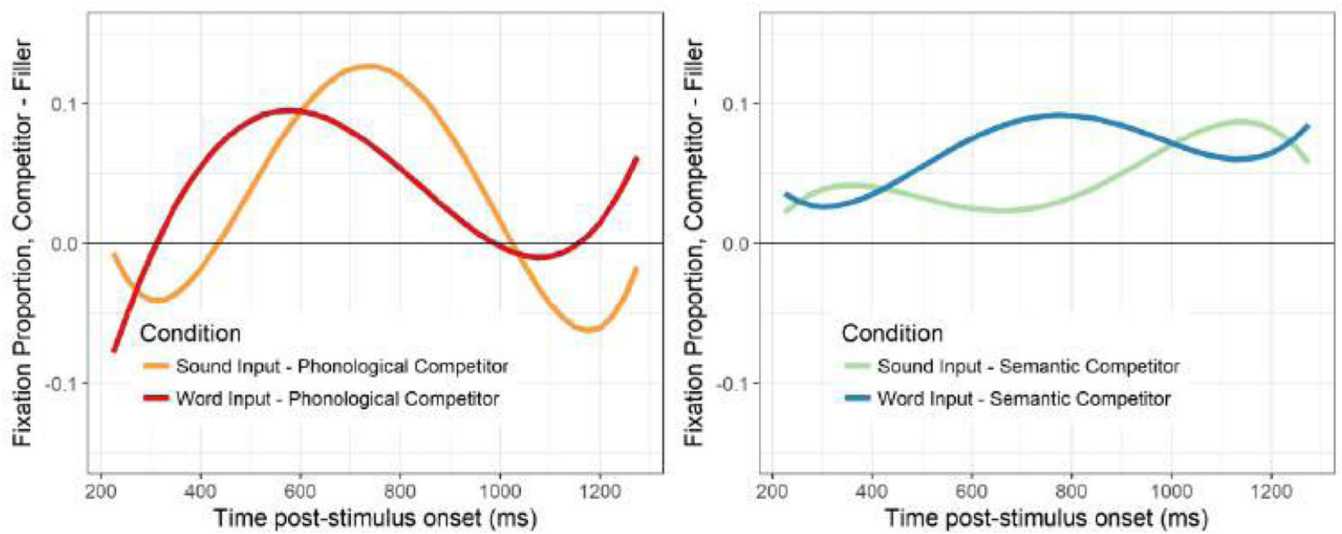
Figure 5.
Relative fixations to phonological competitors (left) following auditory input from sounds (orange) and words (red). Relative fixations to semantic competitors (right) following auditory input from sounds (green) and words (blue). Curves represent growth curve model fits for difference waves comparing target fixations to control baselines when identification time was held constant (set to mean values). Positive values indicate more looks to the competitor than the control.

Table 1.

Cognitive and Linguistic Data

Measures	Characteristic Sounds Participants		Spoken Words Participants		Statistical Comparison (*t*-test)
	Mean	SD	Mean	SD	*p*-value
Age	20.33	2.94	21.93	2.84	0.14
IQ Standard Score (*WASI*)	113.0	9.9	115.1	8.4	0.54
Phonological Memory Standard Score (*CTOPP*)	110.2	9.3	114.6	9.9	0.22
English Vocabulary Standard Score (*PPVT*)	116.6	11.7	120.1	8.4	0.36

Table 2.

Linear mixed effect regression models of accuracy.

Fixed effect	Estimate	SE	df	t	p	
By-subject model						
(Intercept)	97.93	0.50	27.00	195.48	<.001	***
AuditoryInput	3.08	1.00	27.00	3.07	0.005	**
Condition	−1.84	0.71	135.00	−2.58	0.011	*
FocalObject1	−0.26	0.76	135.00	−0.34	0.733	
FocalObject2	3.10	0.87	135.00	3.55	0.001	***
AudIn:Cond	2.54	1.43	135.00	1.78	0.077	
AudIn:FocalObject1	0.79	1.51	135.00	0.53	0.600	
AudIn:FocalObject2	−4.18	1.75	135.00	−2.39	0.018	*
Cond:FocalObject1	−2.24	1.51	135.00	−1.48	0.141	
Cond:FocalObject2	6.21	1.75	135.00	3.55	0.001	***
AudIn:Cond:FocalObject1	6.89	3.03	135.00	2.28	0.025	*
AudIn:Cond:FocalObject2	−11.82	3.50	135.00	−3.38	0.001	***
By-item model						
(Intercept)	97.77	0.46	96	211.75	<.001	***
AuditoryInput	3.22	0.92	96	3.49	0.001	***
Condition	−1.88	0.93	96	−2.03	0.046	*
FocalObject1	−0.40	0.98	96	−0.41	0.681	
FocalObject2	3.53	1.13	96	3.12	0.002	**
AudIn:Cond	2.54	1.86	96	1.37	0.174	
AudIn:FocalObject1	1.18	1.96	96	0.60	0.549	
AudIn:FocalObject2	−4.84	2.26	96	−2.14	0.035	*
Cond:FocalObject1	−2.36	1.97	96	−1.20	0.234	
Cond:FocalObject2	6.41	2.28	96	2.82	0.006	**
AudIn:Cond:FocalObject1	6.89	3.94	96	1.75	0.084	
AudIn:Cond:FocalObject2	−11.82	4.55	96	−2.60	0.011	*

Note. See Appendix for full model description. AuditoryInput = spoken word or characteristic sound; Condition = phonological or semantic; Focal Object1 = target or competitor/control; Focal Object2 = competitor or control. Contrasts for each factor were centered, and *t*-tests used the Satterthwaite approximation for degrees of freedom. Follow-up tests were Tukey-adjusted for multiple comparisons.

***
  p<.001

**
  p<.01

*
  p<.05

Table 3.

Accuracy, Target Sound, and Competitor Image for items in the Semantic Sound condition.

Accuracy	Target Sound	Competitor Image
64.3%	<door slam>	hinge
78.6%	<cow moo>	milk
78.6%	<rain patter>	umbrella
78.6%	<can opening>	beer (in glass stein)
85.7%	<chicken cluck>	egg
92.9%	<jackhammer>	sidewalk
100.0%	<dog bark>	bone
100.0%	<sheep baa>	yarn
100.0%	<toilet flush>	sink
100.0%	<owl hoot>	tree

Table 4.

Linear mixed effect regression model of response time on correct trials.

Fixed Effects	Estimate	SE	df	t	p	
(Intercept)	1232.58	49.17	37.40	25.07	<.001	***
Condition	8.98	50.03	16.00	0.18	0.860	
AuditoryInput	93.49	86.97	27.00	1.08	0.292	
FocalObject1	30.47	20.02	1497.10	1.52	0.128	
FocalObject2	−95.57	23.17	1497.30	−4.13	<.001	***
Cond:AudIn	−64.87	38.03	1497.20	−1.71	0.088	
Cond:FocalObject1	−3.67	40.28	1497.20	−0.09	0.927	
Cond:FocalObject2	−48.16	46.55	1497.20	−1.04	0.301	
AudIn:FocalObject1	−42.07	40.10	1497.20	−1.05	0.294	
AudIn:FocalObject2	18.92	46.40	1497.30	0.41	0.684	
Cond:AudIn:FocalObject1	83.75	80.65	1497.20	1.04	0.299	
Cond:AudIn:FocalObject2	−19.21	93.22	1497.20	−0.21	0.837	

Note. See Appendix for full model description. AuditoryInput = spoken word or characteristic sound; Condition = phonological or semantic; Focal Object1 = target or competitor/control; Focal Object2 = competitor or control. Contrasts for each factor were centered, and *t*-tests used the Satterthwaite approximation for degrees of freedom.

***
p<.001

**
p<.01

*
p<.05

Table 5.

GCA model of target/competitor fixations by auditory-input, condition, and Focal Object

Fixed Effect	Estimate	SE	df	t	p
(Intercept)	0.17	0.03	17.90	6.77	< .001 ***
Linear	0.37	0.04	17.50	8.75	< .001 ***
Quadratic	−0.29	0.05	16.50	−5.58	.002 **
Cubic	−0.02	0.03	17.20	−0.76	1.000
Quartic	0.07	0.03	13.70	2.61	.960
AuditoryInput	−0.01	0.00	3023	−2.26	1.000
Condition	0.08	0.05	17.90	1.64	1.000
FocalObject	0.25	0.00	3018	51.30	< .001 ***
Linear:FocalObject	0.70	0.03	3018	22.28	< .001 ***
Quad:FocalObject	−0.26	0.03	3018	−8.32	< .001 ***
Cubic:FocalObject	−0.17	0.03	3018	−5.24	< .001 ***
Quartic:FocalObject	0.01	0.03	3018	0.29	1.000
Linear:AudInput:FocalObject	0.13	0.06	3018	2.04	1.000
Quad:AudInput:FocalObject	0.09	0.06	3018	1.49	1.000
Cubic:AudInput:FocalObject	−0.02	0.06	3018	−0.32	1.000
Quartic:AudInput:Focus	0.00	0.06	3018	0.04	1.000
Linear:Cond:Focus	−0.06	0.06	3018	−0.93	1.000
Quad:Cond:Focus	−0.33	0.06	3018	−5.15	< .001 ***
Cubic:Cond:Focus	0.19	0.06	3018	3.05	.127
Quartic:Cond:Focus	0.12	0.06	3018	1.84	1.000
AudInput:Cond:Focus	0.07	0.02	3018	3.65	.016 *
Linear:AudInput:Cond:Focus	−0.09	0.13	3018	−0.74	1.000
Quad:AudInput:Cond:Focus	0.42	0.13	3018	3.26	.064
Cubic:AudInput:Cond:Focus	0.26	0.13	3018	2.05	1.000
Quartic:AudInput:Cond:Focus	−0.43	0.13	3018	−3.34	.050

Note. See Appendix Table B1 for full model output including effects of Identification Time. The contrasts for Auditory-input (sound: −.5, word: +.5), Condition (phonological: −.556, semantic: +.444), and Focal Object (competitor: −.5, target: +.5) were centered, and *t*-tests used the Satterthwaite approximation for degrees of freedom with Holm-Bonferroni-adjusted p-values.

***
p<.001

**
p<.01

*
p<.05.

Table 6.

Relative fixations to phonological and semantic competitors by words and sounds

Fixed Effects	Estimate	SE	df	t	p
Phonological Competitors					
(Intercept)	0.03	0.02	7.90	1.30	1.000
Linear	−0.12	0.15	7.80	−0.79	1.000
Quadratic	−0.31	0.09	7.90	−3.45	.098
Cubic	0.14	0.10	8.00	1.41	1.000
Quartic	0.15	0.06	7.70	2.44	.334
AuditoryInput	0.00	0.01	658.40	−0.09	1.000
Linear:AudIn	0.08	0.05	654.40	1.59	.786
Quad:AudIn	0.27	0.05	661.30	5.06	< .001 ***
Cubic:AudIn	0.16	0.05	655.90	2.98	.042 *
Quartic:AudIn	−0.15	0.05	650.10	−2.87	.051
Semantic Competitors					
(Intercept)	0.06	0.02	10.00	3.57	.076
Linear	0.11	0.07	9.50	1.56	1.000
Quadratic	−0.05	0.15	9.70	−0.32	1.000
Cubic	0.00	0.06	9.00	−0.05	1.000
Quartic	0.01	0.07	9.30	0.09	1.000
AuditoryInput	0.02	0.01	807.90	3.27	.018 *
Linear:AudIn	−0.01	0.04	810.10	−0.34	1.000
Quad:AudIn	−0.18	0.04	808.30	−4.73	< .001 ***
Cubic:AudIn	0.03	0.04	814.70	0.71	1.000
Quartic:AudIn	0.13	0.04	810.60	3.28	.018 *

Note. See Appendix Tables B3 and B4 for full model outputs including effects of Identification Time. The contrasts for Auditory-input were centered (sound: −.5, word: +.5), and *t*-tests used the Satterthwaite approximation for degrees of freedom with Holm-Bonferroni-adjusted p-values.

***
 p<.001

**
 p<.01

*
 p<.05.