# Words and non-speech sounds access lexical and semantic knowledge differently

**Peiyao Chen[1] (pchen@u.northwestern.edu)**
**James Bartolotti[1] (j-bartolotti@u.northwestern.edu)**
**Scott R. Schroeder[2] (scott.r.schroeder@hofstra.edu)**
**Sirada Rochanavibhata[1] (siradarochanavibhata2020@u.northwestern.edu)**
**Viorica Marian[1] (v-marian@northwestern.edu)**

[1]Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208 USA
[2]Department of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, NY 11549 USA

**Abstract**

Using an eye-tracking paradigm, we examined the strength and speed of access to lexical knowledge (e.g., our representation of the word *dog* in our mental vocabulary) and semantic knowledge (e.g., our knowledge that a dog is associated with a leash) via both spoken words (e.g., "dog") and characteristic sounds (e.g., a dog's bark). Results show that both spoken words and characteristic sounds activate lexical and semantic knowledge, but with different patterns. Spoken words activate lexical knowledge faster than characteristic sounds do, but with the same strength. In contrast, characteristic sounds access semantic knowledge stronger than spoken words do, but with the same speed. These findings reveal similarities and differences in the activation of conceptual knowledge by verbal and non-verbal means and advance our understanding of how auditory input is cognitively processed.

**Keywords:** speech comprehension; sound processing; lexical competition; semantic competition; eye-tracking

## Introduction

The human auditory system receives and processes different types of input from the environment. Sounds that come from an entity typically provide information about a specific member of a group: the sound of a dog barking usually reveals that particular dog's size and location. Words, in contrast, often refer to a category without providing information about the specific member of that group. For example, the spoken word "dog" provides no information about either the dog's size or location. Models of auditory processing take into account words' and sounds' unique features, and propose that these two types of input access conceptual knowledge via different routes (Chen & Spence, 2011). In the current study, we directly examine and compare the timecourse of semantic and lexical activation by spoken words and characteristic sounds.

During speech processing, individual words' lexical form and semantic meaning are rapidly accessed (Connolly & Phillips, 1994; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Hearing "dog" activates the representation of the word *dog* in the mental lexicon, as well as semantic features associated with the concept of dog (e.g., "barks" and "has fur"). Evidence of lexical and semantic activation cued by spoken words can be observed in the form of spreading activation to related words or concepts. Eye-tracking studies have shown that upon hearing a spoken word (e.g., "dog"), people often briefly look at pictures representing words that are lexically related (e.g., *doctor* which shares its onset with the word *dog*) or semantically related (e.g., *cat* which belongs to the same semantic category as *dog*, or *leash* which is associatively related to *dog*) (Allopenna, Magnuson, & Tanenhaus, 1998; Huettig & McQueen, 2007; Yee & Sedivy, 2006).

Characteristic sounds, similar to spoken words, have been shown to also trigger access to semantic information (Chen & Spence, 2011, 2013; Edmiston & Lupyan, 2015). Hearing a characteristic sound, like a dog's bark, activates an entity's semantic features and facilitates picture identification (Chen & Spence, 2011) and visual search (Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010). However, direct evidence of how sounds access lexical information is lacking. Furthermore, while current speech processing models, such as TRACE (McClelland & Elman, 1986), can be adapted to incorporate non-speech sounds, empirical data comparing word and sound processing are needed to inform modeling efforts.

The aim of the current study is to directly compare the strength and rate with which spoken words and characteristic sounds provide access to information associated with a concept. In a visual world eye-tracking experiment, we assessed spreading activation from auditorily-presented targets (a spoken word, e.g., "dog" or characteristic sound, e.g., <bark-bark>) to their lexical and semantic competitors. In lexical 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 an associative semantic competitor was present on the screen (e.g., a picture of a *bone* when the target was the word "dog" or a <bark-bark> sound). Access to lexical/sematic information is indexed by visual fixation patterns to lexical/semantic competitors.

Our predictions are based on the multisensory framework proposed by Chen and Spence (2011), an extension of Glaser and Glaser's reading-naming interference model (1989). Chen and Spence propose that spoken words and characteristic sounds cue access to concepts via different intermediaries. Spoken words have a direct connection to phono-lexical representations, whereas characteristic sounds connect directly to semantic representations. The phono-

lexical and semantic representations are interconnected, allowing for words and sounds to each access lexical and semantic information. Based on this framework, we predict that spoken words will activate lexical representation stronger and/or faster than characteristic sounds. Likewise, characteristic sounds will activate semantic representation stronger and/or faster than spoken words.

# Method

## Participants

Thirty monolingual English speakers participated in the study. These participants were randomly assigned to the characteristic sound condition (n = 15, 14 female) or the spoken word condition (n = 15, 13 female). Eye-tracking data for one participant in the characteristic sound condition was lost due to equipment error. The remaining participants in the sound and word conditions did not differ in age, 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).

## Materials

Fifteen sets of stimuli were created for each competitor type, lexical and semantic. The 15 lexical 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., *chicken*), an associative semantic competitor (e.g., *egg*), and a control (e.g., *snowman*). In each group of 15 sets, the target word, the lexical/semantic competitor, and the control did not differ from each other in word frequency (SUBTLEXUS; Brysbaert & New, 2009), phonological and orthographic neighborhood size (CLEARPOND; Marian, Bartolotti, Chabal, & Shook, 2012), familiarity, concreteness, or imageability (MRC Psycholinguistic Database; Coltheart, 1981).

These sets were used to create 240 trials; in 50% of these trials, the target picture was absent from the display. Sixty of these target-absent trials comprised the set of experimental trials; analyses were limited to target-absent trials to ensure that competitor activation was caused by the auditory stimulus itself, instead of only the pictures on the screen (see Chabal & Marian, 2015). In 30 competitor trials, each competitor (e.g., *cloud* or *egg*) appeared in a display with three unrelated pictures. In 30 control trials, the competitor was replaced with a control object (e.g., *lightbulb* or *snowman*) in the same location. The 180 filler trials were designed to mask the experimental manipulation and to balance the number of times each picture was viewed in the experiment (i.e., targets, competitors, controls, and other unrelated items).

Pictures were black and white line drawings from the International Picture Naming Database (Bates et al., 2000) or independently normed by 20 English monolinguals using Amazon Mechanical Turk (http://www.mturk.com). These pictures were positioned in the four corners of a 3 x 3 invisible square grid. Pictures in the same display were similar in saturation (i.e., none of the pictures were darker than the others) and line thickness. Participants were seated approximately 80 cm away from a computer screen (2560 x 1440 resolution) while their eye-movements were tracked using an Eyelink 1000 eye-tracking system recording at 250 Hz. The words representing the 30 target items were recorded by a Midwestern female speaker of Standard American English. Word and sound stimuli were amplitude normalized and played through closed-back headphones. Spoken word durations ($M$ = 731.7 ms, $SE$ = 4.93, $Range$ = [502, 1066]) were shorter than characteristic sounds ($M$ = 1545.4 ms, $SE$ = 28.53, $Range$ = [329, 3868]), $t(29) = 5.33$, $p < .001$, due to the fact that many continuous sounds do not have a fixed ending point, as words do. Note that duration was not correlated with response times ($R^2 = .001$, n.s.). To account for any potential effects of auditory recording length on visual fixations, duration was included as an additional predictor in all models.

## Procedure

A fixation cross was shown on the screen for 1500 ms, followed by the four-object display. The display was shown for 500 ms before the participants heard either a characteristic sound or a spoken word. After the onset of the auditory input, the objects remained onscreen for 4500 ms before they disappeared. Participants were instructed to 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. Before the experiment, participants completed a set of practice trials.

## Data Analysis

Accuracy was analyzed using linear mixed effects regression. By-subject and by-item averaged models were created; with fixed effects of Auditory-input (word, sound), Condition (lexical, semantic), and Competition (competitor, control) and their interactions, as well as a random intercept of either subject or item (mixed effects logistic regression with subject and item random effects was not possible due to multicollinearity of fixed effects). Response times were analyzed for correct trials only, and outliers (greater than the condition mean plus two standard deviations) were replaced with M+2SD (4.72% of trials). The RT model included fixed effects of Auditory-input, Condition, and Competition plus their interactions, as well as random intercepts of both subject and item. Significance of fixed effects were obtained using $t$-tests and the Satterthwaite approximation for degrees of freedom. Follow-up pairwise comparisons used the Tukey correction for multiple comparisons.

The time course of visual fixations to semantic and lexical competitors was analyzed using growth curve analysis

(Mirman, Dixon, & Magnuson, 2008). Visual fixations were analyzed in 25 ms bins for correct trials only, averaged by items. Fixations were analyzed from 200 ms post-word onset (the time required to plan and execute an eye movement, Viviani, 1990) until each condition's average RT. Level-1 models used fourth-order orthogonal polynomials to capture the rise and fall of visual fixations over time. Level-2 models included all time terms and random effects of item on all time terms, plus additional fixed effects of each variable of interest. The difference between fixations to competitors and controls was analyzed separately for each combination of Auditory-input (word, sound) and Condition (lexical, semantic). All models included each item's auditory duration (scaled score) on all time terms, as adding auditory duration significantly improved each model's fit (*p*s < .001). Parameter *p*-values were obtained using the Satterthwaite approximation for degrees of freedom.

## Results

### Eye movements

**Competitor fixations.** We found a significant effect of lexical competition in response to spoken words on the intercept ($\beta$ = -0.030, *SE* = 0.005, *t*(1245) = -6.39, *p* < .001), linear ($\beta$ = 0.077, *SE* = 0.031, *t*(1245) = 2.47, *p* < .05), and cubic terms ($\beta$ = -0.161, *SE* = 0.031, *t*(1245) = -5.17, *p* < .001). These effects captured a larger, earlier fixation peak for lexical competitors compared to controls (Figure 1, top-left), indicating rapid lexical access by spoken words. Duration interacted with competition on the intercept ($\beta$ = 0.013, *SE* = 0.005, *t*(1245) = 2.85, *p* < .01) and quadratic terms ($\beta$ = -0.119, *SE* = 0.031, *t*(1245) = -3.81, *p* < .001); longer words activated lexical information less strongly.

There was a significant effect of lexical competition in response to characteristic sounds on the intercept ($\beta$ = -0.023, *SE* = 0.005, *t*(1200) = -5.19, *p* < .001), quadratic ($\beta$ = 0.174, *SE* = 0.029, *t*(1200) = 5.92, *p* < .001), and quartic ($\beta$ = -0.131, *SE* = 0.029, *t*(1200) = -4.47, *p* < .001) terms. These effects captured a late divergence between competitor and control fixations (Figure 1, top-right), indicating delayed lexical access by sounds. Duration interacted with competition on the intercept ($\beta$ = 0.013, *SE* = 0.005, *t*(1200) = 2.95, *p* < .01) and quadratic terms ($\beta$ = -0.060, *SE* = 0.029, *t*(1200) = -2.04, *p* < .05). As with words, sounds with longer durations activated lexical information less strongly.

There was a significant effect of semantic competition in response to spoken words on the intercept ($\beta$ = -0.023, *SE* = 0.004, *t*(1305) = -5.57, *p* < .001), quadratic ($\beta$ = 0.155, *SE* = 0.028, *t*(1305) = 5.54, *p* < .001), and quartic ($\beta$ = -0.067, *SE* = 0.028, *t*(1305) = -2.39, *p* < .05) terms. These effects captured a large competitor peak above a steady control baseline in the middle of the analysis window (Figure 1, bottom-left), indicating late semantic access by spoken words. Duration had a significant effect on the cubic term ($\beta$ = -0.040, *SE* = 0.017, *t*(25.1) = -2.34, *p* < .05), and interacted with competition on the intercept ($\beta$ = 0.054, *SE* = 0.004, *t*(1305) = 12.93, *p* < .001), linear ($\beta$ = 0.056, *SE* = 0.028, *t*(1305) = 2.00, *p* < .05), and quadratic

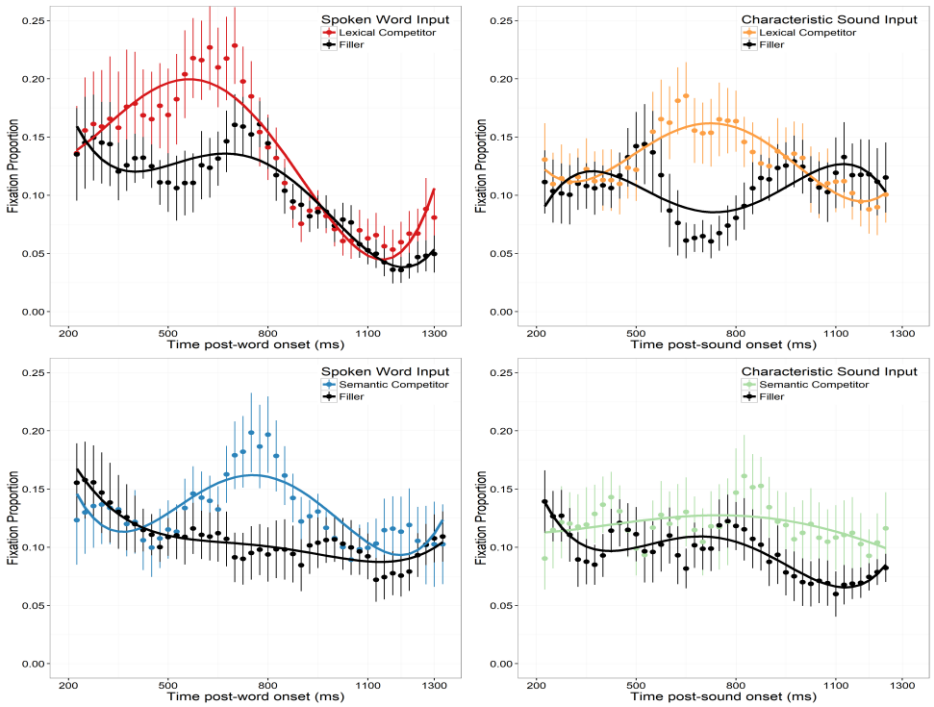


Figure 1: Activation of lexical and semantic competitors in response to spoken words and characteristic sounds. Lines represent model fits for fixations to competitors (color) and controls (black). Dots and vertical lines indicate observed values and standard error, respectively.

terms ($\beta$ = -0.100, $SE$ = 0.028, $t$(1305) = -3.58, $p$ < .001). These effects captured decreased semantic activation in response to longer words.

Finally, there was a significant effect of semantic competition in response to characteristic sounds on the intercept ($\beta$ = -0.024, $SE$ = 0.004, $t$(1185) = -5.86, $p$ < .001), linear ($\beta$ = -0.076, $SE$ = 0.027, $t$(1185) = -2.84, $p$ < .01), and quartic ($\beta$ = 0.058, $SE$ = 0.027, $t$(1185) = 2.15, $p$ < .05) terms. These effects captured a small, steady increase in competitor fixations early on, followed by a large late divergence between competitor and control fixations (Figure 1, bottom-right), which indicates sustained access to semantic information by sounds. Duration had a significant effect on the quadratic term ($\beta$ = 0.065, $SE$ = 0.029, $t$(15) = 2.27, $p$ < .05), and interacted with competition on the intercept ($\beta$ = 0.036, $SE$ = 0.004, $t$(1185) = 8.72, $p$ < .001), linear ($\beta$ = -0.110, $SE$ = 0.027, $t$(1185) = -4.09, $p$ < .001), and cubic ($\beta$ = -0.115, $SE$ = 0.027, $t$(1185) = -4.30, $p$ < .001) terms. These effects capture a large effect of duration across the observed time window. Longer sounds only activate semantics at a very late stage, whereas shorter sounds activate semantic information at both early and late stages.

**Comparing word and sound access to lexical and semantic information**. To facilitate comparisons across conditions, difference curves were calculated by subtracting control fixations from competitor fixations for each of the four levels of Auditory-input by Condition. A linear mixed effects regression model was designed, including fixed effects of Auditory-input (word, sound), Condition (lexical, semantic), and duration plus their interactions on all time terms, as well as a random effect of item. Crucially, there was an interaction between Auditory-input and Condition on the quadratic term ($\beta$ = 1.015, $SE$ = 0.16, $t$(2021) = 6.35, $p$ < .001), which is followed up in two analyses, one comparing activation of lexical information by words vs. sounds, and the other comparing activation of semantic information by words vs. sounds.

For lexical activation, Auditory input had a significant effect on the intercept ($\beta$ = -0.073, $SE$ = 0.02, $t$(1119) = , $p$ < .001) and quadratic terms($\beta$ = -0.271, $SE$ = 0.12, $t$(816) = , $p$ < .05); duration had an effect on the intercept ($\beta$ = -0.076, $SE$ = 0.02, $t$(1042) = -4.60, $p$ < .001) and quadratic terms ($\beta$ = -0.324, $SE$ = 0.11, $t$(689) = -2.99, $p$ < .01), and interacted with auditory input on the intercept ($\beta$ = -0.121, $SE$ = 0.03, $t$(1194) = -4.10, $p$ < .001) and quadratic ($\beta$ = -0.815, $SE$ = 0.19, $t$(943) = -4.19, $p$ < .001) terms. The combined effects captured the earlier peak of lexical activation for words (Figure 2, left, dark red) compared to sounds (Figure 2, left, light orange). These results suggest that words access lexical information faster than sounds.

For semantic activation, Auditory input also had a significant effect on the intercept ($\beta$ = -0.086, $SE$ = 0.02, $t$(1090) = -5.58, $p$ < .001) and quadratic terms($\beta$ = 0.713, $SE$ = 0.10, $t$(1034) = 6.94, $p$ < .001); duration had an effect on the intercept ($\beta$ = -0.072, $SE$ = 0.01, $t$(1060) = -5.74, $p$ < .001) and quadratic ($\beta$ = 0.704, $SE$ = 0.08, $t$(995) = 8.43, $p$ < .001) terms, and interacted with auditory input on the intercept ($\beta$ = -0.129, $SE$ = 0.02, $t$(1138) = -5.48, $p$ < .001) and quadratic terms ($\beta$ = 1.162, $SE$ = 0.16, $t$(1079) = 7.43, $p$ < .001). The combined effects manifested differently than lexical activation: For semantic information, words resulted in a peak in the middle of the window (Figure 2, right, dark blue), and sounds peaked closer to the offset of the window (Figure 2, right, light green). While words and sounds activated semantic information at the same rate, sounds had stronger access to semantics with higher peak activation.

## Accuracy.

We found a significant three-way interaction (by-subjects, $t$(81) = 3.38, $p$ < .001; by-items, $t$(84) = 2.74, $p$ < .01). Follow-up pairwise comparisons indicated that the Semantic-Sound Competitor had lower accuracy (86.5%) than all other conditions (all higher than 97.6%, $p$s < .001, by-subjects and by-items); no other comparisons were significant. Most errors (83.3%) in the Semantic Sound condition were caused by clicks on the semantic competitor.

## Response time.

There was a significant main effect of Competition ($\beta$ = -118.72, $SE$ = 19.22, $t$(1635.3) = -6.18, $p$ < .001) indicating that the presence of a competitor slowed down participants' assertion that the target was not present. The

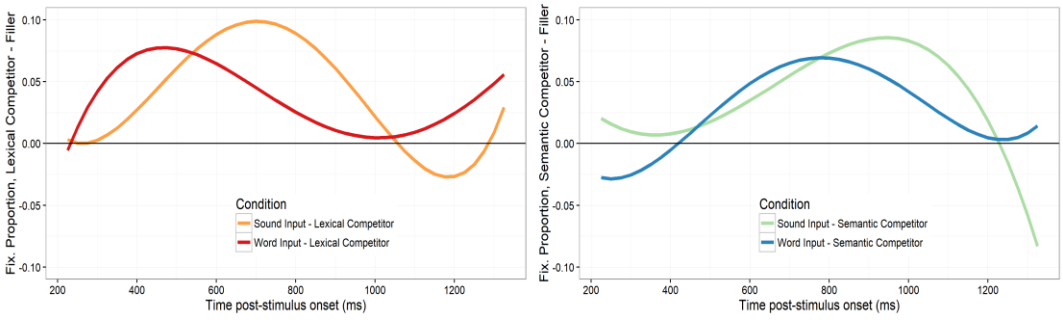


Figure 2: Effect of auditory input on lexical/semantic activation. Left: Words (red) activate lexical information earlier than sounds (orange). Right: Sounds (green) activate semantic information more strongly than words (blue). Curves represent predicted model values when auditory duration is set to a constant value (median word duration).

interaction between Condition and Competition was also significant ($\beta$ = 86.83, $SE$ = 38.44, $t$(1635.3) = 2.26, $p$ < .05). Follow-up pairwise comparisons showed that RTs during trials with semantic competitors were 163.56 ms slower than trials with matched controls, $t$(1636) = 5.96, $p$ < .001 and that RTs during trials with lexical competitors were 77.19 ms slower than trials with matched controls, $t$(1634.8) = 2.86, $p$ < .05.

## Discussion

The current study examined relative activation of word-form (i.e., lexical) knowledge and meaning (i.e., semantic) knowledge while listening to spoken words or characteristic sounds using eye-tracking in a visual world paradigm. While sounds and words are processed similarly in many ways: both are influenced by context, familiarity, and frequency (Ballas, 1993, Edmiston & Lupyan, 2014, Stuart & Jones, 1995), and both are similarly influenced by noise degradation (Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010, Gygi, Kidd, & Watson, 2004), models of auditory processing propose that words and sounds may access conceptual knowledge via different routes (Chen & Spence, 2011). Our aim was to determine whether sounds and words vary in how they provide access to lexical and semantic knowledge. We found different patterns for words and sounds in their access to lexical/semantic information. Specifically, spoken words were found to access lexical information earlier than sounds, but with similar intensity. In contrast, characteristic sounds were found to access semantics more strongly than words, but at a similar rate.

By comparing the shape and timecourse of visual fixations to lexical and semantic competitors, we discovered privileged access by spoken words to lexical information, and by characteristic sounds to semantic information. While lexical competition can be activated by both a spoken word and a characteristic sound, participants fixated the lexical competitor several hundred milliseconds earlier when cued by a word compared to a sound. This result supports the auditory processing model of Chen and Spence (2011), which states that spoken words first activate a lexical representation, whereas sounds first activate a semantic representation, which then spreads to the lexicon. These direct and indirect lexical pathways are reflected in the staggered timing of activation peaks observed in our study. Our results also demonstrate that non-linguistic sounds alone can provide fast access to lexical information, potentially via the concepts they activate.

A different pattern was observed for activation of semantic information. Once again, both a spoken word and a characteristic sound created semantic competition. This finding is consistent with results from cortical processing of semantic violations, where words and sounds were found to evoke similar cortical responses using event-related potentials (Hendrickson, Walenski, Friend, & Love, 2015). However, while both words and sounds started to increase semantic activation at the same rate, words reached an earlier and lower peak than sounds did. Chen and Spence's model proposes that characteristic sounds first activate semantic representations, which then feed forward to lexical representations. Our results partially support this proposal, as we find stronger activation of semantics by sounds, but we do not find a sound advantage in rate – in fact, words reach earlier peak activation than sounds.

This apparent departure from the model may be resolved when we consider differences in the nature of the semantic representation that is primarily accessed by words and sounds. Words, particularly concrete nouns as used in the current study, activate prototypical sematic concepts: "bird" typically makes one think of a songbird animal, rather than an ostrich or penguin (Hampton, 2016). Characteristic sounds, on the other hand, are closely linked to their original source and specific matching referents (Edmiston & Lupyan, 2013, 2015). In the context of the current study, the spoken word cue may have first accessed a lexical representation, followed by a prototypical semantic concept, which spread activation to related semantic concepts. The characteristic sound cue may have first accessed a representation for a specific referent that closely matched the source sound; this specific representation then spread to the prototypical semantic concept, and from there to related semantic concepts (i.e., the semantic competitor). This additional specific-to-general step for sounds may have contributed to the slower rate of competitor activation.

Our results also demonstrate the influence of the duration of an auditory signal on information access. Changes in the duration of either words or sounds had the same effects, where shorter durations increased lexical and semantic activation relative to longer durations. This consistent duration effect may be related to continuous auditory input processing. Speech processing models posit that as a spoken word unfolds, all lexical items that are consistent with the partially received input become activated, and start to decline as they diverge from the input (McClelland & Elman, 1986; Shook & Marian, 2013). During the partially-produced stage, activation is spread diffusely among multiple representations, which decreases the level of any individual item. It is possible that non-speech sound processing follows a similar pattern where multiple representations are initially activated and then pruned, leading to the same duration effect for sounds that we observe for words.

We elected to use separate targets to examine lexical and semantic competition in order to minimize priming effects, and due to the constraints inherent in selecting identifiable picture pairs with recognizable characteristic sounds. Now that distinct lexical and semantic effects have been established, it will be informative to directly compare them using target – lexical competitor – semantic competitor triplets (e.g., clock-cloud-radio). In addition, the issue of different word and sound durations should be controlled in future work. Note that the longer sound durations likely increased the ecological validity of this study, as many environmental sounds are continuous, compared to spoken words' fixed ending points.

In conclusion, we have identified similarities and differences in how humans process two types of auditory input – linguistic spoken words and non-linguistic characteristic sounds. The observed preferential access to lexical information by spoken words, and to semantic information by non-speech sounds, reveals features of the cognitive architecture used to process sounds. These results highlight the interconnectivity of the mind, with interactions observed among linguistic and non-linguistic processing, auditory and visual processing, and lexical and semantic processing.

## Acknowledgments

The authors thank the members of the Northwestern University *Bilingualism and Psycholinguistics Research Group* for helpful comments and input. This work was supported in part by grant NICHD 2R01 HD059858.

## References

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

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 Kucera 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.

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General*, *144*(3), 539–550.

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.

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.

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

Connolly, J. F., & Phillips, N. A. (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.

Dunn, L. M. (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.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology. General*, *118*(1), 13–42.

Hampton, J. A. (2016). Categories, prototypes, and exemplars. In N. Reimer (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.

Huettig, F., & McQueen, J. M. (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.

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.

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.

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

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280.

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

Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism (Cambridge, England)*, *16*(2), 304–324.

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*.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. *Reviews of Oculomotor Research*, *4*, 353–93.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). The comprehensive test of phonological processing. Austin, TX: Pro-Ed.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14.