INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

The Effect of Background Noise on the Activation of Phonological and Semantic Information during Spoken-Word Recognition

Florian Hintz [1], *Odette Scharenborg*[1,2]

[1] Centre for Language Studies, Radboud University Nijmegen (NL)
[2] Donders Institute for Brain, Cognition, & Behavior, Radboud University Nijmegen (NL)
f.hintz@let.ru.nl, o.scharenborg@let.ru.nl

Abstract

During spoken-word recognition, listeners experience phonological competition between multiple word candidates, which increases, relative to optimal listening conditions, when speech is masked by noise. Moreover, listeners activate semantic word knowledge during the word's unfolding. Here, we replicated the effect of background noise on phonological competition and investigated to which extent noise affects the activation of semantic information in phonological competitors. Participants' eye movements were recorded when they listened to sentences containing a target word and looked at three types of displays. The displays either contained a picture of the target word, or a picture of a phonological onset competitor, or a picture of a word semantically related to the onset competitor, each along with three unrelated distractors. The analyses revealed that, in noise, fixations to the target and to the phonological onset competitor were delayed and smaller in magnitude compared to the clean listening condition, most likely reflecting enhanced phonological competition. No evidence for the activation of semantic information in the phonological competitors was observed in noise and, surprisingly, also not in the clear. We discuss the implications of the lack of an effect and differences between the present and earlier studies.

Index Terms: listening in noise, phonological competition, semantic competition, eye-tracking

1. Introduction

In every-day life, humans comprehend spoken language in different situations. These include situations where speech is accompanied by relevant visual input [1] or situations where the speech signal is suboptimal (e.g., partially masked by background noise). Crucially, in all situations, recognizing spoken words is the key to successful comprehension.

Previous research on the processes and mechanisms underlying spoken-word recognition has established two important characteristics: First, as a word unfolds, the spoken input is continuously mapped onto phonological representations stored in the mental lexicon, which results in candidate words that partially overlap with the incoming signal competing for recognition [2-5]. Previous studies have used eye-tracking to examine such phonological competition. [2] recorded participants' eye gaze as they listened to instructions such as "Pick up the beaker. Now put it above the diamond" while looking at a display featuring four geometrical shapes and four pictures. One was a depiction of the target word (*beaker*), while the names of two pictures overlapped with the target in word onset (*beetle*) and offset (*speaker*), respectively. The fourth picture was an unrelated distractor. The results showed that participants' likelihood of fixations to both the picture of a beaker and the picture of a beetle increased as the word "beaker" started to unfold. As the acoustic information from "beaker" started to mismatch with the phonological information of "beetle", the likelihood of looks to the beetle decreased as the likelihood of looks to the beaker continued to rise. In addition, looks to the picture of a speaker started to increase as the end of the word "beaker" acoustically unfolded. Interestingly, using a similar set-up, it has been shown that the time to resolve visually-induced phonological competition was elongated when the speech was masked by noise [6-8], suggesting that adverse listening conditions affect the phonological competition dynamics underlying spoken-word recognition. Relative to word recognition in the clear, listeners are assumed to interpret the acoustic signal with more flexibility, thereby entertaining competing lexical items for a longer period of time rather than quickly eliminating all lexical competitors [6].

A second important characteristic of spoken-word recognition concerns the flow of activation within the levels of mental representations. It has been argued that information from the acoustic signal cascades to higher (e.g., semantic) levels before processing at lower (e.g., phonological) levels is completed [9]. Specifically, during a word's unfolding, listeners activate its semantics [10,11] and spuriously that of its phonological competitors [11,12].

The present experiment investigated how noise-induced change in competition dynamics at a phonological level affects the subsequent flow of activation to semantic levels in phonological competitors. Put differently, does an increase in phonological competition due to the presence of noise amplify or reduce semantic activation in similar sounding words? To address this question, native Dutch listeners took part in an eye-tracking experiment consisting of spoken sentences that each contained a target word. The sentences were paired with three types of displays, either featuring a picture of the target word, or a picture of a phonological onset competitor, or a picture of a word semantically related to the onset competitor, each along with three unrelated distractors. The sentences were presented to the participants either in the clear or masked by noise at two different signal-to-noise ratios (SNRs). Participants' eye movements to the various pictures were analyzed starting at the spoken onset of the target words, assumed to reflect processing of the concurrently unfolding linguistic input. For the clear condition, we predicted to replicate previous findings. That is, participants should fixate on the target and on the onset competitor shortly after word onset. While the likelihood of looks to the target was expected to rise as the spoken word further unfolds, the likelihood of looks to the onset competitor was expected to decrease after the speech signal had disambiguated the target from the onset competitor [2]. Similarly, we predicted a bias in fixations, compared to the unrelated distractors, to the semantic competitor for the time period of the phonological overlap between target and onset competitor [11]. In line with the findings discussed above, we predicted that perceiving the target words in noise would result in

http://dx.doi.org/10.21437/Interspeech.2016-882

elongated phonological competition, which would manifest itself in delayed fixations to the target and elongated fixation biases to the onset competitor. It is currently unclear whether such gaze behavior reflects an extension of the set of activated word candidates or a slower elimination of the depicted phonological competitor. Consequently, increased phonological competition could impact semantic activation in two ways: 1) due to an increase in number of activated word candidates, semantic activation may be too weak to surface as an eye movement to the respective semantic competitor, or 2) semantic activation may be amplified as a function of maintaining one (or few) word candidates for a longer period of time.

2. Experimental set-up

2.1. Participants

Forty-six members (mean age = 24, *SD* = 4, 15 male) of the participant pool of the Radboud University, all native speakers of Dutch, were paid for their participation. All participants had normal hearing and normal or corrected-to-normal vision. All participants gave written consent beforehand. The study was approved by the ethics board of the university. Due to extensive track loss the data of one participant had to be excluded from the analysis.

2.2. Materials

Sextuples of words were selected for 66 experimental trials. Each consisted of a critical target word (e.g., *zwaan*, swan), a phonological cohort competitor that overlapped with the target in onset and was otherwise unrelated (e.g., *zwaard*, sword), a word that was semantically related to the onset competitor and unrelated to the target (e.g., *schild*, shield), and three distractors that were visually, phonologically, and semantically unrelated to target, onset and semantic competitors (e.g., toilet, cucumber, wine bottle; see Figure 1, for examples of the three different displays).

To ensure that the phonological and semantic competitors and the three distractors were semantically and visually unrelated to the target words, 36 native Dutch participants (mean age = 23, SD = 4, eight male), none of which took part in the main experiment, provided semantic and visual similarity ratings. To that end, each target word was paired with three types of displays, containing either a picture of the target, or a picture of the phonological competitor or a picture of the semantic competitor, and the three distractors. The four objects in one display were arranged on a virtual 2x2 grid (Figure 1); the positions were randomized. Participants were instructed to read the target words (e.g., swan, positioned above the grid) and, in the visual similarity rating study, judge how similar the typical visual shape of the concept denoted by the printed word was to the physical shape of the referents of the depicted objects, ignoring any similarity in meaning. In the semantic similarity rating, participants were asked to judge meaning similarity while ignoring shape similarity. A rating scale ranging from 1 (no similarity) to 10 (identical) was used in both tasks. The results of the visual similarity rating confirmed that the target objects depicted the concepts invoked by the written words (mean rating = 9.36, *SD* = 1.14) while phonological (mean rating = 1.47, *SD* = 1.14) and semantic (mean rating = 1.51, *SD* = .65) competitors were visually unrelated. The semantic similarity rating confirmed that the target objects matched the semantic representations invoked by the written words (mean rating = 9.86, SD = 0.25); phonological (mean rating = 1.19, *SD* = .45) and semantic (mean rating = 1.16, *SD* = .33) competitors were semantically unrelated. The mean distractor score in the visual similarity rating was 1.61 (SD = 0.89); in the semantic similarity rating it was 1.25 (SD = 0.46).

The critical words and the unrelated distractors were matched for frequency using the Subtlex-NL database [13] ($F(5,383) = .722$, $p > .6$). The average phoneme overlap between target and onset competitor was 2.5. Target and onset competitor were additionally matched for number of syllables ($t(130) = 1.233$, $p = .22$), number of letters ($t(130) = .842$, $p = .41$), number of phonological neighbors ($t(130) = .558$, $p = .58$), and the phonological neighbors' frequency ($t(116) = -.812$, $p = .42$). The semantic relationship between the onset and the semantic competitor was deemed fairly strong by a native speaker of Dutch. A free association database [14] was used to determine the forward association strength between onset competitors (cues) and semantic competitors (responses), which was .062 (range: .003-.287; 18 were not listed as responses). Admittedly, this was not very high but note that eye movements to semantic competitors in the visual world can be driven by semantic feature overlap or category membership as well [10].

The target words were embedded in neutral carrier sentences (e.g., for *zwaan*, swan, *Hij dacht direct aan een zwaan toen Bob over ganzen begon te praten,* 'He thought immediately about a swan when Bob started talking about geese') and could not be predicted from the sentential context.

A further 22 quadruples of words were selected for filler trials. These sets included a target word that was placed in a neutral sentence context (like those used in experimental trials) and three unrelated distractors. The filler trials ensured an equal number of target-present and target-absent trials.

The experimental and filler sentences were read by a male native speaker of Dutch. Recordings of these utterances, at a sampling rate of 44.1 kHz with 16-bit resolution, were made in a sound-attenuated booth. The sentences were read with a neutral intonation contour such that, in particular, the critical words were not highlighted. We created two additional versions of each recorded sentence by adding stationary speech-shaped noise (SSN) with SNRs of +3 and -3, respectively, using Praat [15]. To that end, the original recordings were down-sampled to 16 kHz. The intensity in the clear and noise sentences was set to 60 dB.

All words in the experimental and filler sets were picturable. Photographs were selected from existing databases [16,17] or searched on the internet and edited to fit the resolution and size of the other pictures.

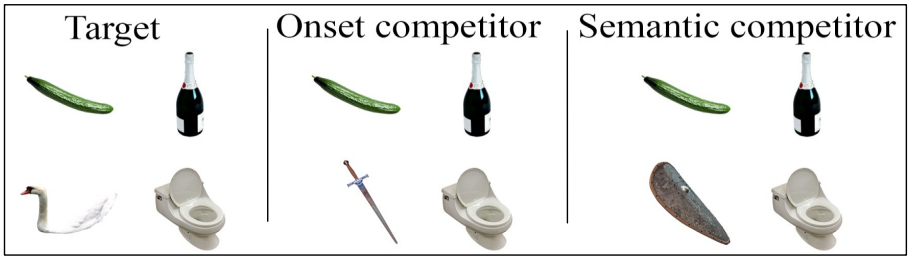

Figure 1: Display configurations for the target word zwaan, swan, featuring a target-present display, an onset competitor display (zwaard, sword) and a semantic competitor display (schild, shield), each along with three unrelated distractors.

2.3. Procedure

The 66 experimental items were rotated across each listening condition (clear, SNR+3, SNR-3) and each display type (target-present, onset competitor-present, semantic competitor-present). The filler items were rotated across the three listening conditions. Nine lists with 88 trials each were generated. On each list, target-present, onset competitor-present and semantic competitor-present trials occurred equally often. Trials were blocked according to their listening condition. One block on each list always contained 30 trials; the other two blocks contained 29 trials. The order of blocks varied between lists. The order of trials within a block (experimental and filler) was randomized.

Participants were randomly assigned one list. They were seated at a comfortable distance from the computer screen and asked to put their chin on a chin rest. Eye movements were monitored with an SR Eyelink remote eye-tracking system, sampling at 500 Hz. Spoken sentences were presented to the participants through headphones. The parameters of each trial were as follows. First, a central fixation point appeared on the screen for 2 s, which was followed by the four objects (each object had a size of 120 x 120 pixels) belonging to a trial. The start of the playback of the sentence was timed such that participants had exactly 3 s to preview the objects before the target word occurred in the spoken sentence. The positions of the pictures were randomized across four fixed positions. Interest areas (250 x 250 pixels) were defined around each object. Participants were not asked to perform any explicit task, but instructed to listen to the sentences carefully. They could look at whatever they wanted to and should not take their eyes off the screen.

The experiment, including calibration, took approximately 15 minutes. The data from participants' left or right eye (depending on the quality of the calibration) were analyzed in terms of fixations, saccades, and blinks, using the algorithm provided in the EyeLink software. Fixations were coded as directed to the target, onset competitor, semantic competitor, to one of the three unrelated distractors, or elsewhere.

3. Results

Less than one percent of all trials had to be excluded due to track loss. The remaining data contributed to the time course graphs in Figure 2, plotting the fixation proportions to targets, onset competitors, semantic competitors and unrelated distractors from the onset of the spoken target words up until one second post-onset for the clear, SNR+3 and SNR-3 listening conditions.

For the purpose of analyzing the data, we defined a trial as starting at 200 ms after target onset, because it takes a minimum of about 180 ms to program and launch a saccadic eye movement [18]. Thus, the 200 ms post-onset do most likely not reflect linguistic processing. To obtain information about the time course of participants' fixations to the various objects, we divided the trial into eight 100-ms windows (200–1000 ms after target onset) and conducted separate planned comparisons on each window.

3.1. Target trials

As reported in numerous eye-tracking studies before (see the Introduction), in the clear speech condition participants shifted their overt visual attention to the target objects shortly after perceiving the initial target word phonemes (see also the top row panels in Figure 2). At the window beginning 200 ms after target onset, fixations on the target were significantly more likely than fixations on the unrelated pictures ($t_1(44) = 2.955$, $p = .005$; $t_2(65) = 2.602$, $p = .01$). This difference remained significant throughout the trial. The same pattern was observed when the target words were masked by background noise at an SNR of +3 (200 ms after onset: $t_1(44) = 2.027$, $p = .05$; $t_2(65) = 2.506$, $p = .01$). At an SNR of -3, the target bias became fully significant at the time window starting 300 ms after word onset ($t_1(44) = 3.041$, $p = .004$; $t_2(65) = 3.14$, $p = .002$). The latter finding is in line with our hypothesis, suggesting that noise delayed fixations to the target.

3.2. Onset competitor trials

We also replicated previous studies that reported a bias in looks to phonological onset competitors for the period spanning the phonological overlap (see also the middle panels in Figure 2). In the clear condition, at the time window starting 400 ms after word onset (also spanning the adjacent time bin), the participants looked significantly more to the onset competitors than to the unrelated distractors ($t_1(44) = 2.955$, $p = .005$; $t_2(65) = 2.828$, $p = .006$). In contrast to our predictions, we only observed a mild trend towards a phonological bias in the SNR+3 condition, showing at the window beginning 500 ms after word onset ($t_1(44) = 1.364$, $p = .09$; $t_2(65) = 1.217$, $p = .11$; both one-tailed). However, we did observe evidence for increased phonological competition in the SNR-3 condition: As hypothesized, compared to the clear condition, participants biased the onset competitor later and for an extended time period. The effect reached statistical significance at the window starting 700 ms after word onset and stayed reliable for the remainder of the analyzed region, i.e. until 1000 ms post-onset ($t_1(44) = 2.006$, $p = .05$; $t_2(65) = 1.482$, $p = .07$; the latter one-tailed).

3.3. Semantic competitor trials

Even though visual inspection may suggest weak evidence for semantic competition in the clear condition, our analyses did not confirm this. We did observe a significant semantic bias in the SNR+3 condition, spanning two time windows (200-400 ms after word onset; 200-300: $t_1(44) = 2.206$, $p = .033$; $t_2(65) = 1.791$, $p = .08$). However, this bias could be due to the fact that participants' likelihood of fixating the semantic competitor was already greater at word onset than their likelihood of fixating the unrelated distractors. This was unexpected and, as the same semantic competitors in the clear and SNR-3 listening conditions did not yield such a bias at word onset, is not easily explained. We will thus not further discuss this effect. Similar to the clear condition, there was not a hint towards a semantic bias in the SNR-3 condition.

4. General discussion

Using an eye-tracking paradigm, the present study investigated the effect of background noise on the activation of phonological and semantic information during spoken-word recognition in noise. Specifically, we tested whether increased phonological competition, induced by the presence of background noise, amplifies or reduces the likelihood of semantic competition in phonological cohort competitors. We replicated earlier studies that showed a phonological bias during the time period of overlap between target and cohort competitor in the clear [2] and studies that reported an elongated phonological bias when the target words were presented in noise [6,7]. In contrast to previous research [11], our data do not support the notion that semantic information is activated in phonological cohort competitors either in clean or in noise.

This is surprising given that even five-year olds have been shown to exhibited gaze behavior reflecting transient semantic competition in phonological cohort competitors in the clear [19]. We are confident to rule out that the lack of an effect is connected to statistical power as the present material set contained more items

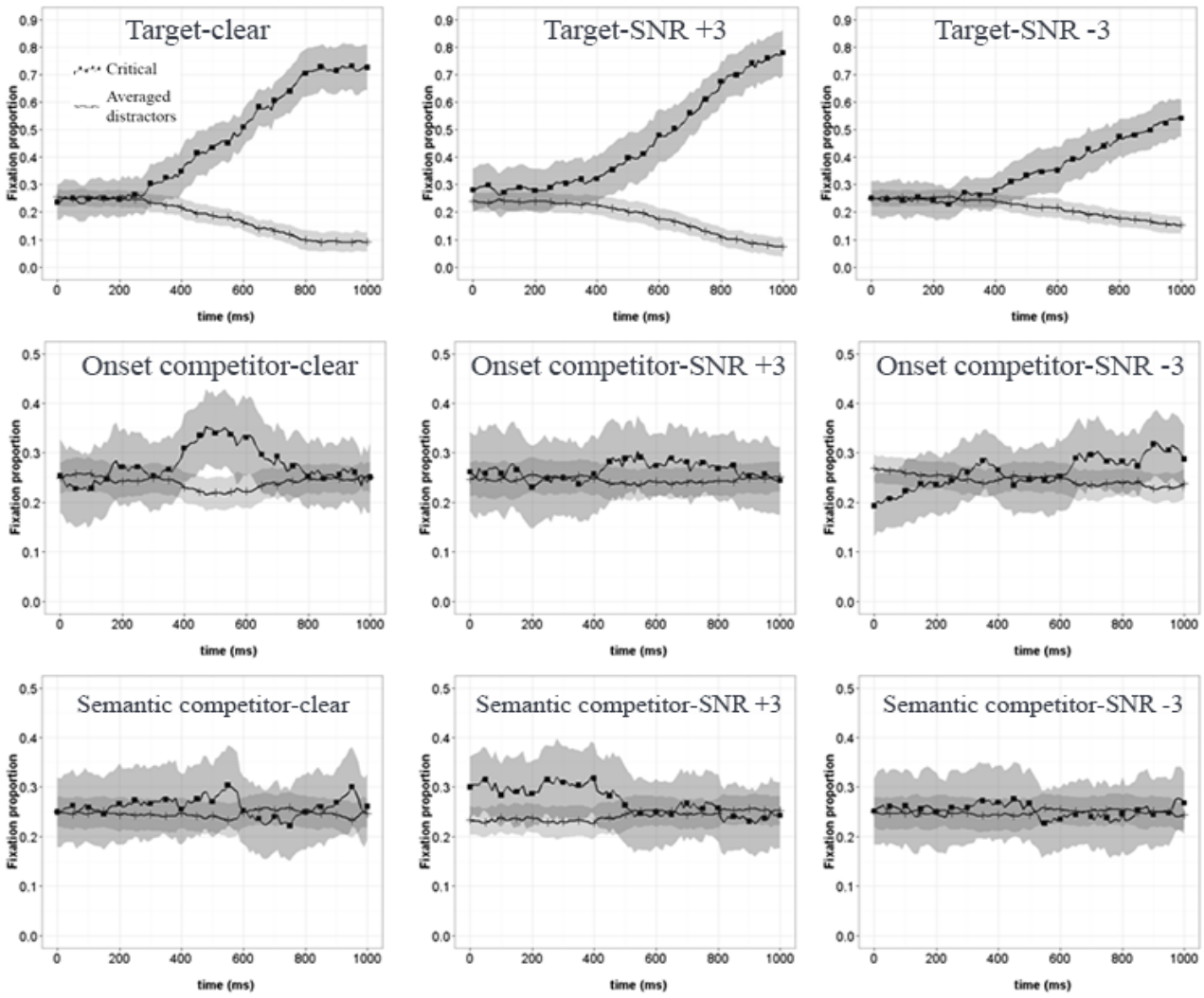

Figure 2: Time course graphs plotting fixation proportions to targets, onset and semantic competitors and averaged distractors from spoken target word onset up until 1000 ms post-onset for the clear, SNR+3 and SNR-3 listening conditions.

than most of the previous, comparable studies. A striking difference between the present and previous studies is that we chose an experimental design where the semantic competitor was presented along with three unrelated distractors as compared to being presented with either a depiction of the spoken target or a picture of the phonological competitor [11,19]. In the latter case, semantic activation in the phonological competitors could benefit from the visual input such that seeing a swan and a shield or seeing a sword and a shield in the same display prior to the critical spoken target ("swan") could provide a head start for activation to spread to semantic levels of representations.

A further explanation for the lack of a semantic bias in the clear could be that the sheer presence of "noise trials" during the experiment, even though they were blocked, affected participants' gaze behavior. We are currently exploring this possibility in a follow-up eye-tracking experiment consisting of "clear trials" only.

As stated above, one hypothesis with regard to the influence of noise on semantic competition in onset competitors was that increased phonological competition would amplify potential semantic competition effects. If this had been the case we should have seen differences in gaze behavior between the clear and the noise conditions which we did not. Comparing the left-most and the right-most plots in the bottom panel of Figure 2 suggests that increased phonological competition has no effect on the likelihood of semantic competition in phonological cohort competitors (or at best a slightly reductive one). However, such a conclusion is hard to draw given the lack of an effect in the clear condition.

On a general note, one may ask whether semantic competition in phonological cohort competitors is a mechanism that is routinely active in all situations of language comprehension. Its fragile nature (see also, e.g., [11]) may suggest that it is particularly potent in situations with relevant visual present. Future research needs to determine whether the same is true for situations of language comprehension were pictorial input is absent.

5. Acknowledgements

This research was funded by a Vidi-grant from the Netherlands Organization for Scientific Research (NWO; grant number: 276-89-003) awarded to OS. The authors would like to thank Ferdy Hubers for lending his voice to the spoken stimuli, as well as Marloes Graauwmans and Tijn Schmitz for assistance in preparing and running the eye-tracking experiment.

6. References

[1] F. Huettig, J. Rommers, and A. S. Meyer, "Using the visual world paradigm to study language processing: A review and critical evaluation," *Acta Psychologica,* vol. 137, pp. 151-171, 2011.

[2] P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *Journal of Memory and Language,* vol. 38, pp. 419-439, 1998.

[3] J. M. McQueen, D. Norris, and A. Cutler, "Competition in spoken word recognition: Spotting words in other words," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 20, pp. 621-638, 1994.

[4] D. Norris, J. M. McQueen, and A. Cutler, "Competition and segmentation in spoken-word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 21, pp. 1209-1228, 1995.

[5] D. W. Gow Jr and P. C. Gordon, "Lexical and prelexical influences on word segmentation: Evidence from priming," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 21, pp. 344-359, 1995.

[6] S. Brouwer and A. R. Bradlow, "The Temporal Dynamics of Spoken Word Recognition in Adverse Listening Conditions," *Journal of Psycholinguistic Research,* pp. 1-10, 2015.

[7] B. M. Ben-David, C. G. Chambers, M. Daneman, M. K. Pichora-Fuller, E. M. Reingold, and B. A. Schneider, "Effects of Aging and Noise on Real-Time Spoken Word Recognition: Evidence From Eye Movements," *Journal of Speech, Language, and Hearing Research,* vol. 54, pp. 243-262, 2011.

[8] J. M. McQueen and F. Huettig, "Changing only the probability that spoken words will be distorted changes how they are recognized," *Journal of the Acoustical Society of America,* vol. 131, pp. 509-517, 2012.

[9] F. Huettig and J. M. McQueen, "The tug of war between phonological, semantic and shape information in language-mediated visual search," *Journal of Memory and Language,* vol. 57, pp. 460-482, 2007.

[10] F. Huettig and G. T. M. Altmann, "Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm," *Cognition,* vol. 96, pp. B23-32, 2005.

[11] E. Yee and J. C. Sedivy, "Eye movements to pictures reveal transient semantic activation during spoken word recognition," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 32, pp. 1-14, 2006.

[12] K. S. Apfelbaum, S. E. Blumstein, and B. McMurray, "Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems," *Psychonomic Bulletin & Review,* vol. 18, pp. 141-149, 2010.

[13] E. Keuleers, M. Brysbaert, and B. New, "SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles," *Behavior Research Methods,* vol. 42, pp. 643-50, 2010.

[14] S. de Deyne, D. J. Navarro, and G. Storms, "Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations," *Behavior Research Methods,* pp. 1-19, 2012.

[15] P. P. G. Boersma, "Praat, a system for doing phonetics by computer (Version 5.1.19) [Computer program]," *Available from http://www.praat.org/,* 2002.

[16] F. de Groot, T. Koelewijn, F. Huettig, and C. N. L. Olivers, "A stimulus set of words and pictures matched for visual and semantic similarity," *Journal of Cognitive Psychology,* vol. 28, pp. 1-15, 2016.

[17] M. B. Brodeur, E. Dionne-Dostie, T. Montreuil, and M. Lepage, "The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research," *PLoS ONE,* vol. 5, p. e10773, 2010.

[18] M. G. Saslow, "Latency for saccadic eye movement," *Journal of the Optical Society of America,* vol. 57, p. 1030, 1967.

[19] Y. T. Huang and J. Snedeker, "Cascading activation across levels of representation in children's lexical processing," *Journal of Child Language,* vol. 38, pp. 644-661, 2011.