Contextual constraints on the activation of lexical forms by non-linguistic sounds

Anuenue Kukona

Division of Psychology, De Montfort University

Author Note

Anuenue Kukona https://orcid.org/0000-0003-4377-3057

This study was supported by an Experimental Psychology Society Small Grant to AK. Elzbieta Malogoska and Sandra Wydera are thanked for their assistance with data collection.

Correspondence concerning this article should be addressed to Anuenue Kukona, De Montfort University, The Gateway, Leicester, LE1 9BH, UK. Telephone: +44 (0)116 250 6184 Email: anuenue.baker-kukona@dmu.ac.uk

Abstract

Three visual world experiments investigated the activation of linguistic knowledge during the processing of non-linguistic auditory stimuli. In Experiment 1, participants heard spoken words such as “car” or environmental sounds such as a sound produced by a car while viewing visual arrays with objects such as a car (target), card (phonologically related competitor) and box (unrelated distractor). Interleaved throughout, participants heard both spoken words and environmental sounds. In Experiment 2 and 3, in order to assess contextual constraints on processing, participants only heard environmental sounds while viewing similar visual arrays (targets were included in the former but not the latter). When participants heard environmental sounds interleaved among spoken words (Experiment 1), they fixated competitors significantly more than distractors during both types of auditory stimuli, suggesting that both engage linguistic systems and representations; however, when participants only heard environmental sounds (Experiment 2 and 3), phonological competition was not observed. These results suggest that the activation of linguistic knowledge by environmental sounds is context dependent rather than automatic, differing from spoken words. Implications for theories addressing the mapping of auditory stimuli onto conceptual knowledge are discussed.

Keywords: Cohort competition; Environmental sounds; Eye movements; Phonological competition; Visual world paradigm

Public Significance Statement: Environmental sounds are non-linguistic sounds that are produced by everyday entities and events, such as the sounds produced by planes, trains and automobiles. Focusing on the retrieval of linguistic information, this study highlights both similarities and differences between the processing of language and environmental sounds.

Introduction

The degree to which linguistic phenomena recruit language specific vs. domain general mechanisms has attracted considerable attention in cognitive science (e.g., see classic examples such as Chomsky, 1959). The current research investigated the mapping of auditory stimuli onto conceptual knowledge. For example, how do individuals hearing spoken words such as “car” or non-linguistic sounds such as the sounds produced by cars activate their knowledge of cars? Chen and Spence (2011, 2018a) describe a theoretical approach that distinguishes linguistic and non-linguistic auditory stimuli, such that the former are hypothesed to (i.e., uniquely) engage language specific systems and representations. However, a growing empirical literature also reveals striking similarities between the processing of these two types of auditory stimuli. The aim of the current research was to address the mapping of auditory stimuli onto conceptual knowledge by investigating phonological competition (e.g., Allopenna et al.,1998): the activation of phonologically related representations (e.g., card) during the processing of auditory stimuli (e.g., “car”) and corresponding impacts on individuals’ attention and eye movements.

Conceptual knowledge is closely linked to two fundamentally different types of auditory stimuli. For example, the spoken word “car” can be used to refer to knowledge about cars, but the non-linguistic sounds produced by cars are also encompassed by this knowledge. Spoken words are characterised by their hierarchical linguistic structure (e.g., phonetics and phonology) and arbitrary links to meaning. In contrast, characteristic, naturalistic or environmental sounds (i.e., as they are variously known; the latter is adopted throughout) are non-linguistic sounds that are produced (i.e., causally structured) by everyday entities and events.

Chen and Spence (2011, 2018a; see also Glaser & Glaser, 1989) hypothesise that spoken words and environmental sounds recruit two fundamentally different processing streams. Their approach distinguishes conceptual vs. linguistic knowledge: although interlinked, conceptual knowledge is assumed to be stored in semantic memory while (i.e., non-semantic) linguistic knowledge is assumed to be stored in the lexicon. Thus, during the processing of spoken words, auditory stimuli make initial contact with the lexicon (e.g., activating lexical forms), which mediates subsequent contact with semantic memory. In contrast, during the processing of environmental sounds, auditory stimuli make direct (i.e., linguistically unmediated) contact with semantic memory.

Chen and Spence’s (2011, 2018a) approach complements classic models of spoken word recognition. For example, TRACE (e.g., McClelland & Elman, 1986) is an interactive activation model with hierarchical levels for features, phonemes and words. During the processing of spoken words, TRACE’s feature detectors respond to acoustic-phonetic information in the auditory stimuli and feed activation forward to the phoneme and word levels. Thus, spoken words are hypothesised to activate linguistic representations at multiple levels. Correspondingly, Allopenna et al. (1998) captured the time course of spoken word recognition using the visual world paradigm (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). Their participants heard spoken words such as “beaker” while viewing visual arrays with objects such as a target beaker, phonologically related beetle and unrelated carriage. Consistent with models such as TRACE (e.g., see also Marslen-Wilson, 1987; Norris, 1994), they found that both beaker and beetle were fixated more than distractors early in the time course of processing, reflecting the cascade of activation across onset (e.g., “b…”) consistent lexical forms. However, they found that beaker was fixated more than both beetle and carriage later in processing, reflecting the continuing activation of stimulus (e.g., “beaker”) consistent lexical forms. These findings suggest that phonological competitors (e.g., beaker-beetle) are transiently activated during the processing of spoken words. In contrast, TRACE’s architecture implies (i.e., it was designed to account for spoken word recognition) that non-linguistic auditory stimuli are processed via another bottom-up stream: environmental sounds are not composed of phonetic features and phonemes, and thus TRACE’s feature detectors would not be expected to respond to these auditory stimuli.

Prior research also reveals processing differences between spoken words and environmental sounds. Chen and Spence (2011) observed a processing advantage for environmental sounds: using a visual detection task, in which they presented participants with auditory primes before picture targets (e.g., dog), they found that (i.e., corresponding) environmental sounds (e.g., barking) facilitated responses, while spoken words (e.g., “dog”) did not. These findings suggest that environmental sounds are mapped onto conceptual knowledge more rapidly than spoken words, such that the (i.e., linguistically mediated) spread of activation across linguistic features, phonology and words (i.e., and then concepts) may require additional processing time (e.g., see also Chen & Spence, 2018b). However, findings from the literature are mixed. In contrast, Lupyan and Thompson-Schill (2012) observed a processing disadvantage for environmental sounds: using a related matching task, they found that (i.e., corresponding) spoken word primes facilitated responses more than environmental sound primes. Taken together, these findings suggest that processing times may not straightforwardly distinguish these streams: although less direct, participants’ considerable experience with language may allow this stream to function more efficiently, counterintuitively yielding faster processing times. In addition, environmental sounds may activate more idiosyncratic conceptual representations, which may also impact processing (e.g., see also Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015). Finally, Toon and Kukona (2020) investigated the activation of conceptual knowledge using the visual world paradigm (e.g., see also Bartolotti et al., 2020). Their participants heard spoken words such as “puppy” or environmental sounds such as a sound produced by a puppy (e.g., barking) while viewing visual arrays with objects such as a target puppy, semantically related bone and unrelated candle. Although both types of auditory stimuli generated semantic competition, as reflected in fixations to the semantically related bone vs. unrelated candle, these effects were more pronounced for environmental sounds.

Relatedly, prior research has also addressed the activation of linguistic knowledge during the processing of environmental sounds. Although Chen and Spence’s (2011, 2018a) approach assumes direct (i.e., linguistically unmediated) links between environmental sounds and semantic memory, it also includes bidirectional links between semantic memory and the lexicon, enabling environmental sounds to make (i.e., mediated) contact with linguistic knowledge. Correspondingly, priming studies reveal that participants can rapidly link environmental sounds to lexical forms (e.g., Ballas, 1993; Frey et al., 2014; Orgs et al., 2006, 2007, 2008; Van Petten & Rheinfelder, 1995). Using a lexical decision task, Van Petten and Rheinfelder (1995) found that environmental sound primes facilitated responses to semantically related spoken word targets. Using ERP, they also found that spoken word targets elicited an attenuated N400 response following semantically related environmental sound primes. These findings (e.g., which parallel words; Meyer & Schvaneveldt, 1971) suggest that even if environmental sounds are processed via a non-linguistic stream, they can nevertheless rapidly engage linguistic systems and representations.

However, limits on the activation of linguistic knowledge during the processing of environmental sounds have also been observed. Iordanescu et al. (2011) investigated experiential associations: using a visual search task, their participants viewed either visual picture or word arrays while hearing auditory stimuli. They found that both picture and word search times were facilitated by hearing (i.e., corresponding) spoken words, while picture (e.g.,see also Iordanescu et al., 2010; Iordanescu et al., 2008) but not word search times were facilitated by hearing environmental sounds. These findings suggest that experience constrains the activation of linguistic knowledge: while it is typical to hear an environmental sound and view a corresponding object, and to hear a spoken word and view either a corresponding object or written word (e.g., reflecting picture naming and reading aloud), it is less typical to hear an environmental sound and view a corresponding written word. Complementing Chen and Spence’s (2011, 2018a) approach, these findings suggest that spoken words and environmental sounds engage linguistic systems and (e.g., orthographic) representations differently.

Recently, Bartolotti et al. (2020) investigated the activation of linguistic knowledge using the visual world paradigm. Building on Allopenna et al. (1998), their participants heard spoken words such as “clock” or environmental sounds such as a sound produced by a clock (e.g., ticking) while viewing visual arrays with objects such as a target clock, phonologically related cloud and/or unrelated lightbulb. They report observing “robust” phonological competition during both types of auditory stimuli, as reflected in fixations to the phonologically related cloud vs. unrelated lightbulb. These findings suggest that not only are lexical forms activated during the processing of environmental sounds, but that this activation also spreads to phonologically related representations. Thus, phonological competition provides particularly compelling support for the recruitment of linguistic systems and representations. These findings are also compelling in another respect: in contrast to studies such as Van Petten and Rheinfelder (1995) and Iordanescu et al. (2011), which required participants to respond to linguistic stimuli (e.g., word targets) immediately following or alongside environmental sounds, Bartolotti et al.'s (2020) participants were tasked with mapping environmental sounds onto (i.e., non-linguistic) pictures, which could conceivably be accomplished without activating linguistic knowledge.

In summary, prior research reveals both similarities and differences between the processing of spoken words and environmental sounds. On the one hand, differing responses have been observed for spoken words and environmental sounds (e.g., Chen & Spence, 2011; Iordanescu et al., 2011; Lupyan & Thompson-Schill, 2012; Toon & Kukona, 2020), consistent with the hypothesis that these auditory stimuli recruit differing mechanisms (e.g., linguistically mediated vs. direct processing streams, respectively; Chen & Spence, 2011, 2018a). On the other hand, Bartolotti et al. (2020) found that both types of auditory stimuli generated "robust" phonological competition (e.g., fixations to a phonologically related cloud vs. unrelated lightbulb when hearing either "clock" or the sound of a ticking clock). These latter findings suggest that both spoken words and environmental sounds may automatically and obligatorily engage linguistic systems and representations. In other words, the mapping of linguistic and non-linguistic stimuli onto conceptual knowledge may be surprisingly similar. However, the pervasiveness of these effects (e.g., across contexts) has not been investigated.

The aim of the current research was twofold: first, to conceptually replicate the phonological competition effects observed by Bartolottiet al. (2020); and second, to investigate contextual constraints on these effects. While Bartolotti et al.'s (2020) visual world findings provide preliminary support for the activation of linguistic knowledge during the processing of environmental sounds, they report a single experiment with a limited sample (e.g., N = 15 heard environmental sounds), suggesting that replication is important. Relatedly, while Toon and Kukona's (2020) visual world findings (e.g., fixations to a semantically related bone vs. unrelated candle when hearing barking) provide support for the activation of conceptual knowledge during the processing of environmental sounds, their findings do not resolve whether these auditory stimuli activate lexical forms. In other words, their participants may have activated their conceptual knowledge of puppies and bones, but not the words "puppy" and "bone". Moreover, the environmental sounds literature has focused considerable attention on context effects (e.g., Gygi & Shafiro, 2011; Krishnan et al., 2013; Leech et al., 2009; Orgs et al., 2007, 2008). For example, in the context of a visual detection task, Chen and Spence (2011) found that environmental sounds but not spoken words facilitated responses to corresponding pictures; however, in the context of a visual identification task (e.g., requiring linguistic responses), they found that both types of auditory stimuli did so. Relatedly, spoken words and environmental sounds are often interleaved within experimental contexts, such that participants may be required to engage with both linguistic and non-linguistic auditory stimuli within or across trials. Importantly, interleaving linguistic and non-linguistic stimuli within experimental contexts may prime participants to activate linguistic knowledge throughout (i.e., even when hearing environmental sounds).

In Experiment 1, participants heard spoken words such as "car" or environmental sounds such as a sound produced by a car while viewing visual arrays with objects such as a car (target), card (phonologically related competitor) and box (unrelated distractor). In Experiment 2 and 3 (in which visual arrays either included targets or not), participants only heard environmental sounds while viewing similar visual arrays. If environmental sounds activate linguistic knowledge automatically and obligatorily, more fixations to competitors (e.g., card) than distractors (e.g., box) are expected across experiments. If the activation of lexical forms is primed by the context (e.g., interleaved linguistic stimuli), weaker (or no) phonological competition is expected in Experiment 2 and 3 vs. 1.

Experiment 1

In order to test for the activation of linguistic knowledge during the processing of auditory stimuli, participants were presented with visual arrays with objects such as a car (target), card (phonologically related competitor), box (unrelated distractor) and hat (other unrelated object) and environmental sounds such as a sound produced by a car or spoken words such as “car”.

Method

Participants. Forty-eight undergraduates from De Montfort University (age *M* = 20.21, *SD* = 3.35; 39 females, 9 males) participated for course credit. All participants were native English speakers with normal or corrected-to-normal vision. The effect under focus was the difference in fixations to phonologically related (i.e., cohort) competitors vs. unrelated distractors; in the case of spoken words, Allopenna et al. (1998) observed an effect size of d_z = 2.60, while Huettig and McQueen (2007) observed an effect size of d_z = 0.62. An analysis using the *pwr* package in *R* revealed that the sample enabled detection of an effect size approximately 33% smaller than that in the latter via a paired sample *t*-test (d_z = 0.41, power = .80, α = .05). The study received research ethics committee approval.

Design. Both object type (target, competitor and distractor) and sound type (environmental sound and spoken word) were manipulated within participants.

Norming. Following Toon and Kukona (2020; see also De Groot et al., 2016), perceptual, conceptual and linguistic properties of the experimental stimuli were assessed via a separate norming study. Fifteen separate native English speakers participated for course credit or payment (£9/hour). Eighteen stimulus sets were created, which each included a target object (i.e., reflecting the spoken words in Experiment 1; e.g., car), phonologically related (i.e., cohort) competitor object (e.g., card), unrelated distractor object (e.g., box) and other object (e.g., star). The other object was a rhyme competitor in the norming (e.g., car-star); however, the other objects were rearranged in Experiment 1 (i.e., and 2) so that targets did not appear alongside rhyme competitors. Corresponding pictures were black and white line drawings drawn largely from the International Picture-Naming

Project (Szekely et al., 2004). The full list of target, competitor and distractor objects is reported in Table A1 of Appendix A. The norming procedure was identical to Toon and Kukona (2020), such that participants provided ratings on an 11-point scale of how much the competitor and distractor objects had to do with the target objects (semantic relatedness), how much the competitor and distractor objects looked like the target objects (visual similarity) and whether the target, competitor and distractor objects were associated with a specific sound or set of sounds that typically would allow them to be identified (associated sounds).

Table 1 reports competitor and distractor means and standard deviations, alongside paired sample *t*-tests and non-parametric Wilcoxon matched pairs tests, for each of the following: numbers of phonemes, letters and syllables; frequencies (Kučera & Francis, 1967); phonological similarities to the target based on the proportion of phonemes that also occurred in the target; target relations based on latent semantic analysis (LSA; Landauer & Dumais, 1997); and ratings of semantic relatedness to the target, visual similarity to the target and associated sounds.

Table 1.

Norming: Properties (*M*, *SD*) of the competitor vs. distractor objects. Phonological similarity, LSA, semantic relatedness and visual similarity are in relation to the target objects.

			t-test		Wilcoxon	
Fixed effect	Competitors	Distractors	*t*	*p*	*Z*	*p*
Phonemes	4.11 (1.13)	4.00 (0.77)	0.32	.75	-0.05	.96
Letters	4.28 (0.75)	4.17 (0.92)	0.40	.70	-0.26	.79
Syllables	1.06 (0.24)	1.00 (0.00)	1.00	.33	-1.00	.32
Frequencies	19.31 (12.24)	26.94 (28.07)	-1.09	.29	-0.57	.57
Phonological similarity	0.74 (0.15)	0.12 (0.15)	11.83	< .001	-3.73	< .001
LSA	0.05 (0.05)	0.09 (0.09)	-1.09	.29	-1.26	.21
Semantic relatedness	0.85 (0.98)	0.81 (0.84)	0.16	.88	-0.46	.65
Visual similarity	0.12 (0.15)	0.28 (0.40)	1.58	.13	-1.08	.28
Associated sounds	2.73 (2.08)	1.67 (1.05)	1.87	.08	-2.03	< .05

Importantly, the competitor and distractor objects differed significantly on the phonological similarity dimension, such that competitor (i.e., vs. distractor) objects were more similar to target objects, as expected. The competitor and distractor objects also differed by a small but significant amount in the non-parametric analysis of associated sounds. However, this result was linked to a single outlier item (17) with a large competitor vs. distractor mean difference (5.73 vs. 0.64); exclusion of this outlier yielded a non-significant difference between competitor (*M* = 2.55, *SD* = 2.00) and distractor (*M* = 1.73, *SD* = 1.05) objects on this dimension, *t*(16) = 1.50, *p* =.15, *Z* = -1.78, *p* = .08. In Experiment 1, analyses including all items are reported; however, the stability of these patterns was also confirmed following removal of this outlier item. In addition, it is conceivable that stronger competitor sound associations may generate weaker competition effects (i.e., the opposite of phonological prediction), because these competitors will be particularly at odds with the (i.e., target) sounds presented to participants. Finally, targets were confirmed as having very strong associations with specific sounds (*M* = 9.59, *SD* = 0.84).

Materials. Eighteen experimental visual arrays were created based on the norming. A target (e.g., car), a phonologically related (i.e., cohort) competitor (e.g., card), an unrelated distractor (e.g., box) and another unrelated object (e.g., hat) were depicted in each visual array. Mirroring Toon and Kukona (2020), visual displays were 1,024 × 768 pixels, images were 200 × 200 pixels, images were arranged in the four corners of the visual display centered 15% from each side and these images defined the fixation interest areas. Each visual array was presented with a corresponding environmental sound (e.g., a sound produced by a car) or spoken word (e.g., “car”). Also mirroring Toon and Kukona (2020), auditory stimuli were from online sound repositories and dictionaries and their peak amplitudes were normalised. The durations of environmental sounds (*M* = 498 ms, *SD* = 185) and spoken words (*M* = 482 ms, *SD* =79) did not differ significantly from each other, *t*(17) = -0.42, *p* =.68. The full list of experimental objects is reported in Table A1 of the Appendix.

Two counterbalanced lists were created by dividing the 18 visual arrays into two groups and rotating them through the two sound conditions in a Latin Square. Each visual array appeared once on each list, and each list included nine environmental sounds and nine spoken words. Eighteen filler visual arrays were also created, half presented with an environmental sound and half with a spoken word. Filler visual arrays included a target and three unrelated objects (i.e., no phonologically related competitors).

Procedure. The procedure was based on Toon and Kukona (2020). Auditory stimuli were presented via headphones, such that their onset followed the onset of the visual array by 1,000 ms. Participants were instructed to click with the computer mouse on the visual stimulus that corresponded to the auditory stimulus. Trials ended after participants made a response following auditory stimulus offset. Participants' eye movements were tracked throughout using an SR Research EyeLink 1000 Plus sampling at 500 Hz. The experiment began with two practice trials followed by 18 experimental and 18 filler trials, 50% presenting environmental sounds and 50% presenting spoken words. Participants were randomly assigned to lists, the same randomised ordering of experimental and filler trials was presented to all participants (e.g., in which no more than two experimental trials appeared back-to-back) and the location of visual stimuli was randomised.

Results

Mean accuracy was 99.77% (*SD* = 1.60) for environmental sounds and 99.07% (*SD* = 3.10) for spoken words. Inaccurate trials were excluded from further analysis. Mean reaction time was 1660 ms (*SD* = 327) for environmental sounds and 1570 ms (*SD* = 268) for spoken words. Spoken word trials were significantly faster, $t(47) = 2.74$, $p < 0.01$. Average proportions of fixations to targets, competitors and distractors during the processing of environmental sounds and spoken words are plotted in Figure 1A and B. The plot spans auditory stimulus onset to 500 ms following mean offset, with fixations (re)synchronised to both auditory stimulus onset and offset. As expected, a marked target advantage was observed by auditory stimulus offset (e.g., see the SEs in Figure 1).

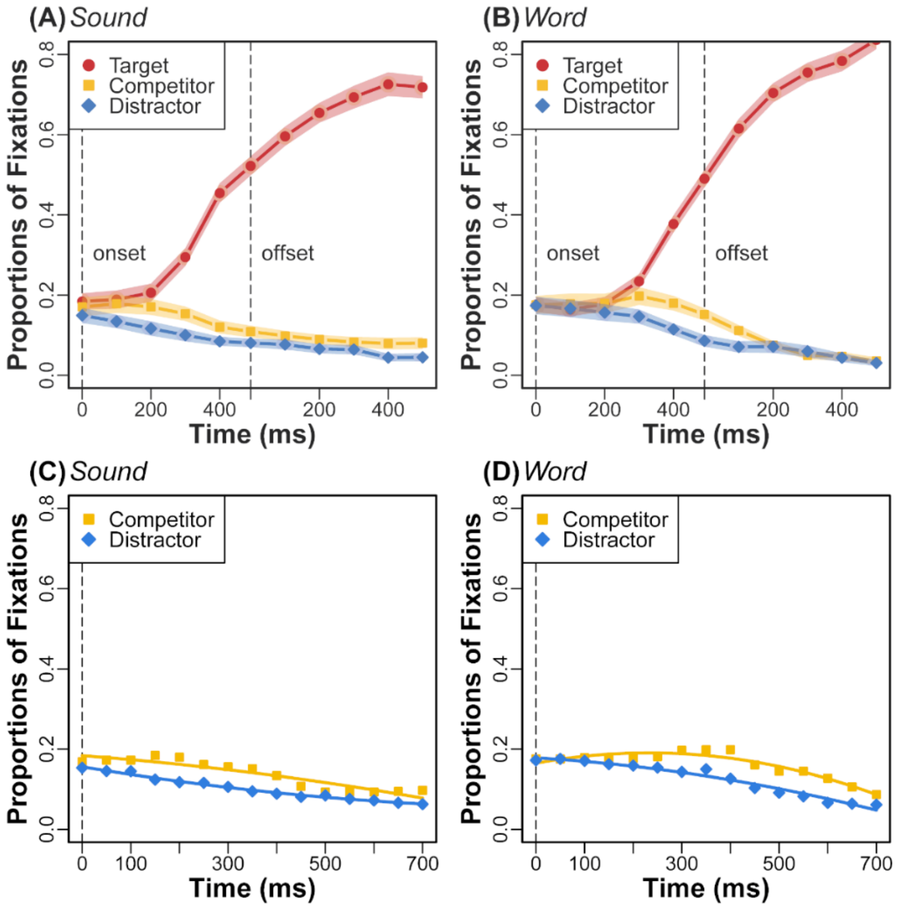

Figure 1. Experiment 1: Average (shaded bands show SEs) proportions of fixations to targets (e.g., car), phonological competitors (e.g., card) and unrelated distractors (e.g., box) between auditory stimulus onset and mean offset (+500 ms) for environmental sounds (e.g., a sound produced by a car; A) and spoken words (e.g., “car”; B), and growth curve analyses (curves indicate model fits and points indicate averages) of competitor vs. distractor fixations for environmental sounds (C) and spoken words (D).

As in Toon and Kukona (2020), growth curve analysis (e.g., Mirman, 2014; Mirman, Dixon, & Magnuson, 2008) was used to model differences in the time course of fixations to competitors vs. distractors (i.e., reflecting the activation of linguistic knowledge). The proportions of fixations to objects between auditory stimulus onset and 200 ms following mean offset (e.g., accommodating the typical lag in eye movement responses) were aggregated by participants into 50 ms bins and submitted to models (*lme4* in *R*; Bates et al., 2015) with orthogonal intercept, linear and quadratic polynomial terms. The focus throughout was on the intercept, which reflected mean differences in height and yielded the simplest interpretation (i.e., more vs. less fixations on average across time). In addition, the linear term reflected mean differences in linear change and the quadratic term reflected higher order differences. Although linear models were suitable in most cases, quadratic models are nevertheless reported throughout in order to confirm the (e.g., non) significance of higher order effects. Importantly, inclusion of these higher order effects did not affect the (i.e., orthogonal) pattern of results among the lower order terms.

First, environmental sounds and spoken words were analysed separately. Analyses modelled the time course of fixations to competitors and distractors and included fixed effects of object type (competitor = −0.5; distractor = 0.5) on each term. Following Mirman (2014), analyses first included participant-by-object and participant random effects on all terms, but these were simplified to participant-by-object random effects on all terms and participant random intercepts throughout due to convergence issues (*R* formula: Fixations ~ (Linear + Quadratic) * Object + (Linear + Quadratic | Participant:Object) + (1| Participant)). Results are reported in Table 2 and depicted in Figure 1C and D. The significant effect of object type on the intercept for both environmental sounds and spoken words indicates that there were more fixations to competitors than distractors on average across the analysis window. Likewise, simplified analyses (i.e., using paired sample *t*-tests; e.g., Allopenna et al., 1998) of the average proportions of competitor vs. distractor fixations between auditory stimulus onset and mean offset (lagged by 200 ms) revealed significant effects for both environmental sounds (competitor *M* = 0.12; *SD* = 0.07; distractor *M* = 0.09, *SD* = 0.05), *t*(47) = -2.66, *p* < .05, d_x = 0.38, and spoken words (competitor *M* = 0.16, *SD* = 0.09; distractor *M* = 0.11, *SD* = 0.08), *t*(47) = -2.94, *p* < .01, d_x = 0.42.

Second, environmental sounds and spoken words were compared directly. Difference curves were generated by participants by subtracting the average proportions of fixations to distractors from competitors within each time bin. Analyses modelled these difference (i.e., competitor minus distractor) curves, and included fixed effects of sound type (environmental sound = −0.5; spoken word = 0.5) on each term and participant-by-sound random effects on all terms (*R* formula: Fixations ~ (Linear + Quadratic) * Sound + (Linear + Quadratic | Participant:Sound)). Results are reported in Table 3 and depicted in Figure 2. The significant intercept indicates that there were more fixations to competitors than distractors on average across the analysis window for both environmental sounds and spoken words. However, there were no significant effects of sound type (e.g., including on the intercept), indicating that environmental sounds and spoken words did not differ significantly from each other.

Table 2.

Experiment 1: Growth curve analysis of competitor vs. distractor fixations for environmental sounds and spoken words (*Est*., *SE* and *95% CI* x 10^{-2}).

Fixed effect	Environmental Sounds				Spoken Words			
	Est. (SE)	*95% CI*	*t*	*p*	*Est. (SE)*	*95% CI*	*t*	*p*
Intercept	11.99 (0.71)	[10.59, 13.38]	16.82	< .001	14.39 (0.93)	[12.56, 16.21]	15.44	< .001
Linear	-11.84 (2.57)	[-16.87, -6.80]	-4.61	< .001	-12.45 (2.73)	[-17.81, -7.10]	-4.56	< .001
Quadratic	0.10 (1.64)	[-3.11, 3.31]	0.06	.95	-5.11 (1.71)	[-8.47, -1.75]	-2.98	< .01
Object	-3.50 (1.36)	[-6.17, -0.84]	-2.58	< .01	-3.67 (1.73)	[-7.05, -0.28]	-2.12	< .05
Object x Linear	1.59 (5.14)	[-8.48, 11.66]	0.31	.76	-6.24 (5.47)	[-16.95, 4.48]	-1.14	.25
Object x Quadratic	2.92 (3.28)	[-3.50, 9.35]	0.89	.37	5.03 (3.43)	[-1.68, 11.75]	1.47	.14

Table 3.

Experiment 1: Growth curve analysis of environmental sound vs. spoken word difference (i.e., competitor minus distractor) curves (*Est*., *SE* and *95% CI* x 10^{-2}).

Fixed effect	*Est. (SE)*	*95% CI*	*t*	*p*
Intercept	3.58 (1.08)	[1.46, 5.71]	3.31	< .001
Linear	2.32 (3.79)	[-5.11, 9.75]	0.61	.54
Quadratic	-3.98 (2.69)	[-9.25, 1.29]	-1.48	.14
Sound	0.16 (2.16)	[-4.08, 4.41]	0.07	.94
Sound x Linear	7.82 (7.58)	[-7.04, 22.68]	1.03	.30
Sound x Quadratic	-2.11 (5.38)	[-12.65, 8.43]	-0.39	.69

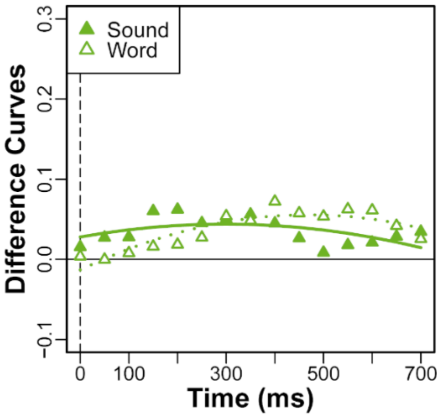

Figure 2. Experiment 1: Growth curve analysis of environmental sound vs. spoken word difference (i.e., competitor minus distractor) curves.

Finally, balancing items' associated sounds (i.e., by removing one outlier item; see the Norming) yielded a similar pattern of results: first, the analysis of competitors vs. distractors revealed significant effects of object type on the intercept for both environmental sounds, *Est*. = -3.43, *SE* = 1.38, *95% CI* = [-6.13, -0.73], *t* = -2.49, *p* < .05, and spoken words, *Est*. = -3.55, *SE* = 1.79, *95% CI* = [-7.05, -0.04], *t* = -1.98, *p* < .05; and second, the analysis of difference curves revealed a significant intercept, *Est*. = 3.49, *SE* = 1.13, *95% CI* = [1.27, 5.71], *t* = 3.08, *p* < .01, but no significant effects of sound type (all *t*s < 1).

Discussion

In Experiment 1, phonological competition was observed during the processing of both environmental sounds (e.g., a sound produced by a car) and spoken words (e.g., "car"), such that targets (e.g., car) were fixated most, but phonological competitors (e.g., card) were also fixated significantly more than distractors (e.g., box). These results confirm that the current linguistic stimuli generate phonological competition (e.g., Allopenna et al., 1998) and that the current visual and auditory stimuli are identifiable (e.g., accuracy >99%). In addition, these results conceptually replicate Bartolotti et al. (2020), suggesting that environmental sounds activate the lexical forms of targets (i.e., beyond mere semantics; e.g., Toon & Kukona, 2020), and that this activation spreads to phonologically related competitors. These results also reveal a strikingly similar pattern of activation between environmental sounds and spoken words, such that competitor (i.e., vs. distractor) fixations did not differ significantly between these two types of auditory stimuli. However, consistent with findings like Lupyan and Thompson-Schill (2012), reaction times were significantly faster for spoken words.

On the one hand, the current results are consistent with the hypothesis that environmental sounds engage linguistic systems and representations automatically and obligatorily. On the other hand, Experiment 1 reflected a context in which participants were required to process linguistic stimuli (i.e., spoken words) on 50% of (i.e., interleaved) trials. This context may prime participants to activate linguistic knowledge even when processing environmental sounds. Thus, participants were only presented with environmental sounds in Experiment 2.

Experiment 2

In order to test for contextual influences on the activation of linguistic knowledge, participants were presented with visual arrays and environmental sounds that were identical to Experiment 1. In contrast to Experiment 1, participants did not hear spoken words.

Method

Participants. Twenty-four undergraduates from De Montfort University (age *M* = 19.42, *SD* = 0.93; 23 females, 1 male) participated for course credit. All participants were native English speakers with normal or corrected-to-normal vision who did not participate in the norming or Experiment 1. The sample was 50% smaller than Experiment 1 due to the simpler design (i.e., sound type included one rather than two levels). The sample enabled detection of an effect size (i.e., competitor vs. distractor fixations) approximately equal to that in Huettig and McQueen (2007), but larger than Experiment 1 (d_z = 0.60, power = .80, α = .05).

Design. Object type (target, competitor and distractor) was manipulated within participants.

Materials. The stimuli were identical to Experiment 1 except that visual arrays were only presented with environmental sounds. In addition, only the practice and filler visual arrays with environmental sounds were presented.

Procedure. The procedure was identical to Experiment 1 except that the experiment began with one practice trial followed by 18 experimental and 9 filler trials, 100% presenting environmental sounds.

Results

Mean accuracy was 100% and mean reaction time was 1585 ms (*SD* = 270) for environmental sounds. Average proportions of fixations to targets, competitors and distractors during the processing of environmental sounds are plotted in Figure 3A. First, the time course of fixations to competitors and distractors was modelled as in Experiment 1. Results are reported in Table 4 and depicted in Figure 3B. There were no significant effects of object type (e.g., including on the intercept), indicating that competitors and distractors did not differ significantly from each other. Likewise, a simplified analysis of the average proportions of competitor vs. distractor fixations between auditory stimulus onset and mean offset (lagged by 200 ms) revealed a non-significant effect (competitor *M* = 0.12; *SD* = 0.05; distractor *M* = 0.13, *SD* = 0.05), *t*(23) = 0.39, *p* = .70, d_x = 0.08. Second, the environmental sound patterns in Experiment 1 and 2 were compared directly. Difference curves were generated by participants by subtracting the average proportions of fixations to distractors from competitors within each time bin. Analyses modelled these difference (i.e., competitor minus distractor) curves, and included fixed effects of experiment (Experiment 1 = −0.5; Experiment 2 = 0.5) on each term and participant random effects on all terms (*R* formula: Fixations ~ (Linear + Quadratic) * Experiment + (Linear + Quadratic | Participant)). Results are reported in Table 5 and depicted in Figure 4. The significant effect of experiment on the intercept indicates that the competitor vs. distractor advantage was larger on average across the analysis window in Experiment 1 vs. 2.

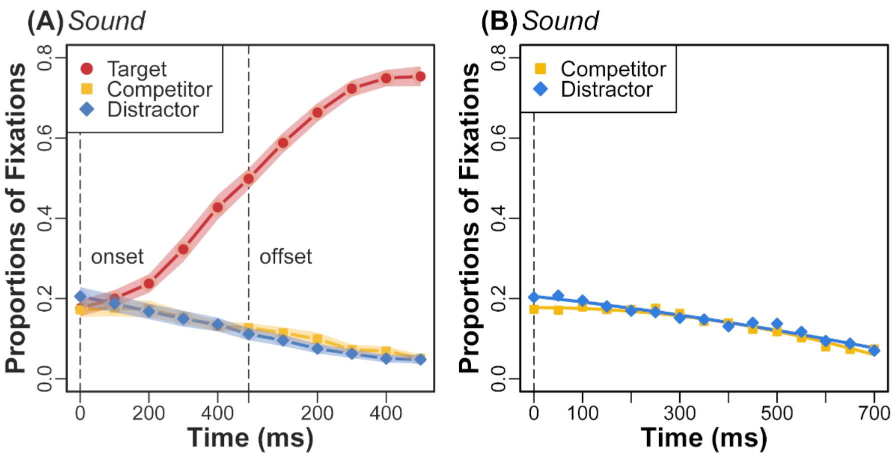

Figure 3. Experiment 2: Average proportions of fixations to targets, phonological competitors and unrelated distractors, and growth curve analysis of competitor vs. distractor fixations (B). Participants heard environmental sounds but not spoken words.

Finally, balancing items’ associated sounds yielded a similar pattern of results: first, the analysis of competitors vs. distractors revealed no significant effects of object type (all *p*s > .10); and second, the analysis of difference curves revealed a significant effect of experiment on the intercept, *Est*. = -4.13, *SE* = 2.10, *95% CI* = [-8.25, -0.01], *t* = -1.97, *p* < .05.

Table 4.

Experiment 2: Growth curve analysis of competitor vs. distractor fixations for environmental sounds (*Est*., *SE* and *95% CI* x 10^{-2}).

Fixed effect	*Est. (SE)*	*95% CI*	*t*	*p*
Intercept	14.21 (0.81)	[12.63, 15.79]	17.64	< .001
Linear	-14.72 (2.45)	[-19.52, -9.91]	-6.00	< .001
Quadratic	-2.58 (1.79)	[-6.09, 0.94]	-1.44	.15
Object	0.92 (1.10)	[-1.24, 3.09]	0.84	.40
Object x Linear	-1.33 (4.90)	[-10.94, 8.28]	-0.27	.79
Object x Quadratic	2.77 (3.59)	[-4.26, 9.80]	0.77	.44

Table 5.

Experiment 1 vs. 2: Growth curve analysis of environmental sound difference (i.e., competitor minus distractor) curves (*Est.*, *SE* and *95% CI* x 10^{-2}).

Fixed effect	*Est. (SE)*	*95% CI*	*t*	*p*
Intercept	1.29 (1.02)	[-0.70, 3.28]	1.27	.20
Linear	-0.13 (3.66)	[-7.31, 7.05]	-0.04	.97
Quadratic	-2.85 (3.02)	[-8.77, 3.07]	-0.94	.35
Experiment	-4.43 (2.03)	[-8.41, -0.45]	-2.18	< .05
Experiment x Linear	2.92 (7.33)	[-11.45, 17.28]	0.40	.69
Experiment x Quadratic	0.15 (6.04)	[-11.69, 11.99]	0.03	.98

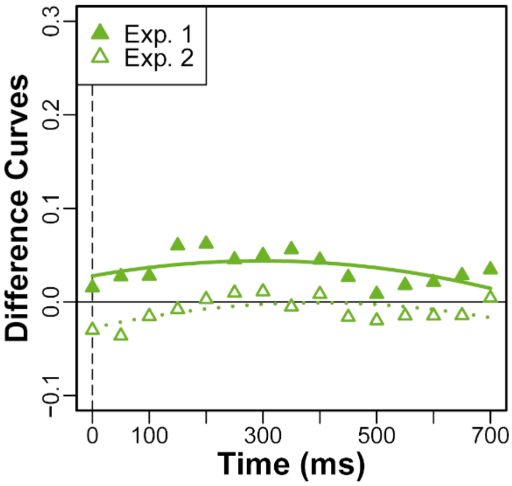

Figure 4. Experiment 1 vs. 2: Growth curve analysis of environmental sound difference (i.e., competitor minus distractor) curves.

Discussion

In Experiment 2, in contrast to Experiment 1, phonological competition was not observed during the processing of environmental sounds (e.g., a sound produced by a car), such that phonological competitors (e.g., card) were not fixated significantly more than distractors (e.g., box). The pattern of competitor (i.e., vs. distractor) fixations also differed significantly between Experiment 1 and 2. These results suggest that environmental sounds do not activate linguistic knowledge automatically and obligatorily. In a context that does not require participants to process linguistic stimuli, participants may not be primed to activate linguistic knowledge during the processing of environmental sounds. However, there was a marked target advantage (e.g., see Figure 3A), which may have masked subtler effects. Relatedly, Toon and Kukona (2020) observed particularly pronounced semantic competition during the processing of environmental sounds when targets were absent from visual arrays (e.g., a mean intercept difference between semantic competitors and distractors of 19.08) vs. present (e.g., a mean difference of 3.85; see also Huettig & Altmann, 2005). Likewise, phonological competition has also been studied using visual arrays in which targets are absent (e.g., Brouwer et al., 2012; Huettig & McQueen, 2007). Thus, using a design that paralleled Toon and Kukona (2020), participants were not presented with targets in Experiment 3.

Experiment 3

In order to test for further subtleties in the activation of linguistic knowledge, participants were presented with visual arrays with objects such as a card (phonologically related competitor) and drug (unrelated distractor) and environmental sounds such as a sound produced by a car. In contrast to Experiment 2, a car (target) was not depicted in the visual array (i.e., in experimental trials). Like Toon and Kukona (2020), the experiment also used visual arrays with two pictures and a look-and-listen task.

Method

Participants. Forty undergraduates from De Montfort University (age *M* = 19.55, *SD* = 1.99; 37 females, 3 males) participated for course credit. All participants were native English speakers with normal or corrected-to-normal vision who did not participate in the norming or Experiment 1 or 2. The sample enabled detection of an effect size (i.e., competitor vs. distractor fixations) approximately 25% smaller than that in Huettig and McQueen (2007), and only slightly larger than Experiment 1 (d_z = 0.45, power = .80, α = .05).

Design. Object type (competitor and distractor) was manipulated within participants.

Materials. The auditory stimuli were identical to Experiment 2. However, the experimental visual arrays were modified in two respects. First, each visual array included only two objects, vertically centered 15% from the left and right sides of the visual display. Second, the 18 experimental environmental sounds from Experiment 1 and 2 were grouped into nine pairs (e.g., the sounds produced by a car and drum), which were used to create nine visual arrays. The targets within each pair were not cohort competitors (e.g., car and drum) and the competitors of these pairs were included in the visual array (e.g., card and drug). Thus, targets were not included in visual arrays and visual arrays were presented (i.e., between participants) with a pair of environmental sounds. In addition, each object (e.g., card) was both a competitor (i.e., when presented with its corresponding environmental sound; e.g., a sound produced by a car) and distractor (i.e., when presented with the other paired environmental sound; e.g., a sound produced by a drum), counterbalancing extraneous linguistic and perceptual properties. Finally, the semantic relation (i.e., based on LSA) between each competitor and its corresponding (*M* = 0.05, *SD* = 0.05) and other paired (*M* = 0.08, *SD* = 0.08) target did not differ significantly from each other, *t*(16) = 1.77, *p* = .10 (note that LSA measures were unavailable for “lager”). The full list of pairs is reported in Table A2 of Appendix B.

Two counterbalanced lists were created by rotating objects through the competitor and distractor conditions in a Latin Square. Each visual array appeared once on each list. Nine filler visual arrays were also created by modifying the materials from Experiment 1 and 2 to include only a target and unrelated object. Thus, visual arrays included targets on one half of trials across the experiment.

Procedure. The procedure was identical to Experiment 1 and 2 except that a look-and-listen task was used. Participants were instructed to look carefully at the visual stimuli and to listen carefully to the auditory stimuli, but to not make an overt response (e.g., experimental trials did not include targets). The experiment began with one practice trial followed by nine experimental and nine filler trials, 100% presenting environmental sounds.

Results

As confirmation that participants were engaging with the look-and-listen task, analysis of the filler trials revealed that the average proportions of fixations to targets (*M* = 0.60; *SD* = 0.14) between auditory stimulus onset and 1,000 ms later was significantly greater than to distractors (*M* = 0.29, *SD* = 0.12), *t*(39) = 9.57, *p* < .001. Average proportions of fixations to competitors and distractors during the processing of environmental sounds are plotted in Figure 5A. The time course of fixations to competitors and distractors was modelled as in Experiment 1 and 2. Results are reported in Table 6 and depicted in Figure 5B. There were no significant effects of object type (e.g., including on the intercept), indicating that competitors and distractors did not differ significantly from each other. Likewise, a simplified analysis of the average proportions of competitor vs. distractor fixations between auditory stimulus onset and mean offset (lagged by 200 ms) revealed a non-significant effect (competitor *M* = 0.43, *SD* = 0.13; distractor *M* = 0.43, *SD* = 0.13), *t*(39) = 0.16, *p* =.87, d_x = 0.03. Finally, the visual arrays in Experiment 3 differed considerably from Experiment 1 and 2, and thus fixations were not compared directly across experiments. Rather, support for the null hypothesis was quantified using Bayes Factors (*BayesFactor*; Rouder et al., 2009). The simplified (i.e., paired sample *t*-test) analysis (difference *M* = 0.01, *95% CI* = [-0.06, 0.07]) yielded a Bayes Factor of 5.79, reflecting moderate support for the null hypothesis. In contrast, corresponding exploratory analyses of the environmental sound patterns in the first two experiments revealed moderate support for the alternative hypothesis in Experiment 1 (difference *M* = -0.04, *95% CI* = [-0.06, -0.01], *BF* = 0.28), and moderate support for the null hypothesis in Experiment 2 (difference *M* = 0.00, *95% CI* = [-0.02, 0.03], *BF* = 4.35).

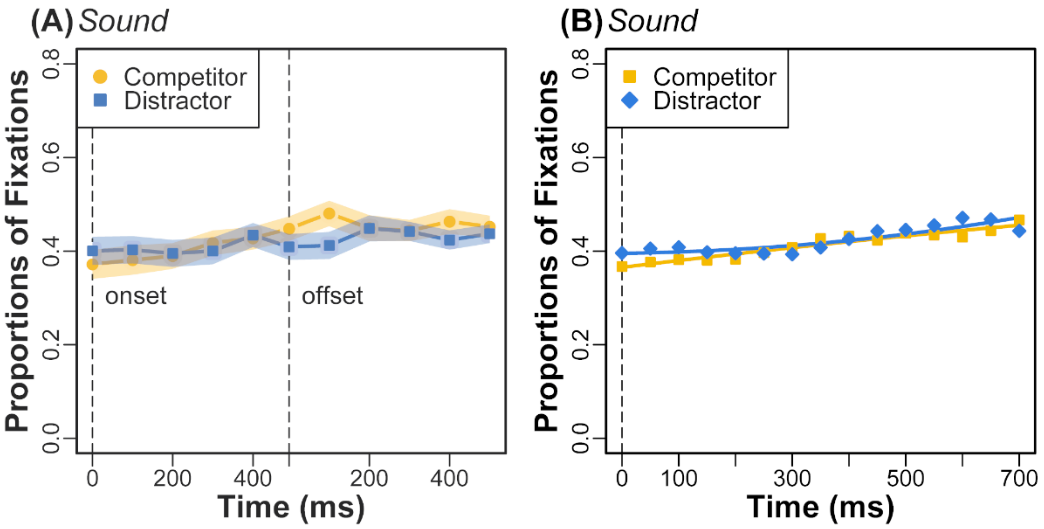

Figure 5. Experiment 3: Average proportions of fixations to phonological competitors and unrelated distractors (A), and growth curve analysis of competitor vs. distractor fixations (B). Participants heard environmental sounds but not spoken words.

Table 6.

Experiment 3: Growth curve analysis of competitor vs. distractor fixations for environmental sounds (*Est*., *SE* and *95% CI* x 10^{-2}).

Fixed effect	*Est. (SE)*	*95% CI*	*t*	*p*
Intercept	41.80 (1.37)	[39.12, 44.49]	30.49	< .001
Linear	9.94 (4.39)	[1.33, 18.55]	2.26	< .05
Quadratic	0.85 (2.07)	[-3.20, 4.90]	0.41	.68
Object	1.07 (2.74)	[-4.30, 6.45]	0.39	.70
Object x Linear	-1.65 (8.79)	[-18.87, 15.57]	-0.19	.85
Object x Quadratic	2.64 (4.13)	[-5.46, 10.74]	0.64	.52

Discussion

In Experiment 3, conceptually replicating Experiment 2 and contrasting with Experiment 1, phonological competition was not observed during the processing of environmental sounds (e.g., a sound produced by a car), such that phonological competitors (e.g., card) were not fixated significantly more than distractors (e.g., drug). In terms of the average fixation difference between competitors and distractors across the analysis window (i.e., intercept), the effect size was near to zero and reflected a small numerical advantage for distractors (*Est*. = 1.07; vs. *Est*. = -3.50 in Experiment 1). Again, these results suggest that environmental sounds do not activate linguistic knowledge automatically and obligatorily.

General Discussion

The current research investigated the activation of linguistic knowledge during the processing of linguistic and non-linguistic auditory stimuli. When participants heard environmental sounds (e.g., a sound produced by a car) interleaved among spoken words (e.g., “car”), they fixated phonological competitors (e.g., card) significantly more than unrelated distractors (e.g., box) during both types of auditory stimuli, conceptually replicating Bartolotti et al. (2020). However, when participants only heard environmental sounds, they did not fixate competitors significantly more than distractors. The current results yield two novel insights into the mapping of auditory stimuli onto conceptual knowledge: first, these results reveal important processing differences between linguistic and non-linguistic auditory stimuli; and second, these results emphasise the important influence of context on processing.

The results of Experiment 2 and 3 are consistent with Chen and Spence’s (2011, 2018a) theoretical approach. They hypothesise that spoken words make initial contact with the lexicon, which mediates subsequent contact with semantic memory, while environmental sounds make direct contact with semantic memory. In other words, linguistic auditory stimuli are processed via the lexicon, while non-linguistic stimuli are not. Likewise, these results reveal that that the processing of environmental sounds does not always lead to phonological activation. For example, the Bayes Factors from Experiment 2 and 3 provided moderate support (i.e., beyond a mere null result) against phonological competition in these experiments. Moreover, while Toon and Kukona (2020) observed particularly pronounced semantic competition when targets were absent from visual arrays, no such phonological effect was observed in Experient 3. These results emphasise an important distinction between spoken words and environmental sounds, such that while the former generate robust phonological competition (e.g., Allopenna et al., 1998; Huettig & McQueen, 2007), the latter do not. These results also complement findings such as Iordanescu et al. (2011), who found that while spoken words facilitated visual word search, environmental sounds did not. We conjecture that the default processing mode for environmental sounds does not involve contact with linguistic knowledge. Rather, the results of Experiment 2 and 3 suggest that environmental sounds are directly linked to semantic memory and can be processed without activating linguistic knowledge, consistent with Chen and Spence's (2011, 2018a) approach. These results also highlight further constraints on their approach: although Chen and Spence (2011, 2018a) assume that semantic memory and the lexicon are connected by bidirectional links, these results suggest that activation does not automatically and obligatorily cascade from the former onto the latter during the processing of environmental sounds.

However, against the backdrop of Chen and Spence's (2011, 2018a) approach, the results of Experiment 1 are more surprising. Consistent with Bartolotti et al. (2020), these results reveal that the processing of environmental sounds can lead to phonological activation. In other words, the mapping of auditory stimuli onto conceptual knowledge is sometimes surprisingly domain general, such that both linguistic and non-linguistic stimuli engage linguistic systems and representations. On the one hand, prior priming research reveals that participants can (e.g., when required by the task) rapidly link environmental sound primes to linguistic targets (e.g., Ballas, 1993; Frey et al., 2014; Orgs et al., 2006, 2007, 2008; Van Petten & Rheinfelder, 1995). Likewise, Chen and Spence's (2011, 2018a) approach includes bidirectional links between semantic memory and the lexicon, facilitating such contact. On the other hand, participants in Experiment 1 and Bartolotti et al. (2020) were tasked with mapping environmental sounds onto (i.e., non-linguistic) pictures, which could conceivably be accomplished without activating linguistic knowledge. In this respect, these results suggest that linguistic knowledge is surprisingly active despite the absence of concurrent linguistic stimuli.

In addition, while Bartolotti et al. (2020) argue that the time course of their effects is consistent with Chen and Spence’s (2011, 2018a) approach, this was not supported by Experiment 1. They report observing earlier phonological competition for spoken words than environmental sounds, such that the spread of activation across semantic memory and then the lexicon for the latter may require additional processing time. However, their spoken words were also shorter in duration than their environmental sounds, which may account for this discrepancy. In addition, the current spoken words and environmental sounds did not differ in duration, and no such time course difference was observed in Experiment 1 (e.g., to the contrary, Figure 2 suggests an early numerical advantage for environmental sounds over spoken words). On balance, this time course invites further study.

Taken together, the current results emphasise the important influence of context on the processing of environmental sounds. While participants only heard and viewed non-linguistic stimuli in Experiment 2 and 3, they explicitly engaged with linguistic auditory stimuli in Experiment 1. We conjecture that the interleaving of environmental sounds among spoken words in Experiment 1 primed participants to activate linguistic knowledge even when hearing non-linguistic auditory stimuli. Two potential mechanisms may underpin this pattern of results. First, building on Chen and Spence’s (2011, 2018a) approach, context may play an important role in constraining the flow of activation between semantic memory and the lexicon during the processing of environmental sounds. In other words, activation may be more likely to cascade from semantic memory onto the lexicon when the context requires participants to otherwise engage with linguistic stimuli, even if not immediately so (e.g., within a trial, as in Experiment 1). Second, participants may also be able to adopt a labeling strategy when hearing environmental sounds. In other words, while it may be possible to map an environmental sound onto a picture without activating linguistic knowledge, it is also possible to do so linguistically by labelling each and linking these labels together. Reflecting a type of carry over, participants may also be more likely to adopt this strategy (i.e., whether consciously or not) when they are otherwise explicitly engaging with labels (e.g., as in the spoken word trials in Experiment 1). In fact, the similarities between environmental sounds and spoken words in Experiment 1 may support this second mechanism (i.e., such that environmental sounds were not making contact with linguistic knowledge via a slower mediated stream). However, these results do not fully distinguish these mechanisms (e.g., or whether this strategy is conscious), reflecting an important direction for future research.

Relatedly, a growing empirical literature reveals various influences of context on the processing of environmental sounds. For example, Gygi and Shafiro (2011) observed an incongruency advantage, such that embedding environmental sounds in incongruous auditory contexts (e.g., hearing the sound of a galloping horse among restaurant sounds) facilitated processing (e.g., see also Krishnan et al., 2013; Leech et al., 2009). With respect to linguistic constraints, not only did Van Petten and Rheinfelder (1995) find that environmental sound primes influenced the processing of semantically related spoken word targets but spoken word primes also influenced the processing of semantically related environmental sound targets. Uddin, Heald, Van Hedger, Klos and Nusbaum (2018) also found that high cloze probability sentence contexts (e.g., “He bought diapers for his…”) facilitated responses to both corresponding spoken words (e.g., “baby”) and environmental sounds (e.g., a sound produced by a baby; see also Uddin, Heald, Van Hedger & Nusbaum, 2018). These findings reveal that like words, the processing of environmental sounds is affected by the immediate lexical and/or sentential context. More generally, this research also highlights an important parallel between environmental sounds and language, such that the processing of both is context dependent (e.g., in the case of language, see classic examples from the visual word paradigm such as Tanenhaus et al., 1995; Chambers et al., 2004). Relatedly, models by Barsalou, (1999), McClelland and Rogers (2003) and McRae et al. (1997) assume that lexical-semantic representations are distributed, featural and sensorimotor, such that perceptual information (e.g., associated sounds) is activated during linguistic processing. Building on these insights, the current results suggest that the processing of environmental sounds is also generically constrained by the linguistic context (i.e., dependent on whether or not participants engage with language at all). However, other aspects of processing may also be automatic and obligatory. Building on Van Petten and

Rheinfelder (1995), Orgs et al. (2008) found that word primes influenced the processing of semantically related environmental sound targets even in the context of a non-semantic task (e.g., detecting whether stimuli were presented to their left or right ear), suggesting that in contrast to linguistic knowledge, environmental sounds may activate conceptual knowledge automatically and obligatorily.

An important limitation of the current research is that it only addresses a single constraint on the processing of environmental sounds. Alongside the interleaving of linguistic and non-linguistic stimuli, we conjecture that other contextual constraints are also likely to prime the activation of linguistic knowledge. In fact, Bartolotti et al.'s (2020) participants only heard environmental sounds or spoken words, not both, paralleling Experiment 2 and 3 (i.e., in which phonological competition was not observed) rather than Experiment 1. However, their experiment diverged from the current experiments in other respects; for example, their participants were presented with hundreds of trials with repeated targets, which may encourage a labelling strategy. Thus, an important issue for future research will be to address related contextual constraints on the activation of linguistic knowledge during the processing of environmental sounds. Two features of Experiment 1 may also be relevant: participants engaged with picture labels (e.g., hearing the spoken word "car" and seeing a picture of a car), and they also switched between (i.e., linguistic and non-linguistic) modes of processing throughout. Thus, whether engaging with language outside a labelling context, and/or whether engaging with task switching, impacts on phonological competition during the processing of environmental sounds remains unresolved. Again, we conjecture that the picture labelling component of Experiment 1 may be critical (e.g., note that in Experiment 2 and 3, participants did engage with some language, such as the task instructions at the beginning of the study), reflecting an important direction for future research. Finally, the effect of context may also change over time (e.g., across trials), also inviting further study.

Finally, two methodological implications are worth highlighting. First, the current results suggest that the visual world paradigm (e.g., Tanenhaus et al., 1995) is a powerful tool for studying the psychology of environmental sounds (e.g., see also Bartolotti et al., 2020; Edmiston & Lupyan, 2015; Toon & Kukona, 2020), building on decades of research using behavioral (e.g., priming) and neurophysiological methodologies. For example, complementing Toon and Kukona (2020; see also Bartolotti et al., 2020), the current results suggest that alongside the activation of conceptual knowledge (e.g., puppy-bone), in some contexts environmental sounds also activate lexical forms (e.g., car-card). Second, the current results also suggest that interleaving linguistic and non-linguistic stimuli within an experimental context is an important design consideration. Relatedly, Dick and Saygin and colleagues (e.g., Dick et al., 2007; Dick et al., 2015; Saygin et al., 2003) highlight the striking similarities between the networks of neural resources recruited by spoken words and environmental sounds (e.g., including language related brain regions in the left hemisphere). On the one hand, the current results suggest that these similarities (e.g., as may reflect the activation of linguistic knowledge) may be primed by the experimental context (e.g., see Experiment 1). On the other hand, neurophysiological research often uses blocked designs (i.e., linguistic and non-linguistic stimuli are interleaved across blocks rather than trials; e.g., Dick et al., 2007); thus, an important issue for future research will be to address related contextual constraints on the activation of linguistic knowledge.

In conclusion, the current visual world results reveal that the activation of linguistic knowledge during the processing of non-linguistic auditory stimuli is context dependent rather than automatic. In the current study, participants engaging with both linguistic and non-linguistic auditory stimuli fixated a card when hearing either “car” or a sound produced by a car (Experiment 1), reflecting the activation of lexical forms (e.g., car-card) across both types of auditory stimuli. However, participants engaging with only non-linguistic stimuli showed no such effects (Experiment 2 and 3), reflecting contextual constraints on processing. These results provide novel insight into the cascade of activation across semantic memory and the lexicon during the processing of auditory stimuli and are interpreted as consistent with theoretical approaches that distinguish linguistically mediated vs. direct processing streams (e.g., Chen & Spence, 2011, 2018a).

References

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

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of experimental psychology: Human perception and performance, 19(2)*, 250-267.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and brain sciences, 22(4)*, 637-660.

Bartolotti, J., Schroeder, S. R., Hayakawa, S., Rochanavibhata, S., Chen, P., & Marian, V. (2020). Listening to speech and non-speech sounds activates phonological and semantic knowledge differently. *Quarterly journal of experimental psychology, 73(8)*, 1135-1149.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software, 67(1)*, 1-48.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of neuroscience, 35(25)*, 9329-9335.

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *Quarterly journal of experimental psychology, 65(11)*, 2193-2220.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of experimental psychology: Learning, memory, and cognition, 30(3)*, 687-696.

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. (2018a). 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.

Chen, Y. C., & Spence, C. (2018b). Dissociating the time courses of the cross-modal semantic priming effects elicited by naturalistic sounds and spoken words. *Psychonomic bulletin & review, 25(3)*, 1138-1146.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language, 35(1)*, 26-58.

De Groot, F., Koelewijn, T., Huettig, F., & Olivers, C. N. L. (2016). A stimulus set of words and pictures matched for visual and semantic similarity. *Journal of Cognitive Psychology, 28(1)*, 1-15.

Dick, F., Krishnan, S., Leech, R., & Saygin, A. P. (2015). Environmental sounds. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 1121-1138). Academic Press.

Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., ... & Pizzamiglio, L. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *Journal of cognitive neuroscience, 19(5)*, 799-816.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition, 143*, 93-100.

Frey, A., Aramaki, M., & Besson, M. (2014). Conceptual priming for realistic auditory scenes and for auditory words. *Brain and cognition, 84(1)*, 141-152.

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.

Gygi, B., & Shafiro, V. (2011). The incongruency advantage for environmental sounds presented in natural auditory scenes. *Journal of experimental psychology: Human perception and performance, 37(2)*, 551.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96(1)*, B23-B32.

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.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta psychologica, 137(2)*, 252-259.

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic bulletin & review, 15(3)*, 548-554.

Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing research, 300*, 46-55.

Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *The Journal of the Acoustical Society of America, 126(6)*, 3147-3155.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of experimental psychology: General, 141(1)*, 170-186.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology, 18(1)*, 1-86.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature reviews neuroscience, 4(4)*, 310–322.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25(1-2)*, 71-102.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of experimental psychology: General, 126(2)*, 99-130.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology, 90(2)*, 227.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. New York: CRC.

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.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52(3)*, 189-234.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and cognition, 62(3)*, 267-272.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory?. *International journal of psychophysiology, 65(2)*, 162-166.

Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience letters, 436(2)*, 133-137.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16(2)*, 225-237.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... & Federmeier, K. (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language, 51(2)*, 247-250.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds Evidence from aphasia. *Brain, 126(4)*, 928–945.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995).

Integration of visual and linguistic information in spoken language comprehension. *Science, 268(5217)*, 1632-1634.

Toon, J., & Kukona, A. (2020). Activating semantic knowledge during spoken words and environmental sounds: Evidence from the visual world paradigm. *Cognitive science, 44(1)*, e12810.

Uddin, S., Heald, S. L., Van Hedger, S. C., Klos, S., & Nusbaum, H. C. (2018). Understanding environmental sounds in sentence context. *Cognition, 172*, 134-143.

Uddin, S., Heald, S. L., Van Hedger, S. C., & Nusbaum, H. C. (2018). Hearing sounds as words: Neural responses to environmental sounds in the context of fluent speech. *Brain and language, 179*, 51-61.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia, 33(4)*, 485-508.

Appendix A

Targets, phonologically related competitors, unrelated distractors and other unrelated objects from Experiment 1 and 2.

Table A1.

Experimental stimuli from Experiment 1 and 2.

	Target	Competitor	Distractor	Other
1	Bell	Belt	Cross	Giraffe
2	Broom	Bruise	Egg	Shell
3	Car	Card	Box	Hat
4	Cat	Cap	Rope	Wig
5	Cow	Couch	Pear	Sleep
6	Dog	Dock	Globe	Truck
7	Drum	Drug	Leaf	Bow
8	Duck	Dump	Kite	Wing
9	Pig	Pin	Map	Brush
10	Saw	Sauce	Bridge	Drain
11	Sheep	Sheet	Beard	Claw
12	Ship	Shit	Bra	Tomb
13	Train	Tray	Slide	Log
14	Flush	Fluff	Tie	Brain
15	Laugh	Lager	Tent	Bum
16	Phone	Foam	Snail	Whip
17	Rain	Rail	Dress	Star
18	Sing	Sink	Nail	Bone

Appendix B

Pairs of targets and phonologically related competitors from Experiment 3.

Table A2.

Experimental stimuli from Experiment 3.

	Target 1	Competitor 1	Target 2	Competitor 2
1	Saw	Sauce	Broom	Bruise
2	Phone	Foam	Sing	Sink
3	Flush	Fluff	Cow	Couch
4	Pig	Pin	Ship	Shit
5	Drum	Drug	Car	Card
6	Bell	Belt	Laugh	Lager
7	Rain	Rail	Sheep	Sheet
8	Dog	Dock	Cat	Cap
9	Train	Tray	Duck	Dump