EXEMPLAR EFFECTS ARISE IN A LEXICAL DECISION TASK, BUT ONLY UNDER ADVERSE LISTENING CONDITIONS

Annika Nijveld[1], Louis ten Bosch[1] & Mirjam Ernestus[1,2]

[1]Radboud University, [2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
a.nijveld@let.ru.nl; l.tenbosch@let.ru.nl; m.ernestus@let.ru.nl

ABSTRACT

This paper studies the influence of adverse listening conditions on exemplar effects in priming experiments that do not instruct participants to use their episodic memories. We conducted two lexical decision experiments, in which a prime and a target represented the same word type and could be spoken by the same or a different speaker. In Experiment 1, participants listened to clear speech, and showed no exemplar effects: they recognised repetitions by the same speaker as quickly as different speaker repetitions. In Experiment 2, the stimuli contained noise, and exemplar effects did arise. Importantly, Experiment 1 elicited longer average RTs than Experiment 2, a result that contradicts the time-course hypothesis, according to which exemplars only play a role when processing is slow. Instead, our findings support the hypothesis that exemplar effects arise under adverse listening conditions, when participants are stimulated to use their episodic memories in addition to their mental lexicons.

Keywords: speech comprehension, exemplar effects, adverse listening conditions, noise

1. INTRODUCTION

Hybrid models of speech comprehension assume two types of lexical representations for the pronunciation of words: abstract representations and exemplars (e.g., [7] and [9]). Abstract representations are strings of sounds symbols that contain no details about the exact pronunciations of words. Exemplars, in contrast, are highly detailed representations of each occurrence of a word, which together form a word cloud. Findings from auditory priming experiments offer support for the representation of words as clouds of exemplars (e.g., [4], [14], [2]). In these studies, listeners recognised the second occurrence of a word more quickly or more accurately when surface details (e.g., speaker voice) of the first (prime) and second (target) occurrence of a word matched compared to when they did not match. However, several studies did not replicate these exemplar effects (e.g., [6], [12]). Because the precise role of exemplars has important consequences for models of speech comprehension, we further investigated under which circumstances exemplars play a role in speech comprehension.

We tested the hypothesis that exemplar effects arise when participants do not only use their mental lexicons to perform a task but also their episodic memories. They may do so because the task in an experiment instructs them so (e.g., old/new judgment) or because relying on episodic memory offers them a processing advantage, for example, under adverse listening conditions.

Exemplar effects indeed arose in previous priming studies where listening conditions were suboptimal: in [10], participants listened to healthy, mildly dysarthric or severely dysarthic speakers, and words were either repeated by the same or a different speaker. Participants were faster to recognise same compared to different voice repetitions, but only when listening to mildly or severely dysarthic speech, and mostly so in the latter case. In [16], words were presented in clear speech, or with increasing levels of white noise (+5 to -5 SNR). The authors found larger exemplar effects on accuracy when the signal to noise ratio was lower.

In these studies, not only were listening conditions suboptimal, but the task (old/new judgment) also instructed participants to use their episodic memories. It is therefore unclear which aspect caused the exemplar effects (and to what extent).

In [8], participants listened to words produced in full or with the first-syllable schwa missing, by the same speaker. Participants were faster to recognise repeated words in the same than in a different pronunciation variant. The variants without schwa occur frequently in casual speech but seldom out of context as in [8]. Their presence in the experiment may therefore have increased participants' processing cost, who were thus stimulated to use their episodic memories. The effect in [11], where participants listened to a talker with a foreign accent, likely also arose for this reason, but was confounded with speaker.

We report two experiments to test our hypothesis that if participants are not instructed to use their episodic memories, exemplar effects especially arise under adverse listening conditions. Experiment 1 is based on the lexical decision experiment by [8] described above, but we used clear speech produced by two different speakers instead of two pronunciation variants produced by one speaker. If the exemplar effects reported by [8] result from the adverse listening conditions, as we hypothesise, our version of the experiment should not show exemplar effects. In Experiment 2, we presented the same stimuli but added speech-shaped noise (+3 dB SNR) to them, which made them harder to understand. We predict that this experiment will show exemplar effects again.

2. EXPERIMENT 1

2.1. Method

2.1.1. Participants

We tested 26 participants, aged between 18 and 25 years (mean: 20 years). Five were male and two were left-handed. None of the participants in this or the other experiment presented in this paper reported any hearing impairment or participated in both experiments. All were paid for their participation.

2.1.2. Materials

The lexical decision experiment was run in Dutch and contained the same words and pseudo words as [8]; all were trisyllabic prefixed infinitives (starting with *be-* or *ver-*, e.g. real word *vertellen* 'to tell' and pseudo word *bekrempen*). All pseudo words were phonotactically legal in Dutch. We repeated 48 real infinitives, which had a mean frequency of occurrence of 3362 per million (range: 456 - 8296 per million; based on [1]).

We divided the experiment in two parts, where each part contained both tokens (prime and target) of half of the repeated words. Each part consisted of two blocks: a familiarization block, with 24 primes, 24 to-be-repeated pseudo word foils and 24 additional foils (12 pseudo words; 12 real words), and a target block, with 24 targets, 24 repeated pseudo word foils and 24 additional foils (12 pseudo words; 12 real words). One real word foil (*besmetten*,'to infect') was accidentally repeated in [8]; we replaced its second occurrence by *bestijgen* ('to ascend').

We recorded all items with a male and a female native speaker of Dutch in a sound-attenuating booth. Target words spoken by the male speaker had an average duration of 688 ms (range: 580 - 843 ms; *SD:* 56 ms), while the targets spoken by the female speaker were 626 ms on average (range: 469 - 760 ms; *SD*: 67 ms).

For the presentation of the trials to participants, we created four lists, in which words occurred in the same pseudo randomised order as in the four master lists of [8]. In the lists, each block started with a foil, primes and targets were always followed by a foil, at most eight real or pseudo words occurred in succession, and primes and targets were separated by maximally 100 trials (average: 67, range: 19 - 100). In each list, an equal number of prime-target pairs were assigned to one of the four possible combinations of speaker voice. Per master list, we created three lists with the same word order but in which the prime-target pairs represented one of the three other possible combinations of speaker voice. In each set of four lists, every prime-target pair represented each voice combination exactly once. In each block of each of the lists, half of the targets and approximately half of the foils were spoken by the male speaker, and the other half were spoken by the female speaker.

2.1.3. Procedure

Participants were tested individually in a sound-attenuating booth. They listened to the stimuli via closed headphones and performed a lexical decision task by means of button presses on a button box (yes-responses with the dominant hand). Per trial, one stimulus was presented and the next trial started one second after a button press or after 3.5 seconds after trial onset. Between the two parts of the experiment, participants took a short break, and one session lasted approximately 15 minutes.

2.1.4. Analyses

We analysed log-transformed response times (RTs) to the target words by means of mixed effects regression models, and accuracy scores of words and pseudo words by means of generalised mixed effects regression models. We used word, participant and speaker (of the target word) as crossed random effects in both analyses. We restricted the RT analysis to trials that received correct responses and whose primes also received correct responses, and we removed trials with response times that differed more than two standard deviations from the grand mean.

For both the RT and accuracy analyses, we tested random slopes for all fixed effects. We only included effects and interactions if they were significant (for this type of data, t-values above 1.96 or below -1.96 imply p-values < .05) and if they significantly improved the statistical model (tested with the `anova()` function from the `R` Statistical Software

[15]). We orthogonalised correlating control variables before they were included in the model (*RT prime* was residualised over *RT preceding trial* in the RT analysis). For the RT analysis, we removed data points with standardised residuals exceeding 2.5 standard deviation units for the best model and refitted the model. Our main predictor was *speaker match*, which reflects whether or not a prime and target were pronounced by the same speaker. We also explored the influence of the control predictors log-transformed response times to the prime *(RT prime)* and to the preceding trial *(RT preceding trial)*, log-transformed *target word duration* and *affix* (whether target words carried the affix *be-* or *ver-*).

2.2. Results

Participants, on average, made errors on 4% of target words and pseudo words. Analysis of accuracy scores did not reveal any effect of *speaker match*.

For the response time (RT) analysis, we excluded one participant (12% errors) and one word (*bekransen* 'to garland': 46% errors). Table 1 shows the statistical model based on the remaining 986 trials. Response times, measured from word onset, were on average 986 ms. The effects of our control predictors reveal that participants responded more quickly to targets when they were quicker on the preceding trial or on the prime, when the word started with *be-* (mean: 952 ms) rather than *ver-* (1006 ms) and when the word was shorter. Crucially, we did not observe a significant effect of *speaker match*, which indicates that participants were not quicker to respond to targets pronounced by the same speaker as the prime than to targets pronounced by the other speaker.

Table 1: Statistical model for the response times to targets in Experiments 1 and 2. Estimated standard deviation is denoted by SD.

	Experiment 1		**Experiment 2**	
Fixed effects	$\hat{\beta}$	t	$\hat{\beta}$	t
Intercept	2.97	9.3	3.05	8.2
RT preceding trial	0.22	9.9	0.15	6.8
RT prime	0.12	5.4	0.14	4.1
Affix *ver-*	0.02	2.5	-	n.s.
Target duration	0.37	8.4	0.42	8.1
Speaker mismatch	-	n.s.	-	n.s.
Speaker mismatch * RT prime	-	n.s.	-0.09	-2.2
Random effects		SD		SD
Word	intercept	0.02	intercept	0.04
Participant	intercept	0.07	intercept	0.06
Speaker	intercept	0.02	intercept	0.02

3. EXPERIMENT 2

3.1. Method

3.1.1. Participants

The participants were 26 native speakers of Dutch, aged between 18 and 29 years (mean: 22 years). Five were left-handed and seven were male.

3.1.2. Materials, procedure and analyses

We used the same materials as in Experiment 1, but superimposed speech-shaped noise at +3 dB SNR to the recordings (energetic masking). An informal pre-test confirmed that this noise level made the words harder to identify while it was still possible to perform the task. We first modified the loudness of the speech to reach the desired SNR level, and subsequently rescaled the speech + noise to the original loudness level of the speech (70 dB). The noise started and ended with a 30 ms ramp.

The procedure was identical to the one of Experiment 1. For the statistical analysis, we used the same method and predictors as in Experiment 1.

3.2. Results

On average, participants made 8% errors on target words and pseudo words. These errors showed no effect of *speaker match*.

For the RT analysis, we again excluded the word (*bekransen* 'to garland': 69% errors) as well as one participant (23% errors). The statistical model based on the remaining 913 trials is summarised in Table 1. Response times to target words were on average 928 ms. The control predictors that played a role in Experiment 1 were again significant and showed similar effects, except for the predictor *affix* that was no longer significant. Importantly, we found an interaction between *speaker match* and *RT prime*, which indicated that participants were quicker to respond to target words repeated by the same speaker than to words repeated by a different speaker, but only for those target words whose primes received quick responses.

The difference between Experiments 1 and 2 in the effect of *speaker match* modulated by *RT prime* is supported by an analysis of the combined datasets, which shows a significant interaction between *speaker match*, *RT prime* and *experiment* ($\hat{\beta}$ = -0.12, t = -2.1). This result confirms that *speaker match*, modulated by *RT prime*, only played a role in Experiment 2.

4. GENERAL DISCUSSION

In this paper, we investigated the hypothesis that when participants are not instructed to use their episodic memories, exemplar effects arise especially under adverse listening conditions. To this end, we reported two experiments in which participants performed lexical decision, a task that does not require participants' use of episodic memory.

In Experiment 1, participants listened to clear speech produced by two different speakers. As hypothesised, we found no exemplar effects: participants were equally fast to recognise repeated words spoken by the same speaker and by the other speaker. This results contrasts with the lexical decision experiment reported by [8], which used the same words. The crucial difference between the two experiments is that [8] presented half of the stimuli with the first-syllable schwa missing. These casual pronunciation variants seldom occur in isolation, as they were presented in [8], and were therefore presumably hard to process. This result confirms our hypothesis that exemplar effects may arise under adverse listening conditions.

In Experiment 2, we re-introduced an adverse listening condition by adding noise to the stimuli of Experiment 1. In contrast to Experiment 1, we did find exemplar effects here: for targets whose primes received quick responses, participants were quicker to respond to words repeated by the same speaker than by the other speaker. Experiment 2 only differed from Experiment 1 in the noise that we added; this result therefore suggests that the exemplar effects in Experiment 2 arose because of the adverse listening condition that was created by the noise. This finding provides further support our hypothesis that adverse listening conditions stimulate participants to use their episodic memories, which evokes exemplar effects.

The exemplar effect we found in Experiment 2 was modulated by *RT prime*, which indicates that exemplars only played a role for those targets whose primes elicited quick responses. RTs for the primes are likely a measure of how easily participants could process the primes [5]. A relatively long RT on a prime may reflect a case in which the participant - even though (s)he gave a correct response - found it hard to identify the prime through the noise. In these cases, participants may have missed details that are needed to build a full episode. Due to the prime's impoverished representation, participants may have been unable to match the target to the prime when they heard the target, and could not benefit from using episodic memory in these cases.

The RTs in Experiment 2 were shorter (928 ms) than in Experiment 1 (986 ms), a result that may be surprising given the enhanced listening difficulty in Experiment 2. Possibly, the more challenging stimuli in Experiment 2 made participants in this experiment more motivated, a factor that can lead to faster responses [3].

The time-course hypothesis ([12], [13]) predicts that exemplar effects only arise when responses are relatively slow, because it takes more time to activate exemplars than abstract representations. Our findings do not support this hypothesis, as we observed exactly the opposite pattern of results: we only found exemplar effects in the experiment with shorter RTs.

The combined results of our study and [8] show that different kinds of adverse listening conditions can lead to exemplar effects in lexical decision experiments: both an experiment that (in part) included casually produced stimuli and an experiment with noise superimposed onto all stimuli showed exemplar effects. Interestingly, the effect in [8] (19 ms) was larger than the one in Experiment 2 (12 ms). Possibly, our experiment showed a smaller exemplar effect because schwa deletion in isolation represents a more severe adverse listening condition than moderate noise does.

Our results have implications for models of spoken word recognition. We found further support for the storage of word pronunciations as clouds of exemplars in Experiment 2, while the absence of exemplars effects in Experiment 1 favours storage as abstract representations. Hybrid models combine both types of representations, but cannot straightforwardly explain why we observed exemplar effects under adverse listening conditions only. Our results are therefore best explained by models that assume that during speech comprehension, the mental lexicon, containing abstract representations, cooperates with domain-general episodic memory, in which clouds of exemplars are represented. Depending on the situation, one or the other plays a more important role.

In conclusion, we investigated the influence of adverse listening conditions on the occurrence of exemplar effects, and only found exemplar effects in the experiment in which the stimuli contained noise. When no noise was used, no exemplar effects arose. This suggests that adverse listening conditions stimulate participants to use their episodic memories (also in the absence of a task that instructs them to do so), which enhances the probability of observing exemplar effects.

5. ACKNOWLEDGMENTS

This work was supported by a VICI grant (277-70-010) from the Netherlands Organisation for Scientific Research (NWO) to the third author.

6. REFERENCES

[1] Baayen, H. R., Piepenbrock, R., Gulikers, L. 1995. *The CELEX Lexical Database. Release 2 (CD-ROM)*. Philadelphia, Pennsylvania: Linguistic Data Consortium, University of Pennsylvania.

[2] Bradlow, A., Nygaard, L., Pisoni, D. B. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept. Psychophys.* 61(2), 206–19.

[3] Capa, R. L., Audiffren, M., Ragot, S. 2008. The interactive effect of achievement motivation and task difficulty on mental effort. *Int. J. Psychophysiol.* 70(2), 144–50.

[4] Craik, F., Kirsner, K. 1974. The effect of speaker's voice on word recognition. *Q. J. Exp. Psychol.* 26(2), 274–284.

[5] Cutler, A., Robinson, T. 1992. Response time as a metric for comparison of speech recognition by humans and machines. *ICSLP* Banff.

[6] Goldinger, S. D. 1996. Words and voices: episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 22(5), 1166–83.

[7] Goldinger, S. D. 2007. A complementary-systems approach to abstract and episodic speech perception. *Proc. 16th ICPhS* Saarbrücken. 49–54.

[8] Hanique, I. A. M., Aalders, E. H., Ernestus, M. T. C. 2013. How robust are exemplar effects in word comprehension. *Phonological and Phonetic considerations of Lexical Processing. Thematic Issue of The Mental Lexicon* 8(3), 269–294.

[9] Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon.* 31, 373–405.

[10] Mattys, S. L., Liss, J. M. 2008. On building models of spoken-word recognition: when there is as much to learn from natural "oddities" as artificial normality. *Percept. Psychophys.* 70(7), 1235–42.

[11] McLennan, C. T., González, J. 2012. Examining talker effects in the perception of native- and foreign-accented speech. *Atten. Percept. Psychophys.* 74(5), 824–30.

[12] McLennan, C. T., Luce, P. A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 31(2), 306–21.

[13] McLennan, C. T., Luce, P. A., Charles-Luce, J. 2003. Representation of lexical form. *J. Exp. Psychol. Learn. Mem. Cogn.* 29(4), 539–553.

[14] Palmeri, T., Goldinger, S. D., Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *J. Exp. Psychol. Learn. Mem. Cogn.* 19(2), 309–328.

[15] RCoreTeam, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[16] Saldaña, H. M., Nygaard, L. C., Pisoni, D. B. 1999. Speechreading by humans and machine: Models, systems, and applications. In: Stork, D., Hennecke, M. E., (eds), *The Handbook of Phonetic Science*. Berlin: Springer-Verlag 275–281.