Available online at www.sciencedirect.com
ScienceDirect
Cognition 105 (2007) 457–465
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Perceptual uniqueness point effects in monitoring internal speech ☆

Rebecca Özdemir [a], Ardi Roelofs [b,*], Willem J.M. Levelt [a]

[a] Max Planck Institute for Psycholinguistics, The Netherlands
[b] Nijmegen Institute for Cognition and Information and F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Spinoza Building B.02.32, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Received 26 January 2006; revised 18 October 2006; accepted 21 October 2006

Abstract

Disagreement exists about how speakers monitor their internal speech. Production-based accounts assume that self-monitoring mechanisms exist within the production system, whereas comprehension-based accounts assume that monitoring is achieved through the speech comprehension system. Comprehension-based accounts predict perception-specific effects, like the perceptual uniqueness-point effect, in the monitoring of internal speech. We ran an extensive experiment testing this prediction using internal phoneme monitoring and picture naming tasks. Our results show an effect of the perceptual uniqueness point of a word in internal phoneme monitoring in the absence of such an effect in picture naming. These results support comprehension-based accounts of the monitoring of internal speech.

Keywords: Self-monitoring; Speech production; Speech comprehension

☆ This manuscript was accepted under the editorship of Jacques Mehler.
* Corresponding author.
E-mail address: A.Roelofs@nici.ru.nl (A. Roelofs).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.10.006

1. Introduction

Speakers monitor their own speech for errors and appropriateness (e.g., Levelt, 1989). There exist different accounts of how this monitoring is achieved (for reviews, see Hartsuiker & Kolk, 2001; Postma, 2000). Probably all models of self-monitoring assume the existence of external monitoring, whereby the speaker monitors self-generated overt speech. This involves the normal speech comprehension process. Self-monitoring models also agree that, in addition, there exist mechanisms for the monitoring of the internal speech plan before it is articulated. However, the models make different claims about the functional locus of the internal monitoring device. One class of model assumes that the internal monitoring device is located inside the production system (e.g., Laver, 1973; Schlenk, Huber, & Wilmes, 1987). Another class of model assumes that internal monitoring is achieved via the speech comprehension system. Such an account has been developed by Levelt and colleagues (Levelt, 1983, 1989; Levelt, Roelofs, & Meyer, 1999; Roelofs, 2004), called the perceptual-loop theory of self-monitoring. According to Levelt et al. (1999), in planning spoken words, a phonological representation is constructed incrementally from the beginning of a word to its end. The phonological word representation is fed into the speech comprehension system as it becomes available over time. This results in sequential activation of the comprehension system, as is the case with the processing of real external speech. The comprehension system is then used to monitor the planned speech.

Because self-monitoring is achieved via the speech comprehension system according to the perceptual-loop theory, it predicts perception-specific effects on internal self-monitoring. One such perception-specific effect is the uniqueness point effect (e.g., Marslen-Wilson, 1990). The uniqueness point of a word is defined as the phoneme in the word where it diverges from all other words in the language, going from the beginning of the word to its end. The uniqueness point influences the speed of spoken word recognition. For example, Marslen-Wilson (1990) observed that listeners are faster in deciding that a spoken item is a word or not (auditory lexical decision) when the uniqueness point is early in a word than when it is late in a word. Moreover, in phoneme monitoring experiments, participants are faster in detecting a target phoneme in a spoken word when the phoneme follows the uniqueness point of the word than when it precedes the uniqueness point (Frauenfelder, Segui, & Dijkstra, 1990). Moreover, if the target phoneme follows the uniqueness point, phoneme monitoring is faster when the distance of the phoneme to the uniqueness point is long than when it is short (Frauenfelder et al., 1990). Whereas the uniqueness point of a word affects spoken word recognition, there is no evidence that suggests that it influences spoken word production.

We report an experiment that examined whether there are perception-specific effects in the monitoring of internal speech, as predicted by the perceptual-loop theory. Participants were presented with pictured objects and they indicated by pressing a button whether the picture name contained a pre-specified target phoneme. We manipulated the position of the target phonemes relative to the uniqueness point of the picture names. This was done to test the critical prediction of the perceptual-loop theory that monitoring latencies should depend on the distance of the phoneme from the uniqueness point of the picture name. Moreover, we manipulated the serial position of the target phoneme within the picture names in order to provide for a replication of the results of Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002), who observed that phonemes at the beginning of a word are detected faster than word-medial phonemes and word-final phonemes. Effects of uniqueness point and serial position should be present in the monitoring of internal speech but not in picture naming. In order to test the latter prediction, participants were also asked to name the pictured objects.

2. Materials

For testing the predictions concerning the uniqueness point, there were 30 critical pictures, all with disyllabic names ending in the target phonemes /l/ or /r/. The uniqueness points of the picture names were determined using a phonetic dictionary of Dutch (Heemskerk & Zonneveld, 2000). There were three distance conditions: no, short, and long. Each condition contained ten pictures, with five names ending in /l/ and five names ending in /r/. In the no-distance condition, the uniqueness point of the picture name was the word-final target phoneme itself, /l/ or /r/ (e.g., "kete*l*" – kettle). In the short-distance condition, the picture name became unique at the phoneme before the word-final target phoneme (e.g., "vog*e*l" – bird). In the long-distance condition, the picture name became unique two phonemes before the word-final target phoneme (e.g., "za*d*el" – saddle). The Appendix A lists the materials. The names in the different conditions were matched on frequency (CELEX database, Baayen, Piepenbrock, & Gulikers, 1995) and number of phonemes.

Additionally, there were 240 filler pictures. The names of 60 of those pictures contained the target phonemes (e.g., /l/) in word-initial ("*l*eraar" – teacher) or in word-medial position ("mo*l*en" – windmill) and 90 filler picture names did not contain the target phoneme. Moreover, there were 90 pictures with the target phonemes /k/, /n/, and /s/, appearing on 15 go- and 15 no-go trials each. The 45 go-trials varied with respect to the target phoneme's serial position. The target phoneme appeared in 15 words in initial position, in 15 words in medial position, and in 15 words in final position. For each serial position, there were five words for each of the three phonemes. The picture names in the different serial position conditions were matched on frequency and length. The Appendix A lists the materials.

3. Pretests

To avoid any confounds due to the between-items design, the pictures were evaluated in two pretests with respect to differences in ease of articulation onset and ease of recognition. Potential differences in ease of articulation were assessed by a delayed naming task, and differences in ease of picture recognition were assessed by a picture recognition task.

3.1. Delayed naming

To test for potential differences in ease of articulation onset, 20 participants (mean age: 21 years) performed a delayed naming task. Pictured objects were presented on a screen for 1 s before a go-signal (a tone). The delay between picture onset and go-signal was 1 s, which was enough to prepare the picture naming response until articulation. Thus, influences prior to articulation onset should not influence the picture naming latencies. Table 1 gives the results (the *SD*s are from the complete data matrix). Statistical analysis of the naming latencies revealed no effect of distance to the uniqueness point, $F_1(2,57) < 1$, $F_2(2,27) = 1.90$, $p = .17$, and no effect of serial position, Fs < 1. There were also no effects on the error rate. Thus, the critical materials do not differ in ease of articulation onset.

3.2. Picture recognition

To test for potential differences in ease of picture recognition, 20 new participants (mean age: 21 years) performed a picture recognition task. A spoken word was presented via headphones. Next, 0.5 s after word offset, a picture was presented on a screen for 1 s. Participants had to indicate by button press whether the word and object referred to the same entity or not. Table 1 gives the results. Statistical analysis of the recognition latencies showed that there was no effect of distance to the uniqueness point, Fs < 1, and no effect of serial position, $F_1 < 1$, $F_2(2,42) = 2.8$, $p = .07$. There were also no effects on the error rate. Thus, the critical pictures do not differ in ease of recognition.

4. Main experiment

4.1. Participants

Thirty-two native speakers of Dutch (mean age: 22 years) participated in the main experiment. None of them had participated in one of the pretests.

Table 1
Mean response latencies in milliseconds (*M*), standard deviations (*SD*), and error percentages (*E*%) in the delayed naming and picture recognition pretests

	Delayed naming			Picture recognition		
	M	*SD*	*E*%	*M*	*SD*	*E*%
Distance to uniqueness point						
No	611	235	3.7	535	153	4.0
Short	613	222	1.7	534	169	5.0
Long	579	223	1.9	550	207	5.5
Serial position						
Initial	576	222	1.6	535	170	4.3
Medial	588	217	1.4	557	183	5.0
Final	592	236	0.8	512	176	2.3

4.2. Design and procedure

The main experiment tested the effect of the crossed factors *task* (picture naming, phoneme monitoring) and *distance to uniqueness point* (no, short, long) and the crossed factors *task* (picture naming, phoneme monitoring) and *serial position* (initial, medial, final). Participants were tested individually in a quiet room. They first performed the picture naming task and then the phoneme monitoring task. Participants received written instructions to name the pictures as quickly and accurately as possible. The pictures were presented in a randomized order. Next, participants received written instructions for the phoneme monitoring task. They were asked to press a button with their dominant hand when the name of the pictured object contained the target phoneme. The target phoneme (/l/, /r/, /k/, /n/, and /s/) changed randomly from trial to trial. The internal phoneme monitoring session consisted of 10 practice trials followed by five experimental blocks containing 54 trials each. There were short breaks between the blocks. Together, the picture naming and phoneme monitoring sessions lasted about 40 min.

Each trial in the picture naming session had the following structure. A picture was presented and stayed on the screen for 1 s. A new picture appeared after 2.5 s. In the internal phoneme monitoring session, the target phoneme was displayed for 0.5 s by means of the corresponding letter (*l*, *r*, *k*, *n*, or *s*). After 1.5 s the picture appeared and stayed on the screen for 1 s. A new target phoneme was displayed after 2.5 s.

4.3. Results and discussion

Trials on which participants made a picture naming or phoneme monitoring error were discarded from the analysis of the response latencies. Table 2 gives the mean response latencies and error rates.

4.3.1. Serial position

Serial position had no effect on the latencies for picture naming, *F*s < 1, but it affected the phoneme monitoring latencies, $F_1(2,93) = 22.73$, $p < .001$,

Table 2
Mean response latencies in milliseconds (*M*), standard deviations (*SD*), and error percentages (*E*%) per manipulation (distance to uniqueness point, serial position) and task (phoneme monitoring, picture naming)

	Phoneme monitoring			Picture naming		
	M	*SD*	*E*%	*M*	*SD*	*E*%
Distance to uniqueness point						
No	960	250	2.8	827	220	2.4
Short	932	219	4.9	860	242	3.5
Long	898	218	4.4	863	231	2.8
Serial position						
Initial	846	219	2.1	915	247	2.9
Medial	1083	329	7.3	966	270	3.6
Final	1092	303	10.5	931	235	4.6

$F_2(2,42) = 28.54$, $p < .001$. Monitoring latencies were longer for medial than for initial phonemes, $t_1(31) = 9.36$, $p < .001$, $t_2(28) = 6.17$, $p < .001$, and longer for final than for initial phonemes, $t_1(31) = 10.51$, $p < .001$, $t_2(28) = 7.83$, $p < .001$, but latencies did not differ between medial and final phonemes, $t_1(31) = 1.15$, $p = .26$, $t_2(28) < 1$, $p = .60$. The analysis of the errors yielded main effects of task, $F_1(1,29) = 183.34$, $p < .001$, $F_2(1,11) = 152.51$, $p < .001$, and serial position, $F_1(2,58) = 19.86$, $p < .001$, $F_2(2,22) = 4.30$, $p = .03$. Moreover, there was an interaction of task and serial position, $F_1(2,58) = 21.15$, $p < .001$, $F_2(2,22) = 8.53$, $p = .002$. Serial position had an error effect in the phoneme monitoring task, $F_1(2,93) = 23.72$, $p < .001$, $F_2(2,42) = 9.25$, $p = .001$, but not in picture naming, Fs < 1. The serial position effect on the phoneme monitoring latencies observed here replicates Wheeldon and Levelt (1995) and Wheeldon and Morgan (2002).

4.3.2. Distance to uniqueness point

The statistical analysis yielded an effect of task, $F_1(1,31) = 18.42$, $p < .001$, $F_2(1,27) = 22.45$, $p < .001$, but no effect of distance to the uniqueness point, Fs < 1. However, task and distance interacted, $F_1(2,62) = 6.86$, $p = .002$, $F_2(2,26) = 3.45$, $p = .047$. For picture naming, there was no effect of distance to the uniqueness point, $F_1(2,93) = 1.84$, $p = .16$, $F_2(2,26) < 1$. However, for phoneme monitoring, there was an effect of the distance to the uniqueness point in the by-participant analysis, $F_1(2,93) = 4.37$, $p = .013$, $F_2(2,26) = 1.93$, $p = .17$. Monitoring latencies were shorter in the long-distance than in the no-distance condition, $t_1(31) = 3.18$, $p < .001$, $t_2(17) = 1.88$, $p < .04$, and monitoring latencies were shorter in the long-distance than in the short-distance condition in the by-participant analysis, $t_1(31) = 1.82$, $p < .04$, $t_2(17) < 1$, $p = .19$. There was no reliable difference in monitoring latency between the no-distance and short-distance conditions, $t_1(31) = 1.57$, $p = .06$, $t_2(17) = 1.16$, $p = .13$. There were no effects on the error rate, except for a main effect of task, $F_1(1,31) = 77.36$, $p < .001$, $F_2(1,27) = 71.55$, $p < .001$.

5. General discussion

Our results show an effect of distance to the uniqueness point in internal phoneme monitoring but not in picture naming. These results support the predictions of the perceptual-loop theory. The interaction of task and distance to the uniqueness point was significant in both the by-participant and by-item analyses, and the same held for the difference between the no-distance and long-distance conditions in the internal monitoring task. Thus, critical effects were significant both by participants and by items. Other effects were significant in the analyses by participants, but not in the by-item analyses. The absence of some of the effects in the item analysis is probably due to the between-items design. Having only ten items per condition reduces the power in the by-item analysis compared to the by-participant analysis. However, it was not possible to find more items meeting the constraints on the materials of the experiment. The trend found in the by-item analysis, however, goes into the same direction as the effects in the by-participant analysis. To conclude, we observed an effect of distance to the uniqueness point in internal phoneme monitoring but not in picture naming, which supports the perceptual-loop theory of self-monitoring.

Under ordinary circumstances, speakers monitor their speech for errors and appropriateness rather than for pre-specified target phonemes, as in the present experiment. This raises the question whether monitoring for errors and monitoring for phonemes in internal speech are accomplished via a single system. Perhaps the monitoring for phonemes is accomplished via the speech comprehension system, as suggested by the present experimental results, whereas the monitoring for errors is not. Note that this dual-system proposal is theoretically less parsimonious than assuming that the monitoring for errors and phonemes are accomplished via the same system. Thus, unless there is specific evidence that demands the division of labor, the single-system view is to be preferred. Moreover, the single-system view is supported by empirical evidence. Dell and Repka (1992) asked participants to detect speech errors in internally and externally generated tongue twisters. They observed that similar types of errors were reported in internal and external speech. However, in internal but not in external speech, the errors occurred predominantly at the beginning of words. This suggests that the participants internally monitored a representation that was incrementally generated from the beginning of a word to its end, and that the internal generation process was stopped after the detection of an error. The effects of serial position observed for internal error monitoring (Dell & Repka, 1992) and for phoneme monitoring in the present experiment suggest that the two types of monitoring are accomplished via the same system.

To conclude, we reported evidence that the uniqueness point of words influences phoneme monitoring in internal speech, but it does not influence picture naming. The most parsimonious explanation of this finding is that the monitoring of internal speech is accomplished via the speech–comprehension system, as maintained by the perceptual-loop theory of self-monitoring.

Appendix A

Target	Distance to uniqueness point		
	No	Short	Long
/l/	ketel (*kettle*)	bijbel (*bible*)	puzzel (*puzzle*)
	appel (*apple*)	wortel (*carrot*)	zadel (*saddle*)
	engel (*angel*)	kachel (*oven*)	stempel (*stamp*)
	tempel (*temple*)	trommel (*drum*)	deksel (*lid*)
	tafel (*table*)	vogel (*bird*)	spiegel (*mirror*)
/r/	kever (*beetle*)	masker (*mask*)	dokter (*doctor*)
	vlieger (*kite*)	boter (*butter*)	spijker (*nail*)
	kikker (*frog*)	koffer (*suitcase*)	wekker (*alarmclock*)
	tijger (*tiger*)	motor (*motorcycle*)	anker (*anchor*)
	halter (*bar-bell*)	vinger (*finger*)	ladder (*ladder*)

Appendix A (*continued*)

Target	Serial position		
	Initial	Medial	Final
/k/	kanon (*canon*)	beker (*cup*)	asbak (*ash tray*)
	kano (*canoe*)	fakkel (*torch*)	rugzak (*backpack*)
	kegel (*bowling pin*)	stekker (*plug*)	zwempak (*swimsuit*)
	ketting (*chain*)	bliksem (*lightning*)	monnik (*monk*)
	kassa (*cash desk*)	sikkel (*sickle*)	handdoek (*towel*)
/s/	sleutel (*key*)	pistool (*revolver*)	vleermuis (*bat*)
	schommel (*seesaw*)	kussen (*pillow*)	infuus (*infusion*)
	snavel (*pecker*)	passer (*pair of compasses*)	cactus (*cactus*)
	schotel (*satellite dish*)	borstel (*brush*)	vleugels (*wings*)
	sigaar (*cigar*)	hamster (*hamster*)	kompas (*compass*)
/n/	nijlpaard (*hippopotamus*)	gondel (*gondola*)	eekhoorn (*squirrel*)
	neushoorn (*rhinoceros*)	honing (*honey*)	citroen (*lemon*)
	nagel (*finger nail*)	printer (*printer*)	kuiken (*chick*)
	navel (*navel*)	magneet (*magnet*)	druiven (*grapes*)
	nijptang (*nippers*)	panda (*panda bear*)	glijbaan (*slide*)

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database. (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium.

Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 237–262). New York: Plenum Press.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance 16*, 77–91.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113–157.

Heemskerk, J. S. M., & Zonneveld, W. (2000). *Uitspraakwoordenboek*. Utrecht: Het Spektrum.

Laver, J. D. M. (1973). The detection and correction of slips of the tongue. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence*. The Hague: Mouton.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148–172). Cambridge: MIT Press.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*, 97–131.

Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: comment on Rapp and Goldrick (2000). *Psychological Review, 111*, 561–572.

Schlenk, K., Huber, W., & Wilmes, K. (1987). "Prepairs" and repairs: different monitoring functions in aphasic language production. *Brain and Language, 30*, 226–244.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34*, 311–334.
Wheeldon, L. R., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes, 17*, 503–535.