ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

POSITIVE AND NEGATIVE INFLUENCES OF THE LEXICON ON PHONEMIC DECISION-MAKING

James M. McQueen, *Anne Cutler* and *Dennis Norris**

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
*MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

ABSTRACT

Lexical knowledge influences how human listeners make decisions about speech sounds. Positive lexical effects (faster responses to target sounds in words than in nonwords) are robust across several laboratory tasks, while negative effects (slower responses to targets in more word-like nonwords than in less word-like nonwords) have been found in phonetic decision tasks but not phoneme monitoring tasks. The present experiments tested whether negative lexical effects are therefore a task-specific consequence of the forced choice required in phonetic decision. We compared phoneme monitoring and phonetic decision performance using the same Dutch materials in each task. In both experiments there were positive lexical effects, but no negative lexical effects. We observe that in all studies showing negative lexical effects, the materials were made by cross-splicing, which meant that they contained perceptual evidence supporting the lexically-consistent phonemes. Lexical knowledge seems to influence phonemic decision-making only when there is evidence for the lexically-consistent phoneme in the speech signal.

1. INTRODUCTION

Lexical knowledge exerts both positive and negative effects on how human listeners make decisions about speech sounds. Positive lexical effects appear, for example, in the phoneme monitoring task (subjects detect target sounds faster in words than in nonwords [1]) and in the phonetic decision task (subjects decide which of two alternative sounds they have heard faster in a word than in a nonword [2]). Positive lexical effects can be accounted for by models of spoken word recognition in which there is feedback from the lexicon to prelexical acoustic-phonetic processing (e.g., TRACE [3]) and by models without feedback, in which the lexicon influences phonemic decisions at a postlexical decision stage (e.g., Merge [4,5]).

Negative lexical effects, however, are harder to obtain. McQueen, Norris and Cutler [4], for example, found that Dutch listeners were slower to decide that the final sound of the Dutch nonword *smep* was a /p/ when the nonword was made by cross-splicing the [smɛ] from the real Dutch word *smet* (stain) with the final [p] from *smep* than when the nonword was made by cross-splicing the [smɛ] from the Dutch nonword *smek* with the final [p] from *smep*. In this case, lexical knowledge (the fact that *smet* is a real word) interferes with the phonemic decision process. McQueen et al. [4] found no such inhibitory effect, however, in a phoneme monitoring task with the same materials. Frauenfelder et al. [1] also failed to find an inhibitory lexical effect in phoneme monitoring: French listeners were able to detect the phoneme target /t/ just as quickly in the nonword *vocabutaire* as in the matched nonword *socabutaire*, in spite of the fact that lexical knowledge (the real French word *vocabulaire*) could have interfered with detection of /t/ in *vocabutaire*. Negative lexical effects have therefore been obtained with the phonetic decision task but not with the phoneme monitoring task.

Negative lexical effects are of particular importance in the debate on feedback in spoken word recognition. Models with feedback predict that there should be positive as well as negative effects of the lexicon on prelexical processing; the absence of such effects has therefore been taken by some [1] as a challenge to models like TRACE [3] which have feedback. On the other hand, models without feedback in which all phonemic decisions to sounds in nonwords are made without lexical involvement (e.g., the Race model [6]) are challenged by the existence of negative lexical effects (because they are effects with nonwords). In fact, the demonstration of negative lexical effects was one of the primary motivations for the new Merge model [4,5]. The present experiments were therefore designed to establish why negative lexical effects have been observed in phonetic decision but not in phoneme monitoring.

We sought to distinguish between two explanations for the failures to observe negative effects in phoneme monitoring. The first is that inhibitory lexical effects may depend on the presence of acoustic-phonetic evidence in support of the lexically-consistent phoneme. In *smep* made by cross-splicing with *smet*, for example, there were strong place of articulation cues signaling a final alveolar consonant. Given bottom-up evidence for a final /t/, the lexicon could then act to bias phonemic decision-making in favor of /t/, thus slowing down the /p/ decision. But the materials in the *vocabutaire* experiments were not made by cross-splicing, so, for example, there was no evidence in favor of /l/ in *vocabutaire*, and thus no inhibitory lexical effect. The other possibility, however, is that negative lexical effects in phonetic decision are simply a consequence of the two-alternative forced choice required in that task. The lexicon might produce an inhibitory effect when there is competition between two response alternatives, but not when there is only one possible response, as in phoneme monitoring. If so, negative lexical effects would reflect task-specific processes, and would therefore be less informative about the mechanisms of speech perception.

We compared phoneme monitoring and phonetic decision performance using exactly the same materials. In Experiment 1, Dutch listeners monitored for target phonemes in Dutch words, and in three types of matched nonword. Subjects monitored for /p/, for example, in the base word *sinaasappel* (orange) and in the control nonword *minaasappel* (differing only in the initial phoneme), and for /k/ in two other matched nonwords *sinaasakkel* (the word-like nonword, differing only on the target phoneme) and its control nonword *minaasakkel* (differing on initial and target phonemes). In Experiment 2, Dutch listeners made phonetic decisions to the same materials. Their task was to decide which of two phonemes occurred in each item.

If negative lexical effects are a task-specific consequence of the forced choice required in phonetic decision, then we would expect a negative lexical effect in Experiment 2 but not in Experiment 1. That is, we would expect slower decisions to /k/ in *sinaasakkel* than in *minaasakkel* due to the activation of the word *sinaasappel* and the resulting competition at the decision stage between /p/ and /k/ in Experiment 2. In Experiment 1, however, where there is no response competition between /p/ and /k/, we would expect phoneme monitoring responses to /k/ in *sinaasakkel* to be just as fast as those to /k/ in *minaasakkel*. If, on the other hand, negative lexical effects depend on evidence in the speech signal supporting the lexically-consistent phoneme, we would expect no negative lexical effects in either experiment, since the materials were natural utterances and involved no cross-splicing. We predicted however that in both tasks there would be positive lexical effects, that is, faster responses to targets in words (e.g., /p/ in *sinaasappel*) than in nonwords (/p/ in *minaasappel*). All target phonemes appeared after the words' Uniqueness Points (UPs; the UP is the point moving left-to-right through a word at which the sequence becomes consistent with only that word). Strong positive lexical effects have been observed for post-UP target phonemes [1].

2. EXPERIMENT 1: PHONEME MONITORING

2.1. Method

Subjects. 48 volunteers from the Max-Planck-Institute subject pool, all native speakers of Dutch, were paid for their participation.

Materials. Thirty Dutch base words with either three or four syllables were chosen. Ten words had a /p/ after the word's UP, contained no other /p/'s, and no /k/'s: *sinaasappel, maatschappij, chimpansee, envelop, europa, olympiade, utopie, antilope, ethiopie, boodschapper*. For each base word, three nonwords were constructed. To make the control nonword matched to the base word, the initial phoneme was replaced with a phoneme differing by several distinctive features (e.g., for the base word *sinaasappel* the control nonword was *minaasappel*). To make the word-like nonwords in each set, the target phoneme was changed from /p/ to /k/ (e.g., *sinaasakkel*), and to make the controls for the word-like nonwords, the target phoneme was changed from /p/ to /k/ and the initial phoneme was changed in the same way as in the other control nonwords (e.g., *minaasakkel*). These items were used in blocks where listeners monitored for either /p/ or /k/. A further ten base words were selected using similar constraints for use in blocks where listeners monitored for either /t/ or /p/ (i.e., the word-like nonwords were made by substituting /t/ with /p/, etc.): *schrijfster, suggestie, minister, resolutie, kabouter, evolutie, informatie, augustus, commentaar, objectief*. The final ten base words were selected using similar constraints for use in blocks where listeners monitored for either /k/ or /p/: *logica, marokko, mirakel, musicus, fabrikant, historicus, risico, vaticaan, inbreker, nootmuskaat*.

Within each block, the ten quadruplets were counterbalanced over two lists, such that in each list there were five base words and the control nonwords for their matched word-like nonwords (e.g., *sinaasappel* and *minaasakkel* in one list) plus the word-like nonwords and the control nonwords for the other five base words (e.g., *sinaasakkel* and *minaasappel* in the other list). In each list, therefore, there were 5 base words and 15 matched nonwords. In each block, half the subjects heard one list and the others heard the other list, and half monitored for each of the two phonemes (e.g., in the /p/-/k/ block, half monitored for /p/ and half for /k/). Thus, no subject heard both a base word and its control nonword, nor both a word-like nonword and its control. Each subject was assigned to the same conditions across all three blocks, such that half of the subjects always had to detect targets in base words and the appropriate control nonwords (to test for positive lexical effects) while the other subjects always had to detect targets in word-like nonwords and the appropriate controls (to test for negative lexical effects).

The same 92 fillers were added to each list within each block. Thus, for example, in the /p/-/k/ block, there were 11 filler words and 11 filler nonwords with /p/ (and no /k/, making 16 words and 16 nonwords in all with the target /p/), 16 filler words and 6 filler nonwords with /k/ (and no /p/, making 32 items in all with the target /k/), and 24 words and 24 nonwords containing neither /p/ nor /k/. For each target phoneme within a block, there were ten words and ten nonwords with medial targets (including the experimental items), and, in the fillers, three words and three nonwords with initial targets and three words and three nonwords with final targets. The proportions of targets in different positions match those in the earlier French experiment [1].

Procedure. The materials were recorded onto DAT tape by a female native speaker of Dutch. Using a speech editor, individual files for each item were made, and the duration from item onset to onset of the target phoneme in each item were measured. The items were played to listeners over closed-ear headphones in random order in three blocks. Order of presentation of the blocks was rotated across subjects.

Each block of trials began with a specification of what the target sound would be for that block, presented in upper-case letters on a computer screen in front of the subjects (P, T or K). Listeners were told that they would then hear a list of words and nonwords. They were told to press the button in front of them as fast as possible if the word or nonword contained the target sound. A warning tone preceded each item. Reaction Times (RTs) were measured from the acoustic onset of each item, and adjusted prior to statistical analysis so as to measure from target phoneme onset. Before the three experimental blocks, listeners performed a short practice block, in which the target phoneme was /s/.

2.2. Results and Discussion

The results are shown in Table 1. All responses faster than 150 ms or slower than 1500 ms after target onset were excluded (2.6% of the data). The data from three quadruplets of items were also excluded from the analysis (one item in each of two sets had been mispronounced; in the third set one item had an error rate over 40%). The remaining data were submitted to Analyses of Variance (ANOVAs) with either subjects (F1) or items (F2) as random factors. In the RT analyses, missing data were replaced with the mean of the available data points for that listener in that condition (or, in the items analyses, for that item in that condition). ANOVAs were performed separately for the comparison of base words with their matched nonwords (to test for positive lexical effects) and for the comparison of word-like nonwords with their matched nonwords (to test for negative lexical effects).

Table 1: Experiment 1: Phoneme monitoring. Reaction Times (RTs, measured from target onset) for correct target detection; proportion of missed targets (Errors); and proportion of false alarms.

	Base word (*sinaasappel*)	Control nonword (*minaasappel*)	Word-like nonword (*sinaasakkel*)	Control nonword (*minaasakkel*)
RTs (ms)	485	616	549	576
Errors (%)	4.4	7.5	3.9	7.2
False Alarms (%)	3.9	7.7	11.5	6.8

There was a positive lexical effect in RTs. Listeners detected targets like /p/ in *sinaasappel* faster than in *minaasappel* (by 131 ms, on average: $F1(1,22) = 78.44$, $p<0.001$; $F2(1,24) = 21.49$, $p<0.001$). The positive lexical effect in error rates was not significant. There were no negative lexical effects. Listeners in fact detected /k/ in *sinaasakkel* slightly faster (by 27 ms, on average) than in *minaasakkel* (a non-significant difference). They were also slightly more accurate (by 3.3%, on average) on targets in word-like nonwords than in the matched control nonword. This difference was significant by subjects only: $F1(1,22) = 5.09$, $p<0.05$; $F2(1,24) = 2.86$, $p=0.1$. No other main effects or interactions were reliable across subjects and items in any of the RT or error analyses.

The false alarm rates were also examined. Listeners monitoring for /p/, for example, could have responded when they heard *sinaasakkel*, as if they had heard the real word *sinaasappel*. The tendency to false alarm more on *sinaasakkel* than on *minaasakkel* (4.7% more, on average) was marginal by subjects ($F1(1,22) = 3.98$, $p=0.06$) but not significant by items ($F2(1,24) = 1.78$, $p>0.15$). There was also a slight tendency to be more accurate on real words than on their matched controls (e.g., fewer /k/ false alarms on *sinaasappel* than on *minaasappel*), but this effect was again only significant by subjects ($F1(1,22) = 5.99$, $p<0.05$; $F2(1,24) = 1.86$, $p>0.15$).

The present results replicate earlier results in French [1], showing a positive but no negative lexical effect in phoneme monitoring. Listeners showed a strong and reliable positive lexical effect, with faster responses to targets in words than to targets in nonwords. They also showed a slight tendency to find monitoring easier in the word-like nonwords than in the control nonwords. Instead of a negative lexical effect, with inhibition from the base words, there was a hint of facilitation in the word-like nonwords.

A stronger demonstration of this facilitation has been described by Wurm and Samuel [7]. In a closely related phoneme monitoring study (with some procedural differences from Experiment 1 and Frauenfelder et al. [1]), Wurm and Samuel found significantly faster responses to targets in word-like nonwords than in control nonwords which were not like any real words. They argued that this facilitation was due to attentional factors which make more word-like nonwords easier to process. Such factors may thus make it harder to observe negative lexical effects in phoneme monitoring. In Experiment 2, however, we asked whether negative lexical effects could be observed with the same materials as in Experiment 1, but using the task which has already produced direct evidence of negative lexical effects: the phonetic decision task [2,4].

3. EXPERIMENT 2: PHONETIC DECISION

3.1. Method

Subjects. 26 native speakers of Dutch, volunteers from the Max-Planck-Institute subject pool, were paid for their participation.

Materials, Design and Procedure. The same 120 experimental items (quadruplets of 30 base words and their three matched nonwords) were used. Since subjects were required to make a two-alternative phonetic decision on every trial, only fillers with target phonemes were used (44 from each of the three Experiment 1 blocks). The quadruplets were again counterbalanced over two lists, such that for a given set the base word and control nonword for the word-like nonword (e.g., *sinaasappel* and *minaasakkel*) appeared in one list, while the word-like nonword and the other control nonword (e.g., *sinaasakkel* and *minaasappel*) appeared in the other list. In contrast to Experiment 1, however, the lists were not presented in separate blocks, with different phoneme targets in each block. Instead, there were two lists of 192 items (60 experimental items and 132 fillers), in random order.

The two appropriate response alternatives (P K, P T, or T K) were presented on a computer screen before each trial. The alternatives were presented as uppercase letters on the left and right of the screen, together with an auditory warning tone. The P always appeared on the left and the K always on the right; only the position of the T varied. Subjects were told that these alternatives corresponded to the sounds /p/, /t/, and /k/, that they would hear a word or a nonword over headphones after each visual presentation of alternatives, and that their task was decide which of the two sounds was present in that word or nonword. They responded by pressing the appropriate response button, positioned below the letters on the screen, as fast as possible. Half the subjects heard one list, and the other half heard the other list. Testing began with a short practice block.

3.2. Results and Discussion

The results are shown in Table 2. All responses outside the 150-1500 ms range were again excluded (4.0% of the data). The data from the same three quadruplets of items that were excluded in Experiment 1 were again excluded here. Similar ANOVAs to those in Experiment 1 were then performed, using the same procedure for replacement of missing data in the RT analyses.

Table 2: Experiment 2: Phonetic decision. Reaction Times (RTs, measured from target onset) for correct target decisions; and proportion of missed targets (Errors).

	Base word (*sinaasappel*)	Control nonword (*minaasappel*)	Word-like nonword (*sinaasakkel*)	Control nonword (*minaasakkel*)
RTs (ms)	557	631	648	662
Errors (%)	5.6	9.4	11.7	12.0

There was again a robust positive lexical effect. Listeners decided faster that there was a /p/ (and not a /k/), in *sinaasappel* than in *minaasappel* (mean difference of 74 ms; $F1(1,24) = 20.90$, $p<0.001$; $F2(1,24) = 11.34$, $p<0.005$). The small advantage for words over nonwords in error rates was only significant by subjects: $F1(1,24) = 6.22$, $p<0.05$; $F2(1,24) = 2.79$, $p=0.1$. There was again no negative lexical effect. As in Experiment 1, listeners were, if anything, faster and more accurate to decide that there was a /k/ (and not a /p/), in *sinaasakkel* than in *minaasakkel*. The mean differences of 14 ms and 0.3% were however not significant. The results of Experiment 2 therefore closely replicate those of Experiment 1.

4. GENERAL DISCUSSION

The results were thus the same in each task. They suggest that even though lexical knowledge (e.g., about the word *sinaasappel*) benefited decisions in either task about lexically-consistent phonemes (/p/ in *sinaasappel*) it did not hinder decisions about lexically-inconsistent phonemes (/k/ in *sinaasakkel*). The absence of negative lexical effects in Experiment 2 suggests that such effects do not reflect response-based competition in phonetic decision. Instead, it would appear that negative lexical effects depend on the availability of evidence in the speech signal in support of lexically-consistent phonemes. In studies which have found clear negative lexical effects [2,4], the materials were made by cross-splicing, so indeed contained such evidence. In *smep* made by cross-splicing with *smet*, for example, there were cues in the vowel consistent with a final /t/ [4]. The present materials, however, were not made by cross-splicing. There was therefore no evidence for /p/ in *sinaasakkel*. But there was of course evidence for /p/ in *sinaasappel*.

We therefore propose that lexical knowledge influences phonemic decision-making (either positively or negatively) only when the speech signal contains evidence for the lexically-consistent phoneme. This claim is consistent with the bottom-up priority rule built into the Merge model [5]. In Merge, lexical activation is only able to influence a phoneme decision unit if that unit has been activated bottom-up by the speech input.

McQueen et al. [4], however, found that there was no negative lexical effect in phoneme monitoring, even with cross-spliced materials. This may be because, although cross-splicing is necessary to generate a negative lexical effect, it is not sufficient. If a task is too easy, listeners can ignore all lexical information [4,5]. Phoneme monitoring is certainly an easier task than phonetic decision. As shown in [4], even in phonetic decision, negative lexical effects could only be obtained with a more difficult version of that task. In Merge, lexical knowledge can only influence phonemic decision-making if task demands encourage the use of lexical knowledge, and only then if the bottom-up priority rule is satisfied. The Merge model can therefore explain both positive and negative lexical effects in phonemic decision-making [5]. It does so while instantiating three important claims about human speech perception: that there is no feedback from the lexicon to earlier stages of processing; that listeners' use of lexical knowledge in phonemic decision-making depends on task demands; and that the lexicon can only influence decisions about phonemes for which there is already evidence in the speech signal.

5. ACKNOWLEDGEMENTS

We thank Mark Scholten for his help in preparing and running these experiments. Correspondence should be addressed to: James M. McQueen, MPI for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands; james.mcqueen@mpi.nl.

6. REFERENCES

1. Frauenfelder, U.H., Segui, J., and Dijkstra, T. "Lexical effects in phonemic processing: Facilitatory or inhibitory?", *J. Exp. Psych.: Human Percept. Perf.*, Vol. 16, 1990, 77-91.

2. Marslen-Wilson, W., and Warren, P. "Levels of perceptual representation and process in lexical access: words, phonemes, and features", *Psych. Rev.*, Vol. 101, 1994, 653-675.

3. McClelland, J.L., and Elman, J.L. "The TRACE model of speech perception", *Cog. Psych.*, Vol. 18, 1986, 1-86.

4. McQueen, J.M., Norris, D., and Cutler, A. "Lexical influence in phonetic decision making: Evidence from subcategorical mismatches", *J. Exp. Psych.: Human Percept. Perf.*, Vol. 25, 1999, 1363-1389.

5. Norris, D., McQueen, J.M., and Cutler, A. "Merging information in speech recognition: Feedback is never necessary", *Behav. Brain Sci.*, Vol. 23, 2000, in press.

6. Cutler, A., and Norris, D. "Monitoring sentence comprehension", in: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, ed. W.E. Cooper and E.C.T. Walker (pp. 113-134), Erlbaum, Hillsdale NJ, 1979.

7. Wurm, L.H., and Samuel, A.G. "Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring", *J. Mem. Lang.*, Vol. 36, 1997, 165-187.