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

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

DO AMBIGUOUS FRICATIVES RHYME? LEXICAL INVOLVEMENT IN PHONETIC DECISION-MAKING DEPENDS ON TASK DEMANDS

James M. McQueen

Max-Planck-Institute for Psycholinguistics
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

ABSTRACT

Lexical involvement in tasks which require phonetic decisions depends upon attentional factors. Attention can be manipulated by varying task demands. Tasks such as phonetic categorization may focus attention on acoustic-phonetic information but tasks such as rhyme decision may encourage use of the lexicon. An experiment is presented in which listeners made rhyme decisions about words and nonwords ending in ambiguous fricatives. Lexical effects were found, even though the materials had previously been used in a categorization experiment where no reliable effects were observed. The results suggest that the demands of the rhyme task do encourage lexical involvement. Attentional shifts due to task demands are more readily accounted for by autonomous than by interactive models of spoken word recognition.

I. INTRODUCTION

If listeners are asked to categorize a continuum of speech sounds (e.g. /s/-/ʃ/) embedded in contexts where only one endpoint forms a word (e.g. *kiss-kish* and *fiss-fish*), they tend to label the most ambiguous sounds in a lexically-consistent way, that is, they categorize ambiguous fricatives (/?/) in the middle of the continuum as /s/ in a /kɪ?/ context and as /ʃ/ in a /fɪ?/ context [1,2,3]. This demonstration of the use of lexical knowledge, along with data from other phonetic judgement tasks showing lexical effects, has been taken as support for interactive models of spoken word recognition, such as TRACE [4]. In this model, top-down connections from the lexicon influence the behaviour of prelexical phoneme detectors. Lexical involvement in tasks requiring phonetic decisions is attributed to these top-down connections. But lexical effects can also be explained by autonomous models with no top-down flow of information, such as the Race model [5]. On this account, lexical effects are due to a race between two procedures: a prelexical procedure, where phonetic decisions are based on prelexical processing; and a lexical procedure, where decisions are based on phonological information stored in the lexicon.

Most of the available data on lexical effects can be explained by both models. One result which is problematic for the interactive TRACE account, however, is the finding in word-final categorization (such as to the /s/-/ʃ/ continuum in *kiss-kish* and *fiss-fish* contexts) that lexical effects are largest in the subjects' fastest responses [2,3]. In TRACE, lexical activation, and hence top-down feedback, builds up over time. The model therefore incorrectly predicts that lexical involvement should be more pronounced for slower responses. In the autonomous Race model, where it can be assumed that the lexical procedure is faster than the prelexical procedure for word-final categorization, lexical effects are correctly predicted to be large for fast responses.

When subjects are asked to identify the initial phonemes of target words and nonwords presented at the end of a neutral phrase, there is no Reaction Time (RT) advantage for word over nonword responses [6]. But when subjects also have to perform a concurrent lexical task (e.g. deciding whether the targets are words or nonwords), reliable lexical effects emerge in the phonetic task [6]. It has also been shown that lexical effects in phoneme monitoring are absent in sentence contexts, even with a concurrent lexical task; they are present in random word lists, but then only with the concurrent task [7]. In sentences, subjects seem to attend to sentential-level processing to make lexical decisions, and to prelexical processing for phonetic decisions. But in meaningless contexts, when subjects must use lexical information in a concurrent task, their attention can be focussed on the lexicon for phoneme monitoring as well [7].

These results, and other attentional effects on lexical involvement in phonetic decision-making, are more in keeping with the Race model than with TRACE [6,7,8]. In the Race model, which has two different procedures for phonetic decisions, attentional shifts can easily be modelled in terms of attention being focussed on the output of one or other procedure. In TRACE, on the other hand, attentional effects must be due to a modification of the amount of top-down facilitation. Attention has to operate as a kind of gain control [8]. This is a much more indirect explanation than that provided by the Race model.

The present experiment tests further whether attentional focus on lexical knowledge, and hence the presence of lexical effects, depends on task demands. Attention was manipulated not by varying a concurrent task, but simply by changing the phonetic task. Do lexical effects on the same materials come and go depending on the task that subjects are asked to perform?

Lexical effects in phonetic categorization are not reliable [2,3]. Lexical involvement in /s/-/ʃ/ categorization of *kiss-kish* and *fiss-fish* continua was only present when the materials were low-pass filtered [2]. Several factors, including stimulus quality, type of phonetic continuum, and position of target phoneme, all contribute to the size and reliability of lexical effects in categorization [3]. In contrast, lexical effects in rhyme-judgement tasks are very robust. Subjects reliably detect words which rhyme with a preceding cue faster than rhyming nonwords [9].

This difference between tasks may be due to different attentional demands. Phonetic categorization requires the listener to attend to the acoustic-phonetic details of specific phonemes, and hence may tend to focus attention on the speech signal. Rhyme judgement, however, requires the listener to make decisions about strings of speech sounds, so may tend to focus attention on stored knowledge about such strings, that is, on the lexicon.

The following experiment addressed this hypothesis. In a rhyme-judgement task, listeners were presented with the ambiguous /s/-/ʃ/ materials from McQueen [2] which failed to yield any lexical effects in phonetic categorization, along with many phonetically-unambiguous words and nonwords. Lexical involvement in this situation would be indicated by a tendency to accept /ʔ/-final materials as rhymes more easily when they form rhyming words (e.g. /kɪʔ/ rhyming with the cue *miss* as the word *kiss*) than when they form rhyming nonwords (e.g. /fɪʔ/ rhyming with the cue *hiss* as the nonword *fiss*). If lexical knowledge is used more extensively in rhyme judgements than in categorization, lexical effects should be detectable in rhyme decisions on items that did not give lexical effects in categorization. For example, subjects should prefer to accept /kɪʔ/ as the word rhyme *kiss* than to accept /fɪʔ/ as the nonword rhyme *fiss*, although in categorization /ʔ/ was recognized equally often as /s/ and /ʃ/ in both the /kɪʔ/ and /fɪʔ/ contexts.

II. METHOD

2.1 Materials and Stimulus Construction.

Four pairs of fricative-final words were selected as targets: *kiss* and *fish*; *press* and *fresh*; *fuss* and *brush*; and *moss* and *squash*. These pairs formed the basis of four sets of targets. Each word target was matched with a nonword target with the opposite final fricative (e.g. *kiss* with *kish*, and *fish* with *fiss*), making 16 unambiguous targets in all, eight words and eight nonwords. These targets were also matched with targets ending in the ambiguous fricative (e.g. *kiss* and *kish* with /kɪʔ/, and *fish* and *fiss* with /fɪʔ/). Sixteen cue words were then selected, each of which was paired with two targets, one unambiguous, and one where the final fricative was replaced with /ʔ/ (e.g. the cue *miss* with the targets *kiss* and /kɪʔ/, *wish* with *fish* and /fɪʔ/, *dish* with *kish* and /kɪʔ/, and *hiss* with *fiss* and /fɪʔ/).

The stimuli for the *kiss/fish* and *press/fresh* sets were taken directly from Experiment 1 in [2], using fricatives 1 (/ʃ/), 5 (/ʔ/) and 9 (/s/) from the continuum made there (5 was the most ambiguous; see Figure 1 in [2]). For the *fuss/brush* set, the items *fush* and *brush* were recorded, and, following the procedure from [2], the /brʌ/ and /fʌ/ segments were spliced onto fricatives 1, 5 and 9 using a speech editor. The *moss/squash* items were taken from the practice block of [2], again using the same three fricatives.

Each subject heard all of the fricative-final materials, including repeated presentation of the matched target strings, once with an unambiguous fricative, and once with the ambiguous fricative. These repetitions required two controls. First, a set of "repetition targets" was made, comprising 16 cues matched with 16 rhyming targets, eight words and eight nonwords, all of which were monosyllables. These cue-target pairs were presented twice, for comparison with the fricative-final targets. The second control involved counterbalancing the order of presentation of the fricative-final targets.

Filler materials consisted of a further 192 additional rhyming monosyllabic targets, 96 words and 96 nonwords, and 256 nonrhyming monosyllables, 128 words and 128 nonwords. The additional targets were paired with 96 cues (as with the other targets each cue was used twice). Across the entire experiment, the number of words and nonwords was balanced, as was the number of items requiring positive and negative rhyme decisions.

2.2 Procedure.

The rhyme decision task was used [9]. Each cue was presented visually, and was followed by two auditory items. A rhyme decision was required on both items in each pair. Targets and fillers were equally likely as the first or second item in a pair, as were words and nonwords. Cue words were presented on the screen of a microcomputer, and auditory items were presented binaurally, over headphones. Subjects were asked to decide if the auditory words and nonwords rhymed with the preceding visual cue word. They were asked to make a decision to both members of every pair, and to respond with a button press as quickly and as accurately as possible, making the response to the first member of each pair before the second one was heard. A key labelled "YES" was held in the preferred hand, and another labelled "NO" in the other hand. RTs were measured from vowel onset in each target.

2.3 Subjects.

Student volunteers from St. John's College, Cambridge, were paid to take part. There were 15 females and 21 males, aged between 17 and 24 years.

III. RESULTS

3.1 Filler Targets.

As with the targets in [9], the filler targets were not repeated. They therefore allow for the most direct comparison with the earlier study. The lexical effect found in that experiment was replicated. Decisions to rhyming words (mean RT, 624 msec) were reliably faster than those to rhyming nonwords (mean RT, 669 msec). This 45 msec difference was reliable: by subjects, $F_1(1,32) = 100.52$, $p<0.001$; and by items, $F_2(1,95) = 21.64$, $p<0.001$.

3.2 Repeated and Unambiguous Fricative-final Targets.

Lexical effects were also measured in several analyses of the repeated and unambiguous fricative-final targets. The means are given in Table 1. In a comparison of the fricative-final targets with the first presentation of the repeated targets, word responses were, on average, 44 msec faster than nonword responses. This difference was reliable: $F_1(1,32) = 11.52$, $p<0.005$; $F_2(1,6) = 25.27$,

	Words	Nonwords	Mean
Unambiguous Fricative-final Targets	(kiss) 653	(fiss) 722	688
Repeated Targets	(cheat)	(speet)	
1st presentation:	728	748	738
2nd presentation:	645	690	668

Table 1. Mean RTs (msec) for positive responses to the unambiguous fricative-final targets and to the two presentations of the repeated targets, with examples. The means are given for word and nonword responses, and overall.

	Words		Nonwords	
Unambiguous				
/s/-final	100	(kiss)	88	(fiss)
/ʃ/-final	97	(fish)	97	(kish)
Overall	99		92	
Ambiguous (/ʔ/-final)				
Rhyming with /s/	65	(/kɪʔ/)	41	(/fɪʔ/)
Rhyming with /ʃ/	81	(/fɪʔ/)	81	(/kɪʔ/)
Overall	73		61	

Table 2. Proportion of positive rhyme decisions (%) to fricative-final targets, with examples. The responses to /ʔ/-final items are shown separately for decisions made after /s/-final cues (rhyming with /s/) and decisions made after /ʃ/-final cues (rhyming with /ʃ/).

$p<0.005$. Responses to e.g. *kiss* and *cheat* were faster than responses to *fiss* and *speet*. Similarly, word responses were 57 msec faster than nonword responses, on average, in a comparison of unambiguous fricative-final targets and the second presentation of the repeated targets. This difference was again statistically significant: $F_1(1,32) = 23.84$, $p<0.001$; $F_2(1,6) = 30.65$, $p<0.005$. The lexical effect was also reliable in a comparison between the first and second presentations of the repeated targets. The advantage of word responses over nonword responses (32 msec, on average) was significant: $F_1(1,32) = 5.82$, $p<0.05$; $F_2(1,3) = 39.21$, $p<0.01$. This analysis also revealed a reliable repetition effect. Rhyme decisions to the second presentation of the repeated targets were, on average, 70 msec faster than to the first presentation: $F_1(1,32) = 40.56, p<0.001$; $F_2(1,3) = 11.26, p<0.05$.

3.3 Fricative-final Targets.

Table 2 shows the proportions of yes responses to fricative-final targets. There were significantly more yes responses to the unambiguous fricative-final items (e.g. *kiss*) than to the /ʔ/-final items (e.g. *kiʔ*): a simple ambiguity effect, $t(35) = 7.21, p<0.001$. There was also a reliable lexical effect in the unambiguous targets, with more correct responses to words *(kiss)* than to nonwords *(fiss)*: $t(35) = 4.31$, $p<0.001$. Importantly, there was a lexical effect in the ambiguous targets: there were more yes responses to the /ʔ/-final words (e.g. /kɪʔ/ after *miss*) than to the /ʔ/-final nonwords (e.g. /fɪʔ/ after *hiss*): $t(35) = 3.46$, $p<0.005$. Subjects were more likely to accept a string with an ambiguous final fricative as a rhyme when that rhyme formed a word than when it formed a nonword.

The lexical effects were asymmetric (see Table 2). There were more yes responses to unambiguous /s/-final words (*kiss*) than to unambiguous /s/-final nonwords (*fiss*) ($t(35) = 4.58$, $p<0.001$), but no differences for unambiguous /ʃ/-final words and nonwords (*fish vs. kish*). Similarly, there were more yes responses to /ʔ/-final words following /s/-final cues (/kɪʔ/ after *miss*) than to /ʔ/-final nonwords following /s/-final cues (/fɪʔ/ after *hiss*): $t(35) = 4.57$, $p<0.001$, but no differences for /ʔ/-final words and nonwords following /ʃ/-final cues. The lexical effect was present only for responses following /s/-final cues.

This bias was also analysed by comparing performance on the same /ʔ/-final strings in different contexts. There were more yes responses to the /ʔ/-final items which were taken to rhyme as words following /ʃ/-final cues (e.g. /fɪʔ/ after *wish*) than to the same items which were taken to rhyme as nonwords following /s/-final cues (e.g. /fɪʔ/ after *hiss*): $t(35) = 5.66$, $p<0.001$ (81% versus 41%). For the other items, the reverse was true: there were more yes responses to the /ʔ/-final items which were taken to rhyme as *nonwords* following /s/-final cues (e.g. /kɪʔ/ after *dish*) than to the same items which were taken to rhyme as words following /ʃ/-final cues (e.g. /kɪʔ/ after *miss*): $t(35) = 2.50, p<0.05$ (65% versus 81%).

The asymmetry in the lexical effect is due to a general palatal bias. The differences between unambiguous /s/-final and /ʃ/-final items were reliable: overall, the items ending in /s/ were less likely to be taken to be rhymes than those ending in /ʃ/ ($t(35) = 2.05$, $p<0.05$). Likewise, targets ending with the ambiguous phoneme /ʔ/ were more often considered to be rhymes of preceding /ʃ/-final cues than to be rhymes of /s/-final cues ($t(35) = 4.82, p<0.001$).

The mean RTs of yes responses to fricative-final stimuli were also analyzed, and are given in Table 3 (the means are different from those in Table 1 since only those subjects who had at least one response in each of the /ʔ/-final conditions were included in this analysis). There was again a reliable lexical effect. Word responses were significantly faster than nonword responses (67 msec on average): $F_1(1,18) = 17.37$, $p<0.001$. The only other significant result was an ambiguity effect. Responses to unambiguous fricative-final targets were faster (79 msec, on average) than responses to the /ʔ/-final targets: $F_1(1,18) = 9.19, p<0.01$. Note that the lexical and ambiguity effects did not interact ($F_1<1$): the RT advantage for word over nonword decisions was statistically equivalent for targets with unambiguous and ambiguous final fricatives (71 msec versus 63 msec). Note also that although these data showed some evidence of the palatal bias (e.g. responses were faster to /ʃ/-final than to /s/-final items, and responses to /ʔ/-final items were faster after /ʃ/-final than after /s/-final cues), these effects were not significant.

IV. DISCUSSION

Rhyme decisions were faster to words than to nonwords in all the item sets. It seems clear that lexical involvement in rhyme decisions is highly robust [9]. Furthermore, it was found that targets such as /kɪʔ/ were more frequently taken as rhymes when following cues such as *miss* than

	Words		Nonwords	
Unambiguous				
/s/-final	611	(kiss)	713	(fiss)
/ʃ/-final	594	(fish)	834	(kish)
Overall	602		673	
Ambiguous (/ʔ/-final)				
Rhyming with /s/	694	(/kɪʔ/)	774	(/fɪʔ/)
Rhyming with /ʃ/	678	(/fɪʔ/)	723	(/kɪʔ/)
Overall	686		749	

Table 3. Mean RTs (msec) for positive responses to fricative-final targets, with examples. The responses to /ʔ/-final items are shown separately for decisions made after /s/-final cues (rhyming with /s/) and decisions made after /ʃ/-final cues (rhyming with /ʃ/).

were targets such as /fɪʔ/ when following cues such as *hiss*. /ʔ/-final strings which rhymed by forming words (/kɪʔ/ rhyming as *kiss*) were more likely to be taken as rhymes than similar targets which rhymed by forming nonwords (/fɪʔ/ rhyming as *fiss*). In phonetic categorization, however, the final ambiguous consonant of these strings was categorized without lexical involvement. It appears that the rhyme task, in contrast to the categorization task, encourages the use of lexical knowledge.

But there was an asymmetry in the data: there was no preference for /fɪʔ/ as a rhyme of *wish* (forming the word *fish*) over /kɪʔ/ as a rhyme of *dish* (forming the nonword *kish*). This asymmetry is due to a palatal bias. /ʔ/-final items were more likely to be taken as rhymes after /ʃ/- than /s/-final cues, and unambiguous /ʃ/-final items were more reliably detected as rhymes than unambiguous /s/-final items. This bias is a consequence of stimulus construction. For all of the fricative-final targets, the final phoneme was always preceded by a vowel taken from a naturally-produced *palatal* context. Therefore, any formant-transition information in the vowels would always cue a palatal place of articulation. This would serve to make the unambiguous /ʃ/ (with consistent information) a better exemplar than the unambiguous /s/ (with conflicing transitional and frication information). It would also cause the ambiguous fricative to be interpreted more readily as /ʃ/ than /s/. This palatal bias appears to mask the lexical effect in some comparisons.

It is nonetheless the case that /ʔ/-final targets were taken as rhymes more often when they formed words than when they formed nonwords, and that yes responses were faster when /ʔ/-final targets rhymed as words than when they rhymed as nonwords. This lexical effect appears to be due to the *rhyme task*: in categorization there were no lexical effects on the same materials. These results therefore support the hypothesis that the demands of the rhyme-judgement task induce attentional focus on the lexicon. Rhyme decisions require listeners to compare phonological strings: two monosyllables only rhyme if all but their initial consonant(s) are identical. Since the lexicon contains stored phonological strings, it would be efficient to use this knowledge, when possible, to make rhyme judgements. Rhyme decision thus encourages the use of lexical information. In contrast, the demands of categorization tend to focus attention on prelexical levels of processing. Subjects have to make fine acoustic judgements to categorize ambiguous phonemes. Phonetic categorization thus discourages use of the lexicon.

Task demands can be manipulated by asking listeners to perform concurrent tasks [6,7], by changing the context in which target items appear [8], or by asking them to perform different tasks with the same materials, as shown here. The effects due to task demands can all be explained by an attentional mechanism. The autonomous Race model can more parsimoniously account for attentional influences than the TRACE model.

In the Race model, when task demands encourage lexical involvement, attention is focussed on the output of the lexical procedure. When demands encourage acoustic-phonetic analysis, attention is focussed on the output of the prelexical procedure. The shifting of attention between these two procedures follows naturally from the architecture of the model. In TRACE, however, attention must operate as a gain control, increasing the amount of top-down facilitation when attention is focussed on the lexicon, and decreasing it when attention is focussed on acoustic-phonetic detail. Although this explanation accounts for the data, it does not follow directly from the model: why should a focus of attention on one or other level of processing result in a modification of the strength of connections between those levels? This explanation is thus rather *ad hoc*. The evidence from the current experiment can therefore be taken as further support for autonomous models, which claim that spoken word recognition is primarily a data-driven process.

ACKNOWLEDGEMENTS

The experimental work was supported by a British Medical Research Council Studentship, and forms part of the author's doctoral dissertation [10].

REFERENCES

[1] W.F. Ganong, "Phonetic Categorization in Auditory Word Perception," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 6, pp. 110-125, 1980.

[2] J.M. McQueen, "The Influence of the Lexicon on Phonetic Categorization: Stimulus Quality in Word-final Ambiguity," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 17, pp. 433-443, 1991.

[3] M.A. Pitt and A.G. Samuel, "An Empirical and Meta-analytic Evaluation of the Phoneme Identification Task," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 19, pp. 699-725, 1993.

[4] J.L. McClelland and J.L. Elman, "The TRACE Model of Speech Perception," *Cognitive Psychology,* vol. 18, pp. 1-86, 1986.

[5] A. Cutler and D. Norris, "Monitoring Sentence Comprehension." In W.E. Cooper and E.C.T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett,* Hillsdale, NJ: Erlbaum, 1979.

[6] P.D. Eimas, S.B. Marcovitz Hornstein, and P. Payton, "Attention and the Role of Dual Codes in Phoneme Monitoring," *Journal of Memory and Language,* vol. 29, pp. 160-180, 1990.

[7] P.D. Eimas and L.C. Nygaard, "Contextual Coherence and Attention in Phoneme Monitoring," *Journal of Memory and Language,* vol. 31, pp. 375-395, 1992.

[8] A. Cutler, J. Mehler, D. Norris, and J. Segui, "Phoneme Identification and the Lexicon," *Cognitive Psychology,* vol. 19, pp. 141-177, 1987.

[9] J.M. McQueen, "Rhyme Decisions to Spoken Words and Nonwords," *Memory & Cognition,* vol. 21, pp. 210-222, 1993.

[10] J.M. McQueen, *Phonetic Decisions and their Relationship to the Lexicon,* Doctoral Dissertation, University of Cambridge, United Kingdom, 1991.