1978

The importance of speech recoding in reading.

Maria Louisa Slowiaczek
University of Massachusetts Amherst

THE IMPORTANCE OF SPEECH RECODING IN READING

A Thesis Presented

By

MARIA LOUISA SLOWIACZEK

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

August, 1978

Psychology

THE IMPORTANCE OF SPEECH RECODING IN READING

A Thesis Presented

By

MARIA LOUISA SLOWIACZEK

Arnold D. Well, Chairman of Committee

Charles E. Clifton, Jr., Member

C. Dalton Jones, Member

Alexander Pollatsek, Member

Marvin W. Daehler, Associate Chair
Department of Psychology

Acknowledgements

I would like to thank Arnold D. Well, Chairman, and the members of my committee, Charles E. Clifton, Jr., C. Dalton Jones and Alexander Pollatsek for many helpful suggestions and hours of moral support.

I am grateful to Alexander Pollatsek for helping with computer programming, and making the computer work when I couldn't.

I am especially grateful to my friend and advisor, Charles E. Clifton, Jr. for setting impossible deadlines, for supplying peanut butter and cookies, and for endless hours of data analysis.

My very special thanks to my adorable little sister, Louisa M. Slowiaczek, for her patience and care in typing this paper.

I would also like to thank Betty Ann Levy for supplying the stimulus materials.

Abstract

The role of speech recoding was examined in reading normal discourse. Subjects were asked to count out loud while reading in order to prevent subvocalization. Performance was measured on tests of both verbatim memory and memory for meaning. The counting task was very disruptive for verbatim tests but not for meaning judgments. In another experiment reading for meaning was tested for stories which required organization of the concepts in order to understand the meaning. In this test, the counting task was more disruptive than in the test for meaning when organization was not required. These results suggest that speech recoding is also helpful in reading for meaning when concepts must be organized in memory.

Table of Contents

List of Tables

By the time a child attempts to read a book, he has already acquired speech and language skills. His knowledge of spoken language reduces the reading task to a translation of the written words into their corresponding spoken words. Once the symbols are in a phonological form, the speech and language processing mechanisms can take over. This is a particularly useful strategy to follow in reading alphabetic languages. For alphabetic languages, not only is there a mapping of written symbols onto spoken words, but also, unfamiliar written symbols can be decoded to spoken words by using spelling to sound correspondence rules. Current reading programs emphasize phonetics and thereby encourage new readers to use their already well-developed speech processing skills in reading.

Although it may be helpful to translate written words to spoken words, this step may only be necessary for the new reader. As the child becomes a skilled reader, he may acquire a direct route from the printed word to its meaning without going through a phonological translation.

In reading prose the child must also learn to combine individual words into phrases and sentences and to integrate sentences with the general context of the discourse. A phonological translation might be used to reduce the difficulty

of the reading task at either the stage of combining words into sentences or sentences into discourse structures. When reading becomes more automatic these stages might also be accomplished without making the translation to a phonological code. As the child develops his reading skill, he may no longer need to translate text into a phonological representation for <u>any</u> stage of reading.

If a direct visual reading system is possible, it must operate in parallel with a phonological system, which would be used in sounding out unfamiliar words. This phonological system might also function as a back-up system, to be used when the reading material is difficult and the processing demands are high. Alternatively, direct visual access to the ultimate comprehension of prose might not be possible, even for the most skilled readers. Although direct visual access might occur at some stages, for example, to the level of individual word meanings, phonological recoding may be necessary for sentence and discourse comprehension. The issue of whether or not phonological recoding is necessary in reading for meaning in prose is addressed in the research discussed here.

Word Encoding

The first stage where a phonological translation might be used is in accessing an individual word. Some studies

show that a phonological code is activated in lexical access but there is considerable controversy over how often it is used in normal reading.

A number of studies have indicated that a phonological code is processed prior to lexical access. Rubenstein, Lewis and Rubenstein (1971) used a lexical decision task, in which some of the nonwords sounded like real words (e.g., Burd, Blud, Groe). Subjects were slower to reject nonwords that sounded like real words than nonwords that did not sound like real words, indicating that the phonological code had been activated. Meyer and Ruddy (1973) found similar results in a semantic categorization task. They presented category members (pear), homophonic nonmembers (pair), and nonmembers (tail), in a task where subjects judged category membership. Subjects were asked to say whether or not the word was spelled like the name of a category member (e.g., Is "pear" a kind of fruit?), or pronounced like the name of a category member (e.g., Does "pair" sound like the name of a fruit?). Subjects were slower to reject homophonic nonmembers than nonhomophonic nonmembers in the spelling task. Other evidence was provided by Baron (1973) in a phrase evaluation task. Subjects made more errors in rejecting phrases such as "It's knot true" where the phrase sounded sensible, than phrases such as "I am kill." (Baron did not find a difference in reaction

time to these two types of phrases, however.) Baron's error data, and the Meyer and Ruddy results, indicate that a phonemic code is activated by a visually presented item and can effect lexical access. However, a visual code must also be operating in lexical access. Irregular words ("through", "bough") can be recognized as real words, even though the pronunciation by rule is misleading. Further, homophones ("pair" and "pear") can be recognized as two distinct words even though they sound identical.

If both phonemic and visual codes are available to use in lexical access, regular words (words which follow regular grapheme to phoneme pronunciation rules) should be encoded more quickly than irregular words, because both codes point to the same lexical entry for regular words. Coltheart, Jonasson, Davelaar, and Besner (1978) compared regular and irregular words in a lexical decision task. They found no advantage for regular words over irregular ones. One possible explanation for their finding is that visual access is faster than phonological access. All decisions must have been made visually, since having the grapheme to phoneme correspondence available provided no advantage. In a detailed review of these results, Coltheart (in press) concluded that visual and phonological access do occur in parallel but that visual access is faster. The Rubenstein et al. (1971) and Meyer and Ruddy (1973) results, in which a string

was processed slowly if it sounded like (but actually was not) a target item, indicate that phonological encoding may be used when visual encoding leads to a "no" response. In normal reading, however, we seldom encounter words or nonword strings whose visual characteristics do not permit a successful encoding, so we do not have to rely on the phonological back-up system for word encoding.

If the lexicon can be accessed visually, the next stage where a phonological code may be useful is in maintaining the word representation in memory. Demonstrations of the Word Superiority Effect using the Reicher-Wheeler (Reicher, 1969; Wheeler, 1970) pattern masking technique indicate that a phonemic code might be useful in maintaining the word representation. Hawkins, Reicher, Rogers and Peterson (1976) studied how flexible this encoding was by varying the proportion of items in which the two choices in a two-alternative forced-choice recognition task were homophones (cent, $^{c}_{s}$ent). Subjects were given either a high homophone list (encouraging a visual strategy) or a low homophone list (encouraging a phonemic strategy). Subjects in the high homophone list condition performed equally well on homophones as on other words, while subjects in the low homophone list condition did not perform as well on the homophones. These results demonstrate that the code used to maintain the stimulus item is flexible, and is under the control of

subject strategy.

Other studies also indicate that use of a phonological code before lexical access is flexible. Forster and Chambers (1973) compared naming time to lexical decision time and found a positive correlation. They reasoned that if naming time was correlated with lexical decision time, a phonemic component was indicated in lexical access. An alternative interpretation of this correlation is that lexical access is involved in naming words. In naming words, it might be faster to access the lexicon visually, and then look up the pronunciation. If lexical access is used in naming, a word frequency effect should occur in naming. High frequency words which are accessed more quickly than low frequency words should also be named more quickly. A word frequency effect was found in naming times, though it was smaller than the word frequency effect in lexical decision. In a similar study, Frederiksen and Kroll (1976) showed that using lexical access to name words is somewhat dependent on subject strategy. They presented words and nonwords in separate blocks or in a mixed list. The word frequency effect was greater on blocked naming, where subjects could use lexical access on all items, than on mixed lists where lexical access could not be used to name nonwords. Subjects seemed to follow a strategy of accessing the lexicon to name a word more often when this strategy was beneficial.

Baron and Strawson (1976) found that subjects differ on how much they rely on using visual access or grapheme to phoneme correspondence rules in naming words. They tested subjects on how well they could judge whether nonwords sounded like real words to see how well they used grapheme to phoneme correspondence rules. They also tested them on detecting spelling errors to determine how much they relied on the visual picture of the word. Subjects who did well on judging nonwords and not so well on the spelling test were classified as "Phonecians" and those who did well on the spelling test but not so well using GPC rules were classified as "Chinese." These subjects were then asked to name regular and exception words. Baron and Strawson found that the difference between the time to name irregular words and the time to name regular words was greater for "Phonecians" than for "Chinese." The "Phonecians" were slower to name exception words because these words cannot be named by using grapheme to phoneme correspondence rules. For the "Chinese" subjects, who were reading each word as a unit, the irregularity in spelling had a much smaller effect.

These naming studies indicate that there are individual differences and flexibility in how lexical access is used in naming words. In reading discourse however, the reader is primarily concerned with finding the meaning of the word.

The phonological component may be more critical in this case. When the task involves processing for meaning, a phonological code does sometimes prove to be more efficient. Meyer and Ruddy (1973) found that subjects were faster to respond "yes" to category members when they were told to make judgments based on pronunciation (Does "pear" sound like the name of a fruit?) than when they made judgments based on spelling (Does "pear" look like the name of a fruit?). In the spelling task, the judgment must be based on visual information since the phonemic code for distractors ("pair") would incorrectly lead to a positive response. In the pronunciation task, either code can be used to make a response. Baron (1973) also found faster response times for genuine phrases in a pronunciation task than in the spelling task. Meyer and Ruddy suggest a race model whereby both phonological and visual codes are processed in parallel. The first one to finish provides a response. Coltheart's suggestion that lexical access is slower via the phonological code than through visual access can be incorporated into this model. For lexical access, the visual code is faster, but as the word is processed for meaning, there is additional time for the phonological code to catch up. Sometimes, the phonological code will provide the word meaning first, so response time will be faster on the average when both codes lead to the cor-

rect response.

Sentence Encoding

Reading for meaning involves more than just understanding isolated word meanings. Sentences must be processed syntactically, so that individual word meanings can be combined into a sensible message. As the syntactic processing becomes more difficult, speech recoding may be more important. Perfetti and Goldman (1976) found persuasive evidence that some components of language processing are used in reading. They compared skilled and poor readers on a digit memory task and on a verbal memory task. Both groups performed equally well on the digit memory task, indicating that poor readers do not have a general memory problem. However, on a task where they were required to recall verbatim sentences that they listened to, good readers were significantly better than poor readers. Perfetti and Goldman suggest that poor readers may be poor language comprehenders. Although these readers may have several problems in comprehension, one possible source of difficulty is in using a speech code.

Kleiman (1975) suggested that speech recoding is important when reading becomes more difficult. He proposed that simple reading tasks can be done visually, but when phrases must be maintained in memory for further processing,

speech recoding is critical. This would occur if the material was syntactically difficult, or when the meaning of a phrase depends on a meaning further along in the discourse. He used a shadowing task to disrupt speech recoding, and asked subjects to perform graphemic, phonemic, category and sentence acceptability tasks on visually presented items. Response times showed that the phonemic and sentence acceptability judgments were greatly disrupted by shadowing though category and graphemic judgments showed very little disruption. He argued that speech recoding is employed when words must be maintained in working memory so that word meanings can be combined, as is needed in a sentence acceptability judgment.

Hardyck and Petrinovich (1970) found evidence that speech processing was more critical for difficult reading materials, as Kleiman would predict. They examined the role of speech recoding by comparing reading when subjects could subvocalize and when subvocalization was suppressed through biofeedback. The readers in the suppression condition performed quite well when they were reading easy material, but they had trouble suppressing subvocalization as well as comprehending when the material was semantically and syntactically more difficult.

Although these results indicate that speech processing is sometimes used in reading, they do not suggest <u>how</u> it

might be used. Betty Ann Levy did a series of studies to investigate how speech recoding is used in reading prose. Levy (1977) had subjects read short passages silently or while counting out loud from one to ten repeatedly. She reasoned that subjects who were not counting were free to subvocalize while the others were not. Interference in the counting condition would be due to the inability to subvocalize. She found poorer performance in a verbatim recognition task when subvocalization was suppressed by counting than when it was not. She also tested subjects in a listening task to see if counting used general attentional capacity rather than disrupting subvocal reading in particular. Subjects listened to the same passages silently or while counting. No interference effect was obtained from counting while listening. Performance in the silent condition was the same for the listening and reading tasks so this result is not due to overall differences in difficulty. She concluded that the process which is being disrupted is specific to the reading task.

In another study, Levy (1977) showed that the effect of suppression (i.e., interference due to counting) is independent of a thematicity effect. She had subjects read sets of related and unrelated sentences, and judge verbatim recognition for sentences which were identical,

changed lexically (one word was replaced by a synonym), or changed semantically (the subject and object words were reversed). For the lexically changed sentences, subjects were no better when the sentences were related thematically than when they were unrelated. For the semantic changes, however, subjects were significantly better when the sentences were related thematically. This thematicity effect occurred to the same extent in both the silent and suppressed conditions though the suppression effect was still obtained. Levy concluded that processing for meaning, which was facilitated by the thematic element in the text, occurred independently of the speech processing which was disrupted by suppression.

More recently, Levy (1978) questioned whether or not the effect of suppression existed only in cases where the verbatim representation of the text needed to be maintained. She asked subjects to make paraphrase judgments, and found no suppression effect. She concluded that the role of speech processing was to hold verbatim information in memory, but that reading could be done visually, without speech recoding, to the level of meaning when exact wording was not needed.

These results are quite compelling but they do not lead directly to the conclusion that reading for meaning does not require speech recoding. Reading for meaning can be broadly

interpreted to include everything from recognizing individual symbols to understanding the implications of a complex philosophical argument. The Levy studies obtained no interference from suppression when reading for meaning was defined somewhere between these extremes.

The passages in the Levy experiments were extremely simple, semantically predictable passages. On close examination of the materials it seemed that the correct paraphrase sentences for many passages could be guessed from the information provided by the title. For example, a passage sentence used was "The concerned policeman approached the worried child.", in a passage titled The Lost Boy. The accurate paraphrase was "The concerned officer approached the upset child." The distractor paraphrase was "The concerned woman approached the carefree child." Given the title, The Lost Boy, the second paraphrase seems less plausible.

Levy also used sets of unrelated sentences. For these passages subjects could not use information from the title since no title was provided. Although the suppression effect was not significant for these passages, the data were extremely variable and the true effect might have been hidden in the noise.

Perhaps the appropriate conclusion from Levy's studies is that reading for meaning is unaffected by speech suppression in highly predictable context conditions. Speech suppression

might be very disruptive under less predictable conditions. The purpose of the present research was to extend Levy's work to other types of reading materials, to see whether speech processing is necessary for other reading tasks.

The Experiments

Experiment 1 was a replication of the Levy paraphrase experiment. Levy's sets of related and unrelated sentences were used as stimuli. The sets of related sentences were short stories presented with a title. They were divided into two sets: 1) Stories in which the correct test sentence could be easily guessed from the title (predictable) and 2) Stories in which the correct test item could not be easily guessed (non-predictable). If the lack of suppression in the Levy paraphrase experiment was due to the greater plausibility of correct test items than of distractor items, then the non-predictable group of stories should show a suppression effect in Experiment 1 and the predicatble group should not. The unrelated sets of sentences were composed of seven sentences from different stories and were presented without a title. Experiment 2 was a replication of the Levy verbatim recognition experiment. In Experiment 3, the reading materials were made more difficult by introducing test items which required inferences from the text.

Experiment 1

Method

Subjects. Eighteen male and fourteen female undergraduate volunteers were used as subjects. They received experimental course credit for one hour of participation.

Materials and Design. The seventy-two reading passages were the same as those used by Levy. Half of the passages were sets of related sentences presented with a title to form a short story (thematic passages). The other passages were sets of unrelated sentences taken from different thematic passages with no title provided (non-thematic passages). The thematic passages were divided into two groups, a group for which the correct paraphrase could be easily guessed (predictable) and a group for which it could not (non-predictable). Stories were assigned to these two groups by a score which was obtained in a normative task. Forty-eight undergraduate volunteers had been presented with the title of each passage and a "correct" and a "distractor" paraphrase test. They were instructed to decide which of the two test sentences would more likely be found in a story with the given title. If at least 75% of the subjects guessed the correct paraphrase ($\overline{X}$ = .88) the passage was assigned to the predictable group. If the correct paraphrase and

distractor were guessed equally often ($\overline{X}$ = .48) it was assigned to the non-predictable group.

All passages were seven sentences long with twelve passages tested on each of sentence positions 4-7. Sentences in positions 1-3 were also tested on the remaining twenty-four passages to ensure that the entire passage was read but these passages were not included in the analysis. All test sentences were seven words long and of the form article/ adjective/noun/verb/article/adjective/noun. Correct paraphrase test sentences were passage sentences with two words changed while keeping the meaning intact. Distractor sentences had two words changed with the meaning changed. Examples of thematic-predictable, thematic-non-predictable, and non-thematic passages along with test sentences are presented in Appendix A.

Subjects were randomly assigned to one of the two groups, the "thematic" group and the "non-thematic" group. In the thematic group, subjects were tested on predictable and non-predictable passages. In the non-thematic group the same predictable and non-predictable test sentences were used although there were no titles or themes in these passages to help guessing.

All subjects were tested in both silent and suppressed reading conditions. In the silent reading condition, subjects read quietly to themselves. In the suppressed

reading condition, subjects counted outloud from one to ten repeatedly as they read. An equal number of predictable and non-predictable passages were randomly assigned to silent and suppressed conditions.[1]

One test sentence was presented at the end of each passage with an equal number of correct and distractor test sentences in each block. Each passage was tested with a correct paraphrase for half of the subjects and a distractor for the other half.

Procedure. Seventy-two passages were presented in six blocks of twelve passages each. Subjects in the thematic group were told that the passages had a theme and that they should try to relate the sentences to the theme. Subjects in the non-thematic group were told to read the sentences and try to remember them.

The blocks were alternated between silent and suppressed reading. All subjects were told whether or not to count at the beginning of each block. They were encouraged to count at a very rapid rate as loudly as they could. The counting was monitored by the experimenter over an intercom system, to be sure that a fast rate was maintained. It took subjects approximately one second to count from one to ten.

The sentences were presented on a Hewlett Packard 2600 A video terminal, at the rate of one sentence every two

seconds. The screen was two feet away from the subject and sentences were written in block letters 6 millimeters by 2.5 millimeters. At the end of each passage, the word "test" appeared on the screen for one second before a single test sentence was presented. Subjects were given ten seconds to respond to the test sentence. Responses were recorded by Hewlett-Packard 2100 A Computer, and subjects were given feedback on the video screen after each response.

The probability of a hit (a "yes" response to a correct paraphrase) and probability of a false alarm (a "yes" response to a distractor) was calculated for each condition. To correct for guessing a derived score justified under the assumptions of the two threshold theory of recognition memory (Kintsch, 1970) was then calculated for each condition. This score (p(hit)-p(FA)) can range from -1 to +1 where 0 is chance performance and +1 is a perfect score.

A 2x2x2 analysis of variance was conducted with suppression, predictability and thematicity as factors. Separate analyses treating subjects and materials as random effects were conducted and the minF' statistic (Clark, 1973) was computed.

Results

Each serial position was tested in this experiment to ensure that subjects read all of the sentences. Test sentences from serial positions 1-3 were used as filler items only and

are not included in these analyses. Since each serial position was not presented equally often in each condition, the data were collapsed over serial positions 4-7.[2] Separate analyses of the serial position effects are presented in Appendix B.

The effects of story type, suppression, and thematicity were examined in a 2x2x2 analysis of variance. The mean scores for the thematic and non-thematic groups in silent and suppressed conditions for both predictable and non-predictable stories are presented in Table 1. This experiment did not replicate the Levy paraphrase experiment. Performance was better for the silent condition (.48) than the suppressed condition (.38). The difference between silent and suppressed conditions, which Levy called the "suppression effect," was significant when subjects were treated as a random factor, ($F(1,30) = 6.13$, $p < .01$), when stories were treated as a random factor, ($F(1,136) = 14.38$, $p < .001$), and when both were treated as random factors ($minF'(1,59) = 4.30$, $p < .05$).

Follow-up contrasts were done to look at the simple effect of suppression in the thematic and unrelated sentence groups separately. These effects were not significant, with subjects as the random variable (thematic group, $t(30) = 1.77$; non-thematic group, $t(30) = 1.74$).

There was a large effect of story-type. The mean (.51) for the predictable stories was significantly greater than

Table 1: p(Hit)-p(FA) for paraphrase tests, Experiment 1

	Thematic		Non-Thematic	
	Predictable	Non-Predictable	Predictable	Non-Predictable
Silent	.63	.34	.50	.39
Suppressed	.53	.29	.39	.29
Difference	.10	.05	.11	.10

the mean (.33) for the non-predictable stories, in the subject analysis ($F(1,30) = 17.05$, $p < .001$), in the story analysis, ($F(1,136) = 9.55$, $p < .002$) and when both subjects and stories were treated as random factors ($minF'(1, 137) = 6.12$, $p < .025$). There was no main effect of group (thematic versus non-thematic) though there was a significant group by story-type interaction when stories were treated as a random factor ($F(1,136) = 6.25$, $p < .014$). A closer look at the means for the thematic and non-thematic groups indicates that the story-type effect appeared mostly in the thematic group, as should be the case if subjects used the thematic information in the predictable stories to guess about the test items. While the story-type manipulation was successful it did not interact with suppression ($F < 1$). Subjects used the predictability of the stories to guess correctly in the silent condition as well as in the suppressed condition.

There was no group x suppression interaction.

Discussion of these results will be deferred until after Experiment 2.

Experiment 2

Method

Subjects. Eight male and eight female undergraduate students were randomly assigned to Experiment 2, which was

run during the same time period as Experiment 1. They received experimental course credit for one hour of participation.

Materials and Design. The seventy-two reading passages were the same as the thematic passages used in Experiment 1. Subjects were tested for verbatim recognition of passage sentences. Test sentences for 48 of the passages were identical to passage sentences, lexically changed, or semantically changed. A lexically changed sentence had one noun in the sentence replaced by a synonym, so that the meaning was the same but the wording was different. A semantically changed sentence had the two nouns in the sentence in reversed positions so the same words were used but the meaning was different. Twenty-four filler test sentences with verb changes were also used to ensure that subjects read the entire passage and not just the nouns. Examples of the test items are presented in Appendix A.

All subjects were tested in both silent and suppressed reading conditions.

Procedure. Seventy-two passages were presented in six blocks of twelve passages each. Half of the blocks were silent reading and half were suppressed with an equal number of predictable and non-predictable passages in each

block. The presentation of the material was the same as in Experiment 1. One test sentence was presented at the end of each passage. Subjects were told to respond "yes" if the test sentence was identical to a sentence in the passage. Six identical and six distractor test sentences were presented in each block. Two of the distractor tests were lexically changed, two were semantically changed tests and two were filler tests. Across blocks, each passage was tested with an identical, a lexically changed and a semantically changed test in both silent and suppressed conditions.[3]

The p(hit)-p(false alarm) measure was calculated for each condition, using only the stories which had been presented in both silent and suppressed blocks.[3] Separate analyses treating subjects and materials as random effects were conducted and the minF' statistic (Clark, 1973) was computed.

<u>Results</u>

As in Experiment 1, the data are collapsed over serial positions 4-7. Separate analyses by serial position are presented in Appendix B.

The mean scores for silent and suppressed conditions are presented separately for predictable and non-predictable stories in Table 2. A preliminary analysis showed

Table 2: p(Hit)-p(FA) for identity test, Experiment 2

	Predictable	Non-Predictable
Silent	.55	.58
Suppressed	.14	.19
Difference	.41	.39

no difference between lexical and semantic test items ($F(1,15) = 1.37$) and no interactions with other factors (all F's < 1.44) so this factor was collapsed in subsequent analyses.

The suppression effect was significant when subjects were treated as a random factor, ($F(1,15) = 30.31$, $p < .001$) when stories were treated as a random factor ($F(1,28) = 28.23$, $p < .001$) and when both were treated as random factors ($minF'(1,43) = 13.54$, $p < .001$). There was no effect of story-type, and no story-type by suppression interaction.

An independent sample t-test was done to compare the results from Experiments 1 and 2. The suppression effect ($\overline{X} = .4$) for the identity condition in Experiment 2 was significantly larger than the suppression effect ($\overline{X} = .09$) for the thematic condition in Experiment 1 ($t(30) = 3.17$, $p < .01$).

Discussion

Although Levy did not find a significant suppression effect in a paraphrase recognition task, the effect was significant in Experiment 1. There was a reliable interference effect when subjects were required to count while reading for meaning. The present experiment was essentially the same as the Levy experiment but presentation timing was more carefully controlled. Levy used flash card

presentation while in the present study, sentences were presented by computer. It may be that the true effect is fairly small and was not detected in the Levy experiment. Additionally, examination of individual subject data from Experiment 1 suggests that there are large differences among subjects. There was a group of eight subjects who showed a substantial suppression effect (mean suppression for the eight subjects was 44% and the range was 25% to 67%). The other 24 subjects showed little or no suppression effect (mean suppression was 3%, and the range was -33% to 17%). Possibly Levy had more subjects who were not affected by suppression.

The original hypothesis that there would be a suppression effect only for the non-predictable stories was not confirmed. There was in fact, a slightly larger suppression effect for the predictable stories. The story-type effect demonstrated that subjects used knowledge of what was plausible in the recognition test, and scores in the thematic group were inflated by this effect. The story-type effect also appeared in the non-thematic group, although it was considerably smaller. Since there were no titles in this condition the effect must have been due to the plausibility of the individual test sentences. Since there was no interaction with suppression, subjects must have used this guessing

strategy to the same degree in the silent condition when subvocalization was not suppressed.

There was no story-type effect when subjects made identity judgments. Thematic information from the title was of no help in judgments about the exact wording of the test sentence.

The suppression effect found for the paraphrase tests was much smaller than the suppression effect found for identity tests. This result, which is consistent with Levy's findings, must still be accounted for. Some reading for meaning can be done while suppressing subvocalization. Maintaining the exact wording of the sentences however, may be difficult or impossible.

There are several lines of evidence which suggest that individual word meanings can be accessed without using speech recoding. Both visual and phonological information is available in lexical access (Baron, 1973; Meyer and Ruddy, 1973) so that if counting disrupts speech recoding, the words can still be processed visually. Also, priming studies (Neely, 1977; Marcel 1974) suggest that individual word meanings can be accessed without attention. Laberge and Samuels (1974) proposed that individual words are learned to a level of automaticity so that in reading, one can focus attention on combining word meanings into meaningful sentences. If speech recoding takes attention, it

must not be necessary for lexical access.

The high level of performance on suppressed trials in the paraphrase experiments may be due to visual access of individual word meanings. The stories in these experiments generally contained one main character and a series of simple events. The paraphrase distractors introduced new concepts which had not been presented in the story. If the meanings of individual words were processed during counting, then subjects could easily choose the correct paraphrase.

Although subjects seemed to be successful in recognizing individual concepts under suppression conditions, they might not have been successful in combining the concepts into meaningful sentences or integrating the sentences with each other. Speech recoding may be critical when organization among concepts is necessary. Experiment 3 was designed to test the effect of suppression on reading stories where organization was required to choose the correct test sentence.

Experiment 3

Method

Subjects. Seven male and nine female undergraduate volunteers were used as subjects. They were given experimental course credit for one hour of participation.

Materials and Design. The passages were sets of related sentences presented with a title to form a story. These stories differed from the stories in Experiment 1, in that each story contained two distinct characters. One of the sentences in each story required reference to another sentence in the story to determine which character was being described. The inference which resulted from combining these two sentences was used as a test item. The stories were constructed so that the tested inference would be of the form $A \rightarrow B$, $B \rightarrow C$, $\therefore A \rightarrow C$ (type A inference) or of the form $A \rightarrow B$, $C \rightarrow A$, $\therefore C \rightarrow B$ (type B inference). Examples of both types of inference are presented in Appendix A. The sentences involved in the inference occurred at four different points in the stories. 1) Both sentences were early in the story with only one sentence intervening (positions 1 and 3); 2) Both sentences were late in the story with only one sentence intervening (positions 5 and 7); 3) Both sentences were from the middle of the story with one sentence intervening (positions 3 and 5); 4) One sentence was early in the story and one late with two sentences intervening (positions 3 and 6). An equal number of stories were constructed for inference type A and B with inference sentences in each of these four conditions.

Subjects were tested on correct inferences, distractor

inferences, correct paraphrases and distractor paraphrases. Inference tests combined two of the sentences in the story. Paraphrase tests were passage sentences with two words changed and the meaning intact. The distractor paraphrase tests mixed characters and events from within the story to make a false statement. Unlike in Experiment 1, new concepts were not introduced in the distractors. An example of a story with each of these test items is presented in Appendix A.

All subjects were tested in both silent and suppressed reading conditions as in Experiments 1 and 2.

Procedure. Seventy-two passages were presented in six blocks of twelve passages each. Half of the blocks were silent reading and half were suppressed. An equal number of the two inference types in each of the four sentence positions was presented in each block. One test sentence was presented after each passage with an equal number of each of the four test types in each block. Across subjects, all test types were presented for all passages in both silent and suppressed conditions.

The sentences were presented as in Experiments 1 and 2. Subjects were told to respond "yes" if the test sentence was true from the information given in the story. They were also told whether or not to count at the beginning of

each block.

The p(hit)-p(false alarm) was calculated for each condition. Analyses of variance were conducted treating subjects and materials as random effects, and the minF' statistic (Clark, 1973) was computed.

Results

The mean scores in the silent and suppressed conditions for both inference and paraphrase test items are presented in Table 3. Performance was better for the silent condition ($\overline{X}$ = .57) than for the suppressed condition ($\overline{X}$ = .36). This effect was significant when subjects were treated as a random factor, ($F(1, 15) = 9.63$, $p < .01$), when stories were treated as a random factor, ($F(1, 56) = 19.24$, $p < .001$) and when both were treated as random factors, (minF' $(1, 30) = 5.89$, $p < .025$).

Paraphrase tests ($\overline{X}$ = .53) were easier than inference tests ($\overline{X}$ = .40). This effect was significant in the subject analysis, ($F(1, 15) = 9.36$, $p < .008$) in the story analysis, ($F(1, 56) = 7.20$, $p < .01$) and when both stories and subjects were treated as random factors (minF' $(1, 55) = 4.07$, $p < .05$). There was no suppression by test-type interaction.

Follow-up contrasts were done to look at the simple effect of suppression for the paraphrase tests and for

Table 3: p(Hit)-p(FA) for Experiment 3

	Inference	Paraphrase
Silent	.51	.63
Suppressed	.29	.44
Difference	.22	.19

the inference tests. For both the paraphrase test ($t(15) = 2.68$, $p < .02$), and for the inference tests ($t(15) = 3.13$, $p < .01$), the suppression effect was significant.

The effect of inference ordering was not significant though performance seemed to be better for Type A inference, ($\overline{X} = .52$) than for the type B inferences ($\overline{X} = .42$). There was no order by suppression interaction. There was no effect of the position of the test sentences and none of the other interactions were significant.

A t-test was done to compare the suppression effect found in this experiment with the effect found in Experiment 1. This difference ($t(30) = 1.12$) was not significant.

Discussion

The results from this experiment indicated a large, reliable suppression effect for both paraphrase and inference test items. When subvocalization is suppressed, reading for meaning is substantially disrupted.

The main effect of test type was significant. The inference tests were more difficult than the paraphrase tests. This result was expected, since inference tests required combining information from two sentences in the passage while paraphrase tests only required relating concepts within one sentence. Inference tests were more difficult than paraphrase tests in the silent condition as

well as the suppressed condition. However, test type did not interact with suppression. Although these results do not show an interaction, they are not necessarily inconsistent with a model which predicts one. Performance in the baseline condition is so different for the inference and paraphrase tests that the absolute amount of disruption may not be a legitimate measure of how much actual interference there is in the two conditions.

If the role of subvocalization was to maintain information from one sentence in memory so that it could be integrated with other sentences, greater temporal separation between inference sentences should have showed greater suppression. The suppression effect did not interact with the sentence position manipulation, however. Performance was the same when there was one or two intervening sentences between the sentences used to make the inference. The difference in separation was only one sentence in this experiment, so that the experiment may not have been sensitive enough to detect a difference in performance. The overall lack of serial position effects does not support a model of using subvocalization to increase memory capacity in general though more careful examination of the memory component is left for further research.

Although the mean suppression effect is twice as large in this experiment as in Experiment 1 the difference

between experiments was not significant. Tremendous variability among the subjects of Experiment 1 may have weakened this comparison. Only some of the individual subjects in Experiment 1 were affected by suppression, while individual subjects in Experiment 3 consistently showed a suppression effect. These data do not provide conclusive evidence that organizing concepts requires subvocal processing, but they do provide evidence that subvocalization is important in some aspect of reading for meaning. The reading comprehension task in the Levy study does seem to be different in a critical way from the reading comprehension task in Experiment 3. Speech recoding was clearly helpful in reading for meaning in this experiment.

General Discussion

These experiments used different types of reading tasks, and the results showed a wide range of dependence on subvocal processing to do the reading. As Levy reported, speech recoding was essential in making identity judgments where the exact wording of the story sentences needed to be maintained in memory. Unlike Levy's findings, the results of Experiments 1 and 3 showed suppression in reading tasks where only the story meaning was needed to make judgments.

Although the data do not provide unambiguous evidence for any distinctions among the types of tests which required

meaning judgments, the mean scores do suggest a difficulty ordering. In Experiment 1, the stories were simple, and test sentences required recognition of concepts which had been introduced in the story. The suppression effect was smallest in this experiment. In Experiment 3, competing concepts were introduced in the stories. A subject had to know the relations among concepts in a story in order to recognize test items successfully. These test items were more difficult, and the suppression effect was larger, than in Experiment 1. For the inference test items, information from two story sentences was required to be integrated together. These test items were the most difficult, though the suppression effect was the same as for the paraphrase tests.

The difference in performance between Experiments 1 and 3 is not just due to the general difficulty of the stories in Experiment 3. Overall performance level in the silent condition was the same for the predictable thematic paraphrases in Experiment 1 and the paraphrases in Experiment 3. Reading the Experiment 3 stories was only more difficult when subvocalization was suppressed.

Several assumptions have been made in interpreting these experiments. First, the counting task is only interesting as a secondary task if it interferes selectively with

subvocalization. If counting uses general processing capacity rather than suppressing subvocalization, than the results from these experiments are not relevant to the role of speech recoding. However, Levy's (1977) listening experiment provided evidence that counting disrupted processing which was specific to reading. Future research should look at other phonological tasks such as shadowing to establish this result more conclusively.

A second assumption is that speech recoding involves articulation and is therefore prevented by suppressing subvocalization. Speech codes may be used at several stages in reading, without using the articulatory apparatus. Allport (Note 1) has suggested that there are two speech stages in reading, one that uses articulation and another which is verbal auditory short term memory. The counting task interferes with only the articulation stage but a task such as shadowing interferes with both. It may be that the second stage is crucial in reading for meaning. Research comparing performance in these two tasks will determine the separate functions of these speech stages.

These data do not suggest a unique interpretation of reading comprehension. They are no doubt consistent with several models of reading. One such model is proposed here.

In normal reading, there are several levels of processing

going on simultaneously. As suggested by Laberge and Samuels (1974), the reader would like to focus attention on the overall meaning of the passage and the relation of new ideas to other things which have been introduced in earlier context. If elementary levels of processing can be done automatically, then the reader can devote full processing capacity to understanding the gist. Early readers learn to process letters automatically, and to combine them into meaningful word patterns.

Once the reader is skilled in recognizing words, word meanings can also be accessed automatically through a visual code, without a phonological translation. When unfamiliar words are introduced, they can still be attended to and translated into a sound pattern, to be matched with a familiar spoken word. This kind of reading is clumsy, and it is probably only used when automatic visual access is unsuccessful.

These word meanings must then be combined together, using some syntactic rules to determine the relations among the concepts. This is the stage where a semantic representation of the sentence is formed. Since the reader is moving swiftly along a page of discourse, the semantic representation for one sentence may not yet be formed when new visual information comes in. If speech recoding takes place at this stage, as suggested by Kleiman (1975)

the syntactic relations can be processed phonologically, clearing the visual buffer for new input. Once the relations within a given sentence have been determined and a semantic representation made available, it can be integrated with the overall story representation in some more permanent memory.

As a story representation is built up, some top-down processing can ease the sentence parsing bottleneck. The story context will have established characters and sets of actions inthe representation. Sometimes concepts in a new sentence can be immediately integrated with the already established semantic representation without going through the full syntactic analysis. If so, speech recoding would not be critical to understand sentences which fit in a framework which was built up from the context of the story. However, at other times, information is introduced which requires building a new structure. In this case, speech recoding should be more important. The effect of context on the use of speech recoding is an interesting question for future research.

Footnotes

1. It was my intention that all passages be presented in both silent and suppressed reading conditions, balanced across subjects. Due to an unfortunate miscalculation in programming this balance did not occur. Since the stories were randomly assigned however, I do not think the obtained effects are due to a difference between the two groups of stories.

2. Complications in the experimental design required balancing of predictability, suppression, response, and serial position. Although all serial positions were represented in all conditions, they were not all represented equally often in all conditions for each subject.

3. Due to the same unfortunate miscalculation noted in Footnote 1, some of the passages in this experiment were presented in both silent and suppressed blocks while others were not. The analyses were done only using the data from the 36 stories which were presented in both conditions.

Reference Note

1. Allport, A. Personal Communication, July, 1978.

References

Baron, J. Phonemic stage not necessary for reading. Quarterly Journal of Experimental Psychology. 1973, 25, 241-246.

Baron, J. & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 386-393.

Clark, H.H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Coltheart, M. Lexical access in simple reading tasks. In: Underwood, G. (Ed.), Strategies of Information Processing. New York: Academic Press, in press.

Forster, K.I. & Chambers, S.M. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 627-635.

Frederiksen, J.R. & Kroll, J.F. Spelling and sound: approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 361-379.

Hardyck, C.D. & Petrinovich, L. Subvocal speech and comprehension level as a function of the difficulty level of reading. Journal of Verbal Learning and and Verbal Behavior, 1970, 9, 647-652.

Hawkins, H., Reicher, G., Rogers, M., & Peterson, L. Flexible coding in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 380-385.

Kleiman, G. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.

Kintsch, W. Learning, Memory and Conceptual Processes. New York: John Wiley & Sons, Inc., 1970.

LaBerge, D. & Samuels, S.J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Levy, B.A. Reading: Speech and meaning processes. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 623-638.

Levy, B.A. Speech processes during reading. In A.M. Lesgold & J.W. Pellegrino (Eds.), Cognitive Psychology and Instruction. New York: Plenum Publishing Co., 1978.

Marcel, A.J. Perception with and without awareness. Presented at the meeting of the Experimental Psychology Society, Stirling, July, 1974.

Meyer, D., & Gutschera, K. Orthographic versus phonemic processing of printed words. Presented at the Psychonomic Society Meeting, Denver, 1975.

Meyer, D. & Ruddy, M. Lexical-memory retrieval based on graphemic and phonemic representations of printed words. Presented at Psychonomic Society Meeting, St. Lous, 1973.

Neely, J.H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 1977, 106, 226-254.

Perfetti, C. & Goldman, S. Discourse memory and reading comprehension skill. Journal of Verbal Learning and Verbal Behavior, 1976, 14, 33-42.

Reicher, G.M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 275-280.

Rubenstein, H., Lewis, S.S. & Rubenstein, M.A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Wheeler, D.D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

Appendix A

Experiments 1 and 2

Predictable Thematic Passage

A Medieval Pest

The only solution was a decisive battle.

The courageous knight confronted the terrible dragon.

The horrible creature had terrorized the countryside.

And all the peasants lived in fear.

The mighty warrior faced the fiery beast.

His well known skills failed the fearful task.

The scaley creature devoured the unlucky hero.

Test Items

Lexically changed - The courageous knight confronted the terrible monster.

Semantically changed - The courageous dragon confronted the terrible knight.

Correct paraphrase - The brave knight confronted the terrible monster.

Distractor paraphrase - The terrified knight confronted the sleeping dragon.

Non-Predictable Thematic Passage

A Flirtatious Employee

The construction company hired the young woman.

The new secretary associated with many crewmen.

The whole group frequented a nearby bar.

An intoxicated worker accosted the shy girl.

This overt act shocked the prudish employer.

The stern boss reprimanded the guilty party.

The disgruntled workman left the unpleasant scene.

Test Items

Correct paraphrase - An intoxicated laborer accosted the timid girl.

Distractor paraphrase - An intoxicated craneman accosted the flirting girl.

Non-Thematic Passage

And the guilty agitators were soon arrested.

The stronger team beat the weak opponents.

The school secretary announced a staff meeting.

The intoxicated worker accosted the shy girl.

The police officer subdued the attacking criminal.

His death caused tension in the underground.

One excited islander grabbed the shining ornaments.

<u>Experiment 3</u>

<u>Type A Inference</u>

A Flirtatious Employee

The construction company hired the young woman.

The whole group frequented a nearby bar.

The craneman was the intoxicated man.

The jovial boss chided the guilty laborer.

The intoxicated man accosted the new secretary.

The shy girl was shocked by his forward behavior.

The other employees were amused by her prudishness.

<u>Test Items</u>

Inference - Yes - The craneman accosted the new secretary.

Inference - No - The jovial boss accosted the new secretary.

Paraphrase - Yes - The drunken worker accosted the new secretary.

Paraphrase - No - The sober employer scolded the new secretary.

<u>Type B Inference</u>

A Pesky Pedlar

The weary housewife answered the ringing doorbell.

The pushy salesman endorsed his company's product.

The busy woman refused the home demonstration.

Meanwhile he began to anger the fatigued lady.

The more persistent one was the busy woman.

Finally she was able to discourage him.

The unsuccessful pedlar mumbled as he left.

Test Items

Inference - Yes - The more persistent one refused the home demonstration.

Inference - No - The more persistent one harassed the fatigued client.

Paraphrase - Yes - The busy housewife refused the saleman's demonstration

Paraphrase - No - The weary woman accepted the sample product.

Appendix B

Experiment 1

The effects of suppression, serial position and thematicity were analyzed in a 2x4x2 factor analysis of variance. The suppression effect was significant when subjects were treated as a random factor, ($F(1,30) = 6.05$, $p < .02$) and when stories were treated as a random factor ($F(1,124) = 8.29$, $p < .005$). In a Clark analysis, this effect was not significant ($minF'(1,79) = 3.50$).

The effect of serial position was also significant in the subject analysis ($F(3,90) = 4.59$, $p < .005$) but not in the story analysis. Effect of thematicity was not significant, and none of the interactions were significant.

Experiment 2

The suppression effect for identity was significant in the subject analysis ($F(1,15) = 30.31$, $p < .001$), in the story analysis ($F(1,30) = 24.46$, $p < .001$) and in the Clark analysis ($minf'(1,40) - 14.62$, $p < .001$). There was no effect of story type or serial position. None of the interactions were significant.