0-315-03570-6

National Library of Canada | Bibliothèque nationale du Canada

Canadian Theses Division | Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

51465

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

MAUREEN LOUISE DOW

Date of Birth — Date de naissance	Country of Birth — Lieu de naissance
23/05/57	CANADA

Permanent Address — Résidence fixe

66 SUNSET BLVD.
WINNIPEG, MANITOBA
R2M 0X4

Title of Thesis — Titre de la thèse

ON THE ROLE OF ORTHOGRAPHY IN EXPERIMENTAL PHONOLOGY

University — Université

UNIVERSITY OF ALBERTA

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

M.Sc.

Year this degree conferred — Année d'obtention de ce grade	Name of Supervisor — Nom du directeur de thèse
1981	DR. BRUCE L. DERWING

Date	Signature
December 17, 1980	Maureen Dow

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

On the Role of Orthography in Experimental Phonology

by

© Maureen Louise Dow

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Science

Linguistics

EDMONTON, ALBERTA

APRIL 1981

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled On the Role of Orthography in Experimental Phonology submitted by Maureen Louise Dow in partial fulfilment of the requirements for the degree of Master of Science.

Bruce L Derwing
Supervisor

Terrance M. Nearey

Tom M S Priestly

John T. Hogan

Date November 27, 1980

Abstract

The possible influence of English orthography on native speakers' phonological judgments was investigated using two tasks. The first was a rhyme task. High school students were asked to provide a rhyming word for each test word given. The test words all had rhymes with alternate spellings (e.g., bear/bare). When the words were both seen and heard (written presentation), the majority of responses matched the spelling of the test word. When the words were just heard (oral presentation), only 50% of the responses matched the spelling of the test word (chance level). The second task was a simple count of the number of "speech sounds" in each test word, by a different group of students. Although presentation mode was included as a variable, there was no significant difference in responses between the oral and written presentation groups. No explanation of the term "speech sounds" had been provided in the instructions, and there appeared to be three groups of subjects, apparently separated by their interpretation of the term "speech sounds", and the resulting criteria for rating the words. These criteria seemed to be: a syllabic criterion, a mixed or unstable criterion, and a phonemic criterion. There was a high (.897) correlation between responses of the phonemic criterion group and a phoneme index of the test words. The experiment was repeated with a different group of students

who were provided with the number of "speech sounds" in sample words, based on phonemes. Again, there was no effect of presentation mode, but the resulting responses correlated even more highly with the phoneme index (.925) than in the second experiment. The phoneme index counted all vowel nuclei as one unit. The results of the latter two experiments supported that analysis and suggested alternate analyses for some consonant + glide clusters that had been analyzed as two units. It was concluded that the influence of orthography on phonological judgments varies depending on the task demands. There was a clear effect of presentation mode in the rhyme task, but not in the speech sounds task. At the level of individual words, there was some evidence to indicate that orthography can influence phonological judgments in both tasks.

Acknowledgments

Many thanks to Dr. B. L. Derwing (who proved you do not have to stay in town to supervise a thesis) for the suggestions, direction, and patient reading of too many final drafts. Almost as many thanks are due to Tracey Derwing for the support, many discussions, and wonderful editing of this thesis. Thanks to Dr. T. M. Nearey for the numerous consultations on data analysis and to the other members of my committee for their suggestions. Finally, thanks to the Social Sciences and Humanities Research Council of Canada and the Province of Alberta for paying for my food and rent the past year and a half.

Table of Contents

List of Tables

List of Figures

I. Background

A. Introduction

The field of linguistics has traditionally belittled the influence of orthography on language and its place in the study of language. Bloomfield (1933) claimed that the only relevance of a writing system for the linguist was as an external recording device, like the use of a phonograph. Further, he pointed out that language is the same regardless of how it is written. To change the orthography would not change the language.

However, Bloomfield was in fact more concerned with the 'unimportant' orthography of a language than might be inferred from the preceding paragraph. He wrote disparagingly of 'ugly' spelling pronunciations and claimed that the language would gain by their elimination. He claimed that such forms were introduced mainly by school-teachers. These were people of a humble class that were put into a position of authority over a generation of standard speakers. Coming from a humble class, the teachers would not always know the standard pronunciation, and when forced into the pretense of knowing it, would often resort to spelling pronunciations as a way out.

Spelling pronunciations have thus caused some changes in the language. Other changes were the result of borrowing of forms from chronologically earlier texts. Bloomfield claimed that 18th century poets revived such words as <u>sooth</u>, guise, prowess, and paramour -- words that had become obsolete in speech but were present in early writings and are again evident in the modern spoken language. Bloomfield also admitted that other orthographic devices can influence the spoken language. The examples he gave were forms that were originally graphic abbreviations, such as prof., lab., quad., and dorm., that are now accepted and widely used in speech.

It is interesting to speculate as to why the written form of English can act as a dialect to be borrowed from into spoken English. One reason, suggested by Hall (1961), is that many, perhaps most, people think that what is written is actually more important and more 'correct' than what is spoken. In fact, people often tend to think of speech as a 'corruption' of what is written (just as many tend to disparage non-standard dialects, perhaps). Bloomfield obviously recognized this perceived primacy of the written form. He claimed that if there is any rivalry between two speech forms, the written form is often favored.

The influence of orthography on language is interesting to linguistics in ways other than language change and borrowings. The relation of speech to writing is important not only in terms of a preferred or standard form of the language, but also in terms of how language knowledge is stored and perceived by speakers -- the domain of psycholinguistic inquiry. The influence of orthography cannot be ignored since literacy is growing. Trager (1972) claimed that it was approaching 90-95% in Western cultures.

A very different view of the role of orthography is presented by Chomsky and Halle (1968). Theirs was a very influential 'statement' of the relationship between English orthography and phonology. They pointed out that an orthography is a system designed for readers who know the language. They further suggested that English orthography is a near-optimal system for the lexical representation of English words. This apparently outlandish suggestion seemed to be based on the assumption that the fundamental principle of any orthographic system is not to represent sounds (as modern alphabetic systems might suggest), but rather to distinguish meaningful units; thus, phonetic variation is not indicated when it is predictable (cf., Hall, 1961). Chomsky and Halle's claim regarding English orthography was therefore not that it related closely to the *speech* of native speakers, but rather concerned its relationship with postulated 'underlying' representations in the lexicon.

In their discussions of the sound pattern of English, Chomsky and Halle repeatedly pointed out that their postulated underlying forms were close to the conventional orthography. One example of this is their discussion of the e-Elision Rule and Cluster Simplification Rule (p. 48). These rules led them to propose underlying forms that were quite abstract, but had justification from independent rules and, in addition, were very close to conventional orthographic representations. Another example (p. 49), involving the words <u>courage</u> and <u>courageous</u>, required postulation of a fairly abstract underlying form that again was quite close to conventional orthography. Because of the consistent resemblance of postulated forms to orthographic representations, Chomsky and Halle suggested that an adequate theory of speech production and perception should be based on a system of representation similar to conventional orthography.

However, some of the statements made by Chomsky and Halle regarding speaker knowledge and use of orthography are suspect. For example, they claimed that readers can produce the correct phonetic forms from orthographic representations by using rules they employ in producing and interpreting speech (p. 49). Whether such rules exist at all, and whether speakers would employ rules from speech in interpreting orthography, are two important but unresolved questions. In addition, there is no evidence that speakers can consistently produce the correct phonetic output for unfamiliar forms. If they could, there would be no 'problem' of spelling pronunciations; no need for dictionaries to specify pronunciations; and no problems for children encountering unfamiliar written words. A familiar example is the common spelling pronunciation of <u>epitome</u> as [ɛ́pətowm] -- quite different from the 'correct phonetic form'.

Another point of contention with Chomsky and Halle's claims about the relation of orthography and underlying forms concerns their reasons for claiming that a conventional orthography may have a very long useful life for a wide range of phonetically divergent dialects. They based this claim on the assumption that very different dialects may have the same or very similar systems of underlying forms and on the "widely confirmed *empirical fact* that underlying representations are fairly resistant to historical change" (p.49, emphasis added). Whether an abstract postulate such as underlying representations can be *empirically confirmed* to remain stable (simply because the postulate can be used in one kind of account for historical change) is highly questionable. The conventional orthography is simply that, a convention, i.e., an arbitrary system of symbols that is accepted by the language community and as such will be resistant to change and useful for various dialects of that language. There is no basis for extravagent assumptions about its relation to underlying forms or for inferring the characteristics of the underlying forms from that assumption, as Chomsky and Halle have done.

Householder (1971) also discussed the possible relation of orthographic shape to phonological shape in the lexicon. Is phonology first specified, then orthography; are both side-by-side in the lexicon; or do we specify the orthographic form in the lexicon and derive the phonological? Although speech is normally acquired prior to a writing system, the intuitions of literate native speakers may well come to favor writing as primary, as already noted. Householder claimed that the question is whether orthography influences pronunciation or whether phonology influences spelling. In terms of economy of rules, he showed quite convincingly that there would be an economical advantage both in terms of lexical representations *and* rules to derive the phonological forms from the orthographic representations.[1]

There is some evidence to suggest that knowledge of the language system influences the phonetic perceptions of native speakers. The most famous example of this is Sapir's (1933) Sarcee speaker's 'illusion' of a stem-final [t] when there was no phonetic evidence of that [t]. The explanation Sapir offered was that knowledge of variant forms that did have the [t] influenced the speaker's perception of the form. In the same way, knowledge of orthography might also influence speaker perceptions.

In an even earlier paper, Sapir (1925) discussed the membership of /ŋ/ in the nasal series (/m/,/n/) in English. He claimed that no native speaker can be "made to feel in his bones" that the /ŋ/ belongs with /m/ and /n/ in the same way that /k/ belongs with /p/ and /t/. One reason Sapir gave was that /ŋ/ is not freely moveable (i.e., it does not occur in initial position in English). He claimed that /ŋ/ "feels" like /ng/ and that orthography was not solely responsible for this 'ng feeling' of /ŋ/.

----

[1] Note that this raises an interesting question for advocates of an 'economy' or 'simplicity' criterion for grammars, since children do acquire phonological forms first.

There are some developmental data that relate to this 'ng feeling'. Read (1973) found that children who invent their own spelling systems before they have learned the standard English orthography do not represent nasals in their writings when they occur after vowels. Interestingly, the *-ing* sequence in English is represented in such writings as *-EG*. Since these children are preliterate, the standard spelling cannot be responsible for these *-G* representations of /ŋ/.

Derwing (1973) reviewed the rider/writer example that is often used in discussions of archiphonemes in phonology (segments involving neutralization of a particular distinctive feature, such as voicing). He concluded that speakers' intuitions were vague in such cases, and their decision as to whether the word had a 't' or a 'd' sound in it seemed to based on the way it was spelt. O'Neil (1972) noted a similar phenomenon. Among linguistically untrained dialectal speakers, there were differences in perceptions of phonetically identical sawed and soared [sɔ:d] which have also been attributed to orthography. O'Neil claimed that children's phonological illusions differ from adults' and suggested that this was because children's phonology is uninfluenced by morphology. Since English orthography reflects morphological relationships between words, at least part of the apparent influence of spelling on adults' phonological judgments might, in this view, be artifactual, the real effect lying in the adults' consideration of word relations.

Hall (1961) also discussed the possible influence of orthography on speakers' sound perceptions. He noted that linguistically untrained English speakers initially had trouble classifying some phonemic diphthongs as such, and claimed that this was because English spelling often represents them as a single letter (e.g., <u>bite</u>, <u>gate</u>). Also, English spelling is likely to be misleading concerning the nature of certain consonant clusters (such as <u>th</u>, <u>ng</u>), which phonetically are single segments (θ,ŋ). With linguistically naive subjects, the interesting question arises as to whether speakers actually perceive such units as clusters or whether they are merely reporting them as such because the orthographic representation is the only one available to them.

However, there is also the other side to the argument. Klima (1972) claimed that orthography reflects phonology and, therefore, phonetic confusion results from knowledge of the sound system, not from orthographic interference. Presumably Chomsky and Halle would also claim this. These two viewpoints -- orthography as a reflection of phonology versus orthography as a distorting influence on phonological intuitions -- provide the basis of the present concern with orthography and phonology. The two tend to get confused and it is thus often difficult to know either the extent or the direction of the influence of orthography, especially in linguistic experimentation. There has been a certain amount of work done concerning phoneme/grapheme correspondences, but most of that was motivated by the desire to optimize reading and spelling instruction, rather than to resolve any of the linguistic questions outlined above.

B. Research into Sound-Spelling Correspondences

Wijk (1966) provided a pronunciation guide for English which consisted of a set of rules for reading the spelling aloud. He noted that although the 'confused and irregular' spelling of English was thought to offer a poor guide to its pronunciation, it was still possible to formulate such rules, since the spellings were systematic for some 90-95% of the total vocabulary. Wijk claimed that the main reason English spelling *seems* so irregular is that 16% of the 1000 commonest words are spelt irregularly. It is interesting to speculate as to the reason for this. Perhaps the less common words are more prone to spelling pronunciations and hence appear more regular, whereas the more common words are more resistant to spelling pronunciations and appear irregular.

An ambitious analysis of phoneme/grapheme correspondences (Hanna, Hanna, Hodges, and Rudorf, 1966) found that 80% of 12,546 phonemes in the 3000 most frequently used American-English words are spelt consistently. The study was designed to analyse statistically the degree and characteristics of correspondence between spoken language (using the phoneme as the unit of sound) and written language (using the grapheme as the unit of encoding). The stated purpose of the study was to help improve school spelling programs. Questions of interest were: the degree to which American-English orthography approximates the alphabetic principle; the relationships between phonological structure of spoken language and its representation in orthography; the degree to which certain factors of the phonological structure underlying the orthography are possible aids to making correct phoneme-grapheme matches; and phonological factors affecting the orthography. Words were classified and grouped according to how much of a phoneme-grapheme match their spellings represented. From this information, presumably, educators could more adequately gear their spelling and reading programs to tie in with children's existing knowledge of the language.

This orientation is also what Hall (1961) preferred. He proposed that we must first know the sounds of English and then determine how and to what extent they are represented by the spelling. For example, there is a well-known correlation between the use of a single vowel letter, one intervening consonant, and another vowel letter (usually 'silent' e) to indicate the 'long' (diphthongal) quality of the first vowel (e.g., hate). Similarly, a single consonant letter at the end of a word or two consonants in the middle indicate 'shortness' of the vowel (e.g., hat vs. hatter). Hall noted that there is a marked lack of economy in graphemic representations for English phonemes. There are both regular and irregular graphemic representations of phonemes and one phoneme often has numerous graphemic representations.[2]

Haas (1970) was concerned with the relation of orthography and phonology in terms of translatability. He claimed that phonemic knowledge is demanded of users of alphabetic script, but such knowledge is not enough to enable one to read and write. He attributed this to two reasons. First, schoolchildren learning to read and write their own language may have already mastered the phonological system of their language (even this is questionable -- see O'Neil, 1972), but this knowledge is only intuitive ('practical'). Haas claimed that analytical knowledge is needed in order to learn sound-grapheme correspondences. The second reason is that the relation between phonemes and graphemes is not a simple one-to-one correspondence. This deviation is known as phonographemic divergence. Haas proposed that the relation by which 'abstract representations' of orthography are related to the sound system is through a type of transformational rule, which must be learned. The rules involved are similar to Wijk's spelling rules. Haas also noted that there are differences in the difficulty of learning the use of certain graphemes. For example, he claimed that the use of th is more readily acquired than that of gh. What would be of

---

[2] An illustration of this, provided by Dr. M. L. Marckworth, is the various spellings for the English phoneme /ɛ/: bet, fete, heifer, leopard, friend, feather, bury, any, Thames, said, says, guest, aesthetic, Oedipus.

interest to linguistics would be how such differential difficulties relate to the language's sound system, i.e., whether they are due to the idiosyncrasies of the orthographic system itself or due to some more basic, underlying knowledge of the sounds that those graphemes represent. For example, the orthographic *th* almost always represents either /θ/ or /ð/ except across syllable boundaries (e.g., cathouse), whereas *gh* does not have a consistent phonetic form (e.g., ghost, fight, tough).

C. Orthography in Linguistics Today

If writing is merely a 'reflection' of language (and an imperfect one, at that), why should a field claiming to study language itself be concerned with the conventional system used to represent it? One important reason is the attitude of native speakers. Since the study of language (a human behavior) must necessarily concern itself with the attitudes of the speakers engaging in that behavior, it is important to know how those attitudes might affect the language.

Another important reason is that English written texts are still English, a part of the language. Householder (1971) pointed out that it is "weird" (his term) to assign all that is read aloud to the province of linguistics, but not things that are merely written down. Not only is this strange, it is probably a major shortcoming of North American linguistics. A field that claims to study language should presumably include all manifestations of language in its domain of investigation. If a written text differs from speech, that should be of interest; it should not be used to trivialize the importance of orthographic language data.

Chomsky and Halle have argued that the orthography of English is a near-optimal system for the underlying phonological representations in the lexicon. Interestingly, a fairly strong case can be made to support underlying forms based solely on orthography without regard for phonology. Householder, in a clever satire of typical transformational argumentation, made just such a case. He claimed that it would be easier (i.e., more efficient) to specify the lexicon as the orthographic representations and have a few spelling rules to relate the underlying forms to phonetic representations, rather than have the phonological component specified and require numerous rules to derive the written form. He pointed out that there are a greater number of ambiguously pronounced words than ambiguously spelt. Therefore, it would be more efficient and 'closer to meaning' to specify the orthographic form in the lexicon. Householder's strongest point was that there is obviously a greater chance of simplicity and exceptionlessness for rules mapping the 109 graphemes to the 30 phonemes of English than the converse.[3] He therefore concluded, both on the basis of native speaker intuitions as to the primacy of the written

----

[3] In addition to the 26 letters of the English alphabet, Householder included orthographic vowel diphthongs (e.g., ee, oo) and consonant clusters (th, ng) in his count of 109 graphemes.

form (discussed above) and from the point of view of economy and plausibility of mapping rules, that writing is logically prior to the spoken form and thus should be represented in the underlying form.

Entertaining as it may be, such argumentation lends little to the question of the influence of orthography on the perceptions of speakers, on language and language change, or on psycholinguistic experimentation. The work done by people interested in teaching reading and spelling provide for spelling rules, but tell us little of the influence of orthography on language. Only if one believes Chomsky and Halle's statement (a difficult thing to do, considering all the criticism) can one dismiss the possibility that orthography is influencing the language, since, according to them, orthography is merely a window to the constant underlying forms. The criticisms to Chomsky and Halle do not really offer alternatives, either. The questions of how, whether, or to what degree orthography influences language and perceptions of native speakers remain.

The question of orthographic influence is of primary concern to investigators of the sound systems of languages. If native speakers' intuitions are of interest (i.e., their language knowledge), then it is important, in drawing conclusions about the spoken language, to know the extent to which the written language might influence those intuitions. In an experimental situation, subjects tend to try to be 'good' subjects and do the 'correct' thing. If their knowledge of the language has included the belief that the written form is the 'correct' one, then much of what is being tapped might be knowledge of prescriptive grammar and spelling. Since literacy in the Western World, especially of English, is approaching 100%, it is foolhardy to ignore the possibility of this influence. It would be more productive to determine the extent of this influence and try to control for its interference in phonological experimentation.

D. Experimental Evidence

A series of related phonological experiments, beginning with Nelson and Nelson (1970), examined the possible confounding influence of orthography on similarity of sound judgments. Nelson and Nelson had subjects rate the similarity of sound of pairs of words on a 7-point scale. The word pairs were presented visually. Words were chosen either to match position of shared letters, or, more interesting for the present purpose, to control for the degree of 'formal' (orthographic) and 'acoustic' (sound) similarity. The latter group of words represented four degrees of similarity: high acoustic-high formal; high acoustic-low formal; low acoustic-high formal; and low acoustic-low formal. Although the sample size for the test of orthographic influence was small (only 24 pairs of words in all, with 12 in the low acoustic-low formal category), the results were interesting. Nelson and Nelson found significant effects of acoustic and formal similarity, as well as a significant interaction of the two variables. They suggested that letter-counting influenced the ratings of similarity of sound, but that the direction of the interaction supported the claim that the ratings were more affected by acoustic than orthographic characteristics. The differences in ratings for pairs differing in formal similarity was less when acoustic similarity was high than when it was low.

The effect of orthography (degree of formal similarity) noted by Nelson and Nelson was confirmed in a study done by Vitz and Winkler (1973). Vitz and Winkler argued that the visual presentation employed in the Nelson and Nelson study increased the likelihood that spelling would influence the similarity of sound ratings, something Nelson and Nelson did discuss with respect to the acoustic-formal similarity groups. Vitz and Winkler carried out an aural replicate of that part of the Nelson and Nelson study concerned with the acoustic-formal similarity groups. There were some procedural discrepencies from the Nelson and Nelson (1970) published account, but these are unimportant to the results to be discussed.

Vitz and Winkler devised a predictor of sound similarity for pairs of words -- the PPD or predicted phonemic distance. They ran four similarity of sound experiments, having subjects rate words compared to a standard word. Correlations of -.94, -.81, -.92. and -.95

for the four experiments (1-4), respectively, were found for similarity judgments compared to predicted phonemic distance. Vitz and Winkler coded the word pairs differing in acoustic and formal similarity from the Nelson and Nelson study as to predicted phonemic distance, and ran a correlation on the similarity judgments that Nelson and Nelson obtained. This correlation was found to be -.70, low with respect to the correlations Vitz and Winkler found in their experiments 1-4. They attributed this to the effect of the words' spellings, and reran the study using oral, rather than written, presentation. The new correlation was -.77. Although Vitz and Winkler claimed that this was low, it is not much lower than the -.81 found for their experiment 2, the only experiment using words with phonemic consonant clusters.

The primary concern of Vitz and Winkler was to predict and explain similarity of sound judgments based on their model of phonemic distance. Of interest in the present discussion, however, is the failure of their model in cases where orthography might be considered an important variable. Vitz and Winkler devised a new index for prediction, the phonemic cluster index, and correlations for both experiment 2 and the Nelson and Nelson replicate improved (-.90 and -.86, respectively) using this index. However, the possible influence of spelling was never really dealt with. It remained for Derwing (1976) to devise a 'grapheme' or orthographic index based on the same procedure Vitz and

Winkler used to calculate predicted phonemic distance scores.

Derwing found that both the predicted phonemic distance and phonemic cluster index from Vitz and Winkler's study correlated highly with ratings on his phonetic similarity task, with correlations of .88 and .89, respectively. However, a similarly high correlation (.82) was obtained for the grapheme index with similarity judgments. Taking partial correlations, Derwing found that the correlations of similarity judgments with phonemic distance and phonemic cluster index were still high when orthographic similarity was factored out (.63 and .68, respectively). When the phonemic similarity was controlled for, however, the correlation between grapheme index and judged similarity was only .34. Thus, although the phonetic similarity judgments seemed somewhat confounded by the orthographic similarity, the phonetic dimension was more important. Derwing suggested that his task should be repeated using oral, rather than written, presentation to control for orthographic interference. Derwing and Nearey (1980) reported that another group of subjects repeated the task with auditory presentation. The correlation between means of each item for the auditory and visual groups was an impressively high .97. A correlation this high suggests that both groups of subjects were performing similarly, but does not dismiss the possibility of orthographic interference in both cases.

Nearey (1979) did a small pilot study using an interesting type of control for possible orthographic interference. When the initial [s] of [sp] clusters is gated out synthetically, the perceptual classifications of the remaining initial unaspirated [p] are generally /b/ rather than /p/. Nearey was interested in how similarity judgments might relate to this result. He ran an aural similarity of sound task, including some pairs of words in which the spelling 'matched' or supported the phonemic analysis and others in which the phoneme of interest was not spelt consistently. Sets of word pairs like <u>pill-spill</u> and comparison pairs like <u>bill-spill</u>, in which the phonemic analysis of the s + stop cluster matched the spelling, were examples of spelling-supported pairs. Pairs like <u>cot-squat</u> and <u>got-squat</u>, in which the first stop was not represented the same way orthographically in both members of the pairs, were examples of non-supported pairs. Of interest was whether the [p] of initial [sp] clusters was perceived as such because of the spelling of the words, or whether speakers really know that it is a /p/, and not a /b/. Nearey found that overall the [p^h]-[sp] comparison means were significantly higher than those for the [b]-[sp] comparisons.

A larger replicate by Derwing and Nearey (1980) confirmed the earlier results. Orthography seemed to be influencing the judgments in that more pairs matching in spelling had significantly different p/sp vs. b/sp comparison means. The suggestion in both cases was that the experiments should be repeated with preliterate or nonliterate subjects to confirm the role that orthography is suspected of playing -- that of influencing phonological judgments.

A study by Fink (1974) did use subjects close to preliterate. Bisyllabic nonsense words containing s + stop clusters (both voiced and voiceless stops) were presented to grade 2 (almost preliterate) and grade 3 children, and to adults, in the form of a spelling test. The adults' spellings showed preference for the voiceless member of the pairs of stops when it occurred after /s/, regardless of place of articulation or presence or absence of voicing. The grade 2 children did not demonstrate the preference and the grade 3 children's responses lay between the grade 2 responses and the adults'. Fink claimed that this difference was due to knowledge of orthography. Adults did not perceive a difference between [b] and [p] (or [d] and [t] or [g] and [k]) after [s] since, for them, a voicing contrast is not possible in that environment. Fink suggested that changes in phonemic categorizations occur as the rules of English spelling are internalized. However, s + voiced stop clusters do occur in English (e.g., <u>disbar</u>, <u>misguide</u>, <u>misdirect</u>). Fink claimed that such clusters only occur for multimorphemic words in English, and the nonsense stimuli that he used were only bisyllabic, not bimorphemic. For these stimuli then, Fink's conclusions seem warranted.

E. The Present Study

The possible confounding influence of orthography has been mentioned in numerous phonological studies and has even been discussed seriously in some, but, as yet, none have concerned themselves with settling the issue. As Derwing and Nearey have both suggested, replications of phonological experiments using pre- or non-literate subjects would certainly be an adequate control for orthographic interference. Unfortunately, such adult subjects are difficult to find (especially native English speakers) and using children confounds the question even more.

Derwing and Nearey's correlation of .97 between oral and written presentation groups on the similarity of sound task suggests that presentation mode alone is not a sufficient control for orthographic interference. It may be, however, that a more subtle task is required to determine the extent of orthographic interference in phonological judgments. If true, this would also suggest that the extent of orthographic interference on phonological judgments is a function of task demands. It may be that there is no 'problem' of orthographic interference with some tasks, whereas others are heavily influenced by knowledge of orthography.

Two tasks were designed to test this hypothesis and to determine the extent of orthographic interference on each task. The first (Experiment 1) was a rhyme task. In order to supply rhyming words for given test items, subjects must access their phonological knowledge. The rhymes that are given should provide some information about influences on that knowledge. The second task was a straightforward count of the number of 'speech sounds' (as a substitute for the technical term 'phonemes') in given words. For both tasks, presentation mode was included as a variable. In each case, the words were heard, to ensure that the intended word was perceived by all listeners. For half of the listeners, the words were presented visually, as well. Thus, the effect of both type of task and presentation mode in phonological judgments was tested. Specific details and predictions for the two experiments are discussed in the relevant sections below.

II. Experimental Work

A. Experiment 1

Generation of rhyme requires the knowledge that words are made up of parts. That knowledge enables one to 'change the first part of the word and leave the last part sounding the same'. In this sense, then, a task requiring one to provide rhymes for given words is in some way accessing phonological knowledge.

A rhyme task seems well-motivated when one examines phonological studies requiring similarity of sound judgments or similar types of ratings. Rhyme has been found to be an important parameter in similarity judgments. Nelson and Nelson (1970) found that words with the last two phonemes (out of three) in common were rated as more similar than those with the first two, the first, or the last phoneme in common. They termed this the 'rhyme effect'. In the Vitz and Winkler (1973) study, post-experimental interviews were held concerning what subjects felt was influencing their similarity of sound judgments. Twenty-five per cent of those reports related to rhyme. Derwing and Nearey (1980) also discussed the rhyme effect with respect to similarity of sound ratings. They noted that differences between consonants in final position were judged to be more important than differences between initials.

For the purposes of the central question, the role of orthography in phonological judgments, rhyme in the English language is ideal. The English orthography has been much-maligned because of its many ways of spelling the same sound, but in this study that aspect proves very useful. One example of this is the alternate spellings of the phonemic sequence /ayt/ (l*ight* vs k*ite*). Spelling alternations such as these are common in the English language (bear/bare, fair/fare, fear/jeer).

In this experiment, subjects were presented with words demonstrating spelling alternations for rhymes and asked to provide a rhyming word for each word given. Responses were written, to avoid any confusion due to homophony. Thus, the written form of the words the experimenter scored were the precise forms intended by the subjects. In one condition, the subjects only heard the words and did not see them. For the other group of subjects, the words were presented both orally and visually. If knowledge of orthography is an important factor in such experiments, one would expect responses to be non-random, with rhymes spelt the same as the given words in significantly more cases than differently spelt rhymes. If, alternatively, orthographic effects vary as a function of presentation mode, one would expect there to be a difference between the oral and written groups. The third possibility would be that orthography does not play any role at all in such a task and responses of both groups would be random with respect to spelling of the rhyme. In such a case, any small trends found might be due only to word frequency.

It may be that spelling ability is a function of orthographic knowledge. If so, there should be some relation between spelling ability and the influence of orthography in a task that demands written responses. It is difficult to predict how this would be reflected in a rhyme task involving alternate spellings. Good spellers may be more aware of different spellings for one sound and be either more or less likely to let the spelling of the given word influence their choice of a rhyme than poor spellers. Certainly if orthographic knowledge and spelling ability are related, there should be some systematic difference in performance on the task with respect to choice of spellings for rhymes between good and poor spellers. To determine the relation of spelling ability to performance on the task, subjects were asked to give a self-rating of their spelling ability.

Method

Subjects. Forty high school students from a school in the Edmonton Public School System participated in this study.[4] All were native speakers of English and their ages ranged from 16 to 18.

Materials. Twenty English rhymes with alternate spellings (e.g., <u>bear</u>/<u>bare</u>) were chosen from a larger list. These will be called 'spelrimes'. The spelrimes were chosen on the basis of the number of rhyming members. Rhyme groups with limited numbers were thereby excluded. From each spelrime, six test words were chosen, three for each spelling. These were all common words with no other familiar homophones.

These 120 words were set up in two blocks of 60 words each (block A and block B; see Appendix A). Both blocks had three samples of each spelrime. The words were randomly ordered except for the constraint that there were at least two words between members of the same spelrime.

Procedure. Twenty of the students received answer sheets that were blank (no words, just blank lines) and twenty received answer sheets with the test words written down. These were, respectively, the oral and written presentation groups. Within each group, half were presented

----

[4] Thanks to Mr. P. Gatto and to the teachers and students at Eastglen Composite High School.

with one order (block A, then block B) and the other half the reverse order. The list of words was read aloud to both oral and written groups, and the written group was instructed to go at the pace set by the experimenter. The exact instructions were as follows:

> Two lists of 60 words each will be read to you. For each word, please provide a rhyming word in the space provided. Do not give words that have already been used. There is no 'right' or 'wrong' answer in this task. Your perception of English is what is of interest.

The students were also asked to indicate, on a five-point scale (excellent, very good, good, fair, poor), a self-rating of their spelling ability. There was no time pressure; students were given as much time as they required and any missed words were repeated at the end of each block. The experiment took about 30 minutes.

The answer sheets were scored by hand, a simple same/different score reflecting whether the rhyme of the word provided by the student was spelt the same or differently from the test word.

Results

A two-way analysis of variance on the 'same' scores for each subject showed that only the effect of presentation type was significant (F=39.1, p<.001 -- see Table 1). The mean 'same' score across subjects with oral presentation was 64.9 (standard deviation of 7.19), and for written presentation, $\bar{X}$=79.9 (st. dev. of 7.74).

A t-test was run to determine whether the mean 'same' score for the oral group was significantly above chance (X=60). The test was significant (t=3.05, d.f.=19, p<.01). Suspicion that a comparison of the mean 'same' score by the expected number due to chance lost information by collapsing the subject data led to the running of a chi-square test of the observed 'same' scores for each subject against the expected score due to chance. The chi-square was not significant (χ^2=24.37, d.f.=19). In the event of such a conflict, precedent dictates using the more conservative statistic, in this case the non-significant chi-square.

Although there were unequal numbers of males and females (12 and 28, respectively), there was no significant difference between the sexes in 'same' scores. The mean 'same' score for males was 71 and the mean for females was 73 (t=0.53, d.f.=38). There did not seem to be any relation of spelling ability to performance on this task, either. There was a near-zero correlation between self-rated spelling ability and 'same' scores (r=-.02). This result was somewhat unexpected and will be discussed in the next

Table 1

Analysis of Variance Source Table

Source	df	SS	MS	F	p
Presentation Type	1	2250.0	2250.0	39.1	.001
Order	1	28.9	28.9	0.5	
PT x O	1	16.9	16.9	0.3	
Error	36	2073.8	57.6		
Total	39	4369.6			

section.

Individual analyses (chi-squares) were run for all of the spelrimes to determine whether one spelling might be 'preferred' and whether any preference might be a function of presentation type. The chi-squares for the test of independence of spelling and presentation, and for the effect of spelling are shown in Table 2. Only one of the spelrimes (oe/ow) had a significant interaction. The *-ow* spelling was the preferred one overall, due mainly to the difference in the oral presentation means. There was almost no difference between the spellings for the written presentation. There was a 'preferred' spelling for both presentation types in eleven of the remaining 19 spelrimes. The preferred forms are underlined in Table 2.

Under the assumption that 'preferred' spellings might simply reflect frequency of occurrence in the language, various measures of frequency of the spelrime forms were compared to the direction of the preference. The first measure, token frequency, was calculated by combining the frequencies of all the spelrime members of each spelling alternation except those words used in the test. The frequencies used were from the count provided by Carroll, Davies, and Richman (1971). The second frequency measure, type frequency, was the number of words contributing to the token frequency measure in each case. The types were obtained using Dolby and Resnikoff's (1967) reverse word list. Words (types) with spellings and rhymes matching the

Table 2

<u>Chi-squares on Spelrimes</u>

	χ^2_{sp}	$\chi^2_{sp \times Pt}$		χ^2_{sp}	$\chi^2_{sp \times Pt}$
1. <u>ail</u> ale	12.13 ***	1.54	11. <u>eek</u> eak	9.07 **	.27
2. air <u>are</u>	10.78 ***	.31	12. ees ese	0	.03
3. ait <u>ate</u>	39.86 ***	.29	13. ot ought	.16	.05
4. <u>aft</u> affed	12.57 ***	2.75	14. off ough	.82	.06
5. ite <u>ight</u>	31.11 ***	1.71	15. oe <u>ow</u>	--	5.73 *
6. ie <u>y</u>	10.47 **	.83	16. ue ew	3.62	.31
7. ied ide	.47	.01	17. oat <u>ote</u>	17.57 ***	.29
8. iar ire	1.22	1.79	18. oot ute	1.36	.16
9. eer <u>ear</u>	8.26 **	.03	19. <u>ool</u> ule	6.23 *	.88
10. een ean	1.64	.003	20. our <u>ower</u>	14.83 ***	3.35

Underlined forms are the preferred forms.

* .02   ** .01,   *** .001
sp - spelling   Pt - Presentation type   sp x Pt - independence

forms in the spelrime were noted and then their frequencies were obtained from Carroll *et al*. The reason for including type frequency in addition to token frequency was to counter the possibility that the token frequency was inflated as a result of one or two particularly common items. The third measure, prototype frequency, was the number of words of each spelling alternation that were of the six most frequent types in the spelrime. The rationale for the prototype frequency was that since six rhymes were required for each spelrime, the most frequent rhymes would more likely be chosen and of the six most frequent, a majority might be of one spelling.

The values for each of the frequencies for each spelrime are shown in Table 3. The combined oral and written 'same' score for each spelling is also given. This score is out of a possible 240 -- six rhyme answers for each spelrime multiplied by 40 subjects. For each spelrime, the 'preferred' form is listed first. If frequency (of some measure) is related to the preference, then the larger value for frequency should be first for all spelrimes. The frequency measures that do not match the direction of the preferred forms (as measured by the size of 'same' scores) are starred.

As can be readily noted from examination of Table 3, none of the frequency measures predict at a level much better than chance. Five of twelve spelrimes do not match for token frequency, a different (but overlapping) five do

Table 3

<u>Frequency Analysis of 'Preferred' Spellings</u>

Spelrime	'same' score	Token f	Type f	Prototype frequency
1. ail	106	1535	15	* tail 624 sail 314 mail 213
ale	61	948	13	whale 281 pale 167 sale 161
2. are	82	* 2893	16	* square 1002 care 886 share 354
air	45	5686	5	air 3673 hair 867 pair 724
3. ate	92	4065	14	state 1474 late 691 ate 440
ait	24	63	3	rate 359 plate 356 gate 292
4. aft	99	239	5	raft 105 shaft 69 draft 57 graft 5 waft 3
affed	55	4	1	graphed 4
5. ight	103	14013	13	rfght 4855 might 2824 light 2415 night 2307
ite	37	13376	12	write 9849 white 2410
6. y	76	29501	20	by 20191 why 4158 try 1958
ie	41	851	4	buy 872 fly 787 eye 707
7. eak	94	* 1197	* 9	* speak 661 weak 245 peak 114
eek	57	1298	9	week 863 creek 177 seek 10
8. ear	78	7698	* 12	year 2338 hear 2159 near 1985 dear 492
eer	46	4912	13	here 4192 deer 332
9. oat	110	* 879	* 7	* coat 393 throat 207 goat 139
ote	56	1798	7	wrote 877 note 717 vote 198
10. ool	82	3524	* 5	school 2745 cool 506 pool 214 stool 58
ule	53	372	7	fuel 262 mule 76
11. ower	98	* 1634	4	* power 1082 flower 308 tower 200
our	51	6945	3	our 5807 hour 914 flour 224
12. ow	93	* 7998	* 11	* know 5677 grow 1423 flow 354
oe	59	21108	11	so 11548 no 8643 joe 684
		5*	5*	6*

not match for type frequency, and six of the twelve do not match for prototype frequency. Some other parameter must be responsible for the evidence of preferred spellings.

Discussion

It seems clear from the results of this experiment that visual presentation of stimuli enhances any influence of orthography. Not only were the means of 'same' judgments for written presentation significantly higher than those for oral presentation, the means for the latter were not significantly different from chance. This would suggest that the basis for choice of a rhyme with the oral presentation was solely the sound of the test word. Either the students did not visualize the spellings, or, if they did, their knowledge of the spelling did not significantly influence the choice of a rhyme word. The students in the written presentation group, on the other hand, must have been somewhat influenced by the spelling of the test word in their choice of a rhyme. Post-experimental interviews supported this conclusion. When told that the purpose of the experiment was to determine whether given rhymes matched the spelling of the test words, a number of students in the written group claimed that they had noticed many of their rhymes matched spellings, whereas none of the oral group seemed to have made any connection between the task and the spelling of the words.

Under the most conservative interpretation, the results indicate that presentation mode is an important variable in this type of elicitation task. A more liberal interpretation would suggest that there is no orthographic interference in this task when the stimuli are presented orally only. Such a conclusion, if warranted, would be very pleasing to many phonological investigators. It would mean one less thing to control for when doing phonological experimentation. However, the results of the various experiments discussed in preceding sections suggest that the influence of orthography is not restricted to written presentation, and they might mean that there is <u>no</u> orthographic interference in either presentation mode: Vitz and Winkler (1973) repeated a visual presentation study using aural presentation with equivalent results in the two studies; Derwing and Nearey (1980) reported a correlation of .97 between oral and written presentation groups on ratings in a similarity of sound task. Under either interpretation, these results, taken together with the present study, indicate that orthographic interference varies with the type of task. The results of the present study confirm that orthography is not an important influence with oral presentation of a rhyme task, but is significant with written presentation. The motivation for the second study, to be discussed further on, is this apparent importance of type of task.

It is interesting (and reassuring) that no significant difference was found between males and females in performance and no correlation was found between self-rated spelling ability and 'same' scores. Both results support the contention that the rhyme task is a legitimate psycholinguistic technique, tapping general language knowledge. If it were not, it would be expected that sex and/or spelling ability would be correlated with the effect of presentation type. It has been found that sex is a significant factor, for example, in certain psychological investigations of reading ability and concept formation (e.g., Maccoby & Jacklin, 1974). The interest in spelling ability, on the other hand, was motivated by the possibility that spelling ability might be related to a sensitivity to the relationship between English sound and spelling, and this might be reflected in a task accessing language knowledge. From the negligible correlation of spelling ability to 'same' score, it could therefore be argued that spelling ability has little to do with language knowledge, but rather is a reflection of memory, or some sensitivity to the English orthographic system itself.

Although there is some evidence (in the results of the chi-squares on each of the spelrimes) that there are 'preferred' spellings for certain rhymes, it is not clear what that preference is based upon. Frequency would be the most obvious explanation, yet none of the token, type, nor prototype frequency are reliable predictors of the preferred spelling of a rhyme. Another possible explanation relates to spelling rules. It may be that certain rules are influential. One example of such a rule is the "vowel+single consonant+final silent 'e' combination triggers the 'long' quality of the first vowel" that was discussed in the introduction. However, the results do not support any influence of such rules. The six spelrimes for which this rule has relevance are clearly split in preference: in three of the cases, the silent 'e' forms are preferred, and in the other three cases, the alternate forms are preferred. There is no apparent systematicity to the preferred forms on the basis of spelling rules.

In fact, it is difficult to discern a trend of any sort in the 'preferred' spellings. An alternative explanation for the presence of preferred spellings might be found within the realm of theoretical phonology and 'underlying forms'. It may be that the 'preferred' forms are more closely related orthographically to the phonological underlying forms in the lexicon (if there are such things). A postulated underlying form for the English word *right* should be relevant here as an example. Chomsky and Halle (1968) justified the underlying form of *right* as /rixt/ on the basis of the pronunciation of the related word *righteous* (p. 233). The results of the preferred spelling of the /ayt/ spelrime indeed show that the *-ight* spelling is preferred. However, all three frequency measures also match the direction of preference in this case. It would be necessary to do detailed phonological analyses of each spelrime and then test those against performance on some task to determine the relationship of preferred spellings to underlying representations. Besides the potential problem of the choice of theory on which to base one's postulations of underlying forms, the fact that no preferred spellings were noted for eight of the twenty spelrimes suggests that any relationship of preferred spelling to underlying forms is tenuous, at best.

There might be 'subgroups' of preferred spellings. For the purposes of this task, educated or prescriptive spellings might have been preferred, for example. Students asked to participate in a linguistic experiment for the university would probably hesitate to give 'uneducated' spellings like nite and lite; therefore the *-ight* spelling would be the preferred form. Another possible explanation is that there are some spelling forms that are 'hard to spell'. One example of this might be the *-our* form for the /a^{w}r̩/ spelrime. The more predictable and 'easier' spelling for that rhyme is probably the **-ower* one, since syllabic /r̩/'s are so commonly represented orthographically in English as *-er*. Some support for this explanation might be found in English spelling errors.

Despite the 'subeffect' of preferred spellings (whatever its cause), the difference in performance between groups with oral versus written presentation was still statistically significant. This suggests that the effect of presentation mode is quite strong. The next question to be explored is whether that effect is true of all tasks tapping phonological knowledge, or whether presentation mode interacts in some way with task demands. This question was the motivation for Experiment 2a.

B. Experiment 2a

The suspected importance of the type of task used to elicit information pertaining to the question of orthographic influence in phonological judgments was discussed in a previous section of this paper. The rhyme task of the preceding experiment is one type of task. The purpose of this second experiment is to examine the role of orthography on phonological judgments in a different task.

The emphasis in phonological and phonetic experimentation has generally been on tasks involving comparison judgments of similarity of sound, discriminations between tokens, and classifications of tokens into phonemic categories (e.g., Liberman *et al.*, 1957). The latter two tasks restrict the scope to the level of single phonemes. The comparison judgments allow the use of whole words and syllables, but are an indirect tapping of phonological knowledge, and phonemicizations can only be guessed at. Naive subjects have been preferred for such experiments to prevent undue influence from knowledge of a particular theory or desired results. Thus, it has been virtually impossible to directly elicit phonemicizations, since some familiarity with phonetic transcription and phonemic theory would be required for such a task. Some investigators have resorted to training their subjects, i.e., a number of training sessions on 'broad phonetic' transcription before the actual experiment. Of course, the question then arises as to whether the subjects can be considered naive, since the experimenter has imposed both his phonetic transcriptions and his segmentation biases on them.

The problem, therefore, becomes one of devising a task that can access the phonological knowledge of naive subjects in a more direct way than comparisons and classifications. One straightforward suggestion that seemed worth trying was to ask subjects simply to write down the number of 'speech sounds' in token words.[5] No definition of the phrase 'speech sounds' was provided in the initial phase of experiment, nor were examples. It was therefore entirely up to the subject to interpret and follow the instructions. The motivation for this was to prevent biasing of responses in the direction of any *a priori* notion of phonetic segmentation.

In order to allow for meaningful comparisons between the present experiment and Experiment 1, the procedural details were kept as similar as possible in the two studies. The same words in the same order were used here as in Experiment 1. In addition, presentation mode was kept as a variable, both for consistency and for the reason discussed below. For one group the words were presented orally and visually, and for the other group, only orally.

Because the instructions were quite vague, it was expected that responses should indicate how the instructions were interpreted, i.e., what 'speech sounds' meant to different people. Presumably, if the interpretation was made in terms of the number of syllables in a word, then the

----

[5] Thanks to Dr. B. L. Derwing for the suggestion.

numbers given should be quite low, only ones and twos, in fact, for the words used, since only monosyllabic and disyllabic stimuli were presented. If, however, the basis of decision was phonemic, or something approximating the phonemic notion of the segment, then the numbers should be in the mid-range (2-5). A third possibility might be that the criterion was letter-counting, in which case the numbers would range slightly higher than for a phonemic count (2-7) and responses would be correlated to a grapheme, rather than a phoneme, index. The possibility of letter-counting as a criterion of decision is the second reason for including presentation mode as a variable. On analogy to the predictions in Experiment [illegible] letter-counting should be more prevalent in the written presentation group than the oral group. If the influence of orthography is important in this type of task, then there should be differences of some kind between the oral and written groups, just as in the first study.

Again, for reasons of consistency, subjects were asked to give an indication of their spelling ability. Although there did not appear to be any relation of spelling ability to task performance in the preceding experiment, it might have been related to the present study, assuming that a more direct accessing of phonological knowledge was involved. (Thus, poor spellers may be more prone to give syllable counts and good spellers more likely to give letter counts, for example.) Whether and whatever the relation of spelling ability to this task, the information is potentially useful.

Method

Subjects. Forty students from a high school in the Edmonton Public School System took part in this study.[6] All were native English speakers, ranging in age from 16 to 18. Twenty of the students were female and twenty were male.

Materials. The word lists (see Appendix A), the orders of presentation, and the experimental design were identical to those of Experiment 1.

Procedure. The procedure was the same as that of Experiment 1, except for instructions and scoring. Twenty students were read the words without seeing them (the oral group) and twenty of the students both saw and heard the words (the written group). Within each group, half heard order 1 (block A before block B) and half heard order 2 (block B before block A). Again, as in Experiment 1, the written group was asked to maintain pace with the experimenter as the words were read and not to get ahead silently. The students were asked to indicate on their answer sheets the number of 'speech sounds' in each word,

----

[6] Although the same school (Eastglen Composite High School) was used in this study as in the previous experiment, the group of students was different.

and no explanation of the term was provided. The exact instructions were as follows:

> Two lists of 60 words each will be read to you. For each word, please provide the number of speech sounds in the space provided. There is no 'right' or 'wrong' answer in this task. Your perception of English is what is of interest.

The students were also asked to indicate, on the same 5-point scale as in Experiment 1, a self-rating of their spelling ability. The entire experiment took about 30 minutes.

The judged number of speech sounds for each word by each student was entered directly into a computer file, coded by word and subject, and subsequent analyses were done on the file.

Results

In contrast to the previous experiment, there were virtually no significant effects when the data were collapsed across subjects. The effect of presentation type was nonsignificant, as was that of order. There was no significant difference between the performance of males and females on the task, and no significant correlation of spelling ability to performance on this task. As in the previous experiment, there was also no relation of spelling ability to sex, as indicated by a *t*-test.

The most noticeable characteristic of the data, when collapsed across subjects, was noise. The mean speech sounds score for subjects ranged from 1.058 to 3.350 for the forty subjects (see Table 4). For each subject, however, there was much less deviation. Subjects with means of around one generally put ones throughout; subjects with means between two and three had some threes and fours, but also numerous ones and twos; means over three represented scores from two to six, with no ones. Because of this separation, a preliminary grouping approximation was done by visual inspection (see Table 5), based on the subjects' mean scores. The apparent cohesiveness of the groups led to the running of a data analytic procedure to determine what groups were in the data. The Clustan hierarchy scheme (Wishart, 1978) was used. The algorithm calculates distance scores pairwise between subjects both based on the shape of the profile and its elevation. In the present case, the profile would be the speech sound scores given for each word. The 'most alike' pair (the lowest distance score) would then be 'linked' and from then on in the algorithm compared to other scores as a unit. The process continues until there is a big 'jump' in the value of the coefficient, signifying a separation of clusters.

There were just three clear clusters using the hierarchical technique (see Figure 1). These clusters were remarkably similar to the preliminary grouping (see Table 6 and compare to Table 5). The smallest group, group 1, had

Table 4

Speech Sound Means per Subject

Oral Group Subject #	Oral Group Mean	Written Group Subject #	Written Group Mean
Order 1			
1	1.392	21	1.508
2	3.167	22	3.167
3	3.350	23	3.092
4	2.042	24	2.600
5	2.242	25	2.125
6	1.308	26	1.158
7	1.800	27	2.294
8	1.233	28	1.725
9	1.225	29	3.275
10	2.667	30	2.383
Order 2			
11	2.250	31	3.017
12	2.175	32	2.483
13	2.400	33	3.242
14	2.975	34	3.100
15	1.642	35	2.733
16	1.875	36	1.328
17	3.158	37	3.133
18	1.825	38	2.075
19	2.983	39	2.942
20	1.058	40	1.092

Table 5

Preliminary Grouping

Group 1		Group 2		Group 3	
S#	Mean	S#	Mean	S#	Mean
1	1.392	4	2.042	2	3.167
6	1.308	5	2.242	3	3.350
7	1.800	10	2.667	14	2.975
8	1.233	11	2.250	17	3.158
9	1.225	12	2.175	19	2.983
15	1.642	13	2.400	22	3.167
16	1.875	24	2.600	23	3.092
18	1.825	25	2.125	29	3.275
20	1.058	27	2.294	31	3.017
21	1.508	30	2.383	33	3.242
26	1.158	32	2.483	34	3.100
28	1.725	35	2.733	37	3.133
36	1.328	38	2.075	39	2.942
40	1.092				

Figure 1

Hierarchical Clustering, Experiment 2a

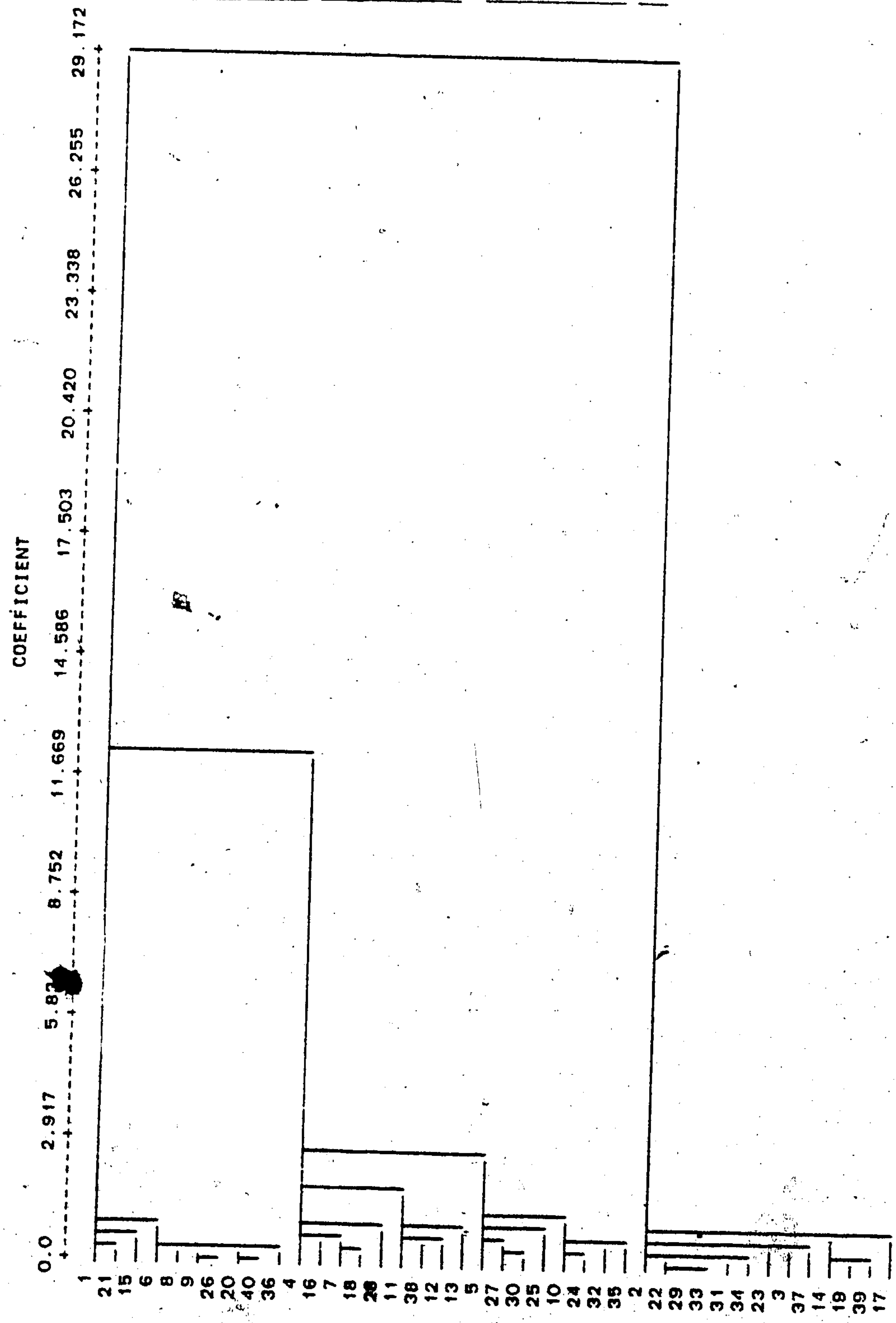


Table 6

<u>Hierarchical Clustering Analysis</u>

Group 1		Group 2		Group 3	
S#	Mean	S#	Mean	S#	Mean
1	1.392	4	2.042	2	3.167
21	1.508	16	1.875	22	3.167
15	1.642	7	1.800	29	3.275
6	1.308	18	1.825	33	3.242
8	1.233	28	1.725	31	3.017
9	1.225	11	2.250	34	3.100
26	1.158	38	2.075	23	3.092
20	1.058	12	2.175	3	3.350
40	1.092	13	2.400	37	3.133
36	1.328	5	2.242	14	2.975
		27	2.294	19	2.983
		30	2.383	39	2.942
		25	2.125	17	3.158
		10	2.667		
		24	2.600		
		32	2.483		
		35	2.733		

means ranging from 1.058 to 1.642. Since the responses given were primarily ones with a few twos, this group was designated as the 'syllabic criterion' group. The next group was the one with the greatest range of means (1.725 to 2.733). Because responses given ranged from a number of ones and twos to threes and fours, this group was designated as the 'mixed criterion' group. The final group was group 3, with means from 2.942 to 3.350. This group was designated as the 'phonemic criterion' group, for reasons to be discussed further on.

Although most of the scores of the 'syllabic' group were ones, there were a few twos. The 'two' scores were given for words such as scour, glower, and liar -- words that might well be thought of as disyllabic. The other words were clearly monosyllabic and were generally rated as such (as indicated by a score of one). There was no effect of presentation mode for this group, and no correlation of spelling ability to score. There were 10 subjects in this group. The mean scores and syllable counts can be found in Appendix B. The correlation of mean number of given speech sounds to number of syllables was .741, which is significant at the .01 level.

Just as for the syllabic group, there was no relation of spelling ability to score and no significant effect of presentation mode for the mixed criterion group. This group showed a greater response range than the other two groups, both between and within subjects. The means for subjects ranged from 1.725 to 2.733 and scores given by an individual subject ranged from 1 to 6. There were 17 students in this group. Because of the number of 'one' and 'two' scores, it was suspected that decisions as to the number of speech sounds might be based on syllables at least some of the time, and on some other basis for the higher numbers (3,4,5). To determine whether a syllabic criterion was used for the first part of the task, and then the decision criterion switched for the latter part (the impression gained from cursory examination of the answer sheets), *t*-tests were run for the first block against presentation mode and compared to those for the second block. However, there were no significant differences in performance on the two blocks on the basis of this test.

For the purposes of experimental phonology, the most interesting group was the one designated as the 'phonemic criterion' group. The basis for decision by this group seemed to be the number of phonemes in the word. This criterion crossed presentation modes. A *t*-test of overall score showed no significant difference between performance of the oral and written presentation groups. Possible differences between presentation groups were also tested at the level of individual words. *T*-tests were run for all 120 words, comparing the oral and written groups. At a .05 decision level, we would expect about one-twentieth or six of those *t*-tests to reach the significance level by chance alone. Only eight of the *t*-tests showed a significant difference between the oral and written groups, certainly not enough to claim a significant difference between presentation groups.

There was also no linear correlation of spelling ability to task, and no difference betweeen males and females. To this degree the phonemic group is similar to the other two groups. However, there was a strong correlation between the number of speech sounds given for each word and the number of phonemes in the word (*r*=.897). The phonemic representations for each word are provided in Appendix A which indicates the type of phonemic analysis utilized. The most notable feature of this analysis is the uniform treatment of all vowel nuclei as single phonemic segments. This analysis was chosen because it was the one most in accord with speakers' judgments.

It might be argued that the criterion was not on the basis of the theoretical notion of phonemes, but the more concrete grapheme. Indeed, the correlation of grapheme count to the number of speech sounds given for each was also relatively high (*r*=.758), and significant at the .01 level. The list of words with speech sound means (for the phonemic group only), phoneme counts, and grapheme counts can be found in Appendix C.

To determine the relative importance of phonology and orthography in these correlations, partial correlations were calculated. The correlation of phoneme count with speech sounds controlling for orthography was *r*=.792. The correlation of grapheme count with speech sounds controlling for phonology was only $r=.432$. Both partial correlations were significant at the .01 level. The substantially higher correlation of speech sounds with the count of theoretical phonemes should be very encouraging to phonological investigators and will be further discussed below.

The Clustan hierarchy scheme applied to the mean speech sound ratings of the words for the phonemic criterion group showed that there were three main clusters or groups of words. One group was composed of words with three clear segments -- two consonants and one vowel (with three exceptions to be discussed below). Another group was made up of words containing consonant clusters, again with two exceptions. The smallest group was a mixed bag, including all those words with two phonemes, and some with three. The groups, the words in each group, and the exceptions (underlined) can be found in Appendix D.

There were certain words that seemed to 'stand out' in some way, both in the correlation of phoneme count to speech sounds and in the hierarchical clustering. Among these were those mentioned in the preceding paragraph. Both <u>cower</u> and <u>bower</u> were in the second or 'consonant cluster' group mentioned above, despite the fact that their rhymes <u>sour</u>, <u>dour</u>, <u>scour</u>, and <u>glower</u> were in the expected groups. The two words also had means different from the number of phonemes as per Appendix A. The first group mentioned had three exceptions, of opposite types. All except those three words were made up of three phoneme segments. One exception was the word row (=/ro/), two segments instead of three, according to the phonemicizations shown in Appendix A. Its rhymes, low and show were also in a group that generally had three segments (see Group 2b, Appendix D). However, the rhymes hoe, go, and foe were in the expected two-segment group (see Group 2a, Appendix D). The second type of exception for the first group were the words quote (=/Kwot/) and mute (=/myut/), which are phonemicized as containing four, not three, segments. It would be expected that these words would be in the consonant cluster group (Group 3, Appendix D).

There also were some obvious 'outliers' in the correlation of speech sounds to phoneme counts, despite its high overall value (r=.897). Again, quote was an exception; the number given for speech sounds differed greatly from the phoneme count (3.385 vs 4). Because quote was an exception in both analyses, and its rhymes boat, dote, tote, gloat, and oat were not, it seems reasonable to suggest that there is something unique about the way the initial /Kw/ cluster of that word is perceived. Unfortunately, there was only the one word with that cluster included in the sample of test items. However, there were other words with initial consonant clusters involving a glide (/y/).

In the clustering technique, cue (=/Kyu/) was in the two segment cluster (Group 2a, Appendix D). In the correlation of speech sounds to phoneme count, the words cue and pew (=/pyu/) differed from their analyses (1.769 and 2.463, respectively, as compared to three phonemes). This might suggest that the /ky/ and /py/ are also perceived as single segments. Yet, the rhyming word view (=/vyu/) had a speech sound score quite close to the phonemic analysis (2.769). The word mute (=/myut/) was both an exception in the clustering technique and had a mean very different from the proposed phonemicization (2.923 compared to 4 segments). This exception supports the notion that consonant + y clusters are generally perceived as a single segment before /u/ (or perhaps the glide is interpreted as part of the vowel). The other words with means that differed greatly from their analyses were staffed, craft, lied, and spite. No explanation will be attempted here for these exceptions, simply because one does not immediately present itself. The exceptions presented so far are of academic interest only at this point: they suggest things about the way theoretical phonology analyzes English words as compared to how people do, but this will be discussed in the next section.

Discussion

The results of this experiment were somewhat surprising. Although caution must be exercised in the interpretation of nonsignificant results and lack of differences, it seems reasonable to conclude that presentation mode is not an important parameter affecting performance in this 'number of speech sounds' task. Comparing the results of this experiment to the first experiment, it seems reasonable to claim that the initial hypothesis that the influence of orthography varies with task type is supported. Neither written presentation nor grapheme count were significantly related to performance on the task used, unlike the earlier study, where presentation mode was the significant factor. In the present experiment the only important parameter was one that can only be inferred -- the decision criterion. This decision criterion clearly separated at least two groups of subjects (syllabic and phonemic groups); the other group seemed to involve a mixed or unstable criterion, such that the basis for decision was inconsistent for individual subjects.

The 'syllabic criterion' group gave ones and twos as the number of speech sounds. Most of the words used as test items were clearly monosyllabic, with some, such as liar, scour, sire, that might be perceived as disyllabic (because of the possibility of a syllabic /r̩/ or /ər/ interpretation). Unfortunately, there were no other polysyllabic words included, so it is impossible to judge whether the criterion used was really syllabic in character. One possibility might be that subjects were counting the number of words or meaning units, also mostly ones for the test items. Twos for words such as liar could also be explained this way; however, it would be difficult to explain twos for words such as scour on this basis. The most likely explanation is the one first suggested, viz., that the basis for decision was the number of syllables. Subject intuitions from post-experimental interviews also support this explanation. Of those subjects offering intuitions about their pe[illegible] many said specifically they were basing their dec[illegible] the number of syllables in the words.

It is impossible to do more than speculate about the mixed criterion group. Its name is for convenience only. The means for both subjects and test items were somewhere between those of the syllabic and phonemic criterion groups, but the precise criterion used by this group is unclear. In fact, there is little to be discussed about either the mixed criterion *or* syllabic group. Despite the relatively high correlation of syllable count and scores for the syllabic group, the basis for decision in either group can only be guessed at and the data are too unsystematic to permit clear conclusions about the particular test items used, either.

In contrast to the other two criterion groups, the phonemic criterion group displayed very interesting characteristics. Although orthography was not a significant factor, and neither sex nor spelling ability were related to test performance, there was one parameter that was highly correlated (phoneme index) with judgements. It is this high correlation (r=.897) that lends credibility to the assertion that the subjects in this group were using a phonemic analysis as their basis for decision, and, further, that their intuitions were largely consistent with what traditional phonemic theory might predict. The high correlation also makes the discussion of exceptions more meaningful. One is more inclined to pay careful attention to 'outliers' of a correlation of .897 between speech sounds and phoneme index than if the correlation were low.

Orthography did not seem to play a very prominent role in the speech sounds task, although there was certainly some correlation of grapheme index to number of speech sounds given. That that correlation is due to the confounding relationship between phonology and orthography is supported by the relatively higher correlation of phoneme index to speech sounds without orthography (r=.792) as compared to the correlation of orthography to speech sounds controlling for the phoneme index (r=.432). Derwing (1976) ran a similar partial correlation test on his data and found that the correlation of a phoneme index with judgments (in this case, similarity of sound) controlling for orthographic similarity was .63, whereas the correlation of grapheme index with judgments controlling for phonological similarity was only .34. The direction of support was the same as the present results and the difference in the partial correlations for the present results is even more striking. One explanation for this difference is that there were a number of words presented in the present experiment with obvious 'silent' letters (e.g., <u>thought</u>, <u>fight</u>). It is unlikely that the *-gh-* would ever influence the judgment of speech sounds, since most native speakers presumably know that combination does not represent the phonemes /g/ and /h/ (or any sound). These kinds of words acted to bring down the correlation of graphemes with speech sounds and probably resulted in an underestimation of the true correlation of orthography with speech sound scores. Derwing used the result to point out the relative importance of phonology with respect to orthography in making similarity of sound judgments. The results of the present experiment also argue for the importance of phonology in the speech sounds task and suggest that orthography does not play a very influential role overall.

The test items that did not seem to fit in, both in the clustering technique and the correlation with phoneme index, were discussed briefly in the preceding section. There was not a large number of exceptions; a total of only eighteen in the two tests, out of a possible 240. The most consistent characteristic of the exceptions was the presence of a glide element. The words <u>cue</u> and <u>quote</u> were exceptions in both the correlation and the clustering technique, and both begin with the k + glide cluster. The word <u>mute</u>, involving an m + glide cluster, was also an exception in both techniques. The other words with an initial consonant + glide cluster (pew /pyu/, view /vyu/) fit in the predicted clusters in the hierarchical technique, but <u>pew</u> was an exception in the correlation.

The other words that might be thought to involve glides are low, row, and show and all three were exceptions in the clustering technique. All were put in the three-segment groups, rather than the two-segment group that their rhymes go, hoe, and foe were in. Orthography might influence the perceptions in these cases, resulting in classifications closer to three than two segments. On analogy, the exceptions of the words bower and cower for both techniques might be a result of orthographic influence. Their rhymes dour, sour, and scour were in the predicted groups (three segments), whereas bower and cower were in the consonant cluster group. The sixth member of that spelrime (glower) was in the consonant cluster group, as predicted by the initial cluster, but may have been there for the same reason bower and cower were. For all of the *ow* words, the orthographic *w* might be influential in judgments of number of speech sounds, as might the extra orthographic vowel of the *er* sequence.

The exceptions discussed above are interesting for what they tell us about how people report (and presumably perceive) sounds of their language as compared to how phonology has traditionally claimed they are perceived. The present results strongly suggest that consonant clusters involving glides that are not represented orthographically are not judged as such by native speakers (e.g., quote, cue), whereas the presence or absence of orthographic glides seems to affect the perception of phonetically identical vowel and syllabic /r/ combinations (bower vs. sour). Such exceptions suggest the need for further phonological research exploring the perception of phonetic and orthographic glide elements. There are a number of related questions of concern to phonologists, as well, such as how the final sound of the English word sing is perceived by speakers. It may be that the [ŋ] is perceived as two sounds, rather than as one. Whether that would be a result of orthography or distributional phonology is a question that remains to be answered. Smith (1980) has convincingly argued, on the basis of distributional data, historical evidence, and evidence from speech errors, that the velar nasal in English is indeed one segment (/ŋ/), as many pre-generative phonologists had treated it. It would further round out his analysis to have some experimental perceptual data such as that collected in this study to clarify the status of this segment, as well.

Although the exceptions were noted and do seem to stand out from the general trends, no formal analysis was carried out to determine whether they were statistically different from the rest of the test items. This was due, in part, to the small sample size of the phonemic criterion group. Although not belabored, the groupings were, of necessity *ad hoc*; moreover, there were only thirteen subjects in the group in question and of those thirteen, only five were in the oral presentation group and eight were in the written group. Because of the small and unequal numbers, it is difficult to draw conclusions on the basis of strict statistical tests. It was therefore decided that more data from a 'phonemic criterion' group must be obtained before trying to make generalizations about how native speakers perceive certain English segments. This was the motivation for Experiment 8b.

C. Experiment 2b

One outcome of the sp sounds task in Experiment 2a was the clear evidence of the importance of strategy in the performance of the task. Leaving the interpretation of instructions to the subjects' discretion resulted in widely disparate, but generally systematic, differences in performance. Responses were apparently based either on a syllabic or phonemic criterion, or else on some third unknown or inconsistent criterion. A graphemic criterion did not seem evident. This dismisses one possibility of orthographic interference, but, in any case, the influence of orthography is by all indications more subtle than letter counting.

Repetition of the speech sounds task of Experiment 2a was necessary to allow generalizations with respect to presentation mode effects and orthographic influences at the individual word level. The group of subjects that seemed to be operating on the basis of a phonemic criterion in the preceding experiment was quite small (n=13) and had members of both presentation groups and orders of presentation. Elimination of some of the procedural discrepencies between subjects was desirable. Since order of presentation did not seem to influence performance at all, that was the variable eliminated in the replicate. Presentation mode was of interest because of its relation to the question of orthographic influence and was therefore retained.

To ensure that subjects were operating under a phonemic criterion, explicit instructions were provided as to what was meant by 'speech sounds' in the second phase of this experiment. A possible difficulty with this method was referred to earlier, viz., the danger of imposing one's own segmentations on the subjects prior to the experiment. For this reason, the examples provided were words that were relatively unambiguous with respect to the number of segments. No words involving glides or syllabic /r/'s were used in the examples, since these were indicated as 'trouble spots' in the preceding experiment and would be of interest here, as well.

For all intents and purposes, then, the responses for this third experiment should have been based on a phonemic criterion. They should have correlated highly with the phoneme index used in Experiment 2a and any lack of correlation for specific items should indicate where the phonemic transcriptions of Appendix A have misrepresented native speaker perceptions. Further, the results of this second speech sounds experiment should closely match those of the phonemic group in the preceding experiment (Experiment 2a). Although the decision criterion there could only be inferred, here it was imposed. If subjects were indeed making their decisions on a phonemic basis in the preceding experiment, a word-by-word analysis comparing the two groups should match closely. Additional analyses of the performance of the subjects in this experiment should also match the performance of the phonemic group of Experiment

2a.

The parameter of most interest was presentation mode, which was nonsignificant in the preceding experiment. However, the phonemic group of Experiment 2a had only five subjects in the oral group and eight in the written group. Here the number in each presentation group was equal and more meaningful statistics could be performed on the parameter of presentation mode. For reasons of consistency, the words presented in this experiment were the same as in the two preceding studies. In addition, the question of spelling ability was retained for information purposes.

Method

Subjects. Twenty grade twelve students from a high school in the Edmonton Public School System participated in the study.[7] All were native speakers of English, ranging in age from 16 to 18. There was an equal number (5) of males and females in each presentation group.

Materials. The word lists (see Appendix A) were presented to all the subjects in the order: block A, then block B. There were two different modes of presentation, oral and written.

---

[7] Thanks to Ms K. Stelck and her class at Harry Ainlay Composite High School.

Procedure. The procedure was similar to that of the preceding experiments. The words were read to all students. Half of the students heard the words without seeing them (the oral group), and half also had the words written on their answer sheets (the written group). The written group was asked to stay at the experimenter's pace, which was controlled by the students. The students were asked to indicate on their answer sheets the number of 'speech sounds' in each word read. No definition of the term 'speech sounds' was provided, but a few examples of words with the number of speech sounds, and their 'speech sound details', were provided. The exact instructions were as follows:

> For each word presented, please write down the number of speech sounds you think it contains. Five examples are given below to serve as a guide. Note that only the number of sounds is required, not the full details given in the examples.

Word	Number of Sounds	Details
cat	3	'k' sound as in 'kite'
		'a' sound as in 'ask'
		't' sound as in 'top'
debt	3	'd' sound as in 'dog'
		'e' sound as in 'egg'
		't' sound as in 'top'
creep	4	'k' sound as in 'kite'
		'r' sound as in 'red'
		'ee' sound as in 'eat'

		'p' sound as in 'pot'
laugh	3	'l' sound as in 'look'
		'a' sound as in 'ask'
		'f' sound as in 'fun'
ghost	4	'g' sound as in 'got'
		'o' sound as in 'oat'
		's' sound as in 'sit'
		't' sound as in 'top'

The instructions were reviewed until the experimenter was convinced that the students understood the task. The students were asked to give an indication of their spelling ability, on the 5-point scale used in the preceding two experiments. The experiment took about 30 minutes.

The number of speech sounds for each word by each student was entered directly into a computer file, coded by word and subject, and subsequent analyses were done on the file.

Results

To determine whether the results of this experiment were comparable to those of the phonemic group of Experiment 2a, *t*-tests were run comparing the means for each word for the phonemic group of the preceding experiment to those of this experiment. Out of a possible 120, 8 words were significantly different between the two groups. At a probability level of .05, one-twentieth of the words (6) would be significantly different by chance alone. No attempt was made to pool the data. However, the number of words scored differently was only slightly above chance, and since the correlation of scores for the phonemic group of Experiment 2a with scores for the present study was significant at the .01 level (r=.626), the two groups could at least be compared informally.

As in the preceding experiment (2a), there was no significant difference in performance between males and females, and no apparent relation of spelling ability to performance on the task. An F-test of overall score for each subject against presentation group was significant (F=6.00, p<.05). Surprisingly, the means for the oral and written groups were identical ($\bar{X}$=390.100), although the standard deviations were substantially different (7.475 and 18.303, respectively). An interpretation of the significance of the F-test would be that the oral group was more consistent in responses (as indicated by the lower standard deviation). The result of the F-test was somewhat suspect in this case, however. The statistic used here (overall score) was originally calculated to separate the syllabic and phonemic groups of Experiment 2a. There, it was relevant. It is questionable how much relevance it has when comparing more homogeneous groups (oral and written presentation groups, presumably using the same strategy). A more meaningful comparison between the two presentation groups would be word-by-word t-tests. With such a comparison, only six words were found to be different at the .05 level between the two groups, which is at the chance level. Because of this, and because of the problems with the overall score statistic for the purposes of comparing presentation groups, it was concluded that the effect of presentation mode was nonsignificant in this experiment, just as it was for the phonemic group of Experiment 2a.

As expected, there was a strong correlation of the number of speech sounds given to the phoneme index (r=.925). The correlation of grapheme index with speech sounds was lower (r=.795). As with Experiment 2a, partial correlations were taken to determine the relative importance of phonology and orthography in these correlations. The correlation of phoneme count with speech sounds controlling for orthography was r=.857. The correlation for grapheme index with speech sounds controlling for phonology was r=.568. Not surprisingly, the two partial correlations are both higher than for Experiment 2a, since the values given for number of speech sounds had a narrower range here (regression to a mean?), and the two indices are correlated. Note that all the above correlations were significant at the .01 level.

Despite the fact that the overall score means were identical and because of the differences in standard deviations discussed above, the characteristics of the two presentation groups were examined more closely. The correlation of the oral group's responses with phoneme index was r=.930, slightly higher than the overall correlation.

Predictably, the correlation of the written group's responses with the phoneme index was somewhat lower than the overall correlation (r=.898, as compared to r=.925). This trend might indicate that phonology is marginally less important with written presentation. The correlations of speech sound means with the grapheme index for the two presentation groups suggests that orthography is relatively more important for the written presentation group. The correlation of speech sound means with grapheme index for the written group was .824, as compared to the correlation of .753 for the oral group. Although there may appear to be a trend here, none of the differences were statistically significant, using Fisher's z_r transformation (Ferguson, 1976).

There were a number of words that had means for number of speech sounds that were quite different from the phoneme count. Appendix E has the list of words with their speech sound means, phoneme counts, and grapheme counts with discrepencies marked with an asterisk. The words cue, pew, and view were problems, as were cower, bower, and glower, as well as quote and mute. These correspond to the exceptions in Experiment 2a and will be discussed further. The other exceptions, buyer, chair, and show, are not systematic in any obvious way and will only be noted. It should also be noted, however, that yule was an exception and was also the first word given in the test, which might explain why it 'stood out' in this experiment only. (Recall that the presentation order was not varied in this experiment.)

Hierarchical clustering of the speech sounds given for each word by each subject showed three main clusters of words, which were similar to those of Experiment 2a. However, the groups were ordered differently from those in the preceding experiment. The first (and smallest) group here had all those words with two segments, with the exception of four items. This group relates to the 'mixed-bag' group of Experiment 2a. The second group had all three-segment words (two consonants and one vowel), with only one exception. The third group had words with consonant clusters and more than three segments, except for three words. The groups, the words in each group, and the underlined exceptions can be found in Appendix F.

In the two-segment group, yule, view, pew, and cue were exceptions. A possible explanation for the presence of yule in this group is the same one offered to explain why it was an outlier in the correlation; it was the first item presented. The other three exceptions are of the same spelrime and all have a glide between the initial consonant and the vowel (C[yu]). The exception in the three-segment group is the word mute. By analogy to the other exceptions, it is not surprising that this word with an initial m + glide cluster is an exception.

The exceptions in the consonant cluster group were buyer, cower, and bower. The word buyer was also an exception in the correlation but was not an exception in

Experiment 2a. The explanation for its being an exception might lie in the presence of the syllabic /r̩/. For similar reasons, the exceptions cower and bower are not unexpected. They were exceptions in the correlation and also in the preceding experiment. Although the third member, glower, was not an exception in the grouping established by the hierarchical clustering technique but was in the correlation, the predicted group for glower was the same as that of bower and cower. Membership in the consonant cluster group for glower might therefore be by virtue of its initial consonant cluster, or for the same reason the other two words are present as exceptions in that group, as discussed below.

There were a number of similarities between the exceptions of Experiment 2a and 2b. In the correlations, the words quote, cue, pew, mute, cower, and bower were all exceptions for both experiments. For the clustering technique, the words mute, cue, bower, and cower were exceptions common to the two experiments. It is for this reason that discussion will concentrate on these words and their traditional phonemicizations.

Discussion

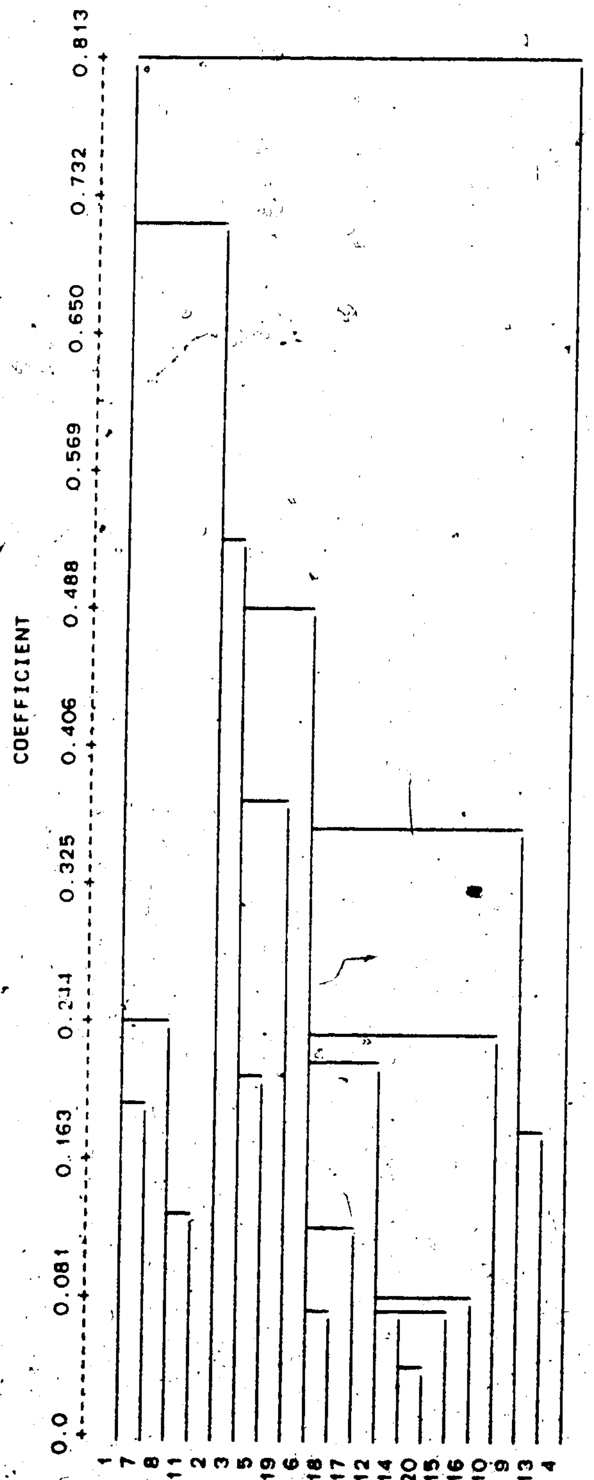
The most obvious aspect of this experiment was to point out the influence of instructions on subjects' performance in this task. All of the subjects obviously understood the instructions and based their decisions as to the number of speech sounds in a word on the intended phonemic criterion. The subjects acted so much alike, in fact, that there was no effect of presentation mode and an attempted clustering scheme refused to group the subjects into anything but a single group (see Figure 2).

Overall, phonology was more important than orthography in this task, as indicated by the partial correlations involving the speech sound scores, the phoneme index, and the grapheme index. In addition, seeing the spelling of the words did not appear to influence the answers given for number of speech sounds (as indicated by the nonsignificance of the effect of presentation mode). However, there were certain words that were rated differently than would be predicted by the given phonemicizations (of Appendix A). These are the 'exceptions' that were noted in the preceding section.

The phonemicizations that were the basis of the phoneme index used in the two speech sounds experiments counted all vowel nuclei as a single segment, thereby including the post-vocalic glide elements of diphthongs with the vowel. The reason for this method of segmentation was that it yielded predictions as to the number of speech sounds in a word that best reflected native speaker responses. The exceptions discussed earlier obviously had been assigned phonemicizations that were not in accord with native speaker intuitions.

Figure 2

Hierarchical Clustering, Experiment 2b



There was one group of exceptions (pew, cue, view) that contain pre-vocalic glides (rather than post-vocalic, as in diphthongs). There was a strong tendency for subjects to claim that there are two, rather than three, speech sounds in these words. The glide element was obviously perceived either as part of the initial consonant, or of the vowel. The two other exceptions that are relevant here are the words mute and quote. Subjects tended to say that there were three, rather than four, speech sounds in these words. The glide for these words must also be 'joined' either to the initial consonant or vowel.

The other exceptions that formed a group were all of the same spelrime (glower, cower, bower). These three words were consistently rated differently from their presumed rhymes, sour, scour, and dour. Rather than counting the $[a^w r]$ part of the words as two segments, subjects evidently rated it as three. Yet, the *-our* spellings were rated as having only two segments for that part of the word. For these words, it may be that spelling in the form of the orthographic *w* affects the perception of the glide element of the diphthong. When occurring in a word spelt without the *w*, the vowel and the glide were perceived as a single segment, but if a word was spelt with a *w*, the glide might have been perceived as separate from the vowel. An alternative explanation for the apparent differences in perception between the *-ower* and *-our* spellings would involve the perception of the syllabic /r/. For words where the /r/ is orthographically *-er*, it may be perceived as [ər] (two segments) and elsewhere as a single segment. This, too, would explain why the *-ower* words were rated as having more speech sounds than the *-our* words.

These two types of exceptions, words with initial consonant + glide clusters and words with a *-ower* ending, consistently 'stood out' in both speech sounds experiments. The two experiments were similar in other ways. Presentation mode was not an important factor in either experiment, nor was sex, or spelling ability. The only important difference between Experiment 2a and 2b was the effect of the instructions. There seemed to be at least two distinct operating criteria in Experiment 2a, syllabic and phonemic. In Experiment 2b, however, there was very little difference in performance between subjects. The high correlation of mean speech sounds for each word with phoneme index for the twenty subjects supports the claim that something like a phonemic criterion was the basis of decision for all subjects in the third experiment. However, there is still the question of what the strategy subjects are using looks like.

When given clear instructions, subjects provide consistent numbers of speech sounds. It has been noticed that when people are asked to provide the details to explain their choice of number, they have great difficulty doing so and often change their answers. It has also been noticed that those changes seem to reflect the orthography more closely than did the original answer.

It may be that an explanation for exceptions in this 'speech sounds' task relates to orthographic influence at the individual spelrime level. One test of this would be to determine whether there is a consistent spelling for nonsense words that reflects spelling rules. Acting on the suggestion that the *-ower* spelling was the productive form, a small pilot study involving nonsense words was run. Eight native English adults were asked to spell eleven words of the /a^{w}r̩/ spelrime. Five of those were real words (scour, glower, bower, sour, flour/flower) and six were nonsense forms (/sna^{w}r̩/, /sta^{w}r̩/, /kla^{w}r̩/, /ma^{w}r̩/, /fa^{w}r̩/, /ǰa^{w}r̩/). The English words were all spelt correctly, with flour the preferred form for the homophonic pair. For the nonsense forms, the majority of the spellings were *-ower* (26 as opposed to 14 of the *-our* spelling). Two of the nonsense forms (/sna^{w}r̩/ and /ma^{w}r̩/) would have English words with a different pronunciation if spelt *-ower* (e.g., snower, mower), so that form was not preferred in those cases. Interestingly, the spellings for those two nonsense words were split between *-our* and *-auer* (on analogy with the trade name Bauer, perhaps). Subjects tended to try the *-ower* spelling first in these cases and then change it. From these results, it seems reasonable to suggest that *-ower* is indeed the productive form.

Of course, the preceding is only anecdotal and impressionistic. It would be interesting to see the results of a task similar to this one but including the requirement that subjects provide phonetic details. The results of the two should be identical, of course, but the impression of this author is that there would be differences. Any differences would certainly point out the importance of instructions and the overwhelming importance of task demands. Caution must be exercised in interpreting any one task as accessing phonological knowledge, since so many variables might be involved. Orthography may be only one of several.

III. Conclusions

In the course of this study, it has become obvious that there are influences other than phonological knowledge on native speakers' phonological judgments. Orthography has definitely been shown to be one of those influences. In the rhyme task of Experiment 1, subjects were significantly influenced in their responses by being exposed to the written, as well as the oral, form of the stimuli. In contrast, presentation mode was not a significant variable in the speech sounds task. This nonsignificance cannot be interpreted to mean that there was no influence of orthography in this task, however. There is evidence at the level of individual spelrimes to suggest that orthography can be an important factor in the decision as to the number of speech sounds in a word. The statistical significance of the partial correlations of orthography with speech sound judgments controlling for phonology (in Experiments 2a and 2b) also suggest the influence of orthography is important in the speech sounds task.

The spelrimes exhibited different characteristics in the rhyme and speech sounds experiments. Not only were there consistent exceptions in Experiments 2a and 2b that were of the same spelrime, but certain spelrimes showed 'preferred' spellings in Experiment 1. An example of this is the /a^{w}r/ spelrime. The *-ower* spellings were consistent exceptions in the speech sounds experiments, and the *-ower* spelling was also the preferred form in Experiment 1. A simple way to account for both of these results is in terms of alternative syllabifications for the two /a^{w}r̩/ spellings. The syllabic /r̩/ is generally spelled as a separate orthographic syllable (-er) at the end of English words. When asked to provide a rhyme for a word ending in /a^{w}r̩/, it is therefore not surprising that subjects usually provided *-ower* words. The *-our* words are probably stored and syllabified as monosyllables. The speech sounds data also support a different syllabification for the two spellings in that the number of speech sounds given for the *-ower* spellings are consistently higher than for the *-our* spellings.

The small pilot study reported in the discussion of Experiment 2b also supports the perception of syllabic /r̩/'s in English as two segments. Subjects' spellings of nonsense forms of the /a^{w}r̩/ spelrime were overwhelmingly *-er* (*-ower* and *-auer* were given for 70% of the answers). It might well be asked at this point what this preference for the *-er* form is based on. It may be that the orthographic *w* causes a different syllabification for *-ower* words (e.g., cow+er vs sour), or it may be that the *-er* is perceived as a pseudo-morpheme, on analogy to the comparative and agentive suffixes in English. One argument against any influence of the orthographic *w* is the the fact that it was the preferred spelling for both modes of presentation (oral and written), and *-ower* words also had a greater number of speech sounds given for both modes of presentation. The results of the nonsense word study argue for the influence of spelling conventions, which is a different type of orthographic interference.

Another example of orthographic influence at the individual word level was for the word quote (=/kwot/). This word was consistently judged as having three, rather than four, speech sounds. In English orthography, the 'q sound' is almost always [kw].[8] There are literally no English words that are spelt with an initial *kw* or *cw* (OED, 1971) and few that are pronounced [kw] and not spelt with an initial *q* (one such exception is choir). Under the influence of such a strong spelling rule, it is hardly surprising that naive speakers claim that there are only three speech sounds in the word quote. For consistency in phonemic analyses, it is thus suggested that the initial cluster of the English word quote be represented as $/k^w/$ rather than as /kw/. Although this might be objected to on phonetic grounds, it more accurately represents native speaker intuitions than does the two-segment analysis.

Three of the other words that had initial consonant + glide clusters were of the *ew/ue* spelrime (pew, view, cue). All of these words stood out as exceptions in both statistical analyses of Experiment 2a and 2b, but there was no preferred spelling (Experiment 1). These words were different from the other members of their spelrime (stew, sue, true) in that they were analyzed phonemically as

---

[8] One rare exception is Qiana, the trade name for a synthetic fabric.

containing a glide (e.g., /Cyu/). They stood out as exceptions in the speech sounds task because speakers analyzed them as containing only two segments, rather than three. This suggests that the glide is perceived as either part of the vowel or of the initial consonant. Evidence from Experiment 1 supports the analysis of the initial consonant + glide cluster as one segment. All native speakers agree that <u>cue</u> and <u>sue</u> rhyme. When asked to provide rhyming words for these words, subjects gave words that had /u/ as the final segment. The only difference was the initial consonant or consonant cluster. Subjects did not give words that ended in /yu/ as a rhyme for <u>cue</u>, which suggests that the initial consonant + glide are perceived as a unit and the vowel as separate from the glide. Combined with the speech sounds data, an analysis of /C^yu/ is supported for these words. The other test item with an initial consonant + glide cluster was <u>mute</u>, another consistent exception in both Experiments 2a and 2b. The argument for <u>mute</u> to be analyzed /m^yut/ is the same as that offered above.

An interesting test of this analysis would be to conduct a speech sounds experiment and/or a rhyme experiment comparing words such as <u>pew/poo</u>, <u>cue/coo</u>, and <u>mew/moo</u>. When asked to make similarity of sound or number of speech sound comparisons between such words, speakers have difficulty providing answers, although they are confident in giving the number of speech sounds for the isolated words. There seems to be some difficulty knowing how to classify the initial consonants of these words that have traditionally been analyzed by linguists as a consonant plus the glide /y/.

Another example of this the the pair of words fuel and fool. In informal interviews, speakers do not agree that those words rhyme with each other, but claim that each rhymes with the words cool, cruel, yule, and stool. The initial /fy/ (or $/f^y/$) of fuel seems to be affecting the perception of the rest of the word only when it is compared directly to fool. Again, it would be interesting to see how experimental data relate to these impressions.

The examples that have been discussed above concern the possible misrepresentation of native speaker perceptions by traditional phonemic analyses. Although not the original motivation for this study, the delineation of such discrepencies between phonological analyses and speaker perceptions should help explain some results that until now have been rather enigmatic. Instead of blaming the "possible confounding influence of orthography" for results that do not fit with established theories and predictions, phonologists should more closely examine their assumptions about speaker knowledge and conduct needed experiments to evaluate any confounding variables.

However, it seems clear that orthography can be a very important influence on phonological judgments, depending on the type of task. In the rhyme task, presentation mode was a significant effect, but in the speech sounds task it was not. This suggests that knowledge of orthography is differentially utilized, dependent on the type of task, and therefore could potentiate or interfere with phonological judgments depending on how much phonographemic divergence there is. In ambiguous cases, where there are no morphological cues, orthography in the form of 'careful pronunciations' might be the deciding factor in making some phonological judgments. A possible example of ambiguity which might involve the need to resort to an orthographically-based "careful pronunciation" is the following dialogue:

Q: Which do you want, the ladder or the pail?

A: The [læɾr]. (ladder or latter?)[9]

Skousen (1980) discussed three possible influences of orthography on adult perceptions of phonetic representations. These were: spelling pronunciations, phonemic overlap or archiphonemes, and interpretation of "reduced" phonetic sequences such as [kæt] (can't). In these types of ambiguous cases, Skousen claimed that orthography was the determining factor in phonological judgments. In a sense, it was only in ambiguous or conflicting cases that orthography seemed to be a factor in the present study, as well. In cases where a phonetic glide is not represented orthographically (e.g., <u>cue</u>, <u>mute</u>, <u>quote</u>) people do not seem to perceive it. For words in which the glide is unambiguously represented orthographically, it is perceived

----

[9] Thanks to Mr. R. Smyth for this example.

as such (e.g., yule).

Ironically, some of the ambiguous cases are the ones of most interest to phonology. They are words that have had underlying representations postulated that differ greatly from their phonetic form. Often these representations were justified on the basis of morphologically related forms, which are also represented by the orthography (e.g., sign, signal). Of interest to the question of abstractness in phonology would be whether there is a /g/ in the underlying representation of sign. However, caution must be exercised in the determination of such questions that orthography is not heavily influencing decisions. Before exploring such abstractness controversies, it would seem reasonable to determine what ambiguities exist in the sound system of English that cause difficulties for speakers making phonological judgments. It is likely that any potential ambiguities (such as whether *q* represents one phonemic segment or two) would be resolved by speakers on the basis of whatever information is available to them. This might include spelling rules, orthographic representations, and/or morphologically related forms.

Discrimination of potential ambiguities by using all means available might also explain differences noted between adults' and children's phonological judgments. O'Neil (1972) noted that children's phonological illusions differed from adults', and attributed this to adults' knowledge of morphology and orthography. Read (1973) found differences between children's and adults' categorizations of certain phonemic sequences in English (e.g., /tr/). Children split classification of /tr/ between [tr] and [čr], whereas adult speakers rarely classified /tr/ as [čr], possibly due to the influence of orthography and spelling conventions. Additional adult/children comparisons were made by Fink (1974) involving *s* + stop clusters in nonsense words by native English speakers. Children spelled (and presumably perceived) [sb] clusters as sb whereas adults would spell them sp. Fink phrased two possible explanations for this, both in terms of the English orthography. The first was that literate speakers know that the letters *b*, *d*, and *g* do not occur after *s* in initial clusters, so they write *p*, *t*, or *k* instead. The second explanation was that an actual change in phonemic categorization occurs as speakers internalize the rules of English spelling, so that adults do not even perceive a difference between word-initial [sb] and [sp].

However, a third explanation also exists for differences between children's and adults' perceptions. Adults have a deeply entrenched set of phonological rules for their language. Even though children may have mastered the phonological rules of their language, they are more receptive to different phonologies. This is most obvious when comparing second language acquisition by children and adults. Children generally speak their second language with no accent, whereas adults rarely manage to do so. This pervasive phenomenon obviously has nothing to do with the orthographies of the languages involved.

Thus, although some differences in phonological judgments that exist between children and adults (especially for those ambiguous cases involving contextual neutralization) might be attributed to knowledge of orthography and spelling rules, others might well reflect real developmental differences in phonological knowledge. It is important, both for control in experimental phonology and research into the representation of the sound system of English, that these differences be recognized and explained.

References

Bloomfield, L. *Language*. New York: Holt, Rinehart & Winston, 1933.

Carroll, J. B., Davies, P., & Richman, B. *Word Frequency Book*. New York: American Heritage Publishing Co., Ltd., 1971.

Chomsky, N. & Halle, M. *The Sound Pattern of English*. New York: Harper & Row, Publishers, 1968.

Derwing, B. L. *Transformational Grammar as a Theory of Language Acquisition*. London: Cambridge University Press, 1973.

Derwing, B. L. Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics*, 1976, 21, 38-66.

Derwing, B. L. & Nearey, T. M. A decade of experimental phonology at the University of Alberta. 1980 (to appear).

Dolby, J. L. & Resnikoff, H. L. *The English Word Speculum (Vol. III)*. The Hague: Mouton & Co., 1967.

Ferguson, G. A. *Statistical Analysis in Psychology and Education*. Montreal: McGraw-Hill, Inc., 1976.

Fink, R. Orthography and the perception of stops after s. *Language and Speech*, 1974, *17* 152-159.

Haas, W. *Phono-graphic Translation*. Manchester: Manchester University Press, 1961.

Hall, R. A. *Sound and Spelling in English*. Philadelphia: Chilton Company, 1961.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H.

Phoneme-Grapheme Correspondences as Cues to Spelling Improvement. Washington: U.S. Government Printing Office, 1966.

Householder, F. W. *Linguistic Speculations*. London: Cambridge University Press, 1971.

Kavanagh, J. F. & Mattingly, I. G. (Eds.) *Language by Ear and by Eye*. Cambridge, Mass.: The Massuchusetts Institute of Technology, 1972.

Klima, E. S. How alphabets might reflect language. In: Kavanagh & Mattingly (Eds.), 1972, 111-116.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *54*, 358.

Maccoby, E. E. & Jacklin, C. N. *The Psychology of Sex Differences*. Stanford: Stanford University Press, 1974.

Nearey, T. M. The psychological reality of phonological representations: Experimental evidence. To appear in: T. Myers, J. Laver, & J. Anderson (Eds.) *The Cognitive Representation of Speech*. Hillsdale, N.S.: Lawrence Erlbaum, 1980.

Nelson, D. L. & Nelson, L. D. Rated acoustic (articulatory) similarity for word pairs varying in number and ordinal position of common letters. *Psychonomic Science*, 1970, *19*, 81-82.

OED: *The Compact Edition of the Oxford English Dictionary*. Oxford: Oxford University Press, 1971.

O'Neil, W. Our collective phonological illusions: young and old. In: Kavanagh & Mattingly (Eds.), 1972, 111-116.

Read, C. Children's judgments of phonetic similarities in relation to English spelling. *Language Learning*, 1973, *23*, 17-38.

Phoneme-Grapheme Correspondences as Cues to Spelling Improvement. Washington: U.S. Government Printing Office, 1966.

Householder, F. W. *Linguistic Speculations*. London: Cambridge University Press, 1971.

Kavanagh, J. F. & Mattingly, I. G. (Eds.) *Language by Ear and by Eye*. Cambridge, Mass.: The Massuchusetts Institute of Technology, 1972.

Klima, E. S. How alphabets might reflect language. In: Kavanagh & Mattingly (Eds.), 1972, 111-116.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and acros phoneme boundaries. *Journal of Experimental Psychology*, 1957, *54*, 358.

Maccoby, E. E. & Jacklin, C. N. *The Psychology of Sex Differences*. Stanford: Stanford University Press, 1974.

Nearey, T. M. The psychological reality of phonological representations: Experimental evidence. To appear in: T. Myers, J. Laver, & J. Anderson (Eds.) *The Cognitive Representation of Speech*. Hillsdale, N.S.: Lawrence Erlbaum, 1980.

Nelson, D. L. & Nelson, L. D. Rated acoustic (articulatory) similarity for word pairs varying in number and ordinal position of common letters. *Psychonomic Science*, 1970, *19*, 81-82.

OED: *The Compact Edition of the Oxford English Dictionary*. Oxford: Oxford University Press, 1971.

O'Neil, W. Our collective phonological illusions: young and old. In: Kavanagh & Mattingly (Eds.), 1972, 111-116.

Read, C. Children's judgments of phonetic similarities in relation to English spelling. *Language Learning*, 1973, *23*, 17-38.

Sapir, E. Sound patterns in language. *Language*, 1925, *1*, 37-51.

Sapir, E. The psychological reality of the phoneme, 1933. In: Makkai, V. B. *Phonological Theory*. New York: Holt, Rinehart and Winston, Inc., 1972, 22-31.

Skousen, R. English spelling and phonemic representation. Unpublished manuscript, 1980.

Smith, I. Abstract phonology and the English velar nasal. Unpublished manuscript, 1980.

Trager, G. L. *Language and Languages*. San Francisco: Chandler Publishing Company, 1972.

Vitz, P. C. & Winkler, B. S. Predicting the judged "similarity of sound" of English words. *Journal of Verbal Learning and Verbal Behavior* , 1973, *12*, 373-388.

Wijk, A. *Rules of Pronunciation for the English Language*. London: Oxford University Press, 1966.

Wishart, D. *Clustan User Manual*. Edinburgh: Edinburgh University Program Unit, 1978.

Appendices

APPENDIX A

Page 1: Block A words and their phonemicizations.

Page 2: Block B words and their phonemicizations.

1. YULE /yul/
2. SHOT /šɔt/
3. CLEAR /klir/
4. CRAFT /kræft/
5. GO /go/
6. MATE /met/
7. CUE /kyu/
8. PEEK /pik/
9. BUYER /bɑ^{y}r/
10. DALE /del/
11. DOFF /dɔf/
12. DARE /der/
13. PIE /pɑ^{y}/
14. WAIT /wet/
15. THOUGHT /θɔt/
16. LIAR /lɑ^{y}r/
17. DOUR /dɑ^{w}r/
18. BEES /biz/
19. PREEN /prin/
20. KITE /kɑ^{y}t/
21. VIE /vɑ^{y}/
22. STAFFED /stæft/
23. FREEZE /friz/
24. LOOT /lut/
25. BEER /bir/
26. BIDE /bɑ^{y}d/
27. FLAIR /fler/
28. STEW /stu/
29. SHALE /šel/
30. FOUGHT /fɔt/
31. OAT /ot/
32. CUFF /kʌf/
33. TIGHT /tɑ^{y}t/
34. SCOUR /skɑ^{w}r/
35. SPIED /spɑ^{y}d/
36. JEER /ǰir/
37. JEAN /ǰin/
38. GREEK /grik/
39. LOW /lo/
40. CHAIR /čer/
41. JUTE /ǰut/
42. TOUGH /tʌf/
43. COWER /kɑ^{w}r/
44. BRIGHT /brɑ^{y}t/
45. BEAK /bik/
46. LEAN /lin/
47. TOOL /tul/
48. HIDE /hɑ^{y}d/
49. AFT /æft/
50. CRATE /kret/
51. SHIRE /šɑ^{y}r/
52. GLOAT /glot/
53. SHOW /šo/
54. TREES /triz/
55. PEW /pyu/
56. TRAIL /trel/
57. TOTE /tot/
58. CRUEL /krul/
59. SCOOT /skut/
60. DRY /drɑ^{y}/

1. CHAFFED /čæft/
2. FEAR /fir/
3. LIED /la^{y}d/
4. SPITE /spa^{y}t/
5. FATE /fet/
6. TIE /ta^{y}/
7. DAFT /dæft/
8. FOE /fo/
9. BITE /ba^{y}t/
10. DOTE /dot/
11. SCALE /skel/
12. GOT /gɔt/
13. TRAIT /tret/
14. FREAK /frik/
15. SKY /ska^{y}/
16. TRUE /tru/
17. RULE /rul/
18. BAIT /bet/
19. BOAT /bot/
20. BRUTE /brut/
21. BROUGHT /brɔt/
22. BOWER /ba^{w}r/
23. FIGHT /fa^{y}t/
24. LAUGHED /læft/
25. SPOOL /spul/
26. SHEER /šir/
27. SPARE /sper/
28. DIRE /da^{y}r/
29. HOOT /hut/
30. CHEESE /čiz/
31. GEAR /gir/
32. SIRE /sa^{y}r/
33. TEAK /tik/
34. QUOTE /kwot/
35. SOUR /sa^{w}r/
36. SHEEN /šin/
37. FOOL /ful/
38. LAIR /ler/
39. NAIL /nel/
40. CHIDE /ča^{y}d/
41. GLARE /gler/
42. MUTE /myut/
43. MY /ma^{y}/
44. JAIL /ǰel/
45. SUE /su/
46. GRUFF /grʌf/
47. ROW /ro/
48. KNEES /niz/
49. FRIED /fra^{y}d/
50. ROUGH /rʌf/
51. REEK /rik/
52. VIEW /vyu/
53. THESE /ðiz/
54. COUGH /kɔf/
55. TROT /trɔt/
56. TEEN /tin/
57. GLOWER /gla^{w}r/
58. HOE /ho/
59. BRIAR /bra^{y}r/
60. GLEAN /glin/

APPENDIX B

Syllable counts and mean number of speech sounds given for the syllabic criterion group, Experiment 2a.

Page 1: Block A words.

Page 2: Block B words.

	Word	Syllables	Sounds
1.	YULE	2	1.7
2.	SHOT	1	1.1
3.	CLEAR	1	1.3
4.	CRAFT	1	1.6
5.	GO	1	1.0
6.	MATE	1	1.1
7.	CUE	1	1.2
8.	PEEK	1	1.3
9.	BUYER	2	1.9
10.	DALE	2	1.7
11.	DOFF	1	1.1
12.	DARE	1	1.4
13.	PIE	1	1.1
14.	WAIT	1	1.3
15.	THOUGHT	1	1.5
16.	LIAR	2	1.9
17.	DOUR	2	1.8
18.	BEES	1	1.2
19.	PREEN	1	1.2
20.	KITE	1	1.2
21.	VIE	1	1.1
22.	STAFFED	1	1.8
23.	FREEZE	1	1.3
24.	LOOT	1	1.1
25.	BEER	1	1.1
26.	BIDE	1	1.2
27.	FLAIR	1	1.4
28.	STEW	1	1.2
29.	SHALE	2	1.5
30.	FOUGHT	1	1.3
31.	OAT	1	1.3
32.	CUFF	1	1.2
33.	TIGHT	1	1.4
34.	SCOUR	2	1.9
35.	SPIED	1	1.6
36.	JEER	1	1.2
37.	JEAN	1	1.0
38.	GREEK	1	1.2
39.	LOW	1	1.0
40.	CHAIR	1	1.2
41.	JUTE	1	1.3
42.	TOUGH	1	1.3
43.	COWER	2	1.9
44.	BRIGHT	1	1.2
45.	BEAK	1	1.2
46.	LEAN	1	1.3
47.	TOOL	2	1.6
48.	HIDE	1	1.1
49.	AFT	1	1.4
50.	CRATE	1	1.4
51.	SHIRE	2	1.9
52.	GLOAT	1	1.5
53.	SHOW	1	1.2
54.	TREES	1	1.2
55.	PEW	1	1.2
56.	TRAIL	2	1.5
57.	TOTE	1	1.4
58.	CRUEL	2	1.7
59.	SCOOT	1	1.3
60.	DRY	1	1.0

	Syllables	Sounds		Syllables	Sounds
1. CHAFFED	1	1.6	31. GEAR	1	1.2
2. FEAR	1	1.2	32. SIRE	2	1.6
3. LIED	1	1.2	33. TEAK	1	1.2
4. SPITE	1	1.1	34. QUOTE	1	1.[illegible]
5. FATE	1	1.2	35. SOUR	2	1.6
6. TIE	1	1.1	36. SHEEN	1	1.0
7. DAFT	1	1.2	37. FOOL	2	1.3
8. FOE	1	1.0	38. LAIR	1	1.2
9. BITE	1	1.3	39. NAIL	2	1.5
10. DOTE	1	1.3	40. CHIDE	1	1.2
11. SCALE	2	1.4	41. GLARE	1	1.1
12. GOT	1	1.1	42. MUTE	1	1.2
13. TRAIT	1	1.1	43. MY	1	1.0
14. FREAK	1	1.1	44. JAIL	2	1.6
15. SKY	1	1.0	45. SUE	1	1.0
16. TRUE	1	1.1	46. GRUFF	1	1.1
17. RULE	2	1.6	47. ROW	1	1.0
18. BAIT	1	1.1	48. KNEES	1	1.0
19. BOAT	1	1.1	49. FRIED	1	1.3
20. BRUTE	1	1.2	50. ROUGH	1	1.0
21. BROUGHT	1	1.2	51. REEK	1	1.1
22. BOWER	2	1.8	52. VIEW	1	1.2
23. FIGHT	1	1.2	53. THESE	1	1.0
24. LAUGHED	1	1.7	54. COUGH	1	1.1
25. SPOOL	2	1.3	55. TROT	1	1.2
26. SHEER	1	1.2	56. TEEN	1	1.1
27. SPARE	1	1.1	57. GLOWER	2	1.9
28. DIRE	2	1.5	58. HOE	1	1.0
29. HOOT	1	1.1	59. BRIAR	2	1.9
30. CHEESE	1	1.1	60. GLEAN	1	1.1

APPENDIX C

Word means, phoneme index (PI), and grapheme index (GI) for the phonemic criterion group, Experiment 2a.

Page 1: Block A words.

Page 2: Block B words.

No.	Word	Mean	PI	GI
1.	YULE	2.615	3	4
2.	SHOT	2.692	3	4
3.	CLEAR	3.692	4	5
4.	CRAFT	4.462	5	5
5.	GO	2.077	2	2
6.	MATE	3.000	3	4
7.	CUE	1.769	3	3
8.	PEEK	2.923	3	4
9.	BUYER	3.231	3	5
10.	DALE	3.000	3	4
11.	DOFF	2.923	3	4
12.	DARE	2.923	3	4
13.	PIE	2.000	2	3
14.	WAIT	3.077	3	4
15.	THOUGHT	3.231	3	7
16.	LIAR	3.385	3	4
17.	DOUR	3.385	3	4
18.	BEES	2.846	3	4
19.	PREEN	3.846	4	5
20.	KITE	2.923	3	4
21.	VIE	2.000	2	3
22.	STAFFED	4.462	5	7
23.	FREEZE	3.846	4	6
24.	LOOT	3.077	3	4
25.	BEER	2.615	3	4
26.	BIDE	3.000	3	4
27.	FLAIR	3.615	4	5
28.	STEW	3.000	3	4
29.	SHALE	3.231	3	5
30.	FOUGHT	3.077	3	6
31.	OAT	2.000	2	3
32.	CUFF	2.923	3	4
33.	TIGHT	3.000	3	5
34.	SCOUR	3.769	4	5
35.	SPIED	3.615	4	5
36.	JEER	2.923	3	4
37.	JEAN	3.000	3	4
38.	GREEK	3.923	4	5
39.	LOW	2.308	2	3
40.	CHAIR	3.000	3	5
41.	JUTE	3.000	3	4
42.	TOUGH	3.000	3	5
43.	COWER	3.615	3	5
44.	BRIGHT	3.692	4	6
45.	BEAK	3.000	3	4
46.	LEAN	3.000	3	4
47.	TOOL	3.077	3	4
48.	HIDE	2.769	3	4
49.	AFT	3.000	3	3
50.	CRATE	3.923	4	5
51.	SHIRE	3.385	3	5
52.	GLOAT	3.846	4	5
53.	SHOW	2.462	2	4
54.	TREES	3.692	4	5
55.	PEW	2.462	3	3
56.	TRAIL	3.923	4	5
57.	TOTE	2.923	3	4
58.	CRUEL	4.154	4	5
59.	SCOOT	3.692	4	5
60.	DRY	2.692	3	3

		Mean	PI	GI
1.	CHAFFED	3.615	4	7
2.	FEAR	2.538	3	4
3.	LIED	2.462	3	4
4.	SPITE	3.232	4	5
5.	FATE	2.769	3	4
6.	TIE	1.846	2	3
7.	DAFT	3.692	4	4
8.	FOE	1.923	2	3
9.	BITE	3.000	3	4
10.	DOTE	2.923	3	4
11.	SCALE	3.769	4	5
12.	GOT	2.923	3	3
13.	TRAIT	3.692	4	5
14.	FREAK	3.769	4	5
15.	SKY	2.769	3	3
16.	TRUE	3.000	3	4
17.	RULE	3.077	3	4
18.	BAIT	3.000	3	4
19.	BOAT	3.000	3	4
20.	BRUTE	3.692	4	5
21.	BROUGHT	3.769	4	7
22.	BOWER	3.846	3	5
23.	FIGHT	3.077	3	5
24.	LAUGHED	3.923	4	7
25.	SPOOL	3.692	4	5
26.	SHEER	3.000	3	5
27.	SPARE	3.615	4	5
28.	DIRE	2.923	3	4
29.	HOOT	3.000	3	4
30.	CHEESE	3.154	3	6
31.	GEAR	2.615	3	4
32.	SIRE	3.077	3	4
33.	TEAK	3.000	3	4
34.	QUOTE	3.385	4	5
35.	SOUR	3.154	3	4
36.	SHEEN	3.077	3	5
37.	FOOL	2.769	3	4
38.	LAIR	2.846	3	4
39.	NAIL	3.077	3	4
40.	CHIDE	3.077	3	5
41.	GLARE	3.615	4	5
42.	MUTE	2.923	4	4
43.	MY	1.923	2	2
44.	JAIL	3.154	3	4
45.	SUE	2.154	2	3
46.	GRUFF	3.769	4	5
47.	ROW	2.385	2	3
48.	KNEES	3.000	3	5
49.	FRIED	3.769	4	5
50.	ROUGH	2.923	3	5
51.	REEK	3.000	3	4
52.	VIEW	2.769	3	4
53.	THESE	3.000	3	5
54.	COUGH	2.923	3	5
55.	TROT	3.769	4	4
56.	TEEN	2.923	3	4
57.	GLOWER	4.308	4	6
58.	HOE	2.077	2	3
59.	BRIAR	4.000	4	5
60.	GLEAN	3.846	4	5

APPENDIX D

Hierarchical clustering output for words for phonemic criterion group, Experiment 2a, exceptions underlined.

Page 1: Cluster 1.

Page 2: Cluster 2.

Page 3: Cluster 3.

Cluster 1 - 3 segments (3 exceptions)

Word #	Word	Word #	Word	Word #	Word
1	yule	113	these	116	teen
2	shot	36	jeer	93	teak
10	dale	83	fight	88	dire
11	doff	102	mute	89	hoot
18	bees	12	dare	40	chair
107	row	20	kite	86	sheer
112	view	70	dote	98	lair
6	mate	72	got	95	sour
26	bide	9	buyer	99	nail
33	tight	94	quote	104	jail
37	jean	14	wait	16	liar
41	jute	30	fought	17	dour
42	tough	15	thought	29	shale
45	beak	90	cheese	51	shire
49	aft	32	cuff	24	loot
69	bite	114	cough	47	tool
76	true	46	lean	77	rule
78	bait	96	sheen	92	sire
79	boat	8	peek	100	chide
108	knees	57	tote	28	stew
111	reek	110	rough		

Cluster 2 - mixed (2 and 3 segments)

Cluster 2a

Word #	Word
5	go
105	sue
118	hoe
13	pie
21	vie
31	oat
103	my
7	cue
66	tie
68	foe

Cluster 2b

Word #	Word
25	beer
48	hide
60	dry
75	sky
91	gear
97	fool
39	low
53	show
55	pew
62	fear
63	lied
65	fate

Cluster 3 - consonant clusters (2 exceptions)

Word #	Word	Word #	Word
3	clear	81	brought
34	scour	115	trot
43	cower	54	trees
19	preen	109	fried
23	freeze	59	scoot
74	freak	85	spool
38	greek	87	spare
50	crate	106	gruff
27	flair	120	glean
44	bright	56	trail
35	spied	58	cruel
71	scale	119	briar
73	trait	4	craft
80	brute	117	glower
101	glare	22	staffed
82	bower	61	chaffed
84	laughed	64	spite
52	gloat	67	daft

APPENDIX E

Word means, phoneme index (PI), and grapheme index (GI) for the subjects of Experiment 2b.

Page 1: Block A words.

Page 2: Block B words.

	Mean	PI	GI
1.* YULE	2.550	3	4
2. SHOT	2.950	3	4
3. CLEAR	3.750	4	5
4. CRAFT	4.800	5	5
5. GO	2.150	2	2
6. MATE	3.000	3	4
7.* CUE	2.050	3	3
8. PEEK	3.000	3	4
9.* BUYER	3.650	3	5
10. DALE	3.000	3	4
11. DOFF	3.000	3	4
12. DARE	3.050	3	4
13. PIE	2.150	2	3
14. WAIT	3.000	3	4
15. THOUGHT	3.350	3	7
16. LIAR	3.250	3	4
17. DOUR	3.300	3	4
18. BEES	3.000	3	4
19. PREEN	3.950	4	5
20. KITE	3.000	3	4
21. VIE	2.000	2	3
22. STAFFED	4.750	5	7
23. FREEZE	3.950	4	6
24. LOOT	3.000	3	4
25. BEER	3.000	3	4
26. BIDE	2.950	3	4
27. FLAIR	3.900	4	5
28. STEW	3.300	3	4
29. SHALE	3.200	3	5
30. FOUGHT	3.300	3	6
31. OAT	2.150	2	3
32. CUFF	3.000	3	4
33. TIGHT	3.250	3	5
34. SCOUR	3.850	4	5
35. SPIED	3.850	4	5
36. JEER	2.950	3	4
37. JEAN	3.000	3	4
38. GREEK	3.850	4	5
39. LOW	2.250	2	3
40.* CHAIR	3.400	3	5
41. JUTE	3.000	3	4
42. TOUGH	3.050	3	5
43.* COWER	3.700	3	5
44. BRIGHT	4.000	4	6
45. BEAK	3.000	3	4
46. LEAN	3.000	3	4
47. TOOL	3.050	3	4
48. HIDE	2.950	3	4
49. AFT	2.950	3	3
50. CRATE	3.900	4	5
51. SHIRE	3.300	3	5
52. GLOAT	3.850	4	5
53. *SHOW	2.550	2	4
54. TREES	3.900	4	5
55. *PEW	2.350	3	3
56. TRAIL	3.900	4	5
57. TOTE	3.000	3	4
58. CRUEL	4.050	4	5
59. SCOOT	3.800	4	5
60. DRY	2.950	3	3

	Word	Mean	PI	GI
1.	CHAFFED	4.200	4	7
2.	FEAR	3.050	3	4
3.	LIED	3.000	3	4
4.	SPITE	3.850	4	5
5.	FATE	3.000	3	4
6.	TIE	2.000	2	3
7.	DAFT	4.000	4	4
8.	FOE	2.000	2	3
9.	BITE	3.000	3	4
10.	DOTE	3.050	3	4
11.	SCALE	3.950	4	5
12.	GOT	3.000	3	3
13.	TRAIT	3.950	4	5
14.	FREAK	3.950	4	5
15.	SKY	2.850	3	3
16.	TRUE	2.900	3	4
17.	RULE	3.000	3	4
18.	BAIT	3.050	3	4
19.	BOAT	3.000	3	4
20.	BRUTE	3.950	4	5
21.	BROUGHT	4.050	4	7
22.	*BOWER	3.900	3	[illegible]
23.	FIGHT	3.100	3	5
24.	LAUGHED	4.000	4	7
25.	SPOOL	3.850	4	5
26.	SHEER	3.350	3	5
27.	SPARE	3.800	4	5
28.	DIRE	3.050	3	4
29.	HOOT	3.000	3	4
30.	CHEESE	3.250	3	6
31.	GEAR	3.000	3	4
32.	SIRE	3.050	3	4
33.	TEAK	3.000	3	4
34.	*QUOTE	3.550	4	5
35.	SOUR	3.150	3	4
36.	SHEEN	3.150	3	5
37.	FOOL	3.000	3	4
38.	LAIR	2.950	3	4
39.	NAIL	3.100	3	4
40.	CHIDE	3.250	3	5
41.	GLARE	3.900	4	5
42.	*MUTE	3.050	4	4
43.	MY	1.950	2	2
44.	JAIL	3.050	3	4
45.	SUE	2.000	2	3
46.	GRUFF	3.750	4	5
47.	ROW	2.300	2	3
48.	KNEES	2.950	3	5
49.	FRIED	4.000	4	5
50.	ROUGH	3.050	3	5
51.	REEK	3.000	3	4
52.	*VIEW	2.600	3	4
53.	THESE	3.200	3	5
54.	COUGH	3.000	3	5
55.	TROT	3.950	4	4
56.	TEEN	2.950	3	4
57.	*GLOWER	4.400	4	6
58.	HOE	2.000	2	3
59.	BRIAR	4.000	4	5
60.	GLEAN	3.950	4	5

APPENDIX F

Hierarchical clustering output for words for Experiment 2b, exceptions underlined.

Page 1: Cluster 1.

Page 2: Cluster 2.

Page 3: Cluster 3.

Cluster 1 - 2 segments (4 exceptions)

Word #	Word
1	yule
112	view
31	oat
55	pew
39	low
107	row
53	show
21	vie
66	tie
68	foe
105	sue
118	hoe
103	my
5	go
13	pie
7	cue

Cluster 2 - 3 segments (1 exception)

Word #	Word	Word #	Word	Word #	Word
2	shot	89	hoot	48	hide
98	lair	91	gear	49	aft
6	mate	93	teak	60	dry
8	peek	97	fool	116	teen
10	dale	111	reek	75	sky
11	doff	114	cough	76	true
14	wait	36	jeer	15	thought
18	bees	70	dote	33	tight
20	kite	102	mute	30	fought
24	loot	104	jail	28	stew
25	beer	26	bide	29	shale
32	cuff	108	knees	40	chair
41	jute	47	tool	51	shire
45	beak	88	dire	16	liar
46	lean	42	tough	90	cheese
57	tote	83	fight	99	nail
63	lied	110	rough	100	chide
65	fate	12	dare	17	dour
69	bibe	62	fear	95	sour
72	got	78	bait	86	sheer
77	rule	92	sire	96	sheen
79	boat	37	jean	113	these

luster 3 - consonant clusters (3 exceptions)

ord #	Word	Word #	Word
3	clear	38	greek
35	spied	50	crate
94	quote	52	gloat
19	preen	54	trees
23	freeze	87	spare
71	scale	59	scoot
73	trait	64	spite
74	freak	85	spool
80	brute	106	gruff
.15	trot	58	cruel
.20	glean	61	chaffed
67	daft	119	briar
.01	glare	117	glower
44	bright	9	buyer
.09	fried	34	scour
81	brought	43	cower
84	laughed	82	bower
56	trail	4	craft
27	flair	22	staffed