The influence of orthography on spoken word recognition in Bangla

MOUMITA MUKHERJEE, *The English and Foreign Languages University*

ABSTRACT

The lexical representation of words constitutes the phonological, orthographic and semantic information about a word, which is accessed together despite the task demanding only one aspect of the information. The role of orthography in word recognition tasks has been validated, though its influence on phonological tasks is lesser known. Recent studies in psycholinguistics have begun to investigate the possible influences of orthography on the auditory processing of words. The present paper reviews studies that have looked at orthographic influence on phonological tasks, and reports findings from a Rhyme-monitoring task in Bangla, to examine the role of orthography in auditory processing.

1 Introduction

What do we hear when we hear a word? What are the levels of information that we access to recognise them? Do the orthographic representations of these words play any role in their recognition? These are some questions that shall be addressed in the present study, specifically with respect to the influence of orthography on auditory word recognition. The target language that we investigate is Bangla.

To recognise words, listeners are required to match speech signals to the phonological representations (of these words) that are stored in the lexical memory (Peereman, Dufour, & Burt 2007). Thus, it would appear that information accessed and retrieved directly pertains to the task at hand i.e. those features which are needed for the task are accessed and other features that are not demanded by the task are ignored. If this were the case, in an auditory processing task only phonological representations would be accessed, and orthographic information would neither be retrieved nor accessed directly or indirectly. However, several studies (Seidenberg & Tanenhaus 1979; Dijkstra, Roelofs, & Fieuws 1995; Castles, Holmes, Neath, & Kinoshita 2003) have shown that orthographic variables which were not demanded by the task actually impacted auditory word processing.

In this paper, we intend to investigate whether this orthographic interference is observed in the auditory word recognition of Bangla speakers as well. We used a rhyme-monitoring test to assess whether certain inconsistencies in Bangla spellings affected the response times (RTs) of the participants. This shall be discussed in detail in the later sections.

The current paper has been organised as follows: Section **2** lists details of the studies that have shown orthographic influence in word processing tasks, followed by the theoretical models that explain auditory processing in Section **3**. A note on Bangla orthography has been provided in Section **4**, and finally, the study on Bangla has been presented in Section **5**.

2 Orthographic influence on phonological tasks

In this section, we report studies which have in some way or the other witnessed or reported orthographic influences on auditory word recognition tasks. We group them under tasks which looked at lexical decision/rhyme detection tasks and phoneme monitoring tasks with reference to the speed of response i.e. reaction time latencies. The experiments are largely done in English, but we report a few in other languages as well.

2.1 Lexical Decision and Rhyme Monitoring tasks

The first work to report orthographic influence during a word recognition task was Meyer, Schvaneveldt and Ruddy (1974), who found that lexical decision was faster when the second words matched the primes in terms of orthography and phonology (*couch–pouch*) than when they matched for orthography and not phonology (*touch–couch*). They explained that once the prime is encoded orthographically and phonologically, it leads to an expectation of a similar word, thus biasing them. In the case of *touch–couch*, such bias needed to be reversed thus leading to a latency.

One of the earliest and most influential sets of studies to support the claim that orthographic access may be an automatic component of lexical access in spoken perception was reported by Seidenberg and Tanenhaus (1979) in rhyme judgement tasks. They presented participants with a visual or auditory cue word, followed by a list of five auditory target words. Participants pressed a button as soon as they detected a word that rhymed with the cue. Perhaps not surprisingly, the results showed that with visual presentation of the cue word, responses were faster when the cue and target were orthographically similar (*stroke–joke*) than when they were not (*stroke–soak*). More importantly, however, a similar pattern was found with the auditory presentation of the cue.

In a further experiment, using a simpler procedure, Seidenberg and Tanenhaus (1979) aurally presented these rhyming pairs. The critical variable was whether the target word was orthographically similar or different from the cue word (*e.g., 'pie–tie' and 'rye–tie'* respectively). They found that monitor latencies to detect orthographically different rhymes were longer than latencies to detect orthographically similar rhymes, whether cue words were presented aurally or visually.

On replicating this orthography effect in the next batch of experiments, they used only auditory presentation of the cue word and a larger sample of items, and found similar results. For the rhymes, averages were 779 ms for orthographically similar pairs, and 878 milliseconds for dissimilar pairs. Hence similar word pairs were judged 99 ms faster than dissimilar pairs. The authors concluded that ''both visual and auditory stimuli may be encoded in terms of both visual and auditory features . . . auditory encoding does not always occur to the exclusion of visual information.'' (Seidenberg and Tanenhaus 1979: 554).

Ziegler and Ferrand (1998) focused on the consistency-regularity issue of words and found that the inconsistency in the spelling-to-sound mapping adversely affects word perception and reading aloud. In a visual lexical decision task in English and French (Stone et al. 1997; Ziegler, Montant, & Jacobs 1997), the authors found that words with phonological rimes that could be spelt in more than one way (e.g., /-ip/ may be spelt -eap or -eep) produced slower correct 'yes' responses and more errors than did words with phonological rimes that could be spelt in only one way (e.g., /-uk / may only be spelt -uck).

Ziegler and Ferrand (1998) thus found that lexical decision tasks were influenced by the orthographic consistencies of words. It was revealed that inconsistent words (words containing rimes that could be spelt in more than one way) led to delayed response time in comparison with consistent words (words containing phonological rimes that could be spelt in only one way).

2.2 Phoneme-monitoring task

One task that is widely used in spoken word recognition is phoneme monitoring, in which subjects push a response button as soon as they identify a target phoneme in a spoken word or nonword. It is assumed that this task is performed using phonetic or phonological representations of words, and should have no use of orthographic information. To test whether an orthographic representation of the words is employed as well, Dijkstra et al (1995) conducted an experiment in which Dutch subjects monitored for phonemes with either a primary or secondary spelling in phonologically matched spoken words and nonwords. Phoneme monitoring times were slower when the phoneme had a secondary spelling than when it had a primary spelling. The effect monitoring times were faster for words than for nonwords.

These findings indicate that an orthographic representation of words is engaged in phoneme monitoring. In another phoneme monitoring task conducted by Castles et al (2003), it was revealed that it was easier for adults to perform phoneme deletion tasks for orthographically related pairs than those which were not. For instance, they found it easier to delete the segment /rə/ from struggle than /wə/ from squabble. Phoneme reversal tasks yielded similar results. When the same experiment was conducted on Grade 5 children, the findings were analogous. In each of these cases, therefore, the influence of orthography on phoneme detection tasks was very pronounced.

3 Theoretical models of lexical access and spoken word recognition

The auditory encoding of words has always played a major role in their recognition, storage, and retrieval; and there has been evidence of auditory encoding in the perception of visual as well as aural stimuli. As Norman opines, "it was (and is) commonly accepted that linguistically based materials — printed words — entered the visual system and then was transformed into an auditory or articulatory form in the short-term memory." (Norman 1972: 277).

Several word recognition models draw heavily on this assumption and we will discuss them in the following subsections.

3.1 Logogen Model

The Logogen Model proposed by J. Morton in 1969 was a model of speech recognition that sought to explain how the human mind processes spoken or written words. As is clear from the name, the basic unit of this model is called the 'logogen'. Morton defines it as "…a device which accepts information from the sensory analysis mechanisms concerning the properties of linguistic stimuli and from context producing mechanisms" (Morton 1969). He claims, when the logogen accumulates more than the threshold amount of information, a response becomes available. This information shapes the definition of the logogen. The information was described by Morton as the members of the sets of attributes that he called *Si*, *Vi*, and *Ai*, referring to the semantic, visual, and acoustic sets, respectively.

According to this theory, therefore, a logogen of a word is said to constitute its meaning, spelling, and pronunciation, or, its semantic, orthographic, and phonological information. The process of word recognition takes place when its logogen passes a threshold and its encoded information becomes available for output. Hence, it can be seen that a single set of logogens is fed both by visual and auditory features.

3.2 Spreading Activation Model

This model of Collins and Loftus (1975) proposes that the process of word recognition entails certain consequences which are subsumed under the notion of what they call "spreading activation". They state that there exist certain interconnected semantic and lexical networks where each node in these networks represents a particular word or feature.

During the process of word recognition, when one such node is 'activated', this activation subsequently spreads along all the interconnected networks. Presentation of one word, therefore, entails the activation of all such words that are semantically or orthographically related to it. A link is established between the semantic, orthographic, and phonological information of words.

This has several important implications. Just as the phonological and semantic codes are rendered available both in the case of auditory as well as visual word recognition, so should the orthographic code, since its means of access is just as similar as the other two codes.

3.3 TRACE Model

A highly interactive model of auditory word processing, the TRACE model was developed by James McClelland and Jeffrey Elman in 1986. It is a top-down word processing model which focuses on context dependency in word recognition, claiming that knowledge of the lexicon aids in the process of acoustic perception.

According to this model, all the elements constituting a word are represented in the form of a network by several connected nodes. These nodes can be affected positively as a result of both lower levels of representation like phonemic or featural properties as well as higher levels such as the sentential information. Consequently, there exists a higher possibility of recognising words that are appropriate to a particular context than those that are not. Further, each node in the networks has different levels of activation and a particular threshold which determines the point that which this level of activation can influence other connected nodes. This influence may be either facilitatory (positive) or inhibitory (negative). There is active competition between all the nodes to get selected. The final word node that dominates all the other nodes is the one that gets recognised.

3.4 Cohort Model

The Cohort Model (Marslen-Wilson & Welsh 1978) proposes that a word's visual and auditory inputs are directly mapped onto a word which exists in the hearer's mental lexicon. Thus, each time a person hears a speech segment, all such words beginning with that segment get 'activated'. As the number of these segments increases, irrelevant words which do not have matching segments get eliminated. Finally, the word which consists of all the segments that correspond to the input signals gets selected.
This entire process takes place in three stages:

• Access: This is the stage at which hearers encounter the first few sound segments of a word which results in the activation of all such words which begin with the same segments in the hearer's mental lexicon. All these possible words are called 'cohorts'.

• Selection: As more and more sound segments are heard, there is a decrease in the activation of those words which no longer match the speech signals. As a result, they get eliminated. This stage is called 'activation and selection' and continues till what is called the 'recognition point', or the stage at which all the competitors have been eliminated, and only one word remains.

• Integration: It is at this stage that the syntactic and semantic information of the words are encoded and integrated into the higher levels of utterance.

The original framework of this model was later reworked to account for the influence that context had on aiding the process of elimination of competitors. Further, it also sought to account for the impact that coarticulation had on activation resulting in minor acoustic mismatches.

What all these studies seem to point at is the fact that there is a continuous interplay between the semantic, phonological, and orthographic contents of words so far as their recognition or retrieval is concerned. Consequently, each of these layers should have an impact on the other. Just as the semantic and phonological codes of a word become available in visual word recognition, so should the orthographic code, when it comes to auditory word recognition. It is this possible influence of orthography on the other levels of information that shall be the focus of this study.

4 Bangla Phonology-Orthography mapping

There exist several writing systems in the world, and each lays out its orthographic structure in different ways, using different phonological/semantic units to which each grapheme maps. While logographic and ideographic systems like Chinese, represent complete words in a grapheme, Korean maps a morpheme, Japanese and Cherokee, a syllable (as a unit, unanalysable into phonemes), and languages like English, German, and Finnish, a phoneme (the alphabetic script). The distinction between writing systems is not watertight as some languages may resort to the use of modified or mixed systems that cannot be strictly classified.

An alpha-syllabary, like Bangla, is one such mixed system. What is significant, however, is the fact that in spite of their diversity, it is assumed that all these orthographic systems tend to influence the perception and cognition of words in literate speakers. It is the extent and nature of this influence that shall be the main concern of this study.

What concerns us, in particular, is the inconsistency in some segments of Bangla spellings and if or/and how they affect auditory word recognition. To this end, some of these inconsistencies have been discussed below.

4.1 Vowels

It has been observed that for vowels, three letters can be read with more than one phonetic form, i.e. অ which represents /ɔ/ and /o/ (1); আ which represents /a/ and /æ/ (2); and এ which represents /e/ and /æ/ (3).

(1) অ read as /ɔ/ with actual values /ɔ o/: অনেক (anek) /ɔnek/ ‘a lot’; অতি /oti/ (ati) ‘very’

(2) আ read as /a/ with actual values /a æ/: আমি (Ami) /ami/ ‘me’; আমেরিকা (AmerikA) /æmerika/ ‘America’

(3) এ read as /e/ with actual values /e æ/: এবং /eboŋ/ ‘and’; এমন /æmon/ ‘like’

This is a case of orthography-phonology inconsistency and should affect reading aloud unless the phonological contexts in which a particular sound value is to be used are specified.

The second kind of inconsistency we observe is a phonology-orthography inconsistency, where a sound can be represented in more than one orthographic form. This is true of the two vowel graphemes ই and ঈ which represent /i/ (4) and উ and উ (5) which represent /u/. This would cause difficulty in spelling.

(4) ই read as /rɔʃʃo i/ with actual values /i/: ইচ্ছে (icche) /iʧʧ^{h}e/ ‘wish’

ঈ read as /dirgho i/ with actual values /i/: ঈগল (Igol) /igɔl/ ‘eagle’

(5) উ read as / rɔʃʃo u/ with actual values /u/: উঠ (uth) /ut^{h}/ ‘camel’

উ read as /dirgho u/ with actual values /u/: উড়ু (Uru) /uru/ ‘thigh’

One of the other confusions relates to the absence of a diacritic representation of the grapheme অ which is the inherent vowel in this language. Hence, they may be articulated as /o/, /ɔ/, or sometimes even without the sound of the inherent vowel. The choice is phonologically driven.

4.2 Consonants

Bangla consonants show a different sort of mapping inconsistency. The first is the presence of consonant sets which are phonemically indistinct i.e. they represent one particular phone. For instance, জ /borgio jɔ/ and য /ɔnt^{h}ost^{h}o jɔ/ which represent /jɔ/, ণ /modd^{h}anno nɔ/ and ন /dɔnto nɔ/ which denote /n/; ত /t̪ɔ/ and ৎ /k^{h}ɔndo /t̪ɔ/ for /t̪/; and স /dɔnto ʃ/, শ /talobbo ʃɔ/, ষ /modd^{h}anno nɔ/, which represent /ʃ/.

The distinctions however remain in orthography — a case of phonology-orthography inconsistency. As discussed earlier, Bangla also has several consonant clusters. Usually, consonant clusters which appear word-finally are always vocalic. They are followed by the sound /o/. For instance:

- শব্দ (shobdo) /ʃɔbdo/ 'words'
- অঙ্ক (onko) /ɔŋko/ 'mathematics'
- ক্লান্ত (klanto) /klanto/ 'tired'

However, in some borrowed words, the final consonant is non-vocalized, as in the following cases:

- বোর্ড (bord) /bord/ 'board'
- বন্ড (bond) /bɔnd/ 'bond'
- পোস্ট (post) /post/ 'post'
- ফার্স্ট (farst) /pʰarst/ 'first'

These variations lead to confusion while reading. What needs to be investigated is whether they lead to delays in word retrieval and recognition as well.

4.3 Consonant clusters

One of the biggest inconsistencies lies in the use of consonant clusters, where the second consonant in orthography is /j/, /m/, or /b/. These consonants seem to have variable effects depending on the position they occur in. In word-final positions, they phonologically geminate the first consonant.

4.3.1 ma-phala

In word-initial positions, for instance, in clusters like স্ম /sm/ and শ্ম /shm/ the sound of /m/ is lost and the preceding character gets nasalised. This can be noticed in words like স্মরণ(smaran) /ʃɔ̃ron/ 'to remember'.

Word medially and word finally, however, the /m/ sound is lost and the consonant sound preceding it gets geminated. This happens in clusters like ত্ম /tm/ and দ্ম /dm/ as well:

- ছদ্ম(chhadma) /ʧʰadda/ 'disguise'
- আত্মা(atma) /ãtta/ 'soul'
- ভীষ্ম(bhishma) /bʰiʃʃo/ 'name of a mythological character'

As a result, the pronunciation of these words is similar to those of words with geminate consonant clusters like দ্দ /d̪d̪/ ত্ত /t̪t̪/ and so on.

4.3.2 ja-phala

The case of ja-phala is also similar. Usually, it can appear in all the word positions. However, this has an effect on the pronunciation of these words. Depending on its place of occurrence, ja-phala has the following pronunciations.

Word initially, there is no change in the utterance of the consonant preceded by the ja-phala. For example, দ্যুত (dyuta) /d̪uto/ ‘gamble’. In word-medial and word-final positions, it can result in three types of utterances

1. Repetition of the preceding sound segment such that the same sound gets geminated. This is observed in words like বাল্য (balyo) /ballo/ ‘child’ and মাল্য (malya) /mallo/ ‘garland’.

2. When preceded by an unvoiced aspirated consonant, the resultant sound unit consists of the sound of the unvoiced unaspirated consonant and its unvoiced aspirated form. For example, মিথ্যে (mithya) /mittʰa/ ‘lie’.

3. A voiced aspirated consonant preceding ja-phala results in a voiced unaspirated consonant and its voiced aspirated form. An example of one such word is অসভ্য (asabhya) /ɔʃobbʰo/ ‘uncivilised’.

4.3.3 ra-phala

Inconsistencies in spellings also occur in words that have consonants with modifiers of the consonant /r/. This happens for words with ra-phala followed by the /i/ sound and the diacritic of ঋ (ৃ) /ri/. Thus, spellings of words like প্রিয় /priyo/ ‘favourite’ and তৃণ /trino/ ‘grass’; both of which have a /ri/ sound, lead to confusion.

/r/ has yet another modifier called reph which appears at word-medial and word-final positions only. For example, in চর্চা (chorcha) /ʧɔrʧa/ ‘practice’, it appears in the word-medial position; and in তর্ক (torko) /tɔrko/ ‘argument’, it appears word-finally.

Full and half consonants: The inconsistency in spelling arises from the fact that the consonant /rɔ/ itself can appear word-medially too, as in words like দরজা (dorja) /dɔrja/ ‘door’. Clearly, this results in spelling variations in the word medial positions.

The consistencies discussed above can be summarised as follows:

(1) Consonant Allographs: জ /borgio jɔ/ and য /ɔntʰostʰo jɔ/ for /jɔ/
ণ /moddʰanno nɔ/ and ন /dɔnto nɔ/ for /n/
ত /t̪ɔ/ and ৎ /kʰɔndo t̪ɔ/ for /t̪/
স /dɔnto ʃ/, শ /talobbo ʃɔ/, ষ /moddʰanno nɔ/ for /ʃ/

(2) Vowel allographs: ই (ি) and ঈ (ী) for /i/ উ (ু) and ঊ (ূ) for /u/

(3) C_1C_2–CC: jo-phala and geminating consonants, as in [গদ্য (goddo) /goddo/ ‘prose’] and [হদ্দ (hoddo) /hɔddo/ ‘limit’]

(4) C_1C_2–C_1C_3: ma-phala and ja-phala, such as [গ্রীষ্ম (grisma) /griʃʃo/ ‘summer’] and [দাস্য (dashyo) /daʃʃo/ ‘servitude’]

(5) Full–half akshara: reph and র /rɔ/ in words like [গর্ত (gorto) /gɔrto/ ‘hole’] and [করত (korto) /korto/ ‘did’]

Keeping in mind these different types of inconsistencies, the task reported below uses a rhyme detection task, where one pair has two words which are orthographically and phonologically similar, and the other pair has ones that are orthographically different but phonologically similar. The second aim was to check whether particular kinds of inconsistency have a differential effect on the rhyme detection task and the speed with which rhymes are detected.

5 The Study

5.1 Objectives

The aim of the present study was to investigate the role of orthography in spoken word recognition in Bangla. The inconsistencies resulting from the factors discussed in the previous section were considered, and the interference resulting from them (if any) was investigated.

The questions that this study wishes to find answers to are the following:

• Do Bangla speakers automatically co-activate orthographic representation when they make judgements on auditory words?

• Does orthographic inconsistency affect their auditory word processing speed?

• Do specific kinds of inconsistency affect their word processing more than others?
What the study wishes to throw light on is the relationship shared between the various levels of information stored in the lexical representation of Bangla words.

5.2 Participants

40 (20=female and 20=male) native speakers of Bangla participated in the experiment. They were all above 18 years with a mean age of 23;0 (SD=2.85). Each participant had had formal education in Bangla and could read, write, and speak the language with fluency. None of them had reported any sort of hearing problems or neurological disorders.

5.3 Task Stimuli

28 pairs of target word pairs and 10 fillers were used for the rhyme detection task. The target words consisted of pairs of disyllabic (CV.CV or CVC.CV) Bangla words which were phonological rhymes. Though the frequency of the words was not controlled, most of the words used in the experiment were very common. It was intuitively felt that about one-seventh of the words were less frequent.

Two versions of the stimuli were designed consisting of 14 target pairs each. There were two conditions:

1. Where the words were phonologically as well as orthographically similar, as in [ছদ্ম–পদ্ম (chhadma–padma) /ʧ^{h}ɔddo–pɔddo/ ‘disguise–lotus’]: [14 pairs]
2. Where the words were phonologically similar but orthographically dissimilar, as in [ছদ্ম–সদ্য (chhoddo–shoddo) /ʧ^{h}ɔddo–s^{h}ɔddo/ ‘disguise–word’]: [14 pairs]

The orthographically dissimilar target words were designed such that the various spelling inconsistencies mentioned earlier could be taken into account.

10 filler non-rhymes were included in each version of the stimuli. These non-rhymes were categorised into two types:

1. Where mismatch occurred in the onset consonant of the rhymes of the words [C_1V_1–C_2V_1] such as বালা /bala/ ‘bangle’ and জামা /jama/ ‘dress’: 5pairs
2. Where the mismatch occurred in the nucleus of the rhymes of the words [C_1V_1-C_1V_2], as in বালি /bali/ ‘sand’ and কালো /kalo/ ‘black’: 5pairs

5.4 Experimental design

The task was a rhyme detection task, where participants listened to a pair of words and said whether the words rhymed or not by pressing YES/NO button on a laptop. The word pairs were manipulated for their orthographic (dis)similarity. The accuracy and time taken to respond were compared. A practice trial consisting of six filler items (3 rhymes and three non-rhymes) was conducted before the main experiment.

Each participant could hear only one version of the stimuli i.e. either orthographically similar targets or dissimilar ones. This was done to avoid practice effects. 10 fillers consisting of non-rhymes were created. The targets and the fillers were randomized. All the stimuli were recorded at a sampling rate of 44kHz. The entire experiment was conducted using the PsychoPy software.

5.5 Results

We discuss the results in terms of accuracy, the response time for the two conditions (orthographically similar and dissimilar), and also with respect to the different types of phonology-orthography mapping inconsistencies in Bangla.

Accuracy: The accuracy rate of rhyme detection was above 95% suggesting that the task which demanded the detection of auditorily presented pairs was done accurately. The accuracy for orthographically similar pairs was 98.21% (550 correct responses out of 560 responses) while that of the orthographically dissimilar pairs was 96.42% (540 correct responses out of 560 responses. Though the rhyme detection accuracy was lower for dissimilar pairs, the difference was not significant. This suggests that orthographic inconsistency in the pairs did not affect accuracy decisions.

Response time: Though the accuracy of rhyme detection does not show a difference between the dissimilar and similar pairs, we found a significant difference in response time i.e. time taken to decide whether the two words in the pair rhyme or not. It was observed that participants took longer (324.1107ms) to identify rhymes that were orthographically dissimilar (Mean=4314.1607, SD=396.821) than those which were similar (Mean=3990.05 SD=288.561). The difference was statistically significant [$F\ (1,78) = 19.1385\ p<0.05$].

Phonology-orthography mapping inconsistency and response time: In the data set we looked at five kinds of inconsistencies, with the objective of examining whether any kind of inconsistency interferes more with rhyme judgements than others.

In the case of the first type, which consisted of words having a difference in spelling arising from the use of consonant diacritics and consonant geminates (C_1C_2–CC), the mean

RT for orthographically similar pairs was 4252.15ms (SD=158.058) while that of orthographically dissimilar pairs was 4615.13ms (SD=350.203), the difference in the RTs being 362.98ms. There was a statistically significant difference between the RTs [$F(1,78) = 35.6984$ $p < 0.05$].

Similarly, dissimilar pairs which differed in the second orthographic consonant in the cluster pairs i.e. $C_1C_2–C_1C_3$ (Mean=4418.15ms SD=476.279) took longer to process (a difference of 601.85ms) than similar pairs (Mean=3816.3ms SD=196.698) and the difference was significant [$F(1,78) = 54.5657$].

In type 3 words, where the spelling variations resulted from the use of half akshara (reph) and the full consonant /r/ in word medial positions, the mean RT for orthographically similar pairs was 4016.13ms (SD=394.197) and that of the orthographically dissimilar pairs was 4403.65ms (SD=368.715). The difference (387.52ms) in this case too was statistically significant [$F(1,78) = 20.5769$ $p < 0.05$].

The next group consisted of words where different consonant allographs led to spelling variations. In this case, the mean RT for orthographically similar words was 4012.39ms (SD=380.222) and for orthographically dissimilar pairs was 4296ms (SD=366.115); their difference being 283.61ms. There was, thus, a significant difference in the RTs [$F(1,78) = 11.5480$, $p< 0.05$).

A table consisting of the mean and SDs of the RTs for each pair type has been provided below:

Type	Similar RT	Dissimilar RT	F	P
Consonant diacritic-Geminate (C_1C_2-CC)	4252.15 (SD=158.058)	4615.13 (SD=350.203)	35.6989	0.0000
Consonant Cluster-Consonant Cluster (C_1C_2-C_1C_3)	3816.3 (SD=196.698)	4418.15 (SD=476.279)	54.5657	0.0000
Half Akshara-Full Consonant	4016.13 (SD=394.197)	4403.65 (SD=368.715)	20.5769	0.0000
Consonant Allographs	4012.39 (SD=380.222)	4296 (SD=366.115)	11.5480	0.0011
Vowel Allographs	3910.5 (SD=318.943)	4085.1 (SD=271.244)	6.9562	0.0101

Table 1. Mean and SD of RTs for each pair type

The given data indicates that the difference in RTs was maximum for type 2 words (C1C2–C1C3) words, i.e. 601.85ms and least for the fifth type which had dissimilarity in vowel diacritics (174.6ms). Further, there is a significant difference between the mean RTs of the first three types of words (Mean=450.78ms, SD=131.407) from that of the last two types (Mean=229.105ms SD=77.081). The difference between these two groups was found to be statistically significant [$F (1,78) = 33.69$, $p <0.05$].

In the case of the fillers, the accuracy rate was 91.25%. It was observed that there was a significant difference of 344ms between the RTs of Type 2 non-rhymes (where differences arose because of the nucleus) from that of Type 1 non-rhymes (which had a different onset) [$F (1,78) = 15.8478$, $p < 0.05$].

Filler Type	Mean RT (in ms)	Standard Deviation
Type 1: C_1V_1- C_2V_1	4091.19	404.7343
Type 2: C_1V_1- C_1V_2	4435.8	368.688

Table 2. Mean and SD of filler items

5.6 Discussion

What the results of the rhyme monitoring reveals is that there was a significant influence of orthography in spoken word processing in Bangla. The source of this orthographic effect is a question that needs to be investigated further.

Seidenberg and Tanenhaus (1979) provided two possible interpretations. According to one view, the influence of orthography can be found at the stage of comparison in cognitive processing. Since this was essentially a phonological task, the subjects had to access the phonological information encoded in the words in order for them to detect the acoustic matches required for rhyming judgement. However, the discrepancies in the RTs for similar and dissimilar pairs indicate that orthographic information was also being accessed. This entails that while making rhyme judgements, subjects accessed both the acoustic and orthographic information of words, and they detected the rhymes by trying to match the targets on both dimensions. When they encountered a mismatch at one level, in this case, the orthographic level, they were compelled to go through yet another processing stage such as checking for an acoustic match.

The other alternative claims that the effect of orthography takes place at the stimulus encoding stage. It suggests that the presentation of the stimuli causes the activation of similar words, as postulated earlier by several word recognition models. When the presented stimuli were similar in their orthography, the target words were primed by the cue words. In case they were dissimilar, no such priming occurred. Consequently, the primed words were detected faster than the unprimed words, leading to a stark difference in the RTs of each case.

However, the influence of orthography is significantly higher for the first three types of words discussed above than those pairs which have variations resulting from differences in the use of consonant allographs or vowel allographs. While this can be explained in terms of the low phonemic density of these words, it is a phenomenon that requires further investigation. This is so because Bangla has undergone several spelling reforms over the years, as a result of which multiple spellings are allowed for words where differences arose from variations in single consonants or vowel allographs. Consequently, speakers have started to accept all the possible variants of these words as acceptable forms, as a result of which these pairs were less affected by the orthographic differences. Variations resulting from the use of different consonant clusters, however, were resistant to such changes resulting from the reforms. Thus, either of these factors could have been responsible for the difference in orthographic interference in the case of some word pairs.

6 Conclusion

Thus, we have concluded that there is an impact of the orthographic level on our speech perception. As discussed above, this effect can manifest itself at two stages: an intermediate stage of mapping between the various levels of lexical information; or at the stimulus encoding stage, which leads to an activation of the connected semantic or orthographic nodes. The question of the locus of orthographic effects in spoken word recognition is not the only interesting issue raised by the results presented in this paper, which future research may further investigate. One very important aspect of the current research and deserving of more attention is that it studies orthographic effects on perceptions of words for literate speakers in a literate society. However, it would be interesting to find out how preliterate children or adult nonliterate speakers responded to the monitoring tests. Logically, if the differences that we found in our results for similar vs. dissimilar/consistent vs. inconsistent pairs were indeed triggered by orthographic effects alone, such patterns should not be found in the case of preliterate children or illiterate speakers because the orthographic information is supposedly absent in their lexical entries. Further, we need to analyse why certain word pairs remained less affected by orthographic influence while others showed a significant influence. Though we have tried to attribute this different trend to the low phonemic density of these words, it has to be substantiated by further research. The factor concerning the acceptance of multiple spellings resulting from the spelling reforms also needs to be considered. Hence, though it may be concluded that orthography plays a significant role in word processing in Bangla, its influence on particular word types and its causes need further investigation.

References

Castles, A., Holmes, V. M., Neath, J., & Kinoshita, S. (2003). How does orthographic knowledge influence performance on phonological awareness tasks? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 56A*(3), 445–467. https://doi.org/10.1080/02724980244000486

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. Canadian journal of experimental psychology = Revue canadienne de psychologie experimentale, 49(2), 264–271. https://doi.org/10.1037/1196-1961.49.2.264

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & cognition*, *2*(2), 309–321. https://doi.org/10.3758/BF03209002

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165–178.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522–536. https://doi.org/10.1037/h0026699

Peereman, R., Dufour, S. & Burt, J.S. Orthographic influences in spoken word recognition: The consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin & Review* **16**, 363–368 (2009). https://doi.org/10.3758/PBR.16.2.363

Seidenberg, M.S., & Tanenhaus, M.K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 546–554.

Stone, G.O., Vanhoy, M., & Orden, G.C. (1997). Perception Is a Two-Way Street: Feedforward and Feedback Phonology in Visual Word Recognition ☆. *Journal of Memory and Language, 36*, 337–359.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. Psychonomic Bulletin & Review(5), 683–89.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*(4), 533–554. https://doi.org/10.1006/jmla.1997.2525