An orthographic effect in phoneme processing, and its limitations

Anne Cutler [1,2,3] * and Chris Davis [3]

[1] Language Comprehension Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[2] Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, Netherlands
[3] MARCS Auditory Laboratories, University of Western Sydney, Sydney, NSW, Australia

In three phoneme goodness rating experiments, listeners heard phonetic tokens varying along a continuum centered on /s/, occurring finally in isolated word or non-word tokens. An effect of spelling appeared in Experiment 1: native English-speakers' goodness ratings for the best /s/ tokens were significantly higher in words spelled with S (e.g., *bless*) than in words spelled with C (e.g., *voice*). Since the tokens were in fact identical in each word, this effect indicates less than optimal evaluation performance. No spelling effect appeared when non-native speakers rated the same materials in Experiment 2, indicating that the observed difference could not be due to acoustic characteristics of the S- versus C-words. In Experiment 3, native English-speakers' ratings for /s/ did not differ in non-words rhyming with words consistently spelled with S (e.g., *pless*) or with words consistently spelled with C (e.g., *floice*); i.e., no effects of lexical rhyme analogs appeared. It is concluded that the findings are better explained in terms of phonemic decisions drawing upon lexical information where convenient than by obligatory influence of lexical knowledge upon pre-lexical processing.

Keywords: speech recognition, phonemes, orthography, spelling, top-down versus bottom-up processing, strategies

INTRODUCTION

Spoken-word recognition researchers agree that the identification of phonemes can be influenced by the lexical context in which they occur, but disagree about how this is to be explained. In the TRACE model (McClelland and Elman, 1986), lexical processes exert top-down control over the logically prior process of phonemic analysis; the activation of words "feeds back" to increase activation of constituent phonemes. Other models assume that information flows strictly bottom-up, at least early in spoken-word recognition (e.g., Cutler et al., 1987; Norris, 1994; Gaskell and Marslen-Wilson, 1997; Luce and Pisoni, 1998; Norris and McQueen, 2008). In bottom-up models, effects of lexical context are accounted for in terms of post-lexical decision-making (Cutler et al., 1987), or by assuming that explicit decisions about phonemes do not arise from the perceptual processing necessary to word recognition, but are performed by a dedicated mechanism drawing simultaneously on incoming phonetic information and on the lexicon (Norris et al., 2000).

Part of what a listener knows about a word is how it is spelled. Further, listeners know the spelling rules of their language, both explicit and implicit. Such knowledge is not acquired from speech, but from instruction, from learning to read, and from reading. Nonetheless, spelling may affect decision-making about speech. Phoneme detection can, under certain task conditions or in certain positions in the word, be harder for phonemes with variable orthographic realization (e.g., English /f/; consider *farm*, *pharmacy*) than for consistently spelled phonemes (e.g., English /b/; Dijkstra et al., 1995; Cutler et al., 2010). Auditory lexical decision is sensitive to both sound and spelling of prior items (Jakimik et al.,

1985; Slowiaczek et al., 2003). As pointed out by Taft et al. (2008), however, experiments on this issue can very often be interpreted as allowing strategic responding.

In this study, we test for the presence of effects of orthography in a phoneme-processing task which certainly does not need orthographic knowledge for its performance. The task, phoneme goodness rating, was developed by Miller and Volaitis (1989) to explore the internal structure of listeners' phoneme categories. Listeners can reliably rate the goodness of renditions of their native phonemes. Their ratings are sensitive to utterance context; thus the best-rated token of a duration-sensitive phoneme such as /p/ is different at a fast versus slower rate of speech (Miller and Volaitis, 1989; Volaitis and Miller, 1992). In altering which exemplars sound best, this rate of speech effect changes the shape of the goodness judgment curve. Lexical status (*peace–beace* versus *peef–beef*) also affects ratings, in a qualitatively different way: decisions at the category boundary shift, but the curve shape does not change (Allen and Miller, 2001). As in these preceding studies, the critical sound in our stimuli is manipulated in steps along a continuum with its midpoint corresponding to a typical exemplar of the category. Listeners then rate how good each token sounds as an exemplar of its category.

We compare ratings for the same phoneme in words in which it has different orthographic realizations. We also assess ratings for the same phoneme in non-words with differently spelled real-word rhyming analogs. In models postulating necessary lexicon-to-phoneme information flow, non-words should automatically evoke analogy effects. If effects of our orthographic manipulation

are observed in words, then non-words for which all lexical analogs have a consistent spelling should be susceptible to such automatic analogies. Lexical effects in non-words, showing lexical influence on the processing of sounds without any completed lexical access having occurred, have been important in the information flow debate. Thus /t/ detection in non-words is faster when the difference from a real word is one phoneme (e.g., *rigament*; cf. *ligament*) rather than more phonemes (e.g., *maffinent*), suggesting that phoneme detection in non-words is sensitive to word-likeness (Connine et al., 1997). Also, a phonetic /g/–/k/ continuum elicits more /g/ choices if presented as *gice–kice*, but more /k/ choices as *gipe–kipe* (Newman et al., 1997). All four endpoints are non-words, without lexical representation, but *–ice* has a larger set of vowel-replacement neighbors beginning /g/ (*goose*, *guess*, *gas*) than /k/, while *–ipe* has more vowel-replacement neighbors beginning /k/ (*cap*, *keep*, *coop*) than /g/; i.e., this difference influenced listeners' judgments such that the non-word outcomes resembled the onset plus coda of more rather than fewer existing words.

The phoneme we test is /s/, which can be realized in English words either with S (*sent*, *seed*, *reseat*, *goose*) or C (*cent*, *cede*, *receipt*, *juice*). The two letters are not equally likely: computations derived from English grapheme-phoneme probabilities (Gontijo et al., 2003) reveal that /s/ in English words is realized in 78% of cases with S alone, in 5.3% with SS, and in 15.5% with C (the remaining 1.2% covers grapheme combinations such as SC, PS). So S is, overall, more than five times as likely as C to represent /s/.

In the case of S and C as orthographic realizations of /s/ as the coda of an English monosyllabic word, there is a restriction that will form part of the implicit knowledge of English-speakers: C can only appear after a tense vowel, including diphthongs: *race*, *piece*, *nice*, *sauce*, *juice*, *voice*. Lax vowels must have S: *mass*, *mess*, *miss*, *moss*, *fuss*. Note that tense vowels may also have S: *case*, *cease*, *close*, *loose*, and *horse* (a singleton coda in non-rhotic varieties such as that of our participants). Note also that the feature vowel tenseness–laxness constrains similarity in lexical activation (see, e.g., Warner, 1999; Cutler et al., 2005), and is hence important for defining lexical analogy.

Certain rimes are orthographically fully consistent in the English vocabulary; *voice*, *choice*, etc., are all spelled with C, and so are *twice*, *vice*, etc. Similarly, *house*, *mouse*, etc., and *press*, *mess* etc. are all spelled with S. Test words were selected from these S- or C-consistent rime sets. We further constructed non-words with word-final /s/ that would inherit the probability of the same rimes represented by the real words. We used tense vowels by necessity in the C case, and in the S case one tense and one lax: the S-consistent non-words were *frouse* and *pless*, the C-consistent *klice* and *bloice*.

One aspect of a comparison between different rimes is that acoustic–phonetic context necessarily varies because the /s/ position was preceded by different vowels. No goodness rating studies have previously examined whether preceding vowel quality might affect ratings. We also compare performance of the task by different listener groups with greater versus lesser command of the English vocabulary (native speakers versus high-proficiency non-native users). Non-native listeners may be less affected by lexical structure than the native listeners, but will be equally susceptible to effects of acoustic–phonetic context.

EXPERIMENTS

Listeners performed goodness ratings on the final segment of real words or non-words. This segment, a fricative on a continuum centered on a typical /s/, was rated for its goodness as /s/. At issue is whether the resulting ratings curve differs with carrier item spelling, actual (in words), or analogous (in non-words). In Experiment 1, native listeners rated /s/ in words spelled with S or C. In Experiment 2, non-native listeners rated /s/ in the same words. In Experiment 3, native listeners rated /s/ in non-words with S- or C-biased real-word analog sets[1].

MATERIALS

Six /s/-final words were chosen: *bless*, *mouse*, *abuse*, *voice*, *twice*, *juice*. Three end in orthographic *s(e)*, three in *ce*. As described, all words that rhyme with *bless* or *mouse* are consistently spelled with S, and all those rhyming with *voice* or *twice* are consistently spelled with C. The remaining pair, *abuse* and *juice*, have a mixed set of rhyming words spelled with S and C (*moose*, *ruse*, *puce*, *adduce*) and provide a control set balanced for the nature of the vowel preceding /s/[2]. Four non-words were further constructed: *ple(s)*, *frou(s)*, *bloi(s)*, and *kli(s)*, rhyming respectively with *bless*, *mouse*, *voice*, and *twice*, and therefore open to analogy with relevant lexical properties of these real words.

To assess whether the differing lexical analogs for the non-words are indeed available to listeners, we conducted a dictation pre-test, in which participants spelled non-words. This test included among the presented non-words five items for each of the S- and C-consistent rimes used in the goodness rating material (*pless*, *dess*, *avess*, *empess*, *iness*; *frouse*, *vowse*, *deglouse*, *terrowse*, *obnowse*; *bloice*, *roice*, *intoice*, *reloice*, *athoice*; *klice*, *chice*, *abice*, *adjice*, *perflice*). Digitized recordings of these non-words were presented over headphones to 43 native English-speaking University of Melbourne undergraduates (tested individually); they typed these items to a computer screen as best they could.

Spellings were scored on whether the /s/ in the rime was represented by S or C. A few responses represented incorrect rimes (e.g., LOYDS for *bloice*). **Table 1** shows responses for the S- and C-consistent sets. Excluding incorrect rimes, and averaging over each rime type, nearly 92% of non-words with S-consistent real-word analogs were spelled with S, while over 67% of non-words with C-consistent analogs were spelled with C, a significant difference. The lax vowel in *pless*, etc., was not spelled with C though the tense vowel in *frouse*, etc., sometimes was. Thus, in choosing a preferred spelling for these non-words, participants indeed have recourse both to their knowledge of spelling rules and to lexical analogy based on the spelling of rhyming real words.

---

[1]We did not compare word and non-word stimuli within subjects, first for reasons of power, and second because within-subject designs are particularly susceptible to inter-condition influences (Poulton, 1982), whereby an observed within-subjects difference may be only fully interpretable with a between-subjects replication.

[2]In the control pair with /u/, the /z/ end of the continuum in each case corresponds to another real word, while the /ʃ/ end does not (*abuseN*, *abuseV*; *juice*, *Jews*). This is true of nearly all words ending with /us/ (e.g., *puce*, *pews*; *moose*, *moos*; *truce*, *trews*; *deuce*, *dues*). Although at least one three-way pair exists (*loose*, *lose*, *louche*), the pattern appears to be a vocabulary feature. The purpose of the /u/ pair is to compare the S-biased versus C-biased cases when the vowel is constant; an independent factor that affects each member of the pair equally should not impinge on this issue.

Table 1 | Non-word pre-test: percentage of non-words spelled with S and with C, as a function of orthographic neighborhood.

Spelling	S-consistent non-words (e.g., *frouse*)	C-consistent non-words (e.g., *bloice*)
S	82.27	30.93
C	7.27	64.19
Incorrect rime	10.47	4.88

Experimental words and non-words were recorded to disk by a male speaker of Australian English. The words were spoken in normal form and also with final /z/ or final /ʃ/ (e.g., *bless, blez, blesh*), the non-words only with final /s/. The duration of the vowel before the final fricative in all recorded items was measured. Using the Stevenson/Repp method (Repp, 1981), we constructed a single continuum from a natural /z/ to a natural /s/, and from this /s/ to a natural /ʃ/. The continuum had 41 steps from each natural endpoint to another, hence 81 steps from /z/ through /s/ to /ʃ/. For each word or non-word, 81 versions were made, one with each fricative step. Vowel duration, which is affected by the nature of syllable codas in English, was varied proportionately. For example, the vowel in *blez* was longer than the vowel in *bless*; the *ble-* preceding step 21 on the fricative continuum therefore had a vowel duration halfway between that of *blez* and *bless*, the *ble-* preceding step 31 had a vowel duration longer than *bless* by one-quarter of the *blez–bless* difference, and so on. Equivalent adjustments were applied to the /s/–/ʃ/ side of the continuum, and separately for each vowel.

PROCEDURE

Listeners heard, over headphones, all versions of the continua in their experiment, in semi-random order (carrier items and phoneme tokens mixed; no two of the same carrier or token in succession). This provided considerably more item power in the present experiment than in some preceding rating studies in which the stimuli were sampled along the continuum; however, it did make for a rather long experiment. In Experiments 1 and 2, the 486 word items were presented once, pre-testing having indicated that the experiment was uncomfortably long with multiple presentations. In Experiment 3, the 324 items were presented twice each. Participants were instructed to rate how good the final sound of each item was as /s/, on a 7-point scale from 1 ("not at all like s") to 7 ("very like normal s").

PARTICIPANTS

Experiment 1: 46 native English-speaking University of Western Sydney undergraduates; Experiment 2: 24 native Dutch-speaking University of Nijmegen undergraduates; Experiment 3: 14 native English-speaking University of Melbourne undergraduates[3]. The Australian students participated to fulfill course requirements; the Dutch students received a small payment.

---

[3]Our study was modeled on those of Volaitis and Miller, who tested 10 and 12 subjects (Miller and Volaitis, 1989; Volaitis and Miller, 1992, respectively). Our plan to replicate their study also in the number of listeners in each experiment fell foul of the new academic year; large numbers of volunteers eager to fulfill course requirements reported for Experiment 1. Experiment 3 had already been run, with 14 subjects.
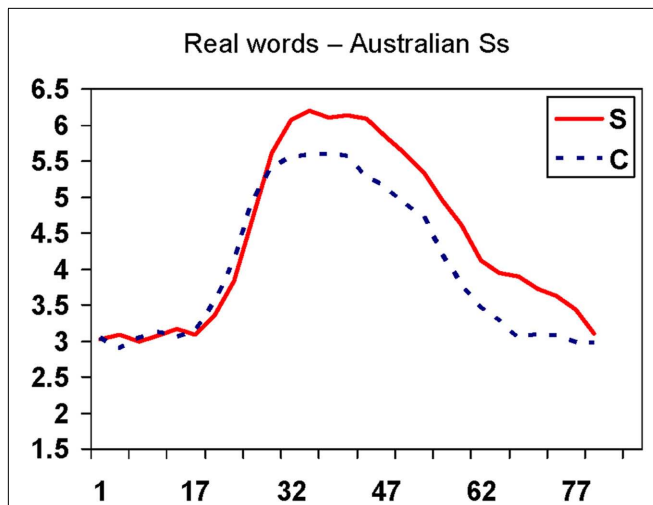

FIGURE 1 | Real words rated by native English-speaking listeners: mean rated goodness as /s/ on a 7-step scale (vertical axis) as a function of step on the 81-step continuum from /z/ at left through /s/ to /ʃ/ at right (horizontal axis), separately for words spelled with S (solid line) and C (dashed line).
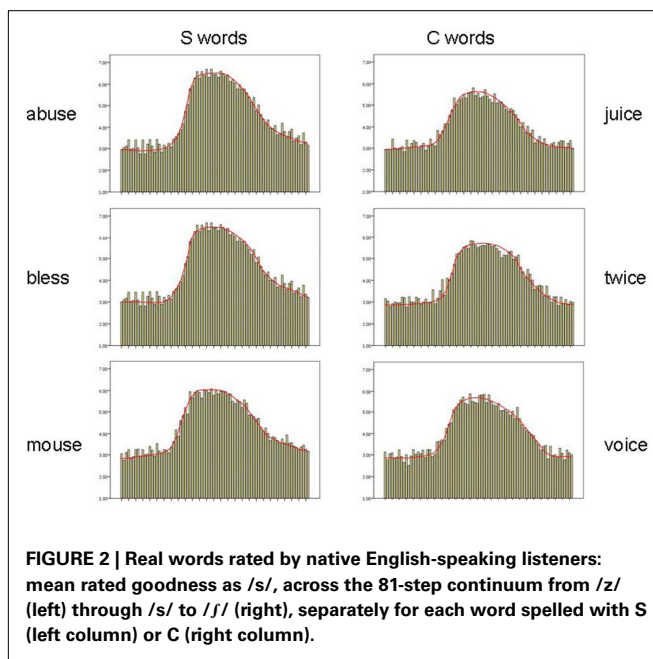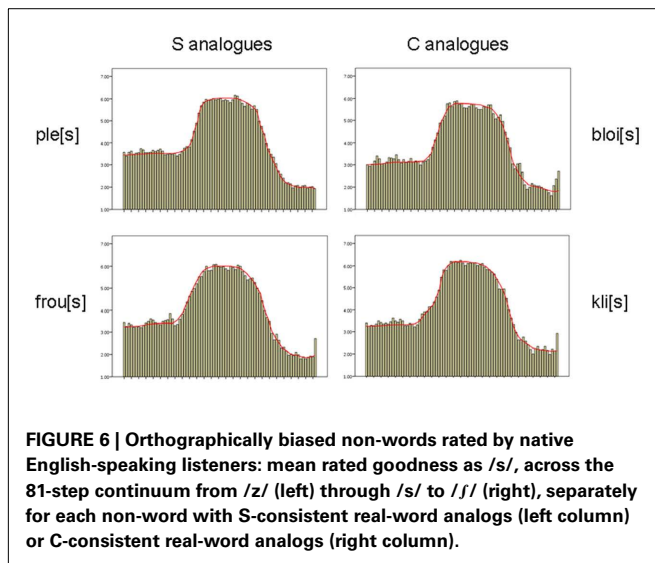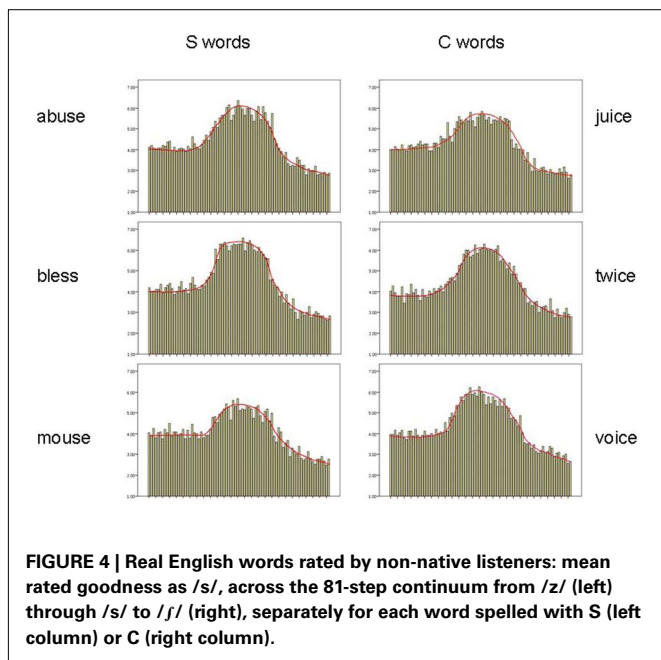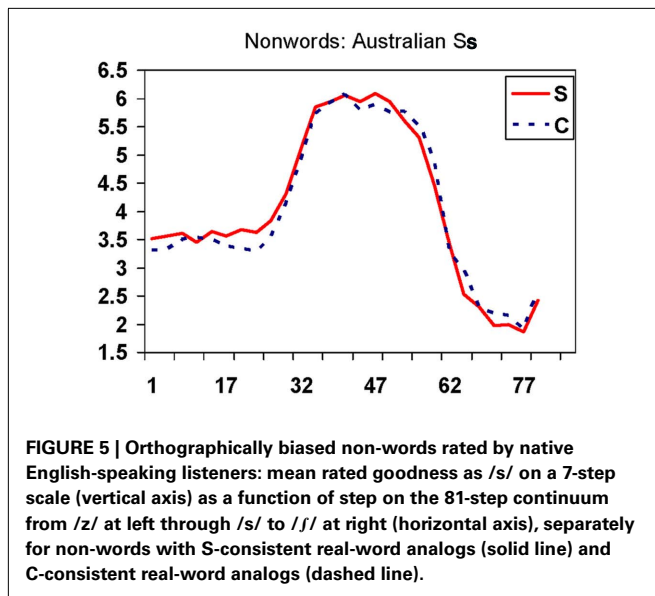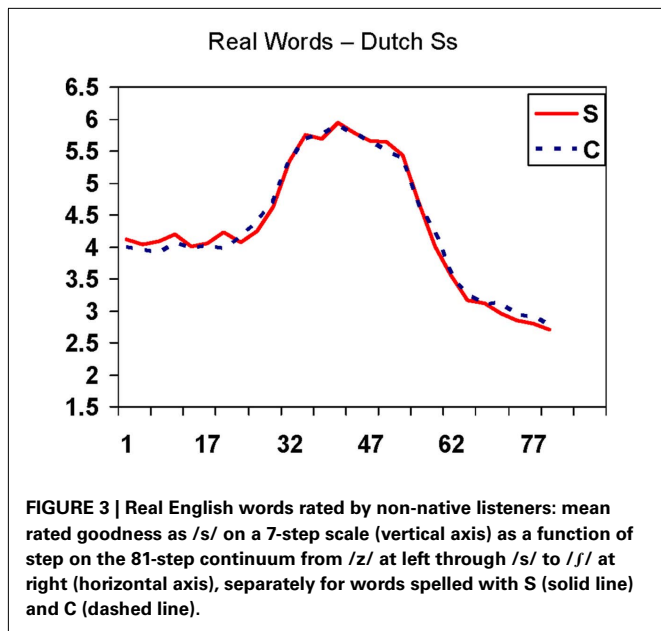

FIGURE 2 | Real words rated by native English-speaking listeners: mean rated goodness as /s/, across the 81-step continuum from /z/ (left) through /s/ to /ʃ/ (right), separately for each word spelled with S (left column) or C (right column).

RESULTS

The mean ratings across the 81-step continuum varied as expected, being significantly lower at the /z/ and /ʃ/ ends of the distribution, significantly higher around the /s/ center, and with the area around the central peak of the target category flattened. Thus the participants in each experiment were performing the task as instructed, and goodness ratings for /s/ are as consistent as the ratings for stop consonants obtained by Miller and colleagues. **Figures 1**, **3**, and **5** present smoothed functions averaged across participants and S-consistent versus C-consistent items; **Figures 2**, **4**, and **6** show

FIGURE 3 | Real English words rated by non-native listeners: mean rated goodness as /s/ on a 7-step scale (vertical axis) as a function of step on the 81-step continuum from /z/ at left through /s/ to /ʃ/ at right (horizontal axis), separately for words spelled with S (solid line) and C (dashed line).

FIGURE 5 | Orthographically biased non-words rated by native English-speaking listeners: mean rated goodness as /s/ on a 7-step scale (vertical axis) as a function of step on the 81-step continuum from /z/ at left through /s/ to /ʃ/ at right (horizontal axis), separately for non-words with S-consistent real-word analogs (solid line) and C-consistent real-word analogs (dashed line).

FIGURE 4 | Real English words rated by non-native listeners: mean rated goodness as /s/, across the 81-step continuum from /z/ (left) through /s/ to /ʃ/ (right), separately for each word spelled with S (left column) or C (right column).

FIGURE 6 | Orthographically biased non-words rated by native English-speaking listeners: mean rated goodness as /s/, across the 81-step continuum from /z/ (left) through /s/ to /ʃ/ (right), separately for each non-word with S-consistent real-word analogs (left column) or C-consistent real-word analogs (right column).

the complete data item by item, i.e., mean ratings across subjects for every step on the continuum for each word and non-word (items with S spelling/analogs are in the left columns, items with C spelling/analogs in the right columns). For the statistical analysis, the 81-step continuum was divided into nine 9-step bins, and ANOVAs were conducted across participants with S/C spelling and bin as independent variables. **Tables 2–4** show the mean ratings in each bin as a function of the spelling factor.

EXPERIMENT 1

The overall analysis revealed significantly higher ratings for /s/ in S-consistent than in C-consistent words $[F(1,45) = 23.37,$

$p < 0.001]$, significant differences between the nine bins $(p < 0.001)$, and a significant interaction between the two factors $(p < 0.001)$, prompting separate analysis of the S/C comparison per bin. The S/C comparison was insignificant in the first three bins (27 steps from the /z/ endpoint in the direction of /s/). Ratings for S-consistent words however became higher from step 29 on the continuum, and stayed higher until step 80; all six remaining bins showed a statistically significant difference at $p < 0.005$ at least, $\eta_p^2 = 0.22$ for bin 4 and 0.45 or above for the rest. Thus exactly the same sounds in the /s/ and /ʃ/ bands of the continuum were rated as being less good exemplars of /s/ for real words spelled with C than for words spelled with S. This difference can easily be seen in **Figure 1**. Furthermore, **Figure 2** shows that it is consistent across items: ratings for all S words peak at higher points than ratings for all words spelled with C. Importantly, the difference also appears in the vowel-matched pair *abuse–juice*, suggesting that it is

Table 2 | Experiment 1: mean ratings by Australian listeners for S- versus C-consistent words across nine 9-step portions of the 81-step continuum from /z/ through /s/ to /ʃ/.

Steps	S-consistent words	C-consistent words
1–9 (most /z/-like)	3.033	2.994
10–18	3.106	3.111
19–27	3.963	4.186
28–36	5.965	5.522
37–45 (most /s/-like)	6.112	5.475
46–54	5.597	4.925
55–63	4.564	3.813
64–72	3.858	3.138
73–81 (most /ʃ/-like)	3.385	3.010

Table 3 | Experiment 2: mean ratings by Dutch listeners for S- versus C-consistent English words across nine 9-step portions of the 81-step continuum from /z/ through /s/ to /ʃ/.

Steps	S-consistent words	C-consistent words
1–9 (most /z/-like)	4.082	3.951
10–18	4.088	4.023
19–27	4.177	4.185
28–36	5.244	5.250
37–45 (most /s/-like)	5.806	5.815
46–54	5.579	5.514
55–63	4.080	4.140
64–72	3.083	3.168
73–81 (most /ʃ/-like)	2.784	2.875

Table 4 | Experiment 3: mean ratings by Australian listeners for non-words with S- versus C-consistent lexical analogs, across nine 9-step portions of the 81-step continuum from /z/ through /s/ to /ʃ/.

Steps	S-consistent non-words	C-consistent non-words
1–9 (most /z/-like)	3.583	3.411
10–18	3.619	3.472
19–27	3.659	3.425
28–36	5.119	4.891
37–45 (most /s/-like)	5.917	5.931
46–54	5.851	5.764
55–63	4.319	4.413
64–72	2.208	2.379
73–81 (most /ʃ/-like)	2.060	2.179

not caused by difference in the vowels of *bless*, *mouse* versus *voice*, *twice*.

Because this participant group was larger than the others, we made three random selections of 14 participants and analyzed each subgroup's responses in the same manner as for the whole group. The three subgroup analyses all patterned almost exactly in the same way as in the overall analysis reported above: the S/C comparison varied across bins (0.001), and in separate ANOVAs the S/C effect was significant in each of bins 5–8 (steps 37–72) at

minimally .05, lowest $\eta_p^2 = 0.038$. It was not significant in bin 9 (steps 73–81) for any subgroup, and was significant in bin 3 (steps 19–23) and bin 4 (steps 28–36) in each case for one subgroup.

EXPERIMENT 2

ANOVAs conducted in the same manner as for Experiment 1 revealed a main effect of bins across the continuum ($p < 0.001$), but no effect of the S/C comparison ($F < 1$) and no interaction between these two factors ($F = 1.1$, n.s). Splitting the Experiment 2 participants into two equal sub-groups of 12 again revealed parallel results for each group, and no interactions. This result further suggests that the significant difference in Experiment 1 cannot be ascribed to acoustic–phonetic factors on which *bless*, *mouse* differ from *voice*, *twice*. One difference across the Experiment 1 and 2 results, easily visible in the figures, is that the non-native listeners' ratings at the /z/ end of the continuum were higher than the native listeners' ratings in Experiment 1; this finding, which **Figure 4** shows to be consistently present in all six items, is considered further in Section "Discussion."

EXPERIMENT 3

The overall ANOVA here included a further factor, comparing first half of the experiment against second. That main effect was insignificant, as were all interactions involving it. The S/C comparison was also insignificant ($F < 1$); the only significant main effect was that of bins along the continuum ($p < 0.001$). However, the bins effect interacted with the S/C effect ($p < 0.01$), so that we again analyzed each bin separately. The S/C comparison was not significant in any bin; the interaction thus presumably represents the fact that the ratings were somewhat higher for one non-word set in about half the bins and somewhat higher for the other non-word set in the other half (see **Table 4** and **Figure 5**). Thus native listeners did not consistently rate these /s/ sounds differently in non-words with C-spelled real-word analogs than in non-words with S-spelled analogs, as had been the case with real words. **Figure 6** also reveals that no individual item shows an aberrant pattern.

A final unequal-N between-group ANOVA compared native listener ratings in Experiments 1 and 3 across the continuum. S–C spelling, Word–Non-word status and continuum bins interacted significantly ($p < 0.001$), and the Word–Non-word versus S/C interaction was significant (minimally .02) from bins 5 to 9 (steps 37–81), otherwise not. The same analysis for the Experiment 3 group versus three randomly chosen Experiment 1 sub-groups of 14 produced in each case the same pattern of significance (0.001 for the overall interaction, minimally .05 for Word–Non-word versus S/C interaction in individual bins); the only difference was that this interaction was not significant in bin 9 for subgroup 2, and was additionally significant (0.04) for subgroup 3 in bin 1 and for subgroups 2 and 3 in bins 2 and 3, reflecting in this /z/ portion of the continuum some S/C separation in Experiment 3 but none in Experiment 1 (see **Figure 5** versus **Figure 1**; **Figure 6** suggests that the effect is largely due to *pless*).

DISCUSSION

The task of rating the goodness of non-standard phoneme realizations requires no recourse to higher-level information, but its

performance can be affected by orthographic knowledge when the phoneme realizations are embedded in real-word carriers. No effect of orthography appears, however, when the same phoneme realizations are embedded in non-words biased toward S versus C spelling (as attested by reliable differences in a dictation test).

In Experiment 1, goodness judgments were significantly higher in /s/-like and /ʃ/-like ranges for carrier words spelled with S than for words spelled with C. This difference appeared even with the vowel-matched pair *abuse–juice*, arguing against an explanation based on acoustic–phonetic differences (e.g., in the preceding vowel). Thus the native listeners' judgments in this task were affected by their knowledge of how the carrier words were spelled. The phoneme /s/ appears more effectively captured by the letter S (which is the usual representation of /s/, and which is used to spell /ʃ/ in English) than by C (which is a rarer /s/ representation, and in fact more often stands for the sound /k/).

Non-native listener ratings (Experiment 2) showed no effect of the S/C comparison. These listeners had all studied English at high school and were proficient in the language. They knew the carrier words, and their spelling, but this knowledge did not affect their ratings. Note that these listeners' ratings would have been susceptible to any acoustic effect of the preceding vowel; the Experiment 2 results provide further evidence against such an effect. As noted, though, ratings at the /z/ end of the continuum were significantly higher in Experiment 2 than in Experiments 1 and 3. The most likely source of this pattern is the Dutch listeners' native-language experience of /s/. Dutch has no /s/–/z/ contrast in word-final position (and indeed, voicing distinctions in Dutch fricatives are weakening word-initially also, to the extent that initial voicing errors are overlooked in making lexical decisions: Ernestus and Mak, 2004). For these listeners, /z/ is apparently a more acceptable version of /s/ than it is for native English listeners, certainly word-finally as here.

Experiment 3 delivered further data suggesting how the difference observed in Experiment 1 should be interpreted. Obligatory and automatic top-down flow of lexical information to pre-lexical processing would make such information available during processing of non-words as well as of words. However, in Experiment 3 the height of the ratings peak for non-words with only S-spelled real-word rhyme analogs (*pless, frouse*) versus with C-spelled analogs (*bloice, klice*) did not significantly differ. The mean peak location was shifted somewhat overall, in comparison to Experiment 1 [in the cross-experiment analysis this appeared as a main effect of Word–Non-word (0.03) in bin 4, but no such difference in the three bins on either side of bin 4]. We attribute this finding to articulatory differences between words and non-words (recall that Allen and Miller, 2001 observed that lexical status affected goodness rating peak height, while effects of the acoustic–phonetic manipulation of speech rate were expressed in peak location.) However, this word-non-word effect in our study was quite independent of the S/C comparison.

How should we interpret the combined results of the present study? Either model type described in the introduction could explain the positive result in Experiment 1: in a top-down model it could be viewed as top-down influence of orthography (lexically stored information) on perception of phonemic information in speech, while in a strictly bottom-up model it could be interpreted

as strategic reference to lexical information in decision-making about the phonemic representations derived from bottom-up processing of the speech.

In either case, we should note that the use of orthography has not resulted in a performance improvement. Exactly the same sounds were rated differently in different items, in defiance of the identical outcome that accurate acoustic–phonetic processing should have delivered. Thus some sounds were interpreted as more like canonical /s/ than they should have been, or other sounds were interpreted as less like canonical /s/ than they should have been, or both. It is difficult to conceive of an ecological advantage for such an outcome, with its implication that words spelled with C will be disfavored in processing. The S/C difference persisted throughout the experiment, even though the same six words were presented repeatedly and no fresh lexical access was appropriate.

A top-down account of the effect, based on automatic lexical-to-phonetic information flow making spelling available during speech perception, would perhaps deal with the lack of benefit by proposing that such availability benefits word recognition often enough, even in languages with many irregular orthographic mappings, that occasional disadvantages do not render it useless. This account would hold, in other words, that although the effect of spelling is in this case not beneficial to task performance, its appearance is evidence of the automatic top-down information channel. In the non-word case, the lack of effect could be ascribed to the weakness of top-down flow from partially over-lapping lexical entries in comparison to that from fully activated entries (with an ancillary account to explain why partially over-lapping entries can facilitate phonetic categorization, as reported by Newman et al., 1997, but not phoneme goodness rating, as here). Alternatively, the lack of effect could be accounted for in terms of contradictory spelling information from multiple over-lapping lexical representations, including those with spellings that would violate the rules applying to the form being heard (e.g., words with tense vowels when the input had a lax vowel, though as noted earlier, such vowel mismatches greatly reduce lexical similarity).

A strategic explanation, in contrast, would need an account of why spelling should be of use for rating phoneme goodness; for instance, it might simplify decision-making in an admittedly long and potentially tedious experiment. Although deciding about /s/ in non-words should then also be based on such a shortcut if there were one, orthographic information would only be available given access to a lexical entry, thus not for non-words; the same tendency to convenience that led to recourse to the lexical entry in the words case would however prevent explicit construction of a proposed spelling in the non-word case.

Thus neither account is definitively ruled out and neither is necessarily entailed by the present results. Is either account preferable on independent grounds? We would argue that the strategic account is preferable for three reasons. First, there is the usual Occam's razor argument: theoretical entities such as top-down flow of orthographic information should not be postulated unless evidence rules out alternatives. Second, by assigning the disadvantage for words in which /s/ is spelled with C to an effect of a short-term benefit to decision-making, a task-specific account removes the implication that such words cause listeners difficulty at all times. Third, such task-specificity is in line with other results

in the literature showing that effects of spelling in spoken-word recognition vary with the nature of the task. Thus congruent spelling (*pie*, *tie* as opposed to *pie*, *guy*) leads to faster responses in rhyme judgment (Seidenberg and Tanenhaus, 1979) but not in repetition (Pattamadilok et al., 2007). Likewise, words with one consistent spelling (e.g., *catch*) produce faster lexical decisions than words with potential alternative spellings (e.g., *deed*, which could also have been spelled like *knead* or *keyed* or *cede*; Ziegler and Ferrand, 1998), and semantic and gender categorization shows the same effect (Peereman et al., 2009); but word repetition only shows such an effect if lexical decision is required as well (Ventura et al., 2004).

In models of spoken-word recognition where information flows obligatorily from the lexical to the pre-lexical level, lexical information should become available irrespective of the lexical status of the spoken input. Such information should have benefit to the listener. In phoneme goodness rating, the decisions are not themselves concerned with lexical status and words play no obvious role in making them. Orthographic knowledge was nevertheless used in such decisions on real words, but it did not lead to improvement of task performance. We argue that a task-based account of orthographic effects in performance of experiments involving speech is preferable, in particular in that it restricts adverse effects to the task-specific situation. Consequently there is no need to answer awkward questions such as: if irregular orthographies hinder speech perception, why do they continue to exist?

REFERENCES

Allen, J. S., and Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate. *Percept. Psychophys.* 63, 798–810.

Connine, C. M., Titone, D., Deelman, T., and Blasko, D. (1997). Similarity mapping in spoken word recognition. *J. Mem. Lang.* 37, 463–480.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (1987). Phoneme identification and the lexicon. *Cogn. Psychol.* 19, 141–177.

Cutler, A., Smits, R., and Cooper, N. (2005). Vowel perception: effects of non-native language vs. non-native dialect. *Speech Commun.* 47, 32–42.

Cutler, A., Treiman, R., and Van Ooijen, B. (2010). Strategic deployment of orthographic knowledge in phoneme detection. *Lang. Speech* 53, 307–320.

Dijkstra, T., Roelofs, A., and Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Can. J. Exp. Psychol.* 49, 264–271.

Ernestus, M., and Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain Lang.* 90, 378–392.

Gaskell, M. G., and Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Lang. Cogn. Process.* 12, 613–656.

Gontijo, P. F. D., Gontijo, I., and Shillcock, R. (2003). Grapheme-phoneme probabilities in British English. *Behav. Res. Methods Instrum. Comput.* 25, 136–157.

Jakimik, J., Cole, R. A., and Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *J. Mem. Lang.* 24, 165–178.

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighbourhood activation model. *Ear Hear.* 19, 1–36.

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.

Miller, J. L., and Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Percept. Psychophys.* 46, 505–512.

Newman, R. S., Sawusch, J. R., and Luce, P. A. (1997). Lexical neighbourhood effects in phonetic processing. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 873–889.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234.

Norris, D., and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395.

Norris, D., McQueen, J. M., and Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behav. Brain Sci.* 23, 299–370.

Pattamadilok, C., Kolinsky, R., Ventura, P., Radeau, M., and Morais, J. (2007). Orthographic representations in spoken word priming: no early automatic activation. *Lang. Speech* 50, 505–531.

Peereman, R., Dufour, S., and Burt, J. S. (2009). Orthographic influences in spoken word recognition: the consistency effect in semantic and gender categorization tasks. *Psychon. Bull. Rev.* 16, 363–368.

Poulton, E. C. (1982). Influential companions: effects of one strategy on another in the within-subjects design of cognitive psychology. *Psychol. Bull.* 91, 673–690.

Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bull. Psychon. Soc.* 18, 12–14.

Seidenberg, M. S., and Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *J. Exp. Psychol. Hum. Learn. Mem.* 5, 546–554.

Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., and Bishop, K. L. (2003). An investigation of phonology and orthography in spoken-word recognition. *Q. J. Exp. Psychol.* 56A, 233–262.

Taft, M., Castles, A, Davis, C., Lazendic, G., and Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: pseudohomograph priming. *J. Mem. Lang.* 58, 366–379.

Ventura, P., Morais, J., Pattamadilok, C., and Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Lang. Cogn. Process.* 19, 57–95.

Volaitis, L. E., and Miller, J. L. (1992). Phonetic prototypes: influences of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.* 92, 723–735.

Warner, N. (1999). "Timing of perception of vocalic distinctive features: implications for vowel system universals," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1961–1964.

Ziegler, J. C., and Ferrand, L. (1998). Orthography shapes the perception of speech. The consistency effect in auditory word recognition. *Psychon. Bull. Rev.* 5, 638–689.