Non-automaticity of use of orthographic knowledge in phoneme evaluation

Anne Cutler[1,2], Chris Davis[2], Jeesun Kim[2]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] MARCS Auditory Laboratories, University of Western Sydney, Australia

`anne.cutler@mpi.nl,a.cutler@uws.edu.au,chris.davis@uws.edu.au,j.kim@uws.edu.au`

Abstract

Two phoneme goodness rating experiments addressed the role of orthographic knowledge in the evaluation of speech sounds. Ratings for the best tokens of /s/ were higher in words spelled with S (e.g., *bless*) than in words where /s/ was spelled with C (e.g., *voice*). This difference did not appear for analogous nonwords for which every lexical neighbour had either S or C spelling (*pless*, *floice*). Models of phonemic processing incorporating obligatory influence of lexical information in phonemic processing cannot explain this dissociation; the data are consistent with models in which phonemic decisions are not subject to necessary top-down lexical influence.

Index Terms: spoken-word recognition, orthography, phonemes, feedback

1. Introduction

Understanding the processes by which perceptual input is recognized is a central goal of the study of cognition. However, speech recognition models notoriously disagree on whether information passes from the prelexical to the lexical level of analysis in a strictly feedforward direction, or whether there is bi-directionality of information flow whereby feedback from the lexical level can actually influence processing at the prelexical level. There is general agreement that phoneme identification can be influenced by lexical context; the debate is about how to explain this within a recognition model.

In the TRACE model [1] lexical influence occurs as a direct result of lexical processes exerting top-down control over the preceding process of phonemic analysis. That is, TRACE accounts for findings showing lexical effects in detection and phonetic categorization by allowing activation of word nodes to feed back to any constituent phonemes. This feedback loop to phonemes from the words they occur in acts to increase their activation faster than the same phonemes in nonwords. By contrast, in the Merge model [2] information flow through the system is entirely bottom-up, from phonetic to lexical processing, never in the reverse direction. Phonemic decisions are made by a dedicated mechanism which is sensitive both to the continuous stream of incoming information from phonetic processing, and to higher-level information, e.g. from the lexicon. Merge captures the fact that phoneme detection and classification decisions show lexical influence by allowing the decision-maker to receive input from both these sources at once.

Striking findings that have influenced this debate are demonstrations of lexical effects in nonwords. A continuum from /g/ to /k/ elicits more /g/ judgements if presented as *gice-kice*, more /k/ judgements if presented as *gipe-kipe* [3]. All are nonwords, but the *–ice* pair has a denser lexical family with /g/ – *goose*, *gas*, etc. – while the *–ipe* pair has more neighbours with /k/ - *cap*, *coop* etc. Similarly, /t/ detection is faster in nonwords which differ from real words by only one phoneme (e.g., *rigament*, like *ligament*) than by more phonemes (*maffinent*) [4]. Such effects in nonwords are important since they show lexical influence on phoneme processing without any completed lexical access having occurred; models like TRACE can explain this using top-down flow of information.

An active current debate in spoken-word recognition research concerns the role of orthographic knowledge in speech processing. This too is of great theoretical interest, because orthographic knowledge is not intrinsic to speech; it is part of language users' stored knowledge about words, acquired from learning to read and from reading. In some languages (e.g., Italian, Dutch), orthography is relatively predictable from the sound pattern of words, while in other languages (e.g., English), it is much less predictable. Especially in the latter case, demonstrations of orthographic effects on spoken word recognition would again be readily captured in the framework of models allowing top-down flow of information.

Such orthographic effects in speech processing have indeed been reported, e.g., in standard speech perception tasks such as phoneme detection or lexical decision. Thus phoneme detection can be influenced by phoneme-to-grapheme mapping, such that phonemes with variable orthographic realisation (e.g., English /f/ – cf. *farm*, *pharmacy*) may be harder to detect than phonemes with consistent spelling (such as English /b/; [5; 6]); and spoken primes which partially match targets orthographically (*bugle – buggy*) facilitate lexical decision to a greater extent than primes which match in sound, but mismatch in spelling (*surgeon – certain;* [7; 8]).

Kouider and Dupoux [9], however, presented evidence that orthographic influence in speech perception may be subject to strategic modulation. Their study used a cross-modal version of the masked repetition priming procedure of Forster and Davis [10] to investigate whether exposure to a written word would automatically and necessarily activate pronunciation. A spoken word was primed by its written form presented under masking. Priming effects were absent unless the written primes were consciously perceived. This suggests independent systems for written and spoken word processing, that influence each other only under conscious conditions.

In the present study we take the debate a step further by combining the approaches described above. We investigate a phoneme processing task, which, like phonetic categorisation [3] and phoneme detection [4; 5; 6], should not necessarily need orthographic knowledge for its performance; and we compare this processing in words versus nonwords, following the logic of [3; 4]. Orthographic effects in listening to words may arise if orthographic information becomes available during lexical access; but orthographic effects in listening to nonwords must have arisen without the intervention of a completed lexical access (since nonwords are not in the lexicon). Moreover, if the emergence of strategic responding is possible in this task as in the priming task of Kouider and Dupoux [9], this too should reveal itself in the form of a difference between the real-word and the nonword conditions.

The task used is phoneme goodness rating. This task was developed by Miller and colleagues [11; 12; 13] to explore the internal structure of listeners' phoneme categories. Listeners

prove to have robust ideas about what is a good versus less good rendition of any phoneme of their language. Moreover, the ratings are sensitive to the goodness of phonemes in utterance context; thus at a fast rate of speech, the best-rated token of a duration-sensitive phoneme such as /p/ differs from the token which receives best ratings at slower rates of speech [11; 12]. We can therefore expect that context within a word or nonword might be taken into account in the ratings listeners make. Further, the task has been shown not to be affected in this automatic way by whether the judgements are consistent or inconsistent with lexical status of the carrier (e.g., in *peace-beace* only /p/ makes a word, whereas in *peef-beef* only /b/ makes a real word; this does not affect phoneme goodness ratings in the way that, e.g., speech rate does [13]). No earlier studies have examined effects of orthography in this task.

Our word stimuli contained the same phoneme, spelled differently; /s/ is realised with S in *house* but with C in *voice*. Our nonword stimuli were chosen such that their lexical neighbourhoods differed in orthographic correspondences. As an example: the nonword *frouse* has lexical neighbours *house*, *mouse*, etc., all of which are spelt with S. In contrast, the nonword *bloice* has lexical neighbours *voice*, *choice* etc. with C. Participants heard word or nonword stimuli in which, as in Miller's studies, the critical sound was manipulated in steps along a continuum with its midpoint corresponding to a typical exemplar of the category. Listeners then rated how good each token sounded as an exemplar of its category.

We examine in a first experiment whether this strictly phonemic processing task shows effects of orthography. Using the phoneme /s/, which varies in orthographic realisation, we compare the same phonemic continuum realised in the context of words normally spelled with S (e.g., *house*) versus C (e.g., *voice*). In a second experiment, we then examine the ratings for the same continuum realised in nonwords differing in lexical neighbourhood, such as *frouse* and *bloice*.

As described above, either top-down models such as TRACE [1] or strictly bottom-up models such as Merge [2] can account for lexically induced effects of orthography when such a task is performed with words. The models differ, though, in whether they actually predict such effects. TRACE does predict them, because phonemic processing of any kind receives top-down information from the lexicon in this model. Merge does not necessarily predict them, because in an exclusively phonemic processing task – and moreover, one which has been shown not to be necessarily influenced by lexical status [13] – there is no necessary reason in Merge for lexical influence to be observed, even with words. With nonwords, TRACE again predicts that top-down information would be available, and any effect seen with words should also be seen with nonwords. Merge predicts that, again, there need be no such effect, and moreover, that effects with words and nonwords can dissociate. Absence of orthographic effects in words, or dissociation of word and nonword effects, would therefore favour Merge over TRACE, while parallel effects with words and nonwords would favour TRACE over Merge.

2. Experiments

Experiments 1 and 2 differed in the lexicality of the materials. In Experiment 1, listeners performed goodness ratings on the final segment of real words ending (in canonical form) in the phoneme /s/, but differing in how /s/ was spelled. In Experiment 2, the same ratings were performed on nonword stimuli, differing in the spelling of lexical neighbours. In other respects the studies shared all features below.

2.1. Participants, stimuli and procedure

The 61 listeners took part as a course requirement; all were native English-speaking undergraduates at either the University of Western Sydney or the University of Melbourne.

Six /s/-final words were chosen as stimuli for Experiment 1: *bless*, *mouse*, *abuse*, *voice*, *twice*, *juice*. Three end in orthographic *s(e)*, three in *ce*. For *bless* and *mouse*, the orthographic neighbourhoods overwhelmingly contain other words with S (*guess, mess, house, grouse,* etc.), while the orthographic neighbourhoods of *voice* and *twice* overwhelmingly contain other words with C (*choice, rejoice, mice, nice,* etc.). The remaining two words, *abuse* and *juice*, have a mixture of neighbours (*deuce, goose, spruce, loose*), and serve as a control set balanced for the nature of the vowel preceding /s/; such balance obviously cannot be achieved when the neighbourhood is deliberately varied.

Four nonwords were chosen for Experiment 2: *ple[s]*, *frou[s]*, *bloi[s]* and *klai[s]*, rhyming respectively with *bless*, *mouse*, *voice* and *twice*, and therefore inheriting the lexical neighbourhood properties of these real words.

All items were spoken in normal form and with final /z/ or final /ʃ/ (e.g., *frouse, frouze, froush*). In general. speakers articulate unfamiliar nonwords more carefully than well-practised known words. However, such factors should not affect S and C items in any different way. A fricative continuum was constructed from the natural /z/ endpoint to the natural /s/ endpoint, and from the /s/ endpoint to the /ʃ// endpoint, using the Stevenson/Repp method [14]. There were 41 steps from each endpoint to the other. A combined continuum from /z/ through /s/ to /ʃ/ therefore had 81 steps. 81 versions of each of the 10 individual words and nonwords were made, one with each fricative version. Vowel duration was varied proportionately for each item with phonemic step (e.g., the duration of the vowel in *frouze* was measured and found to be longer than the vowel in *frouse*; the vowel duration of step 21 on the fricative continuum was therefore presented with a vowel duration halfway between that of *frouze* and that of *frouse*, and the same adjustments were made on the /s/-/ʃ/ side of the continuum, and with each item).

Testing took place in a single session. Listeners heard all versions of all word continua (Experiment 1), or all versions of all nonword continua (Experiment 2), in semi-random order (carrier words and phoneme tokens mixed; no two of the same word or token in succession). They were instructed to rate how good the final sound of the item was as /s/, on a 7-point scale from 1 ("not at all like s") to 7 ("very like normal s").

2.2. Results

The mean ratings across the 81-step continuum all varied as expected, being significantly lower at the /z/ and /ʃ/ ends of the distribution, and significantly higher around the /s/ centre. Thus the listeners responded as intended to the experimental stimuli. Figure 1 shows the mean ratings across subjects for every step on the continuum for each of the six words in Experiment 1; Figure 2 shows the same information for the four nonwords of Experiment 2. Overall, peaks are somewhat sharper for the nonword than for the word continua, as would be expected for a comparison between carefully articulated stimuli (nonwords) versus more easily uttered stimuli (words). The word-nonword difference is however not at issue in our analyses. The crucial comparison, separately for the words and the nonwords, is that between items with orthographic S neighbourhoods (in the left column, in each figure) and items with orthographic C neighbourhoods (in each case in the right column).
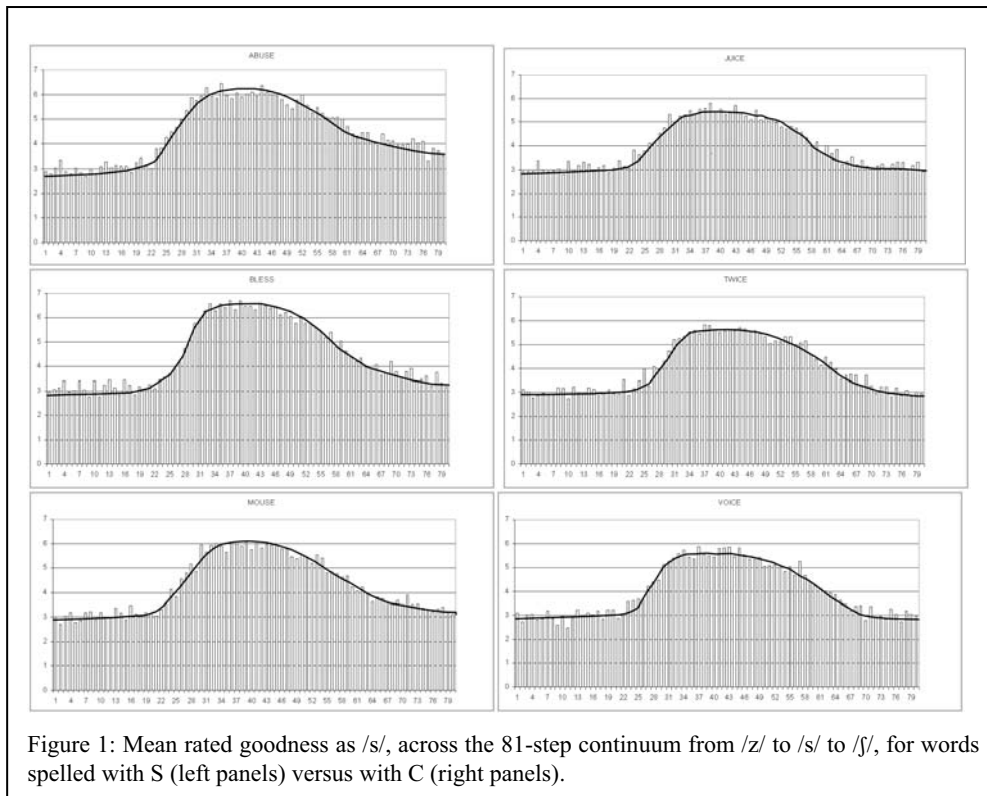
Figure 1: Mean rated goodness as /s/, across the 81-step continuum from /z/ to /s/ to /ʃ/, for words spelled with S (left panels) versus with C (right panels).

2.2.1 Experiment 1: Words

Inspection of the graphs in Figure 1 shows that, as expected, in every case the curves flatten out at their peak. This is around continuum points 30-50; point 40 is overall the best-rated /s/. However, it can also be clearly seen that there is a large orthographic effect: ratings for words spelled with S peak at a higher point than ratings for words spelled with C. This is also seen in the vowel-matched pair *abuse-juice*. An ANOVA was conducted to test the means of the two conditions (C words versus S words) for the ratings from points 30-50. The mean rated goodness across this range for C words was 5.466 (SE 0.035), and for S words 6.049 (SE 0.035). The analysis showed that the difference between these two word sets was significant, $F(1,120) = 134.93$, $p < 05$. That is, exactly the same /s/ sounds were rated as being less good exemplars for words that were spelled with C than for words spelled with S.

2.2.2 Experiment 2: Nonwords

Inspection of Figure 2 reveals overall more well-defined peaks than with the words (as expected for more articulatory care). Further, the curve shape differs from Figure 1; the curves flatten out at a peak between points 40-60 (not 30-50 as for the words). Importantly, no difference can be seen between nonwords with S neighbourhoods, in the left column, and with C neighbourhoods, on the right. An ANOVA was conducted to test the means of the two conditions (C nonwords and S nonwords) for the ratings from 40-60, in the same manner as for Experiment 1. The mean rated goodness for C-biased nonwords in this peak region was 5.717 (SE 0.055) and the mean rating for S-biased nonwords was 5.746 (SE 0.055). The ANOVA showed here that this difference between conditions was not significant ($F < 1$, ns). Thus /s/ sounds were rated no differently in nonwords with word neighbours spelled with C than in nonwords with word neighbours spelled with S.

3. Discussion

The results are clear: orthography indeed affects listeners' judgements of the goodness of individual phoneme tokens; however, its effect appears in real-word contexts and not in nonword contexts. There were articulatory differences between words and nonwords, but these differences impact alike upon S and C items. The shape of the curve, however, differed for S and C words, but did not differ for S and C nonwords. Accordingly, the effect in words does not arise from automatic top-down activation flow from lexical neighbours. In nonwords such as *frouse*, all similar (rhyming) words in the lexicon are spelled with S, but this does not induce the pattern seen with real S-words such as *mouse*; likewise, all lexical neighbours of the nonwords *bloice* are spelled with C, but the goodness rating pattern does not differ from the S-nonwords in the way that the C-word and S-word patterns differ.

As laid out in the introduction, the appearance of effects of orthography with words but not with nonwords is consistent with the phonemic decision-making model Merge [2], but runs counter to the predictions of the top-down model TRACE [1]. In TRACE, top-down connections operate automatically, so such dissociation should not appear. The dissociation indicates that listeners respond strategically, in the case of words.

Previous demonstrations of orthographic influences on speech processing have been explained both with and without top-down processing [6; 5]. For phoneme detection, effects of orthography arise in TRACE by top-down lexical-to-phonetic connections. The same effects are also explicable in Merge, however, despite the model's lack of top-down connections; there, the effects are held to arise at the decision stage.

An important characteristic of the effect of orthography in phoneme detection is that it has proven to be variable – in fact, listeners in general detect inconsistently spelled word-initial consonants just as rapidly in words with minority spelling (e.g., /f/ in *pharmacy*) as in words with majority spelling (/f/ in *farm* [5; 6]).
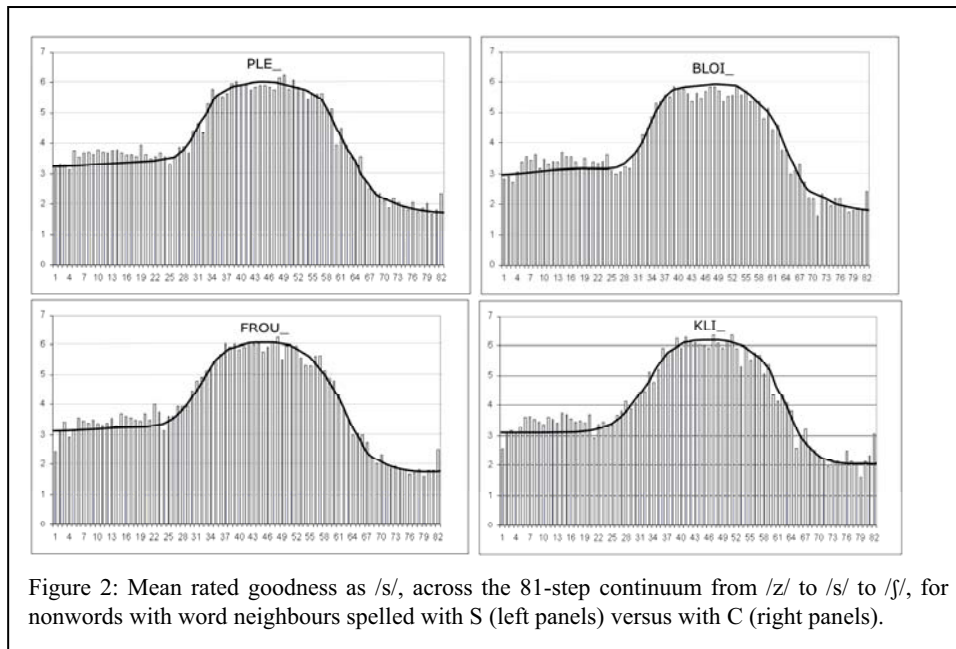
Figure 2: Mean rated goodness as /s/, across the 81-step continuum from /z/ to /s/ to /ʃ/, for nonwords with word neighbours spelled with S (left panels) versus with C (right panels).

Also, listeners detect consonants which are consistently versus inconsistently spelled equally rapidly. Both these results counter-indicate any sensitivity to orthographic consistency. But if an experiment calls attention to the orthographic dimension, such sensitivity appears, with the very same input. In [5], when many fillers had irregular spelling (e.g., *ghastly*, *yacht*, *suit*), listeners then responded more rapidly to the target phonemes with consistent realization (e.g. /b/) than to the inconsistently realized ones (e.g., /f/). The account of this effect in [5] is that orthography, part of the stored knowledge about words, is available to the Merge decision mechanism when words are activated. Salience of orthography in an experiment causes the decision mechanism then to give weight to this information in making phoneme decisions, even though spelling is not normally drawn upon.

This explanation draws on strategic responding in the same way as the explanation given by Kouider and Dupoux [9] for their result from priming. In phoneme detection, frequent occurrence of irregularly spelled words makes orthography salient and suggests that taking it into account may be of value in the detection task. In priming, conscious awareness that a written version of a word precedes the spoken item about which a decision has to be made draws attention to written forms and suggests the availabili

ty of a strategy in which orthography is taken into account.

In the present task, despite its low-level nature, similar factors appear to affect listeners' response patterns. Spelling played a role in how the task was performed, but only when real words such as *abuse*, *house* versus *juice*, *voice* were carriers of the sound being rated, not when the carriers were nonwords such as *frouse*, *bloice*. This information was used strategically in the word case; it was not automatically available from lexical activation flow to phoneme processing, or it would have been usable in the nonword case too.

We conclude, therefore, that orthographic knowledge, while it forms part of the stored knowledge that listeners have about words, does not become available in an automatic manner via top-down lexical-to-phonetic information flow during speech processing.

4. Acknowledgements

5. References

[1] McClelland, J.L. and Elman, J.L., "The TRACE model of speech perception", Cogn. Psychol., 18: 1-86, 1986.

[2] Norris, D., McQueen, J.M., and Cutler, A. "Merging information in speech recognition: Feedback is never necessary", Behav. and Brain Sci., 23: 299-370, 2000.

[3] Newman, R.S., Sawusch, J.R. and Luce, P.A., "Lexical neighborhood effects in phonetic processing", J. Exp.Psychol: Hum. Perc. Perf., 23: 873-889, 1997.

[4] Connine, C.M., Titone, D., Deelman, T. and Blasko, D., "Similarity mapping in spoken word recognition", J.Mem.Lang., 37: 463-480, 1997.

[5] Cutler, A., Treiman R. and Van Ooijen, B. "Strategic deployment of orthographic knowledge in phoneme detection", Lang. and Speech, submitted

[6] Dijkstra, T., Roelofs, A. and Fieuws, S., "Orthographic effects on phoneme monitoring", Canadian J. Exp. Psychol.: 49, 264-271, 1995.

[7] Jakimik, J., Cole, R.A. and Rudnicky, A.I., "Sound and spelling in spoken word recognition", J. Mem. Lang., 24: 165-178, 1985.

[8] Slowiaczek, L.M., Soltano, E.G., Wieting, S.J. and Bishop, K.L., "An investigation of phonology and orthography in spoken-word recognition", Quart. J. Exp. Psychol., 56A: 233-262, 2003.

[9] Kouider, S. and Dupoux, E., "A functional disconnection between spoken and visual word recognition: Evidence from unconscious priming", Cognition, 82: B35-B49, 2001.

[10] Forster, K.I. and Davis, C. "Repetition priming and frequency attenuation in lexical access", J. Exp. Psychol.: Learn Mem.. Cogn., 10: 690-698, 1984.

[11] Volaitis, L.E. and Miller, J.L., "Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing categories", J. Acoust. Soc. Am., 92: 723-735, 1992.

[12] Miller, J.L. and Volaitis, L.E., "Effect of speaking rate on the perceptual structure of a phonetic category", Perc. & Psychophys., 46: 505-512, 1989.

[13] Allen, J.S. and Miller, J.L., "Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate", Perc. & Psychophys., 63: 798-810, 2001.

[14] Repp, B.H. "Perceptual equivalence of two kinds of ambiguous speech stimuli", Bull. Psychonomic Soc., 18, 12-14, 1981.