Title
Match and Mismatch in Phonological Context

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

Authors
Gaskell, Gareth
Marslen-Wilson, William

Publication Date
1993

Peer reviewed

which the uniqueness point occurred early on in the word. According to models such as TRACE, the issue of lexical access should already be resolved here and so little effect of mismatch would be expected. These results showed that just one phoneme deviation is enough to interrupt lexical access. Experiments on word-initial distortion (Moss, Marslen-Wilson, & Spence, 1989; Marslen-Wilson & Zwitserlood, 1989) produced similar results, even when the distortion was restricted to a single phonetic feature.

These experiments indicate that we have a lexical access system that is strongly dependent on bottom-up information and intolerant of mismatch. If this proposition is correct in its simplest form, we must conclude that much natural variation in speech causes mismatch and that listeners either fail to understand much of what is said to them or that they must employ post-perceptual 'recovery' mechanisms in cases of mismatch.

Phonological Variation

Phonological variation is rule-governed variation occurring within conjunctions of sounds in speech. The type of variation we addressed in our experiments is known as place assimilation, which occurs optionally in normal connected speech. In English it occurs syllable-finally, usually at the end of a word, when the previous consonant adopts the place of articulation of the following segment. The effect only occurs for coronal phonemes (e.g. /d/, /t/ n/) which are followed by non-coronals such as labials (e.g. /b/, /p/, /m/) or velars (e.g. /g/, /k/, /ŋ/). In a phrase such as /wɪkɪdprænk/ (wicked prank), the labial place of the /p/ can migrate to the previous consonant, meaning that the sequence is phonetically realised as [wɪkɪbprænk]. However, place assimilation in English is an asymmetric process, so non-coronal phonemes cannot be realised as coronals.

This type of variation is particularly problematic for current models of speech perception for a number of reasons. Firstly, it produces just the kind of deviation that was found to mismatch in the experiments on single words, that is, single feature word-final distortions. It normally occurs across word boundaries and so is difficult to compensate for within a lexicon structured around single word units. Place assimilation is a strong effect producing a completely different phoneme, as opposed to the coarticulation studied by Elman and McClelland (1988), which just shifts the boundary between two phonemes. Finally, it is a regressive effect, giving a left to right processor less time and information to work with.

The problem we need to address is how the human speech recognition mechanism can process phonologically variable speech when the matching process has such a low tolerance level for variation A number of solutions present themselves. The first is that the experiments on which we have based our view of the matching process are unrepresentative. It is feasible that experiments on isolated words allow subjects to act much more critically than is normally possible, exaggerating the intolerance of the matching process. Therefore, one objective of this research was to look at the same kind of deviations embedded in sentential context, where the load on the speech processor is more realistic. However, if we assume for the moment that the earlier research is accurate, there are a number of options.

One hypothesis comes from the phonological theory of radical underspecification (Archangeli, 1988). This theory states that only marked, non-redundant features are represented in a lexical entry. A marked feature is one that is not in the default state whereas a non-redundant feature is a feature which cannot be established using context-sensitive redundancy rules. The application of these principles produces an extremely compact, structured lexicon containing only essential information. Implementing this theory for place assimilation requires the assumption that the default place of articulation is coronal, so velar and labial places are specified in the lexicon, but coronal is not. The susceptibility of coronal phonemes to place assimilation can then be explained. A default rule gives unspecified segments their usual place of articulation, namely a coronal place.

[ ] → [+coronal]

In addition, a context sensitive assimilation rule can give these same segments a labial or velar place.

[ ] → [+velar] /– #[+velar]

[ ] → [+labial] /– #[+labial]

These rules state that the unspecified segment can gain a labial or velar gesture provided the following phonological context is appropriate. However, the assimilation process cannot occur for non-coronal segments since they are already specified for place.

For underspecification to have value in the area of perception, we need to add a matching strategy.

This was suggested by Lahiri and Marslen-Wilson (1991) in their study of vowel nasalization. They proposed three outcomes for a matching process:

1. A lexical *match,* in which the input matches a specified feature in the lexicon.

2. A *mismatch,* in which the input is inconsistent with a specified feature in the lexicon.

3. A *lack of mismatch,* in which the feature in the input is irrelevant since the lexical entry for that feature is unspecified, either through its default status or its redundancy.

The third outcome explains how we could have an intolerant lexical access system but still allow phonological change. Distortions caused by phonological processes alter unspecified features and so do not mismatch the lexical entry. Other types of distortion, however, which alter features which are specified in the lexicon, will cause mismatch with the lexical representation as in outcome 2 above.

An alternative, or maybe additional proposition, is that the speech processor evaluates the identity of phonemes with respect to their phonological context. Variation such as place assimilation necessitates a processor that is less linearly dependent, since each segment of the speech stream must be evaluated in the context of what follows it as well as what precedes it. This approach has been successfully modeled using a simple recurrent network that learns to pick out the phonological context that allows assimilation to take place (Gaskell, Hare, & Marslen-Wilson, 1992). The prediction based on this hypothesis is that a distortion will only mismatch if, *in its phonological context,* it violates a phonological rule.

In summary, our research was an attempt to answer a number of questions derived from the application of three different hypotheses. Firstly, does a single feature distortion produce mismatch when presented embedded in a sentence? Secondly, is there a mismatch effect when the feature being distorted is, according to phonological theory, unspecified in the lexicon? Finally, does the viability of the distortion in its phonological context have any effect on the matching process?

Experimental Data

The technique used to explore these questions was cross-modal repetition priming. The prime word was presented auditorily, embedded in sentential context, and a visual target was presented at the offset of the prime. The task of the subject was to decide whether or not the visual target was a word and press a button accordingly. So, for example, a subject would hear *'I would say you got what you deserved. That was a wickib prank',* and at the offset of *wickib,* the target word WICKED would be presented visually. The subject would then make a lexical decision as quickly as possible.

Table 1 shows the design of the experiment. Six versions of each prime sentence manipulated three variables: The prime-target relationship – related or control, the presence or absence of distortion, and the viability of the distortion given the following context.

The prime was either related to the target, as in *wicked*-WICKED, or unrelated, as in *valid*-WICKED. Note that the manipulations were all carried out on the prime word and the target word was kept constant throughout the six conditions. The viability variable was only manipulated in the related conditions as the unrelated conditions were only to be used as baselines for priming measurements.

The distortions used were all single feature deviations, altering the place of the word-final phoneme of the prime from coronal to either labial or velar. According to the under-specification view, these distortions should all be altering an unspecified feature and therefore cause no mismatch.

The phonological viability of the distortions was manipulated by altering the place of articulation of

Table 1. Examples of the prime words used in the experiments with one word of their following context.

		Distortion	
		Distorted	Undistorted
Sentence Type	Viable	*wickib prank*	*wicked prank*
	Unviable	*wickib game*	*wicked game*
	Control	*valib sentence*	*valid sentence*

Table 2. Reaction times (ms) for Experiment 1 (right context of prime removed). The test sentences are collapsed across the phonological viability variable.

		Distortion		Distortion effect (D-U)
		Distorted	Undistorted	
Sentence Type	Test	587.2	578.7	+8.5
	Control	661.2	644.4	+16.8

the phoneme immediately after the prime word (i.e. the phonological right context). If the place of this context phoneme matched the place of the distortion, as in *wickib prank*, the context was assumed to be viable, since it conforms to the phonological rule for assimilation. The unviable contexts were created by matching a velar distortion with a labial context phoneme and vice versa. Thus, *wickib game* is in the unviable condition because the places of the [b] and the [g] do not match and so the distortion could not have been caused by place assimilation.

Two separate experiments were performed using these stimuli. In the first experiment the sentence up to the offset of the prime word was presented so that we could assess the matching process in the absence of phonological context. In the second experiment the whole of the prime sentence was presented, allowing phonological right context to have an effect. The target word was positioned at the same point in each sentence.

Note that in the second experiment, at the time of presentation of the visual target, the following context has yet to be presented. This was to maximize the chances of any priming effects found being the result of on-line perceptual processes, but it also meant that there was little time for phonological context to have any effect, the limit being about six hundred milliseconds, the mean response time of the subjects.

Because of the complex nature of the stimuli, they were recorded using a non-naive speaker. To make sure there were no confounding effects due to either speaker bias or ease of articulation, the stimulus sentences were pretested by playing the sentences up to the offset of the prime word to a set of 48 subjects. The subjects made a forced-choice response to the prime word, circling either the undistorted prime or an orthographic representation of the distorted prime on a response sheet. For example, a subject may have to choose between WICKED and WICKIB. The differences between the distorted and undistorted tokens of the prime word were minimal but despite this, subjects were generally able to discriminate between the two options. The overall proportion of correct responses was over 90% and a two way analysis of variance (ANOVA) on the error proportions revealed no significant effects.

In Experiment 1, thirty-six subjects were tested on 3 experimental versions. Experiment 2 used 46 subjects in six experimental versions.

Results

Experiment 1

In Experiment 1, the sentences were presented up to the offset of the prime word. As the following context was not presented, we expected no effects of the viability of the phonological context and none were found. The results in Table 2 are therefore collapsed across this variable. The test sentences primed strongly, of the order of 60 ms, with a highly significant effect of sentence type ($minF'[2,124] = 18.05$, $p < 0.001$). However, the effect of the variable distortion was not significant, implying that in this experiment distortion produced no mismatch effect. This result contradicted the findings of the experiments on single words, but it did not contradict any of our three hypotheses. The sentential context in this experiment could have caused an increased tolerance to mismatch. On the other hand, given an underspecified lexical representation, we would expect no mismatch since the distortions were all varying underspecified features. Finally, the effect of phonological context cannot be evaluated here since the sentences were only presented up to the offset of the prime word. To differentiate these hypotheses, we need to look at the results of the second experiment.

Experiment 2

Experiment 2 used the full prime sentences and so allowed us to examine the effect of phonological

Table 3. Reaction times (ms) for Experiment 2 (complete prime sentences presented).

		Distortion		Distortion
		Distorted	Undistorted	effect (D-U)
Sentence Type	Viable	624	625	-1
	Unviable	655	615	+40
	Control	679	651	+28

context (see Table 3). For the prime words in viable context for assimilation, there was no effect of distortion; the tokens with distorted phonemes primed as strongly as the undistorted ones. But when there was an unviable context for assimilation, there was a strong mismatch effect for the distorted words. A two-way ANOVA on the four related conditions (i.e. excluding the two control conditions) revealed a significant interaction between distortion and phonological viability ($F_1 = 4.70$, $p < 0.05$; $F_2 = 5.165$, $p < 0.05$). A Tukey HSD comparison of the distortion effects for each prime type showed that only the 40 ms effect for sentences with unviable contexts was significant at the 5% level.

Discussion

The first question our research addressed concerned the level of distortion of the speech signal needed to disrupt lexical access. Considering the unviable context conditions of the two experiments, the only difference between the sentences in these conditions was in the place of articulation of the word-final consonant of the prime word. In Experiment 1, with no following context this made no difference, but in Experiment 2 it was enough to reduce the priming found by 40 ms. The results of Experiment 2 are therefore more strong evidence for a lexical matching process that needs an extremely good fit to activate lexical representations.

The presence of a mismatch effect confirms that the repetition priming technique is sensitive to small changes in the speech signal and is therefore a valid tool for examining the process of lexical match. Thus, the absence of this effect in Experiment 1, where the sentences were cut at the offset of the prime word, itself requires explanation, given the results of experiments on isolated words where mismatch effects were found across the board. One possibility is that the additional processing load created by the preceding speech forces subjects to process speech with increased tolerance for deviation. But this cannot be a general tolerance for deviation, given the results of Experiment 2, which suggest instead an account based on phonological processes and representations.

Comparison of the two experiments implies that in Experiment 2, phonological context is brought into play very rapidly, in less time than it takes for the subject to respond to the target word. This kind of result is difficult to model using the winner-takes-all architecture of models like TRACE (McClelland & Elman, 1986), since our results suggest that the winner takes *nothing* unless it meets a fairly stringent criterion. Where TRACE falls down is that the only mechanism it has for reducing a word-node activation is by lateral inhibition from other active words. In this experiment the distortion occurred word-finally, at a point where the base word will normally be dominant in terms of word level activation and therefore fairly resistant to inhibition. We argue that our results, if interpreted in TRACE-like terms, require direct inhibition from featural information, so that mismatching information can have strong effects in a short time.

This research has also shown that the definition of what constitutes a mismatch must now be revised. A further question we raised was whether phonological processes play a direct role in the perception of speech. The results of the second experiment indicate that they do. When sentences were presented with a distorted prime in a context in which the distortion was phonologically viable, there was no distortion effect – the target was facilitated as strongly as for the undistorted primes. However, the same distortion, in circumstances where it could not occur naturally, strongly disrupted the priming effect. A matching process that analyses phonemes or features without reference to their neighbours cannot cope with results like this. Neither the distorted phonemes nor the context phonemes *by themselves* create mismatch, only when these two elements combine does mismatch occur.

Our third hypothesis was based on the phonological theory of underspecification. This predicted that the distortions we used would not create mismatch because they were mapping onto coronal segments in the lexicon which are unspecified for place of articulation. Experiment 1, where the phonological context of the distortions was unknown seems to support this prediction since no mismatch effect was found, but the results of the second experiment are more difficult to handle. The application of a simple processing strategy to an underspecified lexicon would predict that the deviations used should not mismatch *whatever* the following context. A more sophisticated processor that can evaluate segments in context is still necessary, irrespective of whether we choose an underspecified or a fully specified lexicon.

We therefore argue that human speech processing mechanisms must be capable of assessing information in its temporal context. But can we say more about the nature of this processor? Research in our laboratory has shown that the perception of phonological processes such as assimilation, deletion, vowel reduction and so on can be modeled using a rule based system, in this case a two-level parser (see Pulman & Hepple, 1993). On the other hand, Gaskell, Hare, & Marslen-Wilson (1992) showed that a simple recurrent network is able to learn the conditions necessary for assimilation, while still maintaining a strong dependence on bottom-up information. There are no doubt other ways of modeling the data and one of the directions of our research at the moment is into methods of constraining the requirements of such a model.

In conclusion, our study was an attempt to discover how listeners cope with natural variation in speech, concentrating on the phonological process of place assimilation. We found that, despite the fact that the lexical matching process can be severely disrupted by the alteration of just one phonetic feature, if the distortion occurs in a phonologically viable context, there is no apparent loss to the system – the distortion is processed as fluently as the undistorted base word. The intolerance of deviation in the access process implies that the assessment of word candidates relies upon both facilitory and inhibitory featural information. The effect of phonological context on the access process leads us to refine our definition of what constitutes a mismatch, and indicates that the human speech system must be sensitive to the phonological processes caught up in the speech stream.

Acknowledgements

This research was supported by a UK SERC studentship awarded to Gareth Gaskell and by UK MRC, SERC and JCI grants awarded to William Marslen-Wilson and Lorraine Tyler.

References

Archangeli, D. 1988. Aspects of underspecification theory. *Phonology* 5:183-207.

Elman, J.L., & McClelland, J.L. 1988. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language* 27:143-165.

Gaskell, G., Hare, M., & Marslen-Wilson, W.D. 1992. Phonological variation in speech perception: A connectionist approach. Paper presented at the Experimental Psychology Society, York.

Lahiri, A., & Marslen-Wilson, W.D. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38:254-294.

Marslen-Wilson, W.D. & Gaskell, G. 1992. Match and mismatch in lexical access. Paper presented at the XXV International Congress of Psychology, Brussels.

Marslen-Wilson, W.D., & Zwitserlood, P. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance* 15:576-585.

McClelland, J.L., & Elman, J.L. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18:1-86.

Moss, H.E., Marslen-Wilson, W.D., & Spence, R. 1989. Perceptual distance and competition in lexical access. Paper presented at the Experimental Psychology Society, London.

Pulman, S.G., & Hepple, M.R. 1993. Two level phonology: A formalism and an implementation. Forthcoming.