Title
Inference Processes in Speech Perception

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Authors
Gaskell, Gareth
Marslen-Wilson, William

Publication Date
1994

Peer reviewed

Inference Processes in Speech Perception

Gareth Gaskell[1] and William Marslen-Wilson[2]
Centre for Speech and Language,
Department of Psychology,
Birkbeck College, University of London,
Malet Street, London WC1E 7HX
ENGLAND
[1]ubjtp69@ccs.bbk.ac.uk
[2]ubjta38@ccs.bbk.ac.uk

Abstract

Cross-modal priming experiments have shown that surface variations in speech are perceptually tolerated as long as they occur in phonologically viable contexts. For example, [klim] (*cleam*) gains access to the mental representation of *clean* when in the context of [klimpaks] (*cleam parks*), since the change is a natural one, reflecting the phonological process of place assimilation. This implies that speech perception involves processes of phonological inference, which recover the underlying form of speech. Here we investigate the locus of these inference processes, using the phoneme monitoring task. A set of stimulus sentences were created containing deviations that were either phonologically viable (as in *cleam parks* above) or unviable. In Experiment 1, subjects monitored for the segment underlying the surface change (in the above example, /n/) and in Experiment 2 the following segment (/p/) was the target. In addition, the lexical status of the carrier word was manipulated (e.g., *clean* vs *threan*), contrasting lexical and non-lexical theories of phonological inference. Both experiments showed strong effects of phonological viability for real words, with weaker effects for the non-word stimuli. These results suggest that phonological inference can occur non-lexically, but that it interacts strongly with the process of lexical access.

Introduction

This research focuses on the effects of natural phonological variation on human speech perception. A common example of such a change is the assimilation of place of articulation, where the place of a word-initial segment migrates to the preceding segment. For example, sweet boy may be produced in connected speech as [swipbɔɪ]. This neutralisation creates surface ambiguity, which must be resolved in the process of speech perception.

Earlier research (Gaskell & Marslen-Wilson, 1993) suggests that this resolution involves a process of *phonological inference*, using phonological context to assess the viability of surface change. Here we employ the phoneme monitoring task to explore this process further. We shall argue that phonological inference can occur non-lexically, but that there are strong lexical interactions with the inference process. These data are interpreted with reference to a connectionist model which learns the mapping from speech input to phonological and semantic form.

Phonological Inference in Lexical Access

Gaskell & Marslen-Wilson (1993) presented evidence that lexical access involves a process of phonological inference. Our study used cross-modal repetition priming in sentential context to examine the effects of phonological changes on lexical access. In one condition, these changes occurred naturally, as the result of assimilation of place of articulation (e.g., 'I would say you got what you deserved, that was a *wickib* prank'). These were compared to sentences with the same phonological changes, but in contexts that violated assimilation rules (e.g., 'I would say you got what you deserved, that was a *wickib* game'). The activation of the underlying base word (e.g., *wicked*) by the phonologically changed prime was reflected in the time taken to make a lexical decision to the base word, which was presented to the subjects visually at the offset of the prime.

We found that when a phonological change occurs as the result of place assimilation, there is no disruption of access to the lexical representation of the underlying word. But the same phonological change in an unviable context for assimilation severely disrupts lexical access. Our discussion of these results contrasted representational and inferential accounts of this phenomenon. A representational account accommodates phonologically lawful variation by either adding to a form-based lexical entry (Harrington & Johnstone, 1987) or reducing the lexical entry to just the abstract and invariant phonological units (Lahiri & Marslen-Wilson, 1991). In contrast, an inferential account assumes that the identities of segments are evaluated using context-dependent inference, based on phonological rules or constraints. We argued that because the phonological changes we examined occurred across word boundaries, they could not be accommodated by a purely representational theory. The viability of a phonological change depends on the comparison of two segments: the phonological change itself and the following segment. Since these segments form part of two different words they cannot be accommodated within a single lexical representation. Instead, lexical access must depend on a process of phonological inference to assess the validity of a phonological change in its segmental context.

The Locus of Phonological Inference

In the current study, we aim to determine the locus of the inference process within the perceptual system. One possibility is that phonological inference is a purely lexical process, consisting of a context-checking mechanism dependent on lexical access. This hypothesis implies that lexical form representations are constructed so that surface variants of a word gain access to the stored information (e.g., Lahiri & Marslen-Wilson, 1991), and that the phonological context of the variant is then checked for its viability.

Alternatively, the inference process could be a compensatory mechanism that applies pre-lexically to some lower-level representation of speech. This hypothesis is made explicit in a connectionist model of phonological inference (Gaskell, Hare & Marslen-Wilson, 1992; Gaskell, 1994), which was trained on the mapping between the surface and underlying forms of phonologically changed speech. The network learned to identify the contextual cues to the presence of surface changes and exploited these cues in a graded manner. Since the model made no use of lexical representations in its responses, a prediction of this approach is that phonological inference should occur even when lexical access fails, in the perception of phonologically changed nonwords.

A second, related, aim is to examine subjects' judgements of the form of phonologically changed speech. We have shown that phonological inference is a vital component of lexical access, but there is more to speech perception than the extraction of the underlying meaning of an utterance. Spoken words produce an auditory *percept*, which involves the sensation of phonological form as well as meaning. Given that phonological inference occurs during word recognition, it is important to examine how this inference process affects perceptual experience. When presented with a surface change such as [swipbɔɪ], subjects may hear a word-final [p], corresponding to the surface form of the segment, or a word-final [t], corresponding to a more abstract underlying form.

Experimental Issues

To examine these issues of phonological inference and abstractness we employed the task of phoneme monitoring, in which subjects listen for segments of speech in sentential context. This allows us to determine the relationship between lexical access and phonological inference, by examining the effects of phonological variation in nonwords. Evidence of inference in nonwords would suggest that it can occur non-lexically, whereas interaction between the effects of phonological inference and lexical status would imply that the inference process is dependent on lexical access.

The phoneme monitoring task encourages subjects to focus on the form rather than meaning of speech. Thus, effects of phonological inference in this task would represent support for the hypothesis that decisions about the phonological structure of speech operate on a relatively abstract code of representation. As such, phoneme monitoring represents a more direct test of the role of phonological inference in the perceptual process than was possible in the earlier priming studies.

We report two experiments looking at these questions. In one experiment, the task was to monitor for the underlying coronal segment of an assimilated word. For example, a subject might have to monitor for the /n/ of *clean* underlying the surface labial segment in 'The city got two awards for its *cleam* parks'. In order to make a response in this situation, a subject must perceive the place of articulation of the target as coronal, even though it is actually produced as a labial segment. If subjects do respond, then a surface representation of speech may be unavailable, with perception of segments operating on a more abstract underlying level. The second experiment involves monitoring for the following word-initial segment, from which the place of articulation migrates (/p/, in the above example). This allowed us to look at the way phonological changes are used in the on-line processing of speech.

In both experiments, we compared sentences containing targets embedded in real words (e.g., *clean*) with ones where the tokens containing the assimilation were nonwords (e.g., *threan*). The nonwords were created by altering the initial consonants of the carrier word, ensuring that lexical access is severely disrupted (Marslen-Wilson, Moss & Van Halen, in press; Connine, Blasko & Titone, 1993; Marslen-Wilson, 1990).

Experiment 1

Forty-eight sets of test sentences, such as (1), were constructed, containing a word-final target embedded in a carrier word.

(1) The city got two awards for its *clean parks*.

Each set consisted of 8 sentences, between which three binary variables were alternated. The target segment (/n/ in the above example) was presented either with unassimilated coronal place (–Phonological Change) or with a non-coronal place (+Phonological Change). In addition, the carrier word (*clean*) and the context word (*parks*) were manipulated. The context word (the word immediately following the target) was varied so that the assimilation was either phonologically viable or unviable, while the carrier word was presented either as a real word (e.g., *clean*) or as a nonword (e.g., *threan*). In all conditions, subjects were required to monitor for the coronal form of the target (see Table 1 for examples of the critical words in each condition).

Table 1: Example critical words (i.e. carrier and context words) for Experiment 1. In this example, the target segment was /n/.

Change	Lexical Status	Viability	Example
+	+	+	cleam parks
+	+	–	cleam guesthouses
+	–	+	thream parks
+	–	–	thream guesthouses
–	+	+	clean parks
–	+	–	clean guesthouses
–	–	+	threan parks
–	–	–	threan guesthouses

The sentences were pre-tested by presenting subjects with the stimulus sentences up to the offset of the carrier word, and a forced-choice task was used to identify the word-final segments. The pre-test found small differences in the clarity of the word-final segments which were included as covariates in the analyses of Experiment 1.

In the experiment, subjects were presented visually with a coronal target segment (either /d/, /t/ or /n/) and asked to monitor for the target in the following sentence. The identity of the target varied from trial to trial. Data were analysed in terms of the proportion of responses elicited (see Table 2) and the time taken to respond. Because of wide variation in the percentage of responses found between conditions, the response times are of limited value and are not discussed here.

Table 2. Mean response rates for Experiment 1.

Change	Lexical Status	Viability	% Response
+	+	+	59.2
+	+	–	34.0
+	–	+	33.0
+	–	–	20.2
–	+	+	77.3
–	+	–	75.6
–	–	+	87.5
–	–	–	82.6

Response proportion analyses showed significant main effects of Phonological Change ($F_1[1,65] = 515.5, p < 0.01$; $F_2[1,31] = 201.1, p < 0.01$), with surface coronal targets (80.4%) provoking far more responses than surface non-coronal targets (36.0%). The effect of Lexical Status ($F_1[1,65] = 16.7, p < 0.01$; $F_2[1,31] = 8.6, p < 0.01$) showed that there were more responses to real words (62.3%) than to nonwords (54.1%) across conditions.

The critical effects, however, are the differences between the viable and unviable conditions for the phonologically changed conditions, since these are a measure of the extent to which the perceived identity of the underlying segment has been affected by its following phonological context. For the real-word carriers, there were 25% more responses when the phonological change was viable than when it was unviable; for the nonwords the difference was 13%. Both these effects were significant at $p < 0.01$ in Newman-Keuls analyses. An analysis of just the phonologically changed conditions showed a significant interaction between Lexical Status and Phonological Viability ($F_1[1,65] = 4.61, p < 0.05$; $F_2[1,31] = 5.31, p < 0.05$), showing that the viability effect for the real words was significantly stronger than for the nonwords. The combination of a viable phonological context and a real word carrier is enough to push the proportion of items provoking a response up to 59%. These are all responses which, on a surface analysis of the speech, would be classified as incorrect. Instead, subjects seem to base their responses on a more abstract underlying code.

These results provide support for a process of phonological inference which acts on both words and nonwords. The following context of the phonological change is used as a cue to the underlying place of articulation of the target segment: if this context licenses the change, subjects are more likely to respond as if they heard a coronal segment. The fact that viability also affects subjects' responses to nonwords means that the process does not require access to lexical information. However, the interaction between viability and lexical status shows that this process will, under normal circumstances, exploit lexical information.

Experiment 2

In Experiment 2, we used phoneme monitoring to examine the way assimilated segments interact with their phonological context as speech is heard. The vast majority of phonological changes that occur in normal speech are viable ones. It is therefore likely that a speech processor that is maximally efficient in the uptake of phonological information will learn to use the presence of phonological change to predict the context that validates the change. Consider the perception of the utterance [klimpaks] (*cleam parks*). By the time the assimilated [m] is perceived, the only lexical derivation for the surface speech is *clean* (since *cleam* is not a word). Unless the word has been mispronounced, the [m] must be an assimilated segment and so must be followed by a labial segment. The labial place of the following /p/ can, therefore, be predicted by the time the [m] is heard.

Experiment 2 used stimulus sentences such as (2) and (3) with the task of monitoring for the word-initial following context of the phonological change (/p/ here).

(2) The film shows a *toad* pouncing on a fly.

(3) The film shows a *groad* pouncing on a fly.

We refer to the italicised word here as the carrier word to maintain continuity with Experiment 1, even though it no longer contains the target segment. In Experiment 2, all the independent variables manipulated this carrier word, which was presented either with no assimilation (e.g., [toʊdpaʊnsɪŋ]; *toad pouncing*), with assimilation made viable by the target segment (e.g., [toʊbpaʊnsɪŋ]) or with assimilation made unviable by the target segment (e.g.,

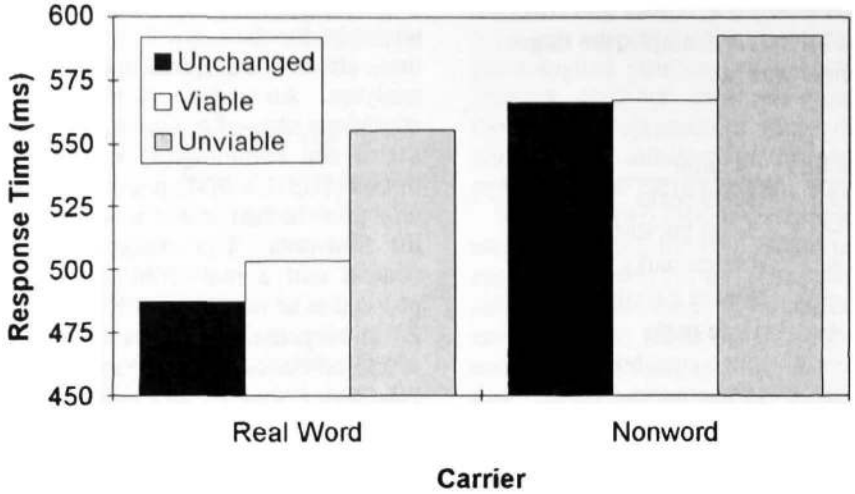

Figure 1: Effects of phonological change on response times in Experiment 2.

[towgpawnsɪŋ]). Both phonologically changed tokens had surface forms which were nonwords. As in Experiment 1, the initial consonant cluster of the carrier word was presented either unchanged (e.g., *toad*) or with word-initial changes, such that in all three conditions the token was a nonword (e.g., *groad*).

A pre-test was carried out on the sentences to check the clarity of the word-final consonant of the carrier word (which in Experiment 2 precedes the target segment), using a forced-choice test. The pre-test found no significant differences between the conditions of the experiment.

Subjects were presented with the visual target segment (/p/, /b/, /k/ or /g/) before each sentence and were instructed to monitor for the target. The results are summarised in Figure 1.

The response time analysis showed a highly significant effect of Lexical Status ($F_1[1,42] = 47.3$, $p < 0.01$; $F_2[1,29] = 16.0$, $p < 0.01$) indicating that the responses to the targets preceded by words were faster (516 ms) than to those preceded by nonwords (575 ms). Across the Lexical Status variable, unviable changes evoked significantly longer response times than both viable changes and unchanged conditions (Newman-Keuls, $p < 0.01$). These viability effects for both real word and nonword conditions were tested separately, with the effect holding for the real word carriers (51 ms difference, $p < 0.05$) but not for the nonword carriers (25 ms difference, $p > 0.1$).

As before, Experiment 2 shows a robust effect of phonological viability on subjects' responses. Here though, the effect shows up in the response times rather than the response proportions, and is also confined to just the unviable conditions. The viable phonological changes neither inhibit nor facilitate responses compared to the unchanged control, which suggests that either the speech processor does not use the presence of assimilation as a cue to the place of the following context, or that the facilitory effect of the cue is cancelled out by conflicting effects of similarity (Dell & Newman, 1980) or the inhibitory effect of the mismatch in the previous word (Foss, Harwood & Blank, 1980).

General Discussion

The Locus of Phonological Effects

These experiments attempted to ascertain the locus of phonological inference by examining the effects of phonological viability in nonwords. Experiment 1 showed that when subjects were presented with a phonologically changed coronal segment embedded in a nonword, they were more likely to make a response when the context of the change conformed to the process of place assimilation. Experiment 2 showed a similar pattern of results in the response times to the following context of an assimilation, although the viability effect for the nonwords did not reach significance.

Our initial analysis of phonological inference hypothesised two possible mechanisms: one in which phonological context is employed to compensate for surface ambiguities in a non-lexical manner, and one in which context checking is employed lexically. The former hypothesis is supported by the observation of phonological viability effects for nonwords in Experiment 1, with a similar trend in Experiment 2. The latter hypothesis gains support from the interaction between lexical status and viability found in Experiment 1, and again from a similar trend in Experiment 2. It seems that the resolution of surface ambiguity depends on both non-lexical and lexical processes.

Abstractness and Phonological Perception

The phoneme monitoring task specifically directs subjects' attention to the phonological form of speech, as opposed to priming studies, which evaluate access to semantic information. Despite this, in Experiment 1, a large proportion of surface labial or velar segments were responded to as coronals. For the viable assimilated segments in real word carriers, this constituted nearly 60% of the responses.

Subjects seem to make use of a mixture of surface and underlying codes in their judgements about phonological form. However, these judgements do not conform in a standard way to lexical and pre-lexical judgements, since subjects are able to make responses to real words based on surface form, as well as using underlying representations for nonwords. The abstractness of subjects' responses depends on a combination of lexical and phonological factors.

Conclusions

The pattern that emerges from these experiments is that phonological inference can occur when lexical access fails, in the perception of nonwords, but that the effects of this inference process are strengthened when lexical access is successful. This interaction suggests that neither a purely pre-lexical nor a purely post-lexical account of the process will suffice.

Although the interpretation of interaction in terms of cognitive architecture is a rather hazardous venture (Norris, 1993; Tabossi, 1993), we feel that the behavior uncovered here can be simply explained as the product of connectionist learning principles applied to the mapping involved in speech perception. The process of word recognition involves the mapping from speech input to both semantic and phonological information. Thus, the retrieval of phonological information (i.e. the underlying form of the words) will normally interact strongly with the retrieval of semantic knowledge. This implies that, under normal circumstances, phonological inference will interact with lexical access, explaining the strong effects of phonological inference for the word stimuli in our experiments. However, when access to lexical information is blocked, by the use of mismatching segments, phonological identification is still possible, allowing weaker inferential effects. By this account, phonological inference can be thought of as lexical, since it is influenced by the retrieval of word-identity information, but it may also be thought of as non-lexical, since it can occur independently of the recognition of words. In both cases, the retrieval of the underlying phonological form of speech makes the best use of the available cues.

Acknowledgements

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

References

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition. *Journal of Memory and Language*, 32, 193-210.

Dell, G. S., & Newman, J. E. (1980). Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*, 19, 608-623.

Foss, D. J., Harwood, D. A., & Blank, M. A. (1980). Deciphering decoding decisions: Data and devices. In R. A. Cole (Ed.), *Perception and Production of Fluent Speech*. Hillsdale, NJ: Erlbaum.

Gaskell, G. (1994). Spoken Word Recognition: A Combined Computational and Experimental Approach. PhD thesis, Birkbeck College, University of London.

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

Gaskell, G., & Marslen-Wilson, W. (1993). Match and mismatch in phonological context. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Harrington, J., & Johnstone, A. (1987). The effects of word boundary ambiguity in continuous speech recognition. In *Proceedings of XIth International Congress of Phonetic Sciences*. Tallin, Estonia.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

Marslen-Wilson, W., Moss, H. E., & Halen, S. van. (in press). Perceptual distance and competition in lexical access.

Norris, D. (1993). Bottom up connectionist models of 'interaction'. In G. Altmann & R. Shillcock (Eds.), *Cognitive Models of Language Processes: Second Sperlonga Meeting*. Hove: Erlbaum.

Tabossi, P. (1993). Connections, competitions and cohorts: comments on the chapters by Marslen-Wilson; Norris; and Bard and Shillcock. In G. Altmann & R. Shillcock (Eds.), *Cognitive Models of Speech Processing: Second Sperlonga Meeting*. Hove: Erlbaum.