Københavns Universitet

Investigating lexical competition and the cost of phonemic restoration

Balling, Laura Winther; Morris, David Jackson; Tøndering, John

Published in:
Acoustical Society of America. Journal

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Investigating lexical competition and the cost of phonemic restoration

Laura Winther Balling,[1,a)] David Jackson Morris,[2,b)] and John Tøndering[2]

[1]*Department of Management, Society and Communication, Copenhagen Business School, Dalgas Have 15, 2000 Frederiksberg, Denmark*

[2]*Department of Nordic Studies and Linguistics, University of Copenhagen, Emil Holms Kanal 2, 2300 Copenhagen, Denmark*

Due to phonemic restoration, listeners can reliably perceive words when a phoneme is replaced with noise. The cost associated with this process was investigated along with the effect of lexical uniqueness on phonemic restoration, using data from a lexical decision experiment where noise replaced phonemes that were either uniqueness points (the phoneme at which a word deviates from all nonrelated words that share the same onset) or phonemes immediately prior to these. A baseline condition was also included with no noise-interrupted stimuli. Results showed a significant cost of phonemic restoration, with 100 ms longer word identification times and a 14% decrease in word identification accuracy for interrupted stimuli compared to the baseline. Regression analysis of response times from the interrupted conditions showed no effect of whether the interrupted phoneme was a uniqueness point, but significant effects for several temporal attributes of the stimuli, including the duration and position of the interrupted segment. These results indicate that uniqueness points are not distinct breakpoints in the cohort reduction that occurs during lexical processing, but that temporal properties of the interrupted stimuli are central to auditory word recognition. These results are interpreted in the context of models of speech perception.
© 2017 Acoustical Society of America. https://doi.org/10.1121/1.5017603

[JFL]   Pages: 3603–3612

I. INTRODUCTION

Phonemic restoration (PR) refers to the auditory phenomenon whereby listeners are able to perceive a word even though specific parts of the word have been replaced or masked with a noise interruptor. This phenomenon was first described within an experimental context by Warren (1970) and has since been explored by Samuel (1981, 1987; Samuel and Ressler, 1986) and others (e.g., Bashford *et al.*, 1992). However, much remains to be learnt about the listening and lexical conditions that drive PR. We investigate these in the experiment reported here, where we consider a range of possible predictors of PR in a lexical decision experiment.

One point on which previous studies of PR are in consensus is that restoration is a consummate perceptual phenomenon whereby participants are usually able to perceive the word, but are unable to reliably identify the position of an interruptor within the word (Warren, 1970; Warren and Obusek, 1971; Warren and Sherman, 1974). Neither are they able to reliably discriminate cases where a noise interruptor replaces a phoneme that has been extracted from the word, from cases where noise is added to the target phoneme (Samuel, 1981; Samuel and Ressler, 1986). Investigations of PR have not been confined to studies using transient interruptors at specific positions within single words. Continuous interruption of longer speech utterances with fully modulated noise has also been investigated as a form of PR, and is sometimes referred to as auditory or temporal induction (Bashford *et al.*, 1992). Investigations have shown significant PR of sentences for normal hearing and listeners with mild hearing impairment using continuous interruptors, where noise duty cycles were half-on and half-off (Başkent, 2010). It is thought that the continuity in this type of interruption plays a facilitatory role in PR by providing a perceptual background upon which restoration can take place (Bashford and Warren, 1987).

One factor that may influence PR is the identity of the affected phoneme. Samuel and Ressler (1986) found that missing nasals elicited less PR than fricatives and vowels, and interpreted this as being due to the larger acoustic mismatch between nasals and the speech-correlated noise that replaced phonemes as the interruptor. The authors attributed the reduction in PR for words where there was considerable acoustic mismatch between the interruptor and the affected phoneme to the attention of the listener being drawn to the noise, and away from lexical processing. There is also other evidence that PR can be influenced by attention being directed to the affected phoneme, namely, effects of cuing and item-specific learning. Samuel and Ressler (1986) showed that PR is reduced when participants are cued with information about the target word and the affected phoneme, but not when only the word, phoneme position, or phoneme identity of the affected phoneme was used as a cue. The same study also showed that when cued with lexical information, in the form of a written version of the upcoming auditory target, participants exhibited lower PR for the

specific target on later trials, but this learning did not generalize to other items.

Investigations of phoneme monitoring indicate that the phonological context, particularly the stress status of the syllable, influences the speed with which a listener can respond to a target phoneme in a word. The detection of target phonemes, when the preceding intonation contour cued stressed monosyllables, has been shown to be faster than when preceding intonation cued unstressed items (Cutler, 1976). These results have been interpreted to suggest that listeners are able to make real-time changes in their attention to speech, so that processing of syllables that are expected to be important based on local-phonological information is selectively enhanced. This work on intra-word attention raises the question of whether the attentional modulation that can occur due to phonological context may also happen due to the content of words as they separate themselves from other candidates in fluent speech.

To investigate this further, the present study examines PR in an auditory lexical decision experiment that also factored in uniqueness points (UPs), as described by Marslen-Wilson (1984). The underlying logic of UPs is that early in a target word many words in the vocabulary are compatible with the input, but as the sounds of the target word become gradually available, fewer and fewer word candidates remain compatible with the emergent signal. The UP occurs when only the target word (and inflected, derived and compound forms starting with that target) is compatible with the input. The later in a word that the UP occurs, the longer a listener takes to recognize that word (e.g., Balling and Baayen, 2008; Marslen-Wilson, 1990). The UP thus marks one endpoint of the lexical competition that unfolds over time in spoken word recognition, and thus a dramatic change in lexical probabilities (Balling and Baayen, 2012) which may give the UP-phoneme a special status. While Balling and Baayen (2008, 2012) showed the relevance of two UPs for morphologically complex words, we focus here on the single UP that is relevant for morphologically simple words.

Lexical uniqueness has been investigated in conjunction with PR by Samuel (1987), but with a different analytical perspective, as we shall argue below, and a different task than in the present study. In his work on PR, Samuel used a decision task where participants were asked to determine whether the noise in the stimulus words was added to or replaced a phoneme. Signal detection theory analysis was used to determine the discriminability of the two types of noise addition methods, with discriminability and bias measures calculated based on rates of misses and false alarms. Samuel defined misses as responses of "added" when the replaced stimuli were presented (i.e., cases where participants reported hearing a speech sound which was not in fact present in the signal), and false alarms as responses of "replaced" when the added stimuli where presented (i.e., cases where participants reported hearing no speech sound when it was in fact present). Stimuli were words that become lexically ambiguous without the affected phoneme and words that remain lexically unique without the affected phoneme, for instance, *legion* which is ambiguous between *legion* and *region* if the initial phoneme is removed,

compared with *lesion* which remains unique even without the initial phoneme. He found that lexically ambiguous words showed lower discriminability than lexically unique words, i.e., participants had more difficulty distinguishing words with noise that replaced the target phoneme from words with noise added to the target phoneme. Samuel interpreted this lower discriminability when the word is ambiguous as evidence of more PR. In contrast, there was a tendency for participants to report lexically unique words as intact more often than lexically ambiguous words. In a second experiment, Samuel compared words with rare first syllables (and hence presumably early uniqueness points, which would typically entail faster recognition) with words with common first syllables (and hence generally later uniqueness points and thus slower recognition), finding more restoration for early unique words than for late unique words. This was observed especially when affected vowels occurred later in the word and was interpreted in terms of lexical influences on speech perception.

The use of the discrimination task in Samuel (1987) allowed investigation of the mechanisms of PR within the context of lexical uniqueness. Our study differs from Samuel (1987), in that we chose a lexical decision task in order to focus on the lexical level of processing, in instances where the UP or the phoneme before the UP was replaced with noise. We used the lexical decision paradigm in order to address two related but different analytical perspectives: the degree of PR attested is informative about how the different predictor variables—specifically those to do with uniqueness points and other temporal aspects of the competition and recognition process—affect word recognition, while the effect of the different variables on the degree of PR helps us understand how PR works. Together, these perspectives have implications both for our understanding of speech perception and word recognition and of the processes involved in auditory comprehension of degraded word-level stimuli. In addition, the use of the lexical decision task allows us to compare words with and without interrupted segments, and thus to investigate the cost of PR in lexical processing, in terms of both response time and the degree of successful recognition of stimuli with interrupted segments.

In addition to the key variables of UP position and interruptor condition (no noise interruptor vs noise interruptor on the UP vs noise interruptor on the pre-UP), we also investigate a range of other variables. These include the duration and position in the word of the removed segment which could affect any effects of interruptor condition and UP. We also consider the phonological identity, following up on the work by Samuel and Ressler (1986) discussed above, and the stress status of the syllable in which the noise occurs, further investigating the stress status effects reported by Cutler (1976).

II. METHODS

An auditory lexical decision experiment was run in order to investigate lexical competition and PR. In auditory lexical decision, participants hear a mixture of real words in the target language, in this case Danish, and phonotactically

legal nonwords. For each item, the participants indicate with a button press, as quickly and as accurately as possible, whether they recognize what they just heard as a word in Danish. The main response variables are reaction time (RT) on word responses to word stimuli and proportion of word responses to word stimuli. The nonwords were included in order to make the lexical decision task work—without nonwords there is no decision to make—but they are not further analysed, except for their contribution to the d' measure. The experiment included words in three different conditions. There were two conditions where a phoneme was replaced with noise: the UP-interruptor condition where the UP-phoneme of the word was replaced, and the PRE-interruptor condition where the phoneme immediately preceding the UP was replaced. These were compared to the third condition, NONE, where no phonemes were replaced. In addition to the condition difference, we also consider several measures relating to the replaced phoneme: its position, given the increase in PR for longer words found by Samuel (1981) and Bashford and Warren (1987) and the better discriminability found for phonemes occurring late in a word by Samuel (1987), duration, syllable stress and degree of match with the interruptor replacing it. Finally, the analyses included a range of corpus-based lexical predictors such as frequencies and neighbourhood density, as well as variables relating to the experimental context such as trial number and reaction time on previous trials (see more detail below, Sec. III A).

A. Participants

The participants were 46 native speakers of Danish recruited on campus by the first author. There were 27 females and 19 males, aged between 19 and 50 yr (mean 22.6, SD 5.4), with no self-reported hearing or cognitive impairments. The participants were volunteers and were not paid for participation. They received information about the procedure and general purpose before the experiment and more detail about the stimuli and specific purpose after their participation. All participants provided signed informed consent before taking part in the experiment.

B. Stimuli

The words used in this experiment were Danish monomorphemic words drawn from Balling and Baayen (2008 and 2012 experiment 2). A 28-yr-old female recorded the stimuli in a sound-treated studio with a high quality recording microphone. The stimuli were di- and tri-syllables and included no monosyllables.

The UP of each target word was determined by querying *KorpusDK*, a large representative corpus of Danish texts (Det Danske Sprog- og Litteraturselskab, n.d.), for spellings that are compatible with the pronunciation of that target. We queried the corpus for increasingly larger chunks of the word, with the UP occurring when the query returned only the target word and inflected, derived and compound continuations of the target. The duration of that phoneme was then measured using Praat, and the UP position in ms was defined as the time from the onset of the word to the temporal midpoint of the UP-segment. This provided a fine-grained

measure of uniqueness and allowed us to investigate the UP-effects in detail. In a similar way, we determined for each nonword the phoneme at which it became uniquely identifiable as a nonword, i.e., a "nonword point," which is the nonword equivalent to the word UP.

We avoided stimulus items where the UP or pre-UP was a sibilant as we anticipated that the acoustic similarity between those segments and the noise interruptors may introduce a confound (cf. Samuel, 1981). Vocalic pre-UPs followed by continuants, and diphthongs over the UP and pre-UP, were deemed too hard to segment and were therefore also avoided. We also considered to what extent the target words are unique at word offset also without the interrupted segments, and tried to only include words that remained unique at word offset both without the UP and without the pre-UP (37 of the 60 target items) and words that were not unique at word offset, neither without the UP, nor without the pre-UP (19 items). We were not able to entirely avoid words that were only unique at word offset in one of the conditions, but the number of such cases were low, only four items.

These selection criteria left 74 words, of which we used 60 as target words and 14 as training words. There was an equivalent number of nonwords, divided in the same way between target and training. The nonwords were constructed by changing between one and three phonemes in the words of the original experiments, and were recorded by the same speaker. Figure 1 shows an example of a single word in all three stimulus conditions (without interruptor, with interruptor replacing the pre-UP or replacing the UP), and sound files of these are available as supplementary material.[1]

To ensure consistency in editing and to mitigate the effect of co-articulation, the following criteria were observed when extracting the UP and pre-UP phonemes from the words and nonwords:

(1) The full closed phases of stops were excised when these occurred.
(2) Stops that were followed by vowels included 1–3 fundamental periods of the following vowel in the excised phoneme.
(3) Long vowels before stops were extracted up until the burst of the stop.
(4) When the excised phoneme was a voiced continuant followed by a vowel, 3–5 fundamental periods of the vowel were also included.
(5) When continuants were followed by voiced liquids, some overlap was also excised, so that the phoneme that was being excised was inaudible.

We examined the durations of the excised segments and found them to be bimodally distributed with a local minimum at 70 ms, local maxima at 40 and 110 ms, and a tail extending to 253 ms. We used the peaks in this distribution to derive the durations of a set of interruptors, while also adding some longer durations to account for the tail. The interruptors were 40, 110, 170, and 230 ms, and these were added to the words based on the duration of the extracted phonemes, so that the 40 ms interruptor was added if the duration of the extracted phoneme was less than 70 ms; the
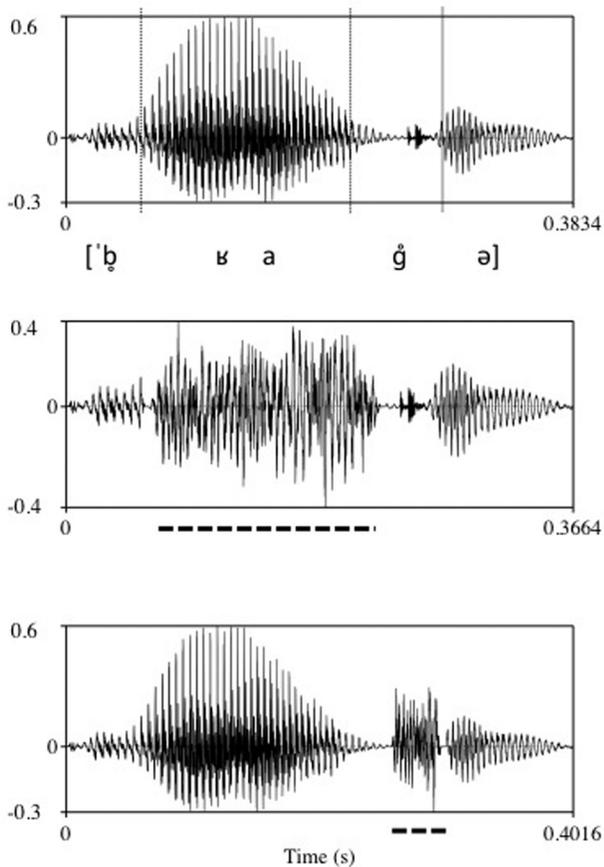
J. Acoust. Soc. Am. **142** (6), December 2017

Balling *et al.* 3605

FIG. 1. Waveforms of the stimulus word "brække" (break). Without interruptor (condition NONE) in the upper panel; interruptor replacing the phoneme immediately before the uniqueness point (condition PRE-interruptor) in the middle panel; and interruptor replacing the phoneme at the uniqueness point (condition UP-interruptor) in the lower panel. Dashed lines indicate the placement of the interruptor. The three stimuli can be heard as supplemental material 1, 2, and 3, respectively.

110 ms interruptor was added if the extracted phoneme was 71–140 ms; the 170 ms interruptor was added if the extracted phoneme was 141–200 ms; and, the 230 ms interruptor was added if the extracted phoneme was greater than 200 ms in duration. In the stimulus set, the range of the durational mismatch between extracted phoneme and interruptor was 0 to 36 ms (mean = 13, SD = 9).

The spectral content of all extracted phonemes was used to derive speech-shaped noise (SSN) that was inserted into each stimulus as an interruptor. Spectral shaping was achieved by calculating the fast Fourier transform (FFT) of all extracted phonemes after which the phase of the frequency components were randomised. The inverse FFT was then calculated to generate SSN that approximated the spectral characteristics of the group of extracted phonemes. The amplitude of the SSN was normalized to that of all the extracted segments. A raised-cosine function was then used to shape the onset and offset of the SSN so that durations of these were approximately proportional to the length of the interruptor. Function durations were 5 ms for the 40 ms interruptor; 10 ms for the 110 ms interruptor; 15 ms for 170 ms interruptor; and 20 ms for 230 ms interruptor. These interruptors were inserted into the stimulus words at zero-crossings

where the UP or pre-UP had been excised, with a 2–4 ms null amplitude portion on either side. This portion was introduced so that the ramping would taper from and to zero, and in order to reduce artifacts that may have arisen due to phase disruption between the speech sounds and the adjacent interruptor. The first and second authors listened to all items and deemed the null amplitude portion to be imperceptible. Generation of the interruptor was carried out in MATLAB, and stimulus editing was performed in PRAAT (Boersma and Weenink, 2012).

C. Procedure

Each of the 60 words in the test phase and 14 words in the training phase occurred in three different conditions: with an interruptor replacing the UP (condition UP-interruptor) or the phoneme before the UP (condition PRE-interruptor) or with no interruptor (condition NONE). Each participant heard each word once and heard the same number of words in each condition (in the test phase; in the training phase, they heard approximately the same number). The 60 nonwords in the test phase and 14 nonwords in the training phase were divided evenly (in the training phase, approximately equally) with one third having an interruptor on the nonword point (parallel to the UP-interruptor condition for the words), one third with an interruptor on the phoneme before the nonword point (parallel to the PRE-interruptor condition for the words), and one third without interruptors. In this way, there were three different experimental lists which varied systematically with regard to the condition in which each word occurred. Sixteen participants completed list 1, and 15 participants each completed lists 2 and 3. The training and test phases were parallel in terms of the balance of conditions, words and nonwords, and there were no differences in procedure.

On arrival, participants were orally instructed about the task. They were told that they would hear a mixture of words in Danish and wordlike strings that were not real words, and that they were to press a button labelled "YES" (with their dominant hand) if they recognized what they heard as a word and otherwise a button labelled "NO" (with their other hand). The two-handed response was used because this tends to give fewer wrong keypresses than when both responses are made with the same hand. In addition to receiving these standard lexical decision instructions, participants were told that some of the stimuli had some noise in them and that they were to decide whether they recognized the word or not regardless of that noise. They were then individually seated at desktop computers in separate quiet rooms. The stimuli were diotically presented over quality circumaural headphones which were connected to computers via soundcards that had identical volume settings. Participants were presented with a screen displaying a written version of the instructions, after which they initiated a training block consisting of 28 randomised items. Half of these were words and half were nonwords and they included interruptors that replaced UP and pre-UP phonemes in a similar ratio to the experimental items, as described above. The experiment itself had 60 word and 60 nonword items in different random orders for each participant, and immediately followed the

training phase with no break or other indication. Each trial was initiated by the decision button press on the previous trial. An attentional cue, which was a white cross, was shown on the computer screen 500 ms prior to stimulus onset. This meant that there was at least 500 ms from the offset of one word to the onset of the next. Reaction time for a word was measured from the onset of that word. The experiment was run in DMDX (Forster and Forster, 2003). The training and test phases, not including instructions and debriefing, took approximately 5 min, while the entire procedure took about 10 min.

III. RESULTS

This section first outlines the variables and analysis procedure used, before presenting two levels of analysis: one including all word test items in all conditions and one including only words in the interruptor conditions PRE-interruptor and UP-interruptor. The former analysis is used as a baseline that establishes the effect of introducing the interruptor, while the latter investigates the properties of the missing segments and interruptors in more detail and thus addresses the core questions of this study. Neither analysis includes the responses to the nonwords.

A. Variables and analysis procedure

As is common in the analysis of lexical decision experiments, we focused on responses to words, primarily reaction times to words recognized as words ("correct" responses) and secondarily the identification of each stimulus presentation as either a word or a nonword. These two dependent variables were analysed using linear mixed regression models in the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2016) packages of the R environment (R Core Team, 2016). Reaction times were log-transformed to reduce skewness as were several of the lexical predictors including frequency. Identification as word was analysed using a generalized linear mixed model with a logistic link function.

Regression models were built using a bottom-up approach in which variables were successively added, starting with the most control-oriented variables and ending with those crucial to our investigation. The more control-oriented variables were included in the following order: variables related to experimental context (trial number, RT, and correctness on previous trial), word length in ms, and lexical predictors that have been significant in previous analyses of Danish lexical decision data (Balling and Baayen, 2008, 2012). The latter were mean letter bigram frequency (as a rough index of mean phoneme bigram frequency, since no spoken corpora of sufficient size and detail are available for Danish), word form frequency, morphological family size and neighborhood density, all extracted from *KorpusDK*, as described by Balling and Baayen (2012). The key variables UP position in ms and interruptor condition, including an interaction between the two, were tested at the end. The analysis of the interruptor conditions (PRE-interruptor and UP-interruptor) included the same variables as well as the following additional variables added at the end of the analysis: the position of interruptor (in phonemes and syllables), the length of the replaced phoneme, the degree of temporal

mismatch between replaced phoneme and interruptor, the phonological category of the replaced phoneme (vowel, stop or obstruent), and the stress status of the syllable in which the interruptor was inserted (stressed or unstressed categorized according to a phonetic and not phonological approach to syllabification). Nonsignificant predictors were excluded from the final models reported. The models for RTs are ones where observations with large standardized residuals were removed, to avoid highly skewed distributions of residuals.

B. Regression analyses: All conditions

The key descriptive statistics from the analysis of all word items are shown in Table I. It can be seen that there is a clear advantage for the NONE condition, both in terms of faster reaction times and fewer identifications of the target as nonword, as expressed as percentages and as d' values calculated from the true positives or hits. The "identification as nonword" is what would traditionally be called the "error rate" in a lexical decision task, but in the interruptor conditions, nonword responses to words may reflect that the interruptor renders the target unrecognizable.

Linear mixed-effects model for RT and nonword responses to words confirm this: responses in the baseline condition with no interruptor are significantly faster than in both interruptor conditions, and there are also significantly fewer nonword responses in the baseline condition (all p-values < 0.0001). This indicates that PR incurs a cost in terms of reaction time. Moreover, PR does not always happen, as indicated by the 14% mean difference between NONE condition on one hand and the mean of the two interruptor conditions considered jointly on the other. For RTs, the difference between the NONE condition and the two interruptor conditions, which can be thought of as the mean cost of PR, is approximately 100 ms. It is also interesting to note that the overall analysis showed a difference between the two interruptor conditions, with faster recognition and fewer nonword responses in the UP-interruptor condition than in the PRE-interruptor condition (p-values < 0.001).[2] What this overall analysis cannot tell us, however, is whether the significant difference between the UP-interruptor and PRE-interruptor conditions is driven by the status of the UP as a breakpoint in the recognition process or by the fact that the UP necessarily occurs later in the word than the pre-UP. In order to investigate that, we analysed the interruptor conditions, including variables that were only available for the items in those two conditions, i.e., variables measuring different characteristics of the interruptors and the replaced phonemes.

TABLE I. Mean reaction time on correct trials, mean rates of identification of words as nonword and d', by interruptor condition. Standard deviations in parentheses.

	RT mean (SD)	Identification as nonword (SD)	d'
UP-interruptor	993 (280) ms	0.148 (0.089)	1.5
PRE-interruptor	1049 (314) ms	0.214 (0.090)	1.1
Mean of interruptor conditions	1020 (298)	0.181 (0.067)	1.3
NONE	917 (220) ms	0.039 (0.035)	2.5

J. Acoust. Soc. Am. **142** (6), December 2017

Balling *et al.* 3607

C. Regression analyses: Interruptor conditions

The analyses of responses in the UP-interruptor and PRE-interruptor conditions are shown in Table II and Fig. 2 (RT) and in Table III and Fig. 3 (nonword responses). The RT analysis is the more sensitive measure of lexical processing and indicates the cost of PR, but the nonword responses to stimuli is also of value as an index of the success of recognition in spite of the interruption of a phoneme.

The first noteworthy feature of this analysis is the absence of a condition effect, once the position in the word of the interrupted phoneme is included in the regression model. In other words, we see a significant effect of how early in the word the interrupted phoneme occurs (both expressed as phoneme and syllable number), with longer RTs when the interrupted phoneme occurs earlier in the word, but no effect of whether the interrupted phoneme is the UP or the pre-UP. This is in contrast to the all-items analysis which showed a clear difference between the two conditions (see Sec. III B). This difference between the two levels of analysis occurs because the analyses of the interruptor conditions alone may include a range of predictors that pertain specifically to the interruptors, and it turns out that these predictors explain the apparent difference between the PRE-interruptor and UP-interruptor conditions in the overall analysis. The effects of condition and interruptor position are illustrated in the top rows of Figs. 2 and 3. It is worth noting that the effect of the position of the interrupted phoneme in the word only has an effect on reaction time, not on recognition success, suggesting that the degree of PR is the same irrespective of the interruptor position, but that it happens faster and thus more automatically later in the word.

In the RT analysis, we see two different effects that relate to position: the position of the interrupted phoneme coded in numerical sequence from the beginning of the word, and the syllable in which the interruptor occurs. These effects are illustrated in the two right panels of the top row of Fig. 2. The later in the word the interruptor occurs, the less it seems to affect RT. This is in contrast to the finding of Samuel (1987) that showed that participants are good at discriminating between added and replaced stimuli for word-final phonemes, but it is in line with his finding that participants are more biased to report stimuli as intact for phonemes occurring later in the word. If we follow Samuel's interpretation of the two measures (discriminability and reporting intactness), then this points to a post-perceptual decision bias driving this aspect of our results. On the other hand, the fact that these variables only affect RTs and not recognition accuracy indicates that they are related to the automaticity of PR rather than exclusively to a post-perceptual bias.

Next, we see that the prosodic content of the word has a bearing on PR, as words with interruptors occurring in stressed syllables have longer RTs (see Fig. 2, left panel of the middle row). In other words, the cost of PR is greater when the interrupted syllable is stressed. This effect is significant even when the duration of the missing segment (with which it covaries) is statistically controlled, indicating that there is an effect of syllable stress over and above the effect of missing segment duration. This extra cost may be due to interruption occurring on an acoustically prominent portion of speech, which is the stressed syllable. According to Altmann and Carter (1989) stressed syllables are more informative than unstressed syllables, so when interruption occurs in a stressed syllable, the information value that is removed is higher and the restoration process becomes harder. It may also indicate that the occurrence of interruptors in stressed syllables is more detrimental to the selection of lexical candidates, as this process may involve matching or modulation according to stress patterns. In this light, the greater cost of PR for stressed phonemes is further confirmation of Cutler's (1976) finding that the prediction of stress location is an important determinant of lexical selection. This would seem to be the case not only for sentences, as she reported, but also for di- and trisyllables presented in isolation.

Another important effect is illustrated in the lower left panel of Fig. 2. We see that, as is commonly observed in auditory lexical decision tasks, words with later UPs have longer RTs. The effect is nonlinear, with no differences for the lowest UP-values, but quite large differences for later UPs. This holds across all three conditions and no interaction

TABLE II. Summary of fixed effects in analysis of reaction time to interruptor conditions. The factors are treatment coded with PRE-interruptor as the reference level for the factor Condition and 1 as the reference level for the factor Interruptor syllable. The model also included random effects of item (SD estimated at 0.0040) and participant (SD 0.0112), random slopes for interruptor phoneme position by participant (SD 0.0006) and random levels for condition (SD 0.0016) and interruptor syllable (syllable 2 SD 0.0007, syllable 3 SD 0.0037) by participant. The residual SD was estimated at 0.0245. n = 1468.

|	Estimate	Std. error	df	t	p
Intercept	6.8860	0.0263	97.1	262.2440	<0.0001
Condition:UP-interruptor	0.0059	0.0196	88.3	0.2990	0.7656
Interruptor phoneme position	−0.0308	0.0148	55.0	−2.0850	0.0417
Interruptor syllable: 2	−0.0497	0.0214	53.1	−2.3200	0.0242
Interruptor syllable: 3	−0.1276	0.0548	49.2	−2.3300	0.0239
Prominence of interruptor syllable: stressed	0.0471	0.0136	763.5	3.4550	0.0006
UP in ms (linear)	−0.2181	0.0846	52.3	−2.5790	0.0128
UP in ms (quadratic)	0.2600	0.0827	52.7	3.1420	0.0027
Missing segment length in ms	0.0451	0.0054	1248.0	8.3880	<0.0001
Temporal mismatch in ms	0.0156	0.0056	908.4	2.7800	0.0056
Log RT on previous trial	0.0222	0.0046	1396.0	4.8060	<0.0001
Word length in ms	0.0362	0.0113	51.4	3.1970	0.0024
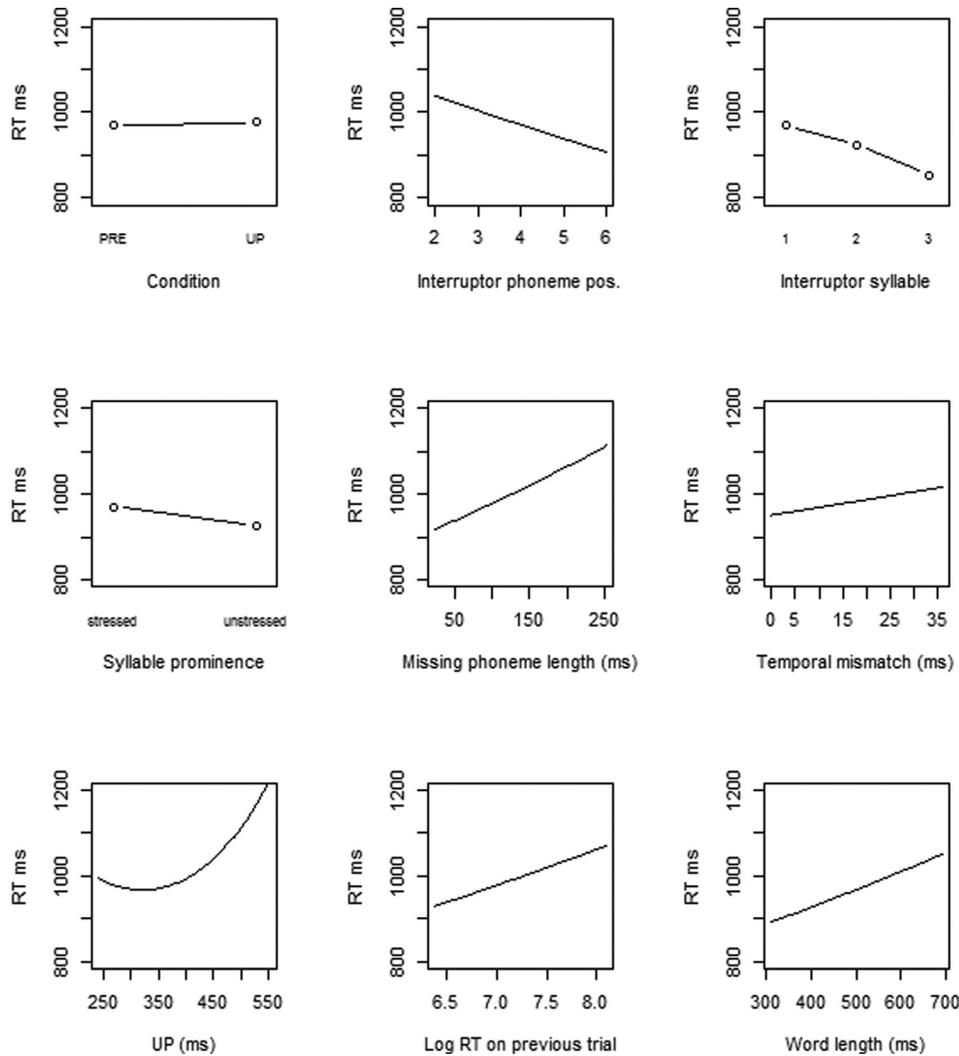
FIG. 2. Fixed effects in the analysis of reaction times to words in interruptor conditions. The plots are partial effects plots, which means that each panel shows the effect of the given predictor when all other predictors are held constant at their median (for co-variates) or reference levels (for factors). For ease of interpretation, the values on the vertical axis are back-transformed from the logarithm used in the analysis.

with condition was observed. Together, the absence of a condition effect and presence of a UP-position effect indicate that although the UP is an important point in the gradual reduction of the lexical competition cohort, it is not a distinct

TABLE III. Summary of fixed effects in mixed logistic regression analysis of nonword responses to words in the interruptor conditions. The factors are treatment coded with PRE-interruptor as the reference level for the factor UP and 1 as the reference level for the factor Interruptor syllable. The model also included random effects of item (SD estimated at 1.2912) and participant (SD 0.4628) and random levels for condition by participant (SD 0.4579). n = 1840.

	Estimate	Standard error	z	p
Intercept	−1.9661	0.2918	−6.7390	<0.0001
Word length in ms	−0.5967	0.1974	−3.0230	0.0025
Log word form frequency	−0.4425	0.2015	−2.1960	0.0281
Interruptor syllable: 2	−0.9410	0.4026	−2.3370	0.0194
Interruptor syllable: 3	−1.8295	1.2115	−1.5100	0.1310
Missing segment length in ms	1.0316	0.1068	9.6630	<0.0001
Temporal mismatch in ms	0.2093	0.1067	1.9630	0.0497
Condition:UP-interruptor	0.0211	0.1928	0.1100	0.9127

breakpoint in the processing. More tentatively it also suggests that the lexical competition process is fundamentally similar across all three conditions.

In the analysis, we also considered the issue of whether the words with interruptors were unique without the removed segments. As mentioned in Sec. II B, we included mostly words that were either unique without the removed segment in both conditions (37 items) or not unique without the removed segment in both conditions (19 items). The remaining four items were only unique at word offset in one of the conditions; to ensure that these four potentially problematic items were not decisive to our results, we reran the main analyses reported here on a reduced dataset that excluded these four items, and found that all conclusions remained the same. In addition, we ran an analysis where we included, in addition to the significant variables listed in Tables II and III, a factor indicating whether a word was unique at word offset or not without the removed segment; this factor had no effect on the response variables and did not affect the other significant results. This further supports the idea that phoneme restoration actually does happen and that the

J. Acoust. Soc. Am. **142** (6), December 2017
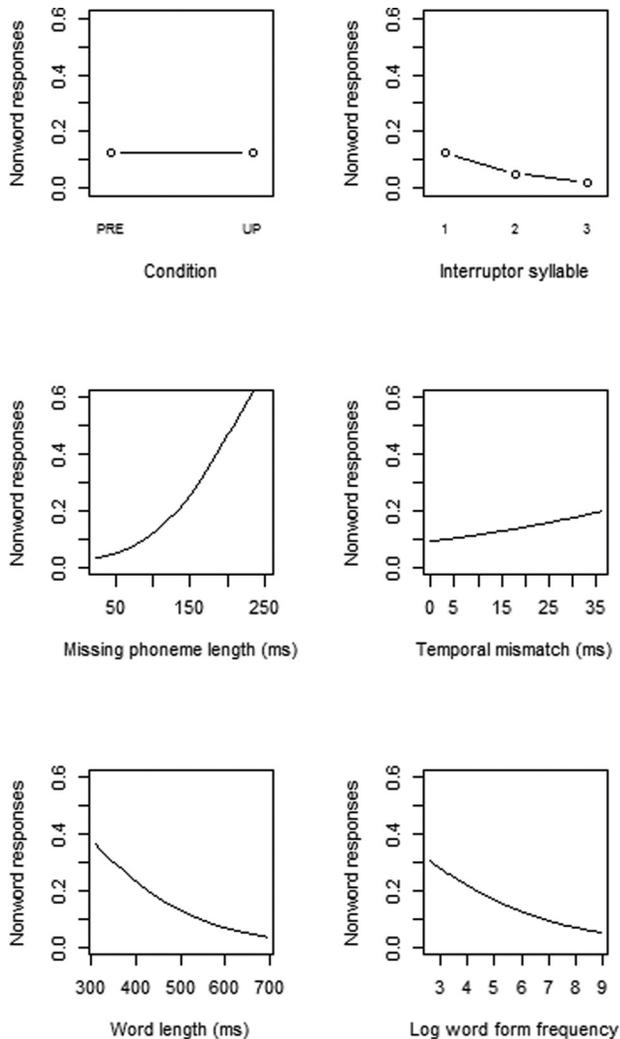
Balling *et al.*   3609

FIG. 3. Partial effects plot of the fixed effects in the analysis of nonword responses to words in interruptor conditions. Log odds are back-transformed to probabilities.

lexical competition process remains similar also for words with interruptors.

A further, very strong effect that is evident is that of the length of the interrupted segment. This means that the longer the interrupted segment, the slower the RT (see middle panel of the middle row of Fig. 2) and the more nonword responses (left panel of the middle row of Fig. 3). In other words, longer interrupted phonemes are associated with less successful PR. We also included a term for the phonological identity of the interrupted phonemes as vowels, obstruents or sonorants (with which the length of the interrupted phoneme varies), but this was not significant. In addition to this, the length difference between the interrupted phoneme and the inserted interruptor also plays a role, with shorter RTs and more successful recognition when the difference was small (see the right panels of the middle rows in Figs. 2 and 3), even though the mismatch was always less than 36 ms. The effects of temporal mismatch are smaller than some of the other effects in the experiment (as can be seen from a comparison of the right panels in the middle rows of Fig. 2 and Fig. 3 with most of the other panels in the respective figures), but systematic enough to be significant for both RT and nonword

identification rate. This suggests that participants are sensitive to even very small mismatches in duration.

The more control-oriented variables included in the model of RTs are shown in the two final panels of Fig. 2. The lower middle panel illustrates the effect of the RT for the previous item which correlates positively with RT on the current item. This means that, despite completing a training block, there were systematic fluctuations in the speed of the responses. These effects are not relevant to the key questions of this paper, but including the effect improves the model. Next, we see the effect of word length in the lower right panel of Fig. 2, with slower responses to longer words, an effect that is commonly encountered in lexical decision experiments. Interestingly, this effect is reversed in the analysis of nonword responses, where there are fewer nonword responses on long words as seen in the lower left panel of Fig. 3. The fewer nonword responses for longer words could be caused by the longer RTs giving more time to recognize the word correctly in spite of the removal of a phoneme.

Another lexical predictor that differs between the two analyses is word form frequency, as found in a large corpus of Danish texts. In assessing candidate models, we found no significant effect of word form frequency on RTs ($t = -1.469$, $p = 0.148$), but a facilitatory effect in the nonword response analysis, with more word (or "correct") responses to stimulus items which are more frequent, as is typically seen in lexical decision experiments. This effect is shown in the lower right panel of Fig. 3.

IV. DISCUSSION

Two main results of the study reported here is that PR comes at a cost and that PR does not always happen, as was observed in the reaction times and nonword identification rates by condition, see Table I. PR is a strong perceptual phenomenon, but our results indicate that it is modulated by a range of stimulus properties, especially the length and the position of the interrupted segment. The use of a lexical decision task and inclusion of a baseline condition allowed us to probe unexplored aspects of PR, but it did not allow us to distinguish between PR as part of the perceptual process and PR as a post-perceptual decision phenomenon, since the task involves both perception and decision. However, previous studies indicate that PR is at least partially an actual perceptual process and that seems also to be the case here: after the experiment, the participants were consistently surprised when told that the noise that they had heard had replaced sounds, rather than been added to the signal. In other words, though we cannot separate the precise contributions of perceptual and decision processes, it seems clear that both play a role.

An important part of our purpose with this investigation was to shed light on the nature of the UP and how this was affected by PR, and vice versa. We have previously described the UP as a dramatic shift in probability (Balling and Baayen, 2012) and we wanted to investigate here whether the shift is so dramatic that the UP acquires a special status in terms of the degree of PR. This is clearly not the case: when the position of a missing phoneme in its

target word was controlled, it did not matter whether the missing phoneme was the UP or the phoneme preceding the UP. The shift in probability that occurs at the UP remains substantial and important, as shown by the effect of UP position in ms in this and other studies, but it is an importance that arises from the gradual reduction of lexical competition, rather than from the UP having an elevated status in lexical processing.

Another categorical difference that may be driven by a continuous temporal effect is that of phoneme category. While Samuel and Ressler (1986) found less PR for nasals than for other phoneme categories, we saw no significant effect of the category of phoneme which was interrupted. This difference may be related to phoneme duration. In our stimuli, there were very systematic differences in the duration of the missing phonemes, with obstruents being the shortest and vowels the longest and sonorants in between. By contrast, phoneme duration does not seem to have been considered in the study by Samuel and Ressler, but may be driving, or may have contributed to, the phoneme category effect they found. In the present study, the length of the missing segment had a large and very highly significant effect on both RT and identification as nonword.

All of the significant effects for words with interrupted segments that we report are in some way temporal. These include both the standard effects such as that of word length, where longer RTs were recorded for longer words, and more novel effects, such as the position and length of the missing segment. The regression analyses also showed that the durational mismatch between the extracted phoneme and the inserted interruptor were significant predictors of the speed and success of PR. In the present study, the mean temporal mismatch between the extracted phoneme and the inserted interruptor was 13 ms, with a maximum mismatch of 36 ms. Despite this degree of temporal fidelity between the extracted phoneme and the inserted interruptor, durational mismatch still emerged as a significant variable, which in turn illustrates the importance of interruptor duration in cuing PR. In terms of other regression variables, we noted that the phoneme categorization was not a significant predictor of performance. *Post hoc* analysis of the removed phonemes according to the classes used in this variable revealed distinctive spectral properties of each class that reflected the correspondence between the acoustic signal and the phonetic unit. Our use of an SSN, the spectral properties of which were derived from the extracted phonemes, may have contributed to the nonsignificant result of removed phoneme category. Alternatively, the regression results could suggest that the temporal match between noise interruptor and removed phoneme influences PR more than the spectral content of the interruptor.

The temporal predictors, particularly the temporal mismatch, are interesting in relation to the cARTWORD model of Grossberg and Kazerounian (2011). This model operates with "item chunks," broadly corresponding to phonemes, which are forward-fed into cognitive working memory, activating units that are referred to as "list chunks," corresponding to higher-level units including words. Once list chunks receive adequate bottom-up confirmatory input, they activate

top-down expectations which in turn may influence lower-level acoustic processing. This network activation of stored items is referred to as a "masking field." The temporal sensitivity of masking fields to the input item chunks is shown in our results, and this durational matching probably precedes interactions between bottom-up distributed features and top-down attention. In other words, our results support the cARTWORD model's assertion that the masking field is "on-centre and off-surround," that is, it is maximally sensitive to input durations that are close to that which is expected, and minimally sensitive to all others. Due to this, slight temporal distortion between expected and stimulus phonemes may inhibit item chunk activation, an effect that is seemingly evident in both the RT and word identification results of the present study.

An advantage of the cARTWORD model in relation to PR is that its feedback loop allows it to account for PR that happens based on acoustic input that arrives after the interrupted segment. The feedback loop may affect the acoustic processing of the interrupted segment, leading to the illusion that the relevant phoneme was actually there. This is in contrast to another prominent model of auditory word recognition, Shortlist B (Norris and McQueen, 2008), which does not allow feedback. This means that Shortlist B can only account for PR that happens based on input that occurs before an interrupted segment, because this prior input may contribute sufficiently to the prior probabilities for that interrupted segment to be perceived. The observation that faster recognition and fewer nonword identifications occur with stimuli where the interruptor is located later in the word is in accordance with this. However, if a segment is only recognizable as that segment based on input following it, the absence of feedback in Shortlist B means that perception of the missing segment, i.e., PR, should in principle be impossible in such cases. Our results show that PR occurs in the PRE condition, where the word is not yet uniquely recognizable, and this supports the possibility of top-down feedback to the perceptual level as posited by the cARTWORD model, but rejected in Shortlist B. However, as previously mentioned, the lexical decision results from our study do not allow us to pinpoint where PR occurs, in terms of perception vs decision. In addition, the apparent feedback channel could perhaps be explained by Shortlist B as the result of a decision mechanism that is relevant with this kind of task, something it may be possible to further investigate in the context of an event-related potential study.

While in the present study we took steps to reduce the impact of co-articulation by selectively removing sections of neighboring vowels and continuants to the interrupted phoneme, co-articulatory cues may still have been partially present in the material. However, any deleterious effect of co-articulation may in turn have been balanced by unintended masking of adjacent phonemes by the SSN interruptor, as it may have masked co-articulatory cues. Temporal masking, where a noise masker adversely affects the perception of phonemes contiguous with those that were extracted and replaced, may be an influence in our and similar PR studies. Forward masking recovery functions with speech targets, that is, the time between masker offset and (near-threshold)

target perception, have not been systematically investigated. There are indications, however, that the time course of recovery functions of broadband maskers, like the SSN interruptor used in our experiment, is likely to be brief. For instance, the threshold of a 4 kHz pure tone target of 20–100 ms, preceded by a 40 dB sound pressure level (SPL) broadband masker with a signal-masker offset of 20 ms was found to be audible at levels below that of the masker (Oxenham, 2001). Also, mean discrimination thresholds when SSN target and masker durations were 30 and 400 ms, respectively, have been shown to fall to between 10 and 15 dB below the masker presentation level of 65 dB SPL, at 16 ms (Grose et al., 2016). These reported recovery times are relatively brief in comparison to the extracted phoneme durations that we employed, indicating that it is unlikely that masking had a considerable bearing on our results. Additionally, the presence of forward masking would not necessarily invalidate the results of this study. The possible net effect of excessive masking may have meant that word position conditions would be shuffled to the right. Even if forward masking was a confound, the absence of a condition effect is still valid, since in the PRE condition, the uniqueness point of the word would be partially or wholly masked, and in the UP condition, the same would occur for the phoneme immediately after the UP (if the UP was not word-final). While this shuffling of effect to the right cannot be excluded, the amplitude gating and normalization of the interruptors were intended to minimize the effect of temporal masking. Moreover, the perceptual deficit whereby listeners cannot identify which phoneme was replaced with noise (for example, Warren and Sherman, 1974) suggests that the contribution of masking to phonemic restoration may be of little consequence in similar experimental contexts.

In summary, our two analytical perspectives inform our understanding of PR and of word recognition. The lexical decision data are informative about the cost that PR incurs on processing time and accuracy, and the degree of PR is informative about word recognition as measured by lexical decisions. Furthermore, we observed that the degree of PR is determined by several temporal predictors related to the interruptor, including the effect of fine-grained temporal mismatch between the interruptor and the removed segment. Along with the effect of UP-position in the word, this confirms the importance and nuance of timing in lexical competition, and more generally in auditory word recognition.

ACKNOWLEDGMENTS

[1]See supplementary material at https://doi.org/10.1121/1.5017603 for examples of one word ("brække," break) in three conditions.

[2]To keep the results as concise as possible, we do not include full summaries of these general mixed-effects analyses, but only reference the effect of condition, as this is of primary interest. The analyses were conducted in the same way as those reported in Tables II and III below and included the following significant variables: RT on previous trial, word length, mean letter bigram frequency, word frequency, neighbourhood density, condition, and UP position in ms.

Altmann, G. T. M., and Carter, D. M. (1989). "Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?," Comput. Speech Lang. 3, 265–275.

Balling, L. W., and Baayen, R. H. (2008). "Morphological effects in auditory word recognition: Evidence from Danish," Lang. Cogn. Process. 23, 1159–1190.

Balling, L. W., and Baayen, R. H. (2012). "Probability and surprise in auditory comprehension of morphologically complex words," Cognition 125, 80–106.

Bashford, J. A., Riener, K. R., and Warren, R. M. (1992). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. 51(3), 211–217.

Bashford, J. A., and Warren, R. M. (1987). "Multiple phonemic restorations follow the rules for auditory induction," Percept. Psychophys. 42, 114–121.

Başkent, D., Eiler, C. L., and Edwards, B. (2010). "Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss," Hear. Res. 260, 54–62.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. 67, 1–48.

Boersma, P., and Weenink, D. (2012). Praat: Doing phonetics by computer [computer program], http://www.praat.org.

Cutler, A. (1976). "Phoneme-monitoring reaction time as a function of preceding intonation contour," Atten. Percept. Psychophys. 20, 55–60.

Det Danske Sprog- og Litteraturselskab (n.d.). KorpusDK. [The Danish Language and Literary Societies KorpusDK], available at http://korpus.dsl.dk/resources.html.

Forster, K. I., and Forster, J. C. (2003). "DMDX: A Windows display program with millisecond accuracy," Behavior Res. Methods, Instrum. Comput. 35, 116–124.

Grose, J. H., Menezes, D. C., Porter, H. L., and Griz, S. (2016). "Masking period patterns and forward masking for speech-shaped noise: Age-related effects," Ear. Hear. 37, 48–54.

Grossberg, S., and Kazerounian, S. (2011). "Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise," J. Acoust. Soc. Am. 130, 440–460.

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2016). lmerTest: Tests in linear mixed effects models, available at https://cran.r-project.org/package=lmerTest.

Marslen-Wilson, W. D. (1984). "Function and process in spoken word recognition," in Attention and Performance X: Control of Language Processes, edited by H. Bouma and D. G. Bouwhuis (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 125–150.

Marslen-Wilson, W. D. (1990). "Activation, competition and frequency in lexical access," in Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives, edited by G. Altmann (MIT Press, Cambridge, MA), pp. 148–172.

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

Oxenham, A. J. (2001). "Forward masking: Adaptation or integration?," J. Acoust. Soc. Am. 109, 732–741.

R Core Team. (2016). R: A language and environment for statistical computing, Vienna, Austria, available at http://www.r-project.org/.

Samuel, A. G. (1981). "Phonemic restoration: Insights from a new methodology," J. Exp. Psychol. Gen. 110, 474–494.

Samuel, A. G. (1987). "Lexical uniqueness effects on phonemic restoration," J. Mem. Lang. 26, 36–56.

Samuel, A. G., and Ressler, W. H. (1986). "Attention within auditory word perception: Insights from the phonemic restoration illusion," J. Exp. Psychol. Hum. Percept. Perform. 12, 70–79.

Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," Science 167, 392–393.

Warren, R. M., and Obusek, C. J. (1971). "Speech perception and phonemic restorations," Atten. Percept. Psychophys. 9, 358–362.

Warren, R. M., and Sherman, G. L. (1974). "Phonemic restorations based on subsequent context," Atten. Percept. Psychophys. 16, 150–156.