Phonetic detail is used to predict a word's morphological composition

Meghan Clayards[ab], M. Gareth Gaskell[c] and Sarah Hawkins[d*]

* Corresponding author: sh110@cam.ac.uk
a. Department of Linguistics, McGill University, Canada
b. School of Communication Sciences and Disorders, McGill University, Canada

c. Department of Psychology, University of York, UK

d. Centre for Music and Science, Faculty of Music, University of Cambridge, UK

Abstract

An eye-tracking experiment tested the hypothesis that listeners use within-word fine phonetic detail that systematically reflects morphological structure, when the phonemes are identical (*dis* in *discolour* (true prefix) vs. *discover* (pseudo prefix)) and when they differ (*re-cover* vs. *recover*). Spoken sentence pairs, identical up to at least the critical word (e.g. *I'd be surprised if the boys discolour/discover it*), were cross-spliced at the prefix-stem boundary to produce stimuli in which the critical syllable's acoustics either matched or mismatched the sentence continuation. On each trial listeners heard one sentence, and selected one of two photographs depicting the pair. Matched and mismatched stimuli were heard in separate sessions, at least a week apart. Matched stimuli led to more looks to the target photograph overall and time-course analysis suggested this was true at the earliest moments. We also observed stronger effects for earlier trials and the effects tended to weaken over the course of the experiment. These results suggest that normal speech perception involves continuously monitoring phonetic detail, and, when it is systematically associated with meaning, using it to facilitate rapid understanding.

Keywords

perception, spoken language understanding, morphological structure, acoustics

1 Introduction

There is increasing evidence that morphological differences in phonemically-identical words or parts of words are reflected in their phonetic realisation (e.g. Hay, 2003, ben Hedia & Plag, 2017; Plag and ben Hedia, 2018; Rose, 2017; Smith, Baker and Hawkins, 2012; Sproat & Fujimura, 1993; Seyfarth et al. 2018; see also Strycharczuk, 2019 for a review). As such evidence accumulates, questions remain unanswered as to whether the observed patterns in production influence listeners' perceptual behavior. This paper addresses perception of a morphological contrast involving prefixes (e.g. *dis*- in *discolour*, *re*- in *re-joined*) and matched non-prefix word-initial syllables that are either phonemically identical to the prefix (e.g. *dis*- in *discover*) or contrast in the vowel phoneme (e.g. *re*- in *rejoiced*). Both types of morphological contrast manifest as two different acoustic-phonetic patterns affecting the entire word-initial syllable and beyond, in prosodically-controlled fluent lab speech (Smith, Baker, & Hawkins, 2012) and in conversational speech in several regional varieties (Hay, Hawkins, Stuart-Smith, Smith and Fromont, in prep). The question asked is whether listeners use the resultant acoustic-phonetic distinction in real time to facilitate word recognition in connected speech. If they do, this would indicate that they use the internal acoustic structure of the first syllable of a prefixed or pseudo-prefixed word to identify its stem, and by implication the lexical item's morphological structure. When the phonemes do not differ, this would amount to identification of bound morphemes from very fine differences in the syllable's internal acoustic structure, and their use to predict lexical identity.

1.1 Phonetics of morphologically complex words

By far the largest body of relevant experimental work examines temporal relationships between acoustic segments or articulatory gestures in suffixed words. Compared with phonemically-matched or similar monomorphemic words, suffixed words have consistently different phonetic characteristics in their suffix, in their stem and at the morpheme boundary. Although a number of studies have demonstrated such properties, several have confounded number of morphemes with number of syllables and/or foot length (e.g. Kemps, Wurm, Ernestus, Schreuder, & Baayen, 2005; Lehiste, 1972). The reported differences between bimorphemic and monomorphemic words can nonetheless be considered robust in that several studies that circumvented these problems showed the same type of results (Sugahara & Turk, 2009; Cho, 2001; Seyfarth et al., 2018). These temporal relationships have also been shown to affect the articulatory gestures of English /l/ at morpheme boundaries resulting in gradient spectral differences in /l/-darkness (Sproat & Fujimura, 1993; Lee-Kim, Davidson, & Hwang, 2013; Strycharczuk & Scobbie, 2016, 2017; Turton, 2017; Mackenzie et al., 2018)

Prefixes have received less attention than suffixes, but Oh and Redford (2012) show durational differences in nasal-nasal sequences dependent on whether the sequence includes a morphological boundary as in *un-named* or a word boundary as in *fun name*. Smith, Baker and Hawkins (2012) and (Hay et al., in prep) document complex, systematic acoustic effects of prefix status for the initial syllables of word pairs such as *discolour* vs. *discover* and *mistypes* vs. *mistakes,* in which the first member of each pair begins with a true (productive) morpheme whereas the second member does not, despite having the same phoneme sequence. Consequently, the initial syllables of *discover* and *mistakes* are termed pseudo prefixes. Prefixes of this type are particularly interesting for models of speech perception because, if their distinctive phonetic detail[1] is processed differently from that of pseudo prefixes, this would suggest that their properties are directly associated with their status as bound morphemes. In contrast, perceptual studies examining the same type of issue for suffixes typically show listeners' sensitivity to differences in the stem, rather than in the suffix itself, as discussed below.

The distinction between true and pseudo prefixes is not completely clearcut, due to a number of interacting influences of quite different types, also discussed below. However, it is possible to control for such influences. Smith, Baker and Hawkins (2012) used criteria that provided tight control over both the type of word and its phonetic and semantic context. They followed Wurm's (1997) strict semantic criteria to select stimulus words: in words such as *discolour* and *mistypes,* the initial syllables, *dis*- and *mis*-, are true prefixes because *colour* and *types* mean roughly the opposite when *dis*- or *mis*- are added. In contrast, the words *discover* and *mistakes* are monomorphemic because *cover* and *takes* do not mean the opposite of *discover* and *mistakes.* Furthermore, each of the prefixed words chosen had a lower frequency than the frequency of its stem, thereby conforming with Hay's (2003) criterion (developed for suffixes) for a relatively strong and unambiguous morpheme boundary.

Smith, Baker and Hawkins (2012) elicited such true and pseudo prefix pairs from speakers of Standard Southern British English (SSBE) in fast, casually-spoken scripted dialogues in which the prosodic and segmental structure of the critical utterances were tightly controlled. Acoustic-phonetic measures supported earlier impressionistic claims (e.g. Hawkins, 2010; Ogden et al., 2000; Whitley, cited by Simpson, 2005) that the first syllables of the true-prefixed words convey a heavier beat in context due to small differences in the acoustic properties of their component segments. As illustrated, for example, in Smith, Baker and Hawkins' (2012) Figure 1, which shows spectrograms and phonological trees for *mistimes* (true prefix) and *mistakes* (pseudo prefix), one very reliable acoustic difference is the duration of aperiodicity for [s] relative to the duration of periodicity of [ɪ]: the [s] takes up a much larger proportion of the syllable in pseudo prefixes. Another is that the second formant frequency of [ɪ] is higher and closer to F3 in the true prefix, suggesting less centralisation. A third is that when a voiceless stop is in the onset of the second syllable of the word, its voice onset time (VOT) is long following the true prefix, but short following the pseudo prefix. In sum, the morphological status is reflected in several phonetic characteristics that affect all segments in the first part of the word. These differences create systematic differences in the overall pattern of relationships between the acoustic segments within the first syllable, termed here its internal acoustic structure, as well as at the syllable juncture and thereafter. The internal structure of the first syllable is such that, though both true and pseudo prefix syllables are metrically weak, true prefixes are more phonetically prominent—indeed, many dictionaries accord some true prefixes a secondary stress, whereas pseudo prefixes are never

[1] See Appendix A for what is meant by 'phonetic detail'.

accorded one. Thus, in any given speech register, the true prefix conveys a heavier rhythmic beat than the pseudo prefix.

While these differences are reasonably distinct for any given speech register, there can be 'gradient' effects in cases where the prefix status of a word is in flux (which occurs for a variety of reasons (cf. Hay et al., 2005) and between different speech registers and modes of data collection (Smith, 2012; Hay, 2018; Zuraw & Peperkamp, 2015). Furthermore, while the majority of prefixes behave as described above, a few do not (Plag, 2014). Words whose prefix status was ambiguous were excluded from the present study.

In summary, systematic phonetic markers of the internal composition of words are embedded in the speech signal and so are potentially available to the listener. These kinds of cues differ from other acoustic-phonetic effects such as those due to assimilation of place of articulation across word boundaries, in that they occur word-internally in a range of speech registers, including careful, clear speech, and so are integral to the identity of words. The goal of this paper is to determine whether listeners are in fact sensitive to such subtle distinctive patterns and use them to build expectations about morphemic structure and hence word identity as they interpret utterances in real time.

1.2 Perception of acoustic cues to word structure

Lexical identification experiments using gating tasks and cross-modal priming (Davis et al., 2002), and eye-tracking with the visual world paradigm (Salverda et al., 2003), show that listeners' early perceptual responses are sensitive to acoustic-phonetic detail that signals word boundaries. They contrast syllables that are either followed by a word boundary or are part of a longer word as in *cap* and *captain* or *ham* and *hamster.* When hearing syllables such as *cap,* listeners were more biased towards a monosyllabic interpretation when *cap* had been spoken as a monosyllabic word rather than as part of a polysyllabic word. While these studies confirm the importance of phonetic detail to lexical identification, they do not examine influences due to morphological structure within words, and, as summarized at the end of this section, they confound a number of linguistic variables which designs using the true vs pseudo prefix distinction can control.

Relatively little work has examined perception of morphological structure (i.e. word-internal junctures). Blazej & Cohen-Goldberg (2014) tested whether the effect of number of syllables found for the *ham* versus *hamster* studies extended to words which are also multi-morphemic by virtue of containing suffixes e.g. *clue* and *clueless.* They found the same pattern as studies that examine shorter words embedded in longer monomorphemic words: listeners anticipated the longer word after hearing a shorter first syllable. A pair of similar studies by Kemps and colleagues (Kemps et al., 2005a; 2005b) using lexical decision and a morpheme decision task (singular or plural) found compatible results for embedded stems in plurals in Dutch and for comparatives (e.g. *stronger*) and agent nouns (e.g. *worker*) in Dutch and English. However because the stimuli in these studies compared monosyllabic, mono-morphemic words with polysyllabic, poly-morphemic words, it is impossible to tell whether their listeners were simply anticipating a longer word (an ability previously demonstrated for mono-morphemic words), or whether they were anticipating the poly-morphemic structure of the longer word, or both.

As described in the next section, the present study of true and pseudo prefix perception circumvents most of these problems. Furthermore, there is intrinsic theoretical interest in distinguishing boundaries within words from those between words, and phonetic support for making that distinction from work on prefixes *un-* and *in-* (Oh & Redford, 2012).

1.3 The present study

The present paper uses eye-tracking in a visual world paradigm to test our hypothesis that listeners can use the internal acoustic structure of the initial syllable of a prefixed or pseudo prefixed word to predict morphological structure that itself predicts word identity. The focus on prefixes allows questions of perceptual sensitivity to the morphological status of a syllable to be assessed without the confounds of morphological complexity, polysyllabicity, and word length which characterize the studies cited in Section 1.2. We compare units that are comparable except in morphological complexity: all words are polysyllabic; all boundaries of interest are word-internal; and no first syllable has an independent lexical meaning—pseudo prefixes, comparable to *cap* in *captain,* convey no meaning independent of the rest of the word, while prefixes are not independent lexical items, so though they convey a meaning, it is only properly interpretable in the context of the meaning of the rest of the word. Furthermore, unlike the critical syllables in the *cap* and *captain* studies, our critical syllables are metrically weak (lack primary stress); weak syllables are often thought to play a subordinate role in lexical identification (e.g. Cutler & Butterfield, 1992).

Our study also differs from studies that have shown effects of phonetic (sub-phonemic or within category) detail on lexical access (e.g. McMurray, Tanenhaus, Aslin & Spivey, 2003; Dahan, Magnuson, Tanenhaus, & Hogan, 2001) in that our phonetic detail contributes primarily to rhythmic and not segmental structure. Furthermore, unlike previous studies that tested phonetic detail using minimal pairs (e.g. McMurray et al., 2003), the present study contrasts word sequences most of whose second and later syllables are not minimal pairs, so listeners in our experiment do not need to use the fine detail of unstressed *mis-* or *dis-* at all in order to distinguish the words or the sentence meanings. Thus ours is a very stringent test of the perceptual salience of phonetic detail: the cues are in weak syllables and they are followed very swiftly by much clearer disambiguating evidence. If we find evidence suggesting that these cues are used despite their not being essential to the task, then we have very strong evidence of the pervasive role of phonetic detail, and rhythmic detail in particular in spoken word recognition. Furthermore, such findings would strengthen the evidence that listeners extract clues to many levels of linguistic structure from the fine phonetic detail in the signal.

We used four types of prefixes: *mis-* and *dis-* as already discussed, and the prefixes *re-* and *ex-* as in *re-peel/repeal* and *ex-trampoliner/extravagance*. The syllables *re-* and *ex-* follow the same patterns as *dis-* and *mis-*, i.e. the true prefixes are rhythmically stronger, but in this case syllabic reduction in the pseudo prefix happens to cross a phoneme category boundary, whereas the same type of syllable reduction does not produce a category change in the *dis-mis-* set. Consequently, for *re-* and *ex-*, the vowel phonemes in the critical syllables differ: *re-peel* /riː'piːl/ but *repeal* /rɪ'piːl/ or /rə'piːl/; *ex-trampoliner* /ɛks'trampəliːnə/ but /ɪks'travəgəns/ or /əks'travəgəns/in SSBE, the regional variety used here (see Smith et al., 2012 for more explanation). Thus while all prefixes differ rhythmically from the pseudo prefixes, the syllables in *re-* and *ex-* differ segmentally as well. This distinction is represented in our design as the independent variable PhonemeChange, with *re-ex-* changing vowel phoneme, and *dis-mis-* not. As segmental differences are uncontroversially part of the lexical representation, we can predict that the segmental cues to morphological structure will be picked up in spoken word recognition. The *mis-* and *dis-* stimuli are a more stringent test of the hypothesis that non-segmental phonetic detail is important for identifying morphological structure because they share the same first four phonemic segments. However, because discrimination of prefixes has not been tested using eye-tracking before, and their relatively abstract meanings necessitate the use of relatively complex visual stimuli, the *re-* and *ex-* stimuli provide a check that our methods are sensitive enough. Then the question is whether nonphonemic morphological divergence will also be exploited. If so, how strong is it compared with the phonemic effect and does it have the same time course?

In sum, our experiment was designed to test whether listeners exploit subtle acoustic cues to morphological structure. In particular, do they identify morphological structure, even to the extent of predicting that they are hearing a morphologically-complex word before they have heard the stem, in good listening conditions when the task does not demand it?

2 Methods

2.1 Design

We used cross-spliced spoken sentences to manipulate whether the acoustic information in critical syllables (i.e. in the true or pseudo prefix) was consistent with the morphological structure of the rest of the word. We presented these stimuli to participants and asked them to choose between two pictures, one representing a situation whose description included the true prefixed word and the other a situation whose description included the matched pseudo prefixed word (see Figure 1 for illustration).

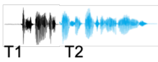
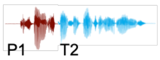
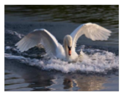
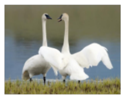
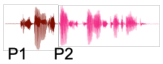
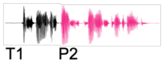
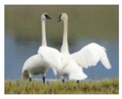
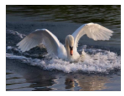
Target word	Match splice type		Target image	Competitor image
	match	mismatch		
displaces	[A swan dis]$_{T1}$ [places water when it lands]$_{T2}$ 	[A swan dis]$_{P1}$ [places water when it lands]$_{T2}$ 		
displays	[A swan dis]$_{P1}$ [plays its plumage to its mate]$_{P2}$ 	[A swan dis]$_{T1}$ [plays its plumage to its mate]$_{P2}$ 		

Figure 1. Construction of match and mismatch stimuli illustrating one pair of sentences. Subscript$_T$ indicates a portion from the original utterance that contained a true prefixed word (e.g. *displaces, discolour*). Subscript$_P$ indicates a portion from the utterance originally containing the pseudo prefixed word (e.g. *displays, discover*). Subscript numbers refer to different sentence recordings. For each sentence the target and competitor images are given.

The words in a pair of such sentences were identical up to the critical syllable. If perceptual behaviour is influenced by the acoustic information available to signal true vs. pseudo prefixes while it is being heard, then we would expect listeners to be delayed in correctly identifying the target word (and therefore the target image) when that acoustic information does not match the morphological structure of the target word/image. Our main analysis therefore compared the looks to target images (defined as the image consistent with the post-splice continuation of the spoken word and sentence) on trials in which the critical syllable contained acoustic information that either matched or mismatched the target image (factor Match). For example, in Figure 1, the image corresponding to the target for *A swan displaces water when it lands* is the image of a swan landing on water, regardless of the prefix status of the cross-spliced critical syllable. Similarly, the competitor is the image consistent with the sentence's pair, in this case the two swans (see Figure 1).

We also considered whether the effect of the critical syllable (Match) depended on a number of factors. The factors of the design are summarized in Table 1. As discussed above, we tested if Match depended on whether the prefix manipulation entailed a phoneme change as in *re-* or *ex-,* or not as in *mis-* or *dis-* (factor PhonemeChange).

Table 1: Summary of factors and their levels in the design.

Factor	Levels	Nesting
Match	match: acoustic information in critical syllable matches the morphological structure of the target word/image mismatch: acoustic information in critical syllable doesn't match the morphological structure of the target word/image	within subjects and items
Interactions with Match		
PhonemeChange	*dis-*/*mis-* *re-*/*ex-*	within subjects between items
Group	M1: match on session 1/mismatch session 2 M2: mismatch on session 1/match on session 2	between subjects within items
TrialNumber	continuous variable from 1 to 99	within subjects and items[2]
PrefixStatus	true: critical syllable spliced from a true prefixed word pseudo: critical syllable spliced from a pseudo prefixed word	within subjects and items

Creating matched and mismatched stimuli is crucial to our experimental manipulation. However, it also disrupts the natural systematic association between the acoustic information in any given critical syllable and its function as a true or pseudo prefix in the word. The fine phonetic detail of interest thus becomes uninformative within the context of the experiment. Because we were worried about the effects of this disruption on the listeners, we presented the matched and mismatched sentences in separate sessions on separate days. In our analyses we tested if our effect of Match depended on whether participants heard all the matches on Day 1 (M1) or on Day 2 (M2) (factor Group, see Counterbalancing section for more details). Furthermore, we tested whether the effect of Match changed over the time course of the experiment (factor TrialNumber—the order that each trial occurred in the experiment for each participant).

Finally, exploratory analyses also considered if the effect of Match depended on whether the acoustics of the critical syllable were from a word that had a true prefix or a pseudo prefix (PrefixStatus). For example, sentences in Figure 1 with a subscript$_{T1}$ for the critical syllable *dis* have a PrefixStatus of true and those with a subscript$_{P1}$ for the critical syllable have a PrefixStatus of pseudo (an example is also illustrated in Figure 3).

2.2 Participants

Participants were 34 native English speakers at the University of York (mean age 21 years, range 18-32, 24 women), with normal or corrected-to-normal vision and no history of speech or hearing problems. Each participated in two sessions on different days at least one week apart. Each session took approximately 45-50

[2] By-item random slopes were not fit due to sparsity of item data for any given TrialNum.

minutes. Two additional participants were discarded due to errors in data collection.

2.3 Auditory Stimuli

All stimuli can be found in the Open Science Framework repository (https://osf.io/dsyxu/ DOI 10.17605/OSF.IO/DSYXU). Sentences were constructed for 32 pairs of target words differing in true vs. pseudo prefix status of their first syllable (e.g. *displaces/displays*). As explained in the Introduction, these comprised two types (factor PhonemeChange). In the *dis-mis-* type (e.g., *dis-*: *discolour/discover* [N = 7]; *mis-*: *mistypes/mistakes* [N = 4]), at least the first four phonemes of each true-pseudo pair were identical. The *re-ex-* type followed the same principle of having identical phonemes into at least the beginning of the second syllable, except that for these words the first syllable's vowel phoneme differed with prefix status (e.g., *re-*: [riː] *re-strings/* [rə] *restricts* [N = 16]; *ex-*: [ɛks] *ex-trampoliner/* [əks] *extravagance* [N = 5]). Primary lexical stress was on the second syllable of each critical word. Each word was placed in a sentence which was identical to that of its pair before the target word, and in some cases after it, and could be illustrated by a picture (e.g. *It was difficult because Sam distrusted/distracted him*).

Sentence pairs that differed after the target word had the same intonational and foot structure, and hence number and stress-pattern of syllables, though not necessarily the same word boundaries within a foot. In one *dis-* and one *ex-* case identical foot structure was achieved by adding an extra syllable because one target word had one less syllable than its pair: *A swan displaces water when it lands* and *A swan displays its plumage to its mate*; *It's a perfect example of ex-trampoliners' sense of balance* and *It's a perfect example of extravagance in public spending.* These additions were made immediately after the target word so as to match the foot structure created by the longer target word, the particular words being chosen to be similar to the longer member of the pair in connected speech e.g. *displays its* vs. *displaces*. All syllable counts were as standardly pronounced in SSBE, as well as in the particular stimuli (e.g. *discourteous* had three syllables, not four: /ˌdɪs'kɜtjəs/. Appendix B shows the complete list.

Because our focus was on matching the phonetic structure of the stems while using word pairs that could occur in sentences that were identical before the critical word and had identical prosodic structure throughout, we could not match target words on frequency. However, this should not bias the results, since over and above the fact that word frequency is not a primary determinant of the morphological distinction itself (Smith et al., 2012, Hay et al., in prep), the key comparisons were to be between cross-spliced stimuli in which the lexical item was the same, the only difference being in the acoustic signal in its first syllable, as described below.

These 32 pairs of sentences were recorded in 6 random orders by a male SSBE speaker. Quality was controlled as follows. To minimize reading effects, the speaker had familiarized himself with the sentences and pictures for some days before the recording, and was encouraged to look at the picture rather than the text while recording. Contrastive stress on the critical words was avoided in that only one picture was displayed at a time, and at least two other sentences separated recording of the two members of each sentence pair. Trained phoneticians checked the stimuli both at the time of recording and afterwards. Errors (utterances that contained disfluencies or that sounded unnatural, unclear, or inappropriate for the intended meaning), including borderline cases, were re-recorded. Two recordings of each sentence were chosen for cross-splicing to create a 'match' and a 'mismatch' version of each member of the pair. Sentences were initially chosen for naturalness and the best impressionistic match of f0, rhythm and loudness. Following this, the chosen pairs of stimuli were inspected acoustically to ensure that they not only sounded acceptable in their original contexts, but that, relative to each other, the internal acoustic structure of each critical syllable conformed to expectations derived from Smith, Baker and Hawkins (2012), primarily using durational criteria. This was necessary in order that any observed behavioural differences could be interpreted in terms of differences in internal acoustic structure. See Section 2.7 for acoustic analyses of the stimuli.

Stimulus construction involved two types of cross-splicing, according to whether the resultant token was a 'match' or 'mismatch' stimulus. Sentences were cut at the end of the critical (target) syllable (just before the burst of the next stop if there was one, as in *mistimes*) and cross-spliced either with the end of an identical sentence for 'match' stimuli or with the end of its pair for 'mismatch' stimuli. Thus each stimulus was constructed from two separate recorded tokens, spliced just after the critical prefix/pseudo-prefix syllable: separate recordings of the same sentence for 'match' stimuli and recordings of different sentences for 'mismatch' stimuli.

Thus, as Figure *1* shows, four versions were created from each sentence pair corresponding to the four combinations of true and pseudo prefixes and continuations: the matches true-true and pseudo-pseudo, and the mismatches true-pseudo and pseudo-true, for a total of 128 test stimuli (32 pairs x 4 conditions). These procedures meant that no perceptually significant acoustic information about the second syllable of the word was present in the first syllable: second syllables all had identical or very similar vowel qualities (see Appendix B), and for *mis-* and *dis-* syllables followed by a stop, Baker (2008) showed that listeners could not predict the following vowel unless they heard the burst and following VOT.

An additional 67 filler sentence pairs were constructed. Of these, 30 pairs had been recorded by the same speaker and used in a previous experiment; in the present study they comprised an independent experiment run at the same time. These 30 consisted of pairs of sentences identical except for one word, differing only in whether it contained an /r/ or /l/ (e.g. *rams* vs. *lambs*), hereafter r-l sentences. Matches and mismatches were created as described above, except that critical words were spliced into the sentences, rather than abutting the first part of one sentence with the second part of another (see Heinrich, Flory, & Hawkins, 2010 for details, and the list of words and sentences). Of the other 37 filler pairs (listed in Appendix C), all but four were designed to mimic the prefix ones in that they contained a word with a true or pseudo morpheme either before or after the target/disambiguating words. For example *You purify water/whisky by distilling it*. The remaining four fillers followed the same semantic and prosodic principles as the others; three of them contained a word beginning *re-* paired in the other sentence with a non-*re-* word. These 37 fillers were recorded twice; the most natural of each was chosen, and not spliced. Six additional filler trials were created in the same way and used in practice blocks.

2.4 Visual Stimuli

A photograph was chosen to represent each sentence from images publicly available on the web, and photographs we took ourselves. Care was taken to ensure that pairs of images were similar in complexity and colourfulness, as judged by six people, the four authors and two research assistants. See the examples in Figure 1. The largest dimension of each image was 600 pixels. As noted in the Results (Section 3), baseline measures of looking preference taken at the moment the auditory stimulus was presented and at the onset of the critical word revealed no systematic preferences for the pictures depicting true versus pseudo prefixes.

2.5 Procedure

Participants were seated in front of a desktop-mounted remote Eyelink 2000 (SR Research) to monitor eye-movements while they performed the task. Auditory stimuli were presented over headphones at a comfortable listening level. Visual stimuli were displayed on a 16”x12” monitor. Each session began with set up and calibration of the eye-tracker followed by two practice blocks of trials, whose structure was identical to the rest of the experiment .

Testing took place on two days (Section 2.6). Each day, participants heard one trial for each of the 99 pairs (32 prefix pairs, 30 r-l pairs and 37 filler pairs). Stimuli were grouped into 33 presentation sets, each containing three pairs of sentences, generally one prefix pair, one r-l pair and one filler pair. Participants were first presented with a block of six familiarization trials in which each of the six sentences from a presentation set appeared, along with its accompanying picture, one at a time, in random order. The sentence was not spoken, but instead printed at the top of the screen. The picture was centered in the middle of the screen, as shown in the left portion of Figure 2. Participants were instructed to read the sentence silently and familiarize themselves with the picture, then click on the picture to continue (the trial did not end before at least 2.5 seconds of viewing).

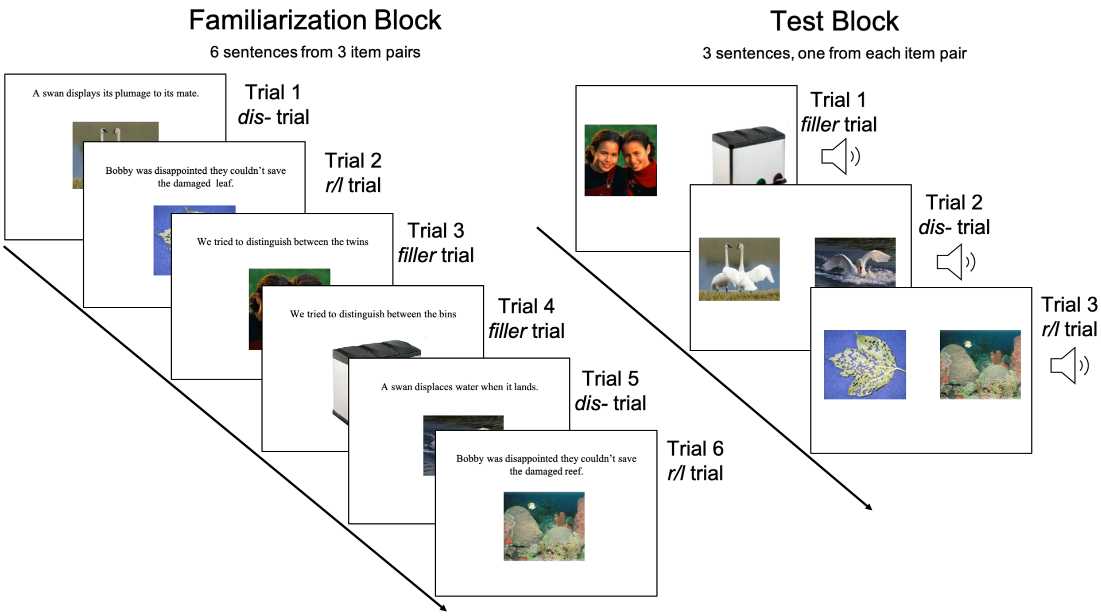


Figure 2: Example of a familiarization block and corresponding test block. Left: Familiarization block, 6 written sentences from each of 3 item pairs. Right: Test block, 3 spoken sentences, one from each of the same 3 item pairs. Colour online.

After the familiarization block came a corresponding block of three test trials (shown in the right half of Figure 2), one from each pair in the 6-item presentation set shown in the left half of Figure 2. Figure 3 shows the structure of an example test trial. Each test trial began with a drift correction for the eye-tracker. One pair of pictures was then presented, one centered in the left half of the screen, the other in the right half. No text was displayed. The side on which the true prefix image was displayed was randomized across trials. After two seconds of preview time, one of the sentences was played. Participants were instructed to click on the picture that matched the sentence as quickly and accurately as possible. Once the participant responded, the pictures stayed onscreen for an additional 0.5 seconds; they were then replaced briefly by a blank screen, after which the next trial began.

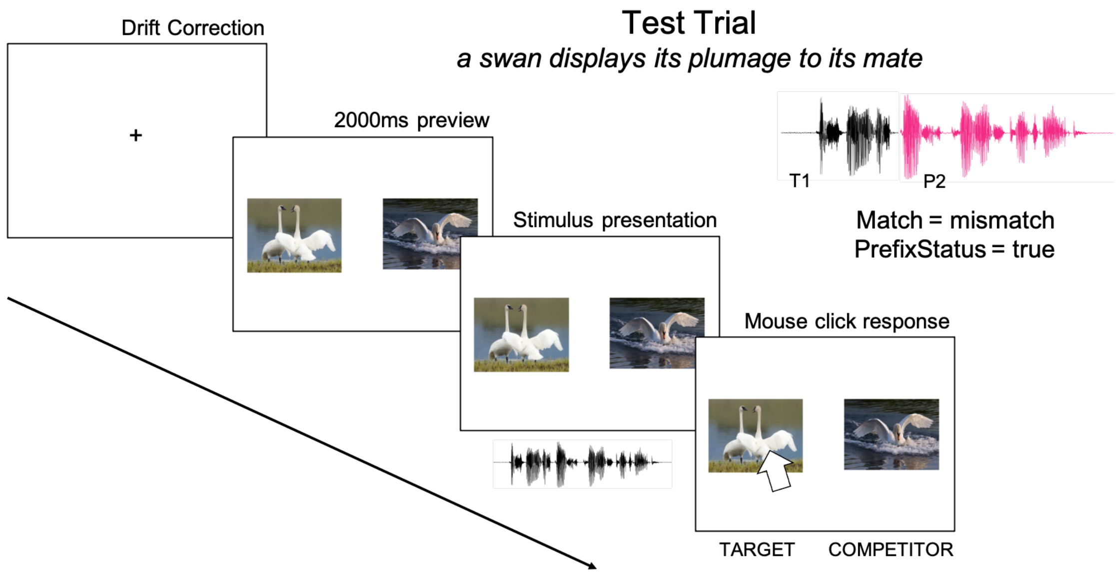


Figure 3: Structure of a single test trial. As this example trial shows, the PrefixStatus of the critical syllable can mismatch the Target sentence. In this example, the acoustics of the critical syllable are from a true prefixed word (*displaces*) but the target word and sentence continuation are the corresponding pseudo prefixed word (*displays (its)*). Colour online.

2.6 Counterbalancing

As discussed above, we were concerned that if the matched and mismatched trials were all presented in the same session, the fine phonetic detail of interest would thus be uninformative within the context of the experiment, and listeners could be expected to quickly learn to ignore it as they have for similarly subtle phonetic information (e.g. Hawkins & Nguyen, 2001, Experiments 2, 3a and 3b). The chosen blocked and counterbalanced design was intended to allow us to assess two things: how the critical phonetic information is used in real time when heard with its normal systematic distribution reflecting morphological status; and to what extent atypical distributions influence recognition behaviour in the shorter term. We thus used a blocked design in which all matched stimuli (r-l and prefix) were presented on one day, and all mismatched stimuli were presented on another, the two sessions being separated by at least a week. The order of match and mismatch was counterbalanced (factor Group): 18 participants heard all matches on Day 1 (M1) and all mismatches on Day 2, while the other 16 heard all mismatches on Day 1, and only matches on Day 2 (M2). Two additional participants were recruited for group M2 but it was later discovered that they had to be excluded due to experimenter error in data collection.

Because the participants would be seeing each pair of images (and hearing one of them described) on Day 1 and again on Day 2, it was important that they not be able to predict which image would be described on the second day. For this reason, a second counterbalancing factor was added. The stimulus pairs were divided into two sets such that for each participant, on one half of trials the same image was described on Days 1 and 2, and on the other half of trials, the opposite image was described on Days 1 and 2. Thus, it was impossible for the participants to predict which picture would be described on any trial. Practice blocks also illustrated this pattern. In all cases (except the 37 unspliced fillers) the stimulus each participant heard was different on the two days, either matching or mismatching depending on the day and the group.

In summary, for critical test trials, each participant heard one of the four stimuli describing each pair on each day; matches and mismatches were never mixed in one session (order was counterbalanced between Groups M1 and M2); and the presentation of spoken sentences within Day 1 and Day 2 was such that predicting which of the two images would be described in the experimental trials should have been at chance. There were four groups created by counterbalancing these conditions. Four additional conditions were created with a different random grouping of stimulus pairs. Trials from the *dis-mis-* and *re-ex-* sets of sentences were evenly distributed throughout all these conditions.

2.7 Acoustics of critical syllables

We extensively analysed the acoustic properties of our stimuli to ensure that they conformed to expectations from the previous phonetic literature and also to ensure that they did not contain unwanted biases. A full report of these analyses is available in the supplemental materials (https://osf.io/dsyxu/DOI 10.17605/OSF.IO/DSYXU). We found that the critical syllables varied according to prefix status as we would expect. We also found no evidence for systematic acoustic differences before those syllables that might bias the interpretation of the critical syllables. Here we briefly report on the duration of the critical syllables themselves.

Figure 4 shows the distribution of syllable durations for each of the prefixes. True prefixes were on average 54 ms longer than pseudo prefixes, with *dis-mis-* syllables 55 ms longer on average than *re-ex-* syllables. However, as Figure 4 shows, these overall observations mask differences within the syllable types that are important for interpreting the eye tracking results. While *mis-* and *dis-* mean durations and standard deviations are similar enough that the two subtypes can be regarded as a roughly homogeneous group, this is not the case for the *re-ex-* set: there are large differences between *re-* and *ex-* syllable durations, with absolute and relative values for *ex-* patterning more like those for *dis-* and *mis-* than those for *re-*. These large differences within the *re-ex-* set are due to the phonetic makeup of the syllables and the consequent degree to which each can be reduced. Whereas *re-* can be severely reduced,inherent durational constraints on English /k/ and particularly /s/ mean that *ex-* is much less open to reduction. Furthermore, *ex-* syllables are the longest, and *re-* syllables the shortest of all four syllable types. Because of this great heterogeneity, it was decided that *re-* and *ex-* should not be treated as a single group; and because there were so few *ex-* tokens it was necessary to exclude the *ex-* stimuli from the analysis.

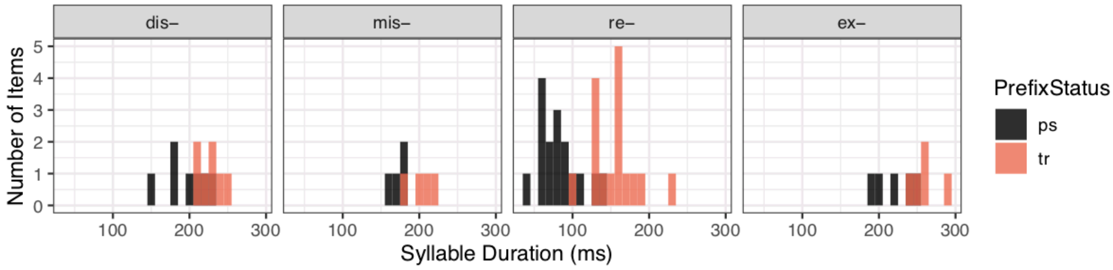


Figure 4: Histograms of critical syllable durations for experimental stimuli. Colour online.

2.8 Eye-tracking data: analysis principles

Responses to filler stimuli were not analyzed. Trials in which the participant clicked on the incorrect image (did not match the continuation of the sentence) were removed (a total of 151 trials or 6% of the data). Eye-movements from all remaining trials were then time-aligned to the start of the critical word.

Because our visual stimuli and sentences were relatively complex, we first established that listeners had no overall preference for the images corresponding to one set of words or the other (true or pseudo prefixed words). The mean proportion of looks to the two types of image was almost identical at the onset of the critical word (plus the expected 200 ms oculomotor delay; Matin, Shao, & Boff, 1993): pseudo = 0.46, true = 0.47; paired sample t-tests p = 0.63 by subjects, p = 0.65 by items. There was likewise no difference at sentence onset (p = 0.98 by subjects, p = 0.94 by items).

For all analyses, proportion of fixations to the target image (as defined by the continuation of the sentence) were computed over a specific time window and these proportions were transformed to log odds for analysis with linear regression. Linear regression with log-odds-transformed proportional data is comparable to logistic regression on data in which each observation is either target or not, but allows for aggregation of data over a given time window.

All statistical analyses were done using mixed model linear regression using the *lmer*() function from the *lme4* package (Bates, Maechler & Bolker, 2015) in R (R Development Core Team). Significance was assessed using the Satterthwaite approximation of degrees of freedom as implemented in the *lmertest* package (Kuznetsova, Brockhoff & Christensen, 2017) in R. All factors (binary categorical variables) were centered by using contrast coding (0.5 vs. -0.5) and continuous variables were centered and scaled. Centering the variables avoids any co-linearity between the effects and their interactions. Factors were Match (match = 0.5, mismatch = -0.5), PhonemeChange (*dis*-*mis*- = 0.5, *re*-=-0.5), Group (M1 = 0.5, M2 = -0.5), and TrialNumber (continuous, scaled).

Random intercepts for subjects and items (each item was a sentence pair) were included in all models. Random slopes were included wherever the design and the data allowed (see below for details).

3 Results

The goal of the first analysis was to determine whether listeners were more likely to look at the target image when the acoustics of the critical syllable matched the morphemic structure of the target word, than when the acoustics mismatched.

3.1 200-800 ms window

3.1.1 Main analysis

Our first analysis aggregated looks within a window from 200 to 800 ms after the onset of the critical word. This window was chosen because it is generally assumed that it takes roughly 200 ms to plan and launch an eye movement (Matin, Shao, & Boff, 1993). Thus the window begins when we would expect to see looks influenced by the onset of the critical target word. The window continues until looks to the target start to asymptote (at which point we expect any effects to disappear). Thus any delay in identifying the target should be seen in this window.

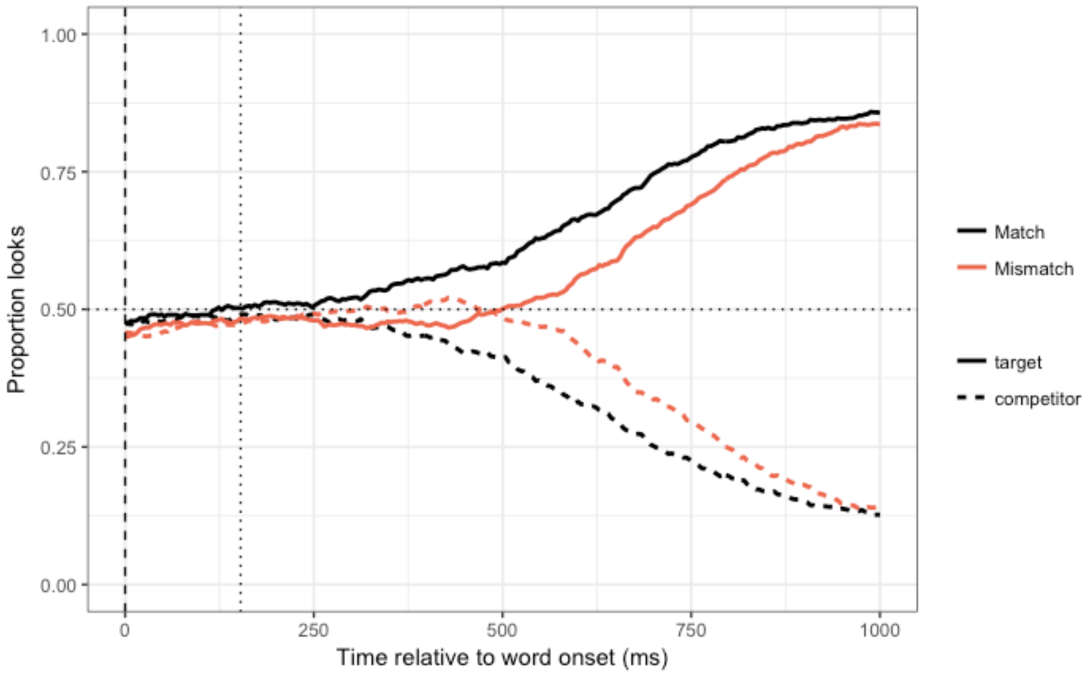


Figure 5. Fixation proportions to the target and competitor image aligned to the word onset for the matching (black lines) and mismatching (red lines) conditions. Solid curves: looks to target image. Dotted curves: looks to competitor image. The dashed vertical line at time = 0 is the alignment point, the beginning of the word. The dotted vertical line indicates average splice point across all stimuli. ‘Target’ is defined as the part of the sentence following the splice point.

Figure 5 shows target and competitor fixations over time for trials in which the critical syllable either matched or mismatched the continuation of the sentence (i.e. the target), aligned at the word onset. Over the course of the trial, participants looked more at the target and less at the competitor and this difference is bigger for matching trials as predicted. We tested the difference in looks to the target by examining the effect of Match in a model that also included PhonemeChange and its interaction with Match to test whether the effect of Match was different for the *dis-mis-* and the *re-* stimuli. This model included random slopes for Match by subject and item as well as random slopes for PhonemeChange and its interaction with Match by subject (i.e. the maximal model). The output of this model is summarized in **Error! Reference source not found.** Table *2*. There was a higher proportion of looks to matched targets than to mismatched targets. Neither PhonemeChange nor the interaction of Match with PhonemeChange significantly affected responses however. See supplemental materials for item and participant variability in effect of Match.

Table 2: Model summary for 200-800ms window.

	Estimate	Std. Error	t	p
Match	0.57	0.21	2.73	0.01
PhonemeChange	0.15	0.30	0.48	0.63
Match:PhonemeChange	-0.01	0.41	0.03	0.97

3.1.2 TrialNumber and Group (200-800 ms window)

We also considered a model that included TrialNumber[3] and Group as well as all the two and three-way interactions with Match and PhonemeChange. These models were considered because, as discussed in the Method (Section **Error! Reference source not found.**, Design), we hypothesized that being exposed to both matching and mismatching stimuli might weaken the relationship between the acoustics and the morphological status. This might lead to a decrease in the Match effect over the course of the experiment, or a decrease in the Match effect just for listeners exposed to mismatches on the first day (Group M2). Furthermore, it may be that only Group M2 would change behavior over the course of the experiment (an interaction between Group and TrialNumber). The model again found a robust effect of Match and no interaction between Match and PhonemeChange. Group was not significant and did not interact with any other effect. TrialNumber, which only approached significance as a main effect (β =-0.17, SE = 0.09, t =1.91, p = 0.06), interacted significantly with PhonemeChange (β =-0.43, SE = 0.18, t = 2.36, p = 0.02) but not with Match (β =-2.6, SE = 0.18, t = 1.43, p = 0.15). This pattern indicates that looks to the target decreased over the course of the experiment, in particular for *dis-mis-* trials. There was also a trend in the data that indicated the effect of Match lessened over the course of the experiment for those *dis-mis-* trials. Analyses including Group and TrialNumber can be found in the supplemental materials.

3.1.3 PrefixStatus (200-800 ms window)

[3] The models reported here did not include any random slopes for TrialNumber as this led to convergence problems, likely due to the sparcity of the data on an individual or item level.

A final set of exploratory analyses examined whether there were any asymmetries in the effects of Match due to PrefixStatus rather than PhonemeChange. Visual inspection of the data suggested asymmetries (as shown in the supplemental materials) with a bigger effect of Match when the critical syllable was taken from true prefixes, especially for group M1 and especially for trials in the first half of the experiment. A model including PrefixStatus, TrialNumber and Group found an interaction between Match, PrefixStatus and TrialNumber ($\beta = 0.93$, $SE = 0.36$, $t = 2.56$, $p = 0.01$) which indicated that the effect of Match was greater when the critical syllable was taken from a true prefix than when it was taken from a pseudo prefix, especially in the beginning of the experiment. Interactions with Group were not significant though numerically the effect of Match was greatest for the true prefixes for group M1.

3.2 Time course of effects

3.2.1 Main analysis

The previous analyses established that listeners spent less time fixating the target image when the acoustic information in the critical syllable mismatched the morphological structure of the target. As noted in the Introduction, we were also interested in how quickly the acoustic information influenced their looking behaviour, i.e. whether listeners used the acoustic information to drive eye-movements predictively, before they heard any disambiguating information. The alternative explanation of the mismatch effect observed above would be that the information is noted, but is not by itself sufficient to guide expectations and hence influence behaviour. Rather it would presumably have a sort of cumulative influence on perceptual decisions that depended on overall acoustic coherence between the first and later syllables in the word, and as such would presumably influence later and not earlier looks. Such a result would be interesting, but would point to a role that is different from our hypothesis of a strong predictive influence that is the main motivation for our work.

To address this issue, we examined the evolution of the effects examined in previous sections over the course of the trial. Following Clayards, Niebuhr & Gaskell (2015) and Kingston, Levy, Rysling & Staum (2016) we binned the eye movements into 100 ms bins and performed the regression model on each bin. We included the same fixed and random effects structure as the main model above (Match, PhonemeChange, and their interaction) as well as TrialNumber and Group and their interactions with the other fixed effects[4]. The estimates for Match, PhonemeChange and their interaction (as well as TrialNumber discussed below) are plotted in Figure 3 in terms of the estimates and p values of the fitted models.

[4] Models run without Trial and Group had the same pattern of results for Match, PhonemeChange and their interaction, see supplemental materials.

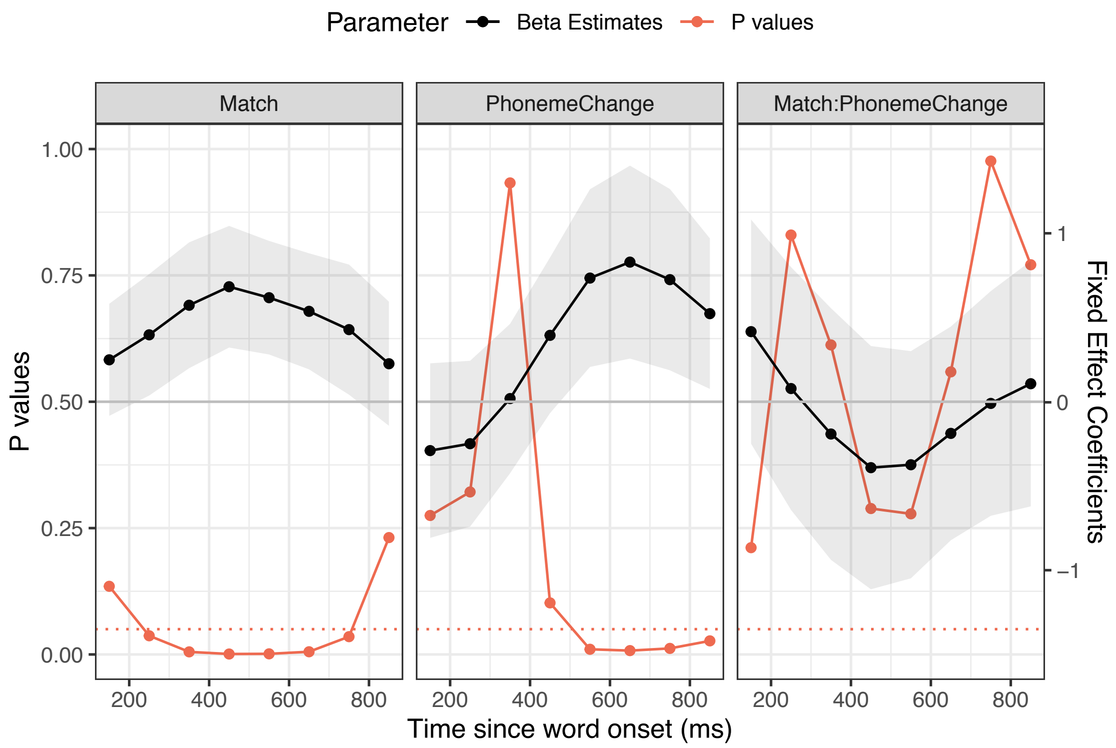


Figure 6. Results of mixed effects regressions over time for Match, PhonemeChange, and their interaction. Shading is two standard errors of the coefficient estimates as calculated by the regression models. On each panel's y axis, p values are shown at the left and beta coefficients at the right. Black curves: beta coefficients; red curves: p values. The dotted red horizontal line indicates p=0.05. The thick grey solid horizontal line indicates Coefficient = 0. Colour online.

Figure 6 (left panel) shows that the beta values for Match increase and then decrease as the trial progresses, asymptoting around 450 ms after critical word onset. At the second time bin, between 200 and 300 ms from the onset of the word, the two-standard error bars just miss touching zero and the p value is 0.04. After that point the effect of Match is clearly below $p = 0.05$ until the last time bin, when it returns to > 0.05. This indicates that the acoustic pattern of the critical syllable affects looks to the target from very early on in the syllable; Matches facilitate correct prediction of the sentence continuation (the target). Assuming the standard 200 ms lag between planning and executing an eye-movement, 200-300 ms after word onset is the earliest possible window for which we might expect to see any effects. The results of that model indicate that at least for many participants and items, there is an effect of Match at this earliest time point. As Figure 4 shows, for most of the items, the critical syllable is longer than the length of this window (100 ms) so the disambiguating information at the splice point has not yet arrived. This strengthens our claim that the acoustic information in the critical syllable is being used to anticipate the target word and looks to the target are delayed when it mismatches.

Although the model on a single large window (Section 3.1) found no effect of PhonemeChange, this time-course analysis shows that PhonemeChange (shown in

Figure 6, middle panel) is significant in the second half of the 200 ms to 800 ms window (from about 500 ms since word onset and after the end of the critical syllable). The coefficient estimates indicate that this was due to more looks to the target for *dis-mis-* items than for *re-* items later in the sentence. As before in the model on a single large window, there is no interaction between Match and PhonemeChange (Figure 6, right panel).

3.2.2 Group and TrialNumber (Time course)

As before, we also included Group (whether the participants heard all matches or all mismatches on day 1) and TrialNumber (the trial order in the experiment) and their interactions. As before we found that neither Group nor any of its interactions had a significant effect at any time point. The model on a single large window had found that looks to the target decreased over the course of the experiment (effect of TrialNumber), especially for the *dis-mis-* stimuli (TrialNumber by PhonemeChange interaction). The models fit every 100 ms found that the effects of TrialNumber as well as its interaction with PhonemeChange were limited to the first few hundred milliseconds after word onset ($p < 0.05$ for the first three time bins, Figure 7). This seems to indicate that over the course of the experiment, participants stopped making early looks to the target, i.e. they stopped anticipating the target. As pointed out in the Introduction, this may be because (within the context of the experiment) it was not essential to pay attention to the early part of the word—the continuation of the sentence typically disambiguated the two images. Participants may have learned this (implicitly or explicitly) as the experiment progressed. There was also a trend for the early effect of Match to get smaller over the course of the experiment (Match by TrialNumber Figure 7). No other two or three-way interactions had any time points with a p-value <0.05 (see supplemental materials for full details).

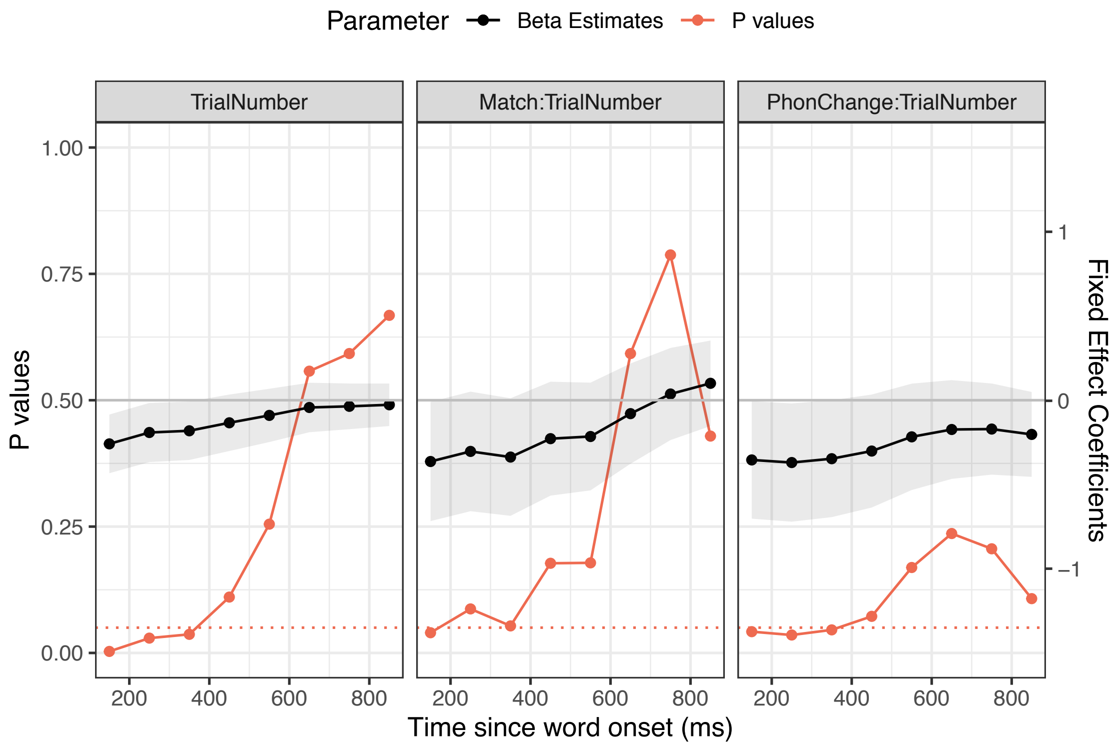


Figure 7: Results of mixed effects regressions over time for TrialNumber and its interactions with Match and PhonemeChange. Shading is two standard errors of the coefficient estimates as calculated by the regression models. On each panel's y axis, p values are shown at the left and beta coefficients at the right. Black curves: beta coefficients; red curves: p values. The dotted red horizontal line indicates p=0.05. The thick grey solid horizontal line indicates Coefficient = 0. Colour online.

3.2.3 Prefix Status (Time course)

A final analysis considered models that included PrefixStatus instead of PhonemeChange and its interactions with Match and Group as well as Match and TrialNumber. Figure 8 (left panel) shows that there was a non-significant trend for an interaction between PrefixStatus and Match in the second half of the 200-800 ms window (after 500 ms from word onset) that indicated that the effect of Match may have been greater for true prefixes, consistent with the numeric trend from the single large-window model. There were also significant interactions between PrefixStatus, Match and TrialNumber during this same later part of the 200-800 ms window, as shown in the right panel of Figure 8. This indicates that the Match x PrefixStatus interaction influenced eye-movements at the beginning of the experiment but became weaker over the course of the experiment, which is probably why it was not statistically significant when aggregated over the whole time-course of the experiment.

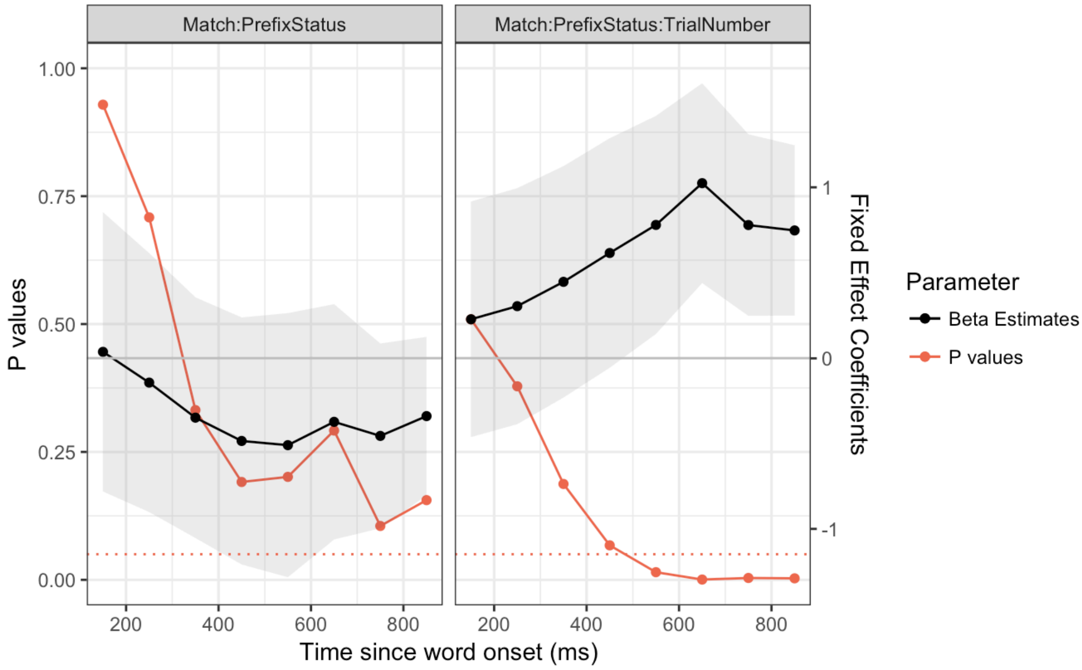


Figure 8: Results of mixed effects regressions over time for the interactions Match x PrefixStatus and Match x PrefixStatus x TrialNumber. Shading is two standard errors of the coefficient estimates as calculated by the regression models. On each panel's y axis, p values are shown at the left and beta coefficients at the right. Black curves: beta coefficients; red curves: p values. The dotted red horizontal line indicates p=0.05. The thick grey solid horizontal line indicates Coefficient = 0. Colour online.

3.3 Summary of results

Overall, listeners spent more time looking at the target image (the one consistent with the sentence continuation) when the acoustic properties of the critical syllable matched those expected for the target word in the associated image (main effect of Match in all models), thus supporting the main hypothesis. This was true both when the phoneme changed (*re-* stimuli) and when only the acoustical pattern within the syllables changed (*dis-mis-* stimuli) (i.e. no interaction with PhonemeChange in the model on a single window from 200 ms to 800 ms).

We also examined patterns over the course of the sentence (successive 100 ms windows between 200 to 800 ms from word onset) and over the course of trials in the experiment (from the first to the last trial for each participant). When we examined the time-course of the sentence, we found that listeners looked to the correct critical syllable from the earliest moments of its being heard. This use of the acoustic-phonetic detail of the critical syllable is reflected in the effect of Match being significant from the 200-300 ms bin. It indicates that the acoustic information was immediately taken up and used predictively by listeners. We also found that a few hundred milliseconds later, listeners had mostly converged on the target for *dis-mis-* stimuli but were looking less consistently at the target for *re-* stimuli (i.e. effect of PhonemeChange starting at the 500-600ms bin). This may be because many of the *re-* stimuli were globally ambiguous, or possibly because the greater acoustic complexity of *dis-mis-* syllables compared with *re-* ones (including the presence of abrupt acoustic boundaries within *dis-mis-* syllables) makes them more auditorily distinctive. Examining the time-course of the experiment, we found that as the experiment progressed, the early looks to the target decreased, especially for the *dis-mis-* stimuli (i.e. PhonemeChange by TrialNumber interaction up to the 300-400 ms bin). This indicates that listeners began to respond differently to (especially) *dis-mis-* critical syllables as they became familiar with the structure of the stimuli. Finally, we found that after about 500 ms from the word onset, effects of Match were greatest when listeners heard a true prefix, though, consistent with the way responses changed as the experiment progressed, this benefit for matched prefixes was only at the beginning of the experiment.

4 Discussion

4.1 The main findings

At a general level, we asked whether listeners are sensitive to phonetic detail – both segmental and rhythmic information – that systematically reflects morphological structure while involving no changes in number of syllables. We further asked whether listeners are likely to use this phonetic detail predictively (in order to help distinguish words that contain true prefixes from those that contain pseudo prefixes) in ordinary listening conditions—that is, in an easy on-line task under good listening conditions in which the aim is response accuracy but not speed.

The three main questions specific to our experiment were whether there is a mismatch effect overall, whether it is independent of the phonemic status of the acoustic difference (i.e. with *dis-mis-* as well as *re-*), and in particular whether the acoustic information within the critical syllable influences perceptual decisions in real time, rather than only being influential in combination with the rest of the word. All three questions are answered affirmatively. Listeners spent more time fixating the target image when the critical syllable matched the continuation of the sentence, whether or not there was a phoneme change, and in real time. Furthermore, while prefix status did not affect the answers to the three main questions, the results suggest that true prefixes may convey more reliable information about their status than pseudo prefixes in some circumstances, as discussed below.

We also made some additional observations. Foremost amongst these is the evidence for rapid learning during the task. As expected, listeners used the internal acoustic structure of the prefixed or non-prefixed syllable predictively. However, as the experiment progressed this prediction effect weakened, presumably as listeners learned that they could wait for the sentence continuation to provide disambiguating information. That they did not wait for the continuation during early trials means that the critical acoustic information is likely to be used predictively in normal listening conditions. This finding has practical as well as theoretical interest: the fact that listeners' behavior changed early in the experiment suggests that future work on this type of distinction should consider trial number as a predictor variable. Furthermore, prefix status (whether the initial syllable came from a true or pseudo prefixed word) seems to affect eye movements. The 200-800 ms single window analysis showed a stronger benefit of Match when the critical syllable was a true prefix, and the interaction with TrialNumber confirmed that this benefit was again especially obvious in the beginning stages of the experiment. As noted in the Introduction, although both prefixes and pseudo prefixes are weak syllables in that they do not carry primary lexical stress and are not normally accented in utterances, prefixes are associated with a degree of stress, or rhythmic prominence, that pseudo prefixes in comparable word structures lack. The present results suggest that the rhythmic emphasis that comes with a true prefix may be more perceptually compelling than the absence of such a focus. This further encourages exploration of the hypothesis that rhythmic properties of the signal are fundamental to speech processing in real time. The next two sections discuss the nature of rhythm and metre, and outline its relevance to a general model of perceptual processing.

4.2 Rhythm and hierarchical metrical structure as organising principles for speech perception

It has been argued that the phonetic detail manipulated in our study is best understood as reflecting differences in rhythmic or metrical structure between prefixes and their equivalent phones in mono-morphemic words (e.g. Smith, 2012; Hawkins, 2001; 2003; 2010). This argument, and our current results, support other suggestions in the literature that rhythmic or metrical structure is an important part of the representation used to recognize speech (cf. Salverda et al., 2003; Brown et al., 2015; Breen et al., 2014). This section first outlines the useful distinction between rhythm and metrical structure, and their interrelationship. Then it explores how the metrical-rhythmic structure of speech might serve as a fundamental organising principle for speech perception, melding multi-modal properties of the physical signal with linguistic and social knowledge to achieve communicative success.

4.2.1 Metre and rhythm

Musical analyses distinguish rhythm from metre. London (2012) expresses the distinction as follows. Rhythm represents a series of physical events having particular relationships with one another. In music, these are largely durational (the inter-onset intervals of notes). Metre, in contrast, is a perceptual phenomenon: an emergent organization involving a degree of periodicity that is constructed by the brain in response to stimuli that are perceived as rhythmic (Fujioka et al., 2012). In hearing metrically, the brain sets up a beat that hierarchically structures the rhythm, focusses attention on the metrical beats, and allows prediction of the time of occurrence of future events (see e.g. Calderone et al., 2014; Lakatos et al., 2005; Lakatos et al., 2008).

This distinction between rhythm and metre can be helpful for speech analysis too. Speech rhythm can often be represented simply in terms of relative durations of similar units in the utterance. Other parameters—f0, amplitude, and sometimes timbre—may contribute to both rhythmic and metrical aspects of speech. In doing so, they can override durational influences on perceived rhythmic and metrical structure (e.g. Dilley, Mattys & Vinke, 2010, Experiment 3a). Thus metre, which is inherently hierarchical, can be represented for speech as the mapping of auditory patterns onto linguistic units, from segments (allophones, phonemes or their psychological equivalents) through syllables to metrical feet and intonational phrases.

Metrical, or beat-based, structure, enables establishment of a metrical hierarchy where faster rhythmic events can happen within slower ones. Faster rhythmic events in speech presumably include syllables (or syllable-like units such as Japanese morae, hereafter not distinguished from syllables). In linguistic terms suitable for languages like English, a beat-based hierarchy of syllabic weight is called stress, the main beats being accented syllables (sometimes called prominence, or primary stress), while less important syllables take secondary stress or are unstressed. Common to both speech and music is that perception of a rhythmic group can change depending on the listener's construal of the wider metrical (e.g. for speech, sentential) structure it occurs in. For speech, preceding meaning and/or rate of speech influence perception (e.g. Pickett and Pollack, 1963; Ernestus, Baayen and Schreuder, 2002; Ernestus, 2014; Ernestus, Hanique and Verboom, 2015; Dilley, Mattys & Vinke, 2010; Heffner, Dilley, McAuley and Pitt, 2013; Morrill, Heffner and Dilley, 2014), while explicit instructions to hear a beat train in ternary or else in binary time can also influence listeners' metrical response, as measured by EEG (Nozaradan, Peretz, Missal and Mouraux, 2011; Nozaradan, Peretz, and Mouraux, 2012).

The experience of rhythm in complex auditory signals such as speech and most music is learned (Mattys et al., 1999; Hannon & Trehub, 2005). Without such (usually implicit) learning from exposure to the relevant signals, which amounts to acculturation, complex sound sequences sound unstructured, even chaotic. For speech, such learning is part of linguistic knowledge. With such knowledge, beat-based listening facilitates prediction of upcoming events, a property increasingly seen as essential to successful communication between individuals (Philips-Silver & Trainor, 2005; Cirelli, Wan & Trainor, 2016).

The principles described here have been used to explore the relationship between speech rhythm and the various frequencies of cortical neural oscillations in the brain which entrain to external stimuli. One of the more complete models, Giraud and Poeppel (2012), privileges rhythm in speech intelligibility. Amongst other things, it identifies low gamma (25-30 Hz, 33-40 ms) and theta (4-8 Hz, 125-250 ms) frequencies as entraining to feature/phoneme-sized and syllable-sized durations respectively. Delta frequencies (1-3 or 4 Hz, 250-1000 ms) are implicated in prosodic processing. However, it seems mistaken to necessarily associate shorter durations with phonological units and longer ones with prosody. English has many instances of long phonological units and short prosodic ones (Hawkins, 2014:1-3).

A stronger argument, relevant to the present study, comes from Mai, Minnett and Wang's (2016) EEG study of Mandarin Chinese. They manipulated sentences of meaningful vs. nonsense disyllabic words, and backwards vs. normal speech. Like Giraud and Poeppel (2012), they concluded that phonological and syntactic-semantic processing engage different neural networks, but identified quite different frequencies: semantic/syntactic processing with fast gamma frequencies, and phonological processing with slower theta and delta frequencies, as well as beta (13-30 Hz, 33-77 ms). These patterns reflect that the syllable, not the phoneme, is the important contrastive unit in Mandarin phonology: consonants strongly determine vowel quality, f0 operates over the entire syllable to change word meaning, as well as in longer prosodies, and syllable stress is relatively invariant.

In sum, perceived speech rhythm involves representation of sequences of linguistic units of the size of syllables or longer. Metre is hierarchical and involves representation of the entire spectrotemporal signal of a phrase or more. That is, metrical structure is constructed or imposed by the listening brain, rather than being a property inherent within the physical signal. So recognition of speech rhythm and metre may entail recognition of entire complex auditory patterns, within which subunits can be discerned. Expected attributes not clearly present in the physical signal but whose presence is implied by and compatible with the overall pattern can be adduced by neural pattern completion, a process attested for both vision (e.g. Meng, Remus and Tong, 2005; Murray, Kersten, Olshausen, Schrater and Woods, 2002) and speech (Shahin, Bishop and Miller, 2009; for a review, see Hawkins, 2014). As such, rhythm and metre offer the possibility of structuring the speech signal such that all its contrasting abstract units are representable in a systematic and economical way that necessarily includes 'top-down' knowledge of the language itself.

4.2.2 Relevance to perception of affixed words

Consistent with the arguments above, every utterance can be described by a metrical (prosodic) hierarchy that partly governs the phonetic detail of segments and syllables within its domain. Autosegmental-Metrical (AM) phonology (e.g. Pierrehumbert & Beckman, 1988; Post, D'Imperio & Gussenhoven, 2007; Cho, 2016), and Firthian Prosodic Analysis (FPA, e.g. Ogden et al., 2000) exemplify two theoretical frameworks based on this approach. In Cho's (2016, p136) words for AM, "Prosodic structure provides a "frame" for articulation based on which abstract phonological representations whose phonetic detail is rather coarsely specified by the phonology of the language are fleshed out with fine-grained phonetic content in both segmental and suprasegmental dimensions…this assumption entails that the prosodic structure of an utterance is phonetically "encoded" into the speech signal and the listener in turn decodes the structural information from the signal and exploits it in speech comprehension." FPA embodies similar principles, and in addition every metrical/prosodic hierarchical structure describing an utterance is linked to its corresponding syntactic hierarchical structure (Ogden et al., 2000). These metrical principles have been extended to perception, particularly of prefixed vs non-prefixed words, with discussion of links to grammar, morphological structure, and lexical items and their associative networks (Hawkins, 2010; Hawkins and Smith, 2001; Hawkins, 2003). The results of the present experiment broadly support claims in the papers cited, but some updating is warranted.

The present study supports four central tenets of Hawkins' theoretical position, made in the papers cited above. The first tenet (which before the present study was a hypothesis) is that the fine phonetic detail that distinguishes prefixes from nonprefix syllables is used by listeners in real time—i.e. their behavior changes as they hear it. This phonetic detail relates to meaning, and does not depend on there being a phonemic contrast to relate it to meaning. By extension, all phonetic detail seems likely to be related to meaning, with no intervening 'levels' of formal linguistic structure likely to be obligatory in the process of relating sound to the talker's meaning. This does not deny the psychological reality of intervening structural 'levels'; it merely says that a given level of analysis does not always have to be accessed for meaning to be understood from the spoken signal. Neuropsychological support for this claim comes from Krieger-Redwood *et al.* (2013), who used TMS to show that processes requiring semantic categorisation (i.e. understanding word meaning) are independent of judgments requiring phonological classification (i.e. phonemic content); semantic judgments can operate when phonological processes are unavailable. This has important consequences: it means that the perceptual system is probably more closely attuned to general, modality-free properties of pattern recognition than is normally assumed by perceptual theories based on the mutually exclusive categories and analytic levels of theoretical linguistics.

The second tenet is that every short pattern of sound (segment or several segments) can only be described and hence perceptually interpreted in terms of its context. The same sound in a different context may be interpreted entirely differently. Our experimental manipulation and analysis addressed this claim implicitly, especially by virtue of tracking decision changes (e.g. to the critical syllable) over time. A listener can map sound directly to meaning, but only for the context in which it is heard. Context is broadly defined. It includes the immediate local context—that is, the prosodic/metrical structure of the utterance that the sound is part of—possibly one or more preceding phrases, and the listener's understanding of the entire communicative situation. Understanding speech is thus inherently situation-specific (see also Hawkins, 2011; Hawkins, 2014).

The third tenet that our results support is that fine phonetic detail will be used to access meaning when it is relevant to the situation at hand. The reduction in early looks to the target for *dis-* and *mis-* stimuli indicates that listeners used the information in the critical syllable at the start of the experiment but quickly adapted to the fact that the task did not demand it—they could wait until later in the sentence for acoustically clearer disambiguation. This rapid adaptation to task requirements is supported by the literature on perceptual learning and adaptation to new accents etc. (e.g. Maye, Aslin & Tanenhaus, 2008; Bradlow & Bent, 2008; Barden & Hawkins, 2013; Nguyen and Hawkins, 2001).

The fourth tenet is that "lack of clear evidence for a particular category, as with the reduced first syllable in *mistakes*, can be informative" (Hawkins, 2010, p486). That is, it was expected that the absence of a perceptual beat on unprefixed forms of the critical syllables *mis-, dis-* and *re-* would help listeners predict monomorphemic word identity. This was broadly the case. However, as noted above, in the early part of the experiment and after about 500 ms from word onset, true prefixes helped word identification more than nonprefixes did. Since the prefix contains a heavier beat than the non-prefix, we interpret this last result as indicating that a perceptual beat is important in driving perceptual decisions, whereas its absence may be less influential. This is consistent with neuroscientific studies showing that a beat-based hierarchy is fundamental to selective attention (Lakatos et al., 2007; Lakatos et al., 2008; Lakatos et al.,, 2009; Arnal & Giraud, 2012). An obvious inference is that the prefix is emphasized because it bears quasi-independent information about meaning that the phonemically-identical nonprefix does not.

If prefixes carry secondary stress to draw attention to them and their meaning, why are suffixes not similarly stressed? One possibility takes us again to the role of prediction in understanding speech. Understood within the natural context of their metrical structure, prefixes herald the beginning of a new lexical item that will be a polysyllabic word with a main lexical stress later than the prefix, and the prefix will change the meaning of the stem. There are numerous subtle changes to segmental durations in polysyllabic words that depend on the word's phonological and acoustic structure (see Hay et al. (in prep.) for information pertinent to prefixes). Suffixes make the word phonologically longer, by adding either a syllable or else a segment that makes the final coda more complex. Word stems are likewise subtly modified by the addition of one or more suffixes, and there are differences between words with different types of suffix, compared with monomorphemic words. For example, Plag, Homann, and Kunter (2015/2017) discuss complexities of regularities found for the various morphemes represented by English /s/ and /z/ (plural, genitive, 3rd person singular, etc).

If the stress on a prefix heralds a polymorphemic word, it seems reasonable that the changes that a suffixed word stem undergo could raise the probability of an upcoming suffix. Perceptual experiments described in Section 1.2 show that they do for both English and Dutch (Blazej & Cohen-Goldberg, 2014; Kemps et al., 2005a; Kemps et al., 2005b). However, as noted there, given that these studies contrasted monosyllabic, mono-morphemic words with polysyllabic, polymorphemic words, it is impossible to tell whether their listeners were simply anticipating a longer word, or anticipating the polymorphemic structure of the longer word, or both.

What is clear, however, is that the suffix itself does not "need" to receive secondary stress because its presence is predictable from the internal acoustic structure of its stem, which would typically take strongest stress in the word. Together with a constraint against stressing word-final syllables of polysyllabic words in English and Dutch, and, presumably, usually some grammatical priming, this might be sufficient to reduce perceptual uncertainty.

In sum, this study supports previous claims and hypotheses that listeners use fine phonetic detail in real time to efficiently access meaning, but only in its appropriate context and if the task makes it relevant and 'cost-effective' to do so.

The previous section concluded by implying that rhythm and metre have the potential to provide the underlying 'glue' of speech communication by focusing attention onto critical events in the speech stream. Those critical events are associated with beats that allow a metrical structure to be created by the listening brain. The metrical structure facilitates prediction and allows meaning to be efficiently accessed. This section concludes by briefly adding two points to that claim. The first is that much of the speech signal is of course crucial to intelligibility but does not receive a metrical beat. How might that 'non-beat' information be processed? Hawkins (2010) suggested that phonetic detail in the entire speech signal is continuously monitored, for a variety of reasons including in order to learn about communicatively significant new patterns. This position is supported by neuroscientific evidence related to that cited above demonstrating creation of beat-based metrical structure. For example, Schroeder and Lakatos (2009) propose that when a stimulus lacks rhythm, lower-frequency neuroelectric oscillations entrained to metrical structure are suppressed and replaced by continuous monitoring (vigilance) that uses higher-frequency oscillations. These systems can operate simultaneously, differing in balance depending on the rhythmicity of the stimulus. This claim is also consistent with experiments that show that lexical activation varies continuously in a way which reflects variation in the acoustic signal as it unfolds over time (e.g. Allopenna, Magnuson, & Tanenhaus, 1998; Gow & McMurray, 2007; McMurray et al., 2003; Warren & Marslen-Wilson, 1987).

The second concluding point is that speech normally takes place as part of meaningful communication between people. Any hypothesis concerning the role of rhythm and metre in understanding speech needs to encompass the interactive and multimodal properties typical of most human communication. The power of multimodal sensory information in facilitating speech intelligibility and spoken communication is well known. The strong correlation between the auditory signal and visual input from the gestural code and facial expression is equally well known. There is ample evidence that rhythm and metrical structure play a crucial role in both multimodal integration of a message from a single talker (e.g. Schroeder et al., 2008) and in the entrainment that occurs during communication between talkers (e.g. Hasson et al., 2012). Detailed discussion goes beyond the scope of this paper, but a general review relevant to rhythm in both spoken and musical interaction can be found in Hawkins, Cross and Ogden (2013).

4.3 Strengths, limitations and extensions of the study

We achieved our aim of demonstrating that a rhythmic distinction of fine phonetic detail, with no phonemic contrast involved, can be used to access meaning and predict lexical identity in real time. We used the morphological prefix distinction because it suited our aims well, since there is relatively good consensus about the meaning of a prefix. Other contrasts could have been used, but few lend themselves as well to our primary question.

Unlike previous studies, we controlled for confounding factors like the number of syllables in the mono-morphemic vs. prefixed words, and prosodic structure of pairs of stimuli. We put much effort into ensuring that our stimuli described and illustrated plausible visual scenes. By tracking our results over the course of the experiment we were also able to show that the experiment itself changed participants' behaviour. This points to an important methodological consideration for future studies. Adaptation to experimental conditions can occur with just a few trials. Researchers should consider this possibility in their analyses before concluding that a manipulation did not affect participants' behaviour, since a weakening of the effect over the course of the experiment can obscure important results relevant to understanding speech in normal, everyday situations. It should also be noted that we found our effect weakened over the course of the experiment, even though the stimuli included a high proportion of filler trials, a subset of which were designed to counteract effects of cross-spliced critical syllables.

We expect these results to generalize to most if not all English prefixes. The phonetic detail will be specific to the prefix, and some words and contexts may differ from the usual pattern. Word-specific influences could include decomposability, and the relative frequency of the prefixed and unprefixed word. But the principle of more vs. less stress (or less vs. more syllable reduction) is expected to be true for all prefixes and pseudo prefixes. We expect some dialect differences in the exact acoustic details: work in progress supports this prediction while confirming that the general patterns hold across dialects of English (Hay et al., in prep). In terms of wider theoretical implications, the same type of reasoning can presumably be generalized to perception of any other audible contrast based on non-phonemic phonetic detail, as long as the auditory contrast has a systematic relationship with distinctions of meaning, broadly defined.

A question for future research is whether the multiple acoustic cues we and others have identified as contributing to the prefix-nonprefix distinction work together in concert, or whether any dominate perceptual responses. Perhaps the most valuable question we can ask pertains to the role of duration in the syllables *dis*- and *mis*. Syllable duration is often shown to be an over-riding perceptual cue (e.g. Salverda, Dahan & McQueen (2003) Experiment 2). But for distinguishing prefix and non-prefix forms of *dis*- and *mis*-, is what matters the duration of the whole syllable, regardless of its internal acoustic structure (e.g. the s:ɪ ratio), or is it the internal acoustic structure regardless (within limits) of the overall syllable duration? We hypothesize that for SSBE, it is likely to be the internal acoustic structure, as carried by the s:ɪ ratio, since that ratio is distinctive in production (Smith, Baker and Hawkins 2012). We suggest this because there is more scope for a relatively long vowel to convey a stronger rhythmic beat (indicative of the prefix form), by virtue of amplitude, f0 and formant spacing, than for the short vowel that tends to accompany the non-prefix form. In sum, the rhythmic hypothesis would be supported if the internal acoustic structure proved more decisive in indicating prefix status than overall syllable duration alone, without any change in s:ɪ ratio. This experiment is planned.

4.4 Concluding summary

This study confirms that phonetic detail associated with prefixes and pseudo prefixes can aid prediction of the upcoming word's identity. Our results, especially from early trials, indicate that such phonetic detail will typically be used predictively in real-world listening conditions. However, we have also shown that over the course of our experiment many of the effects weakened, indicating relatively fast adaptation to the experimental conditions. There was also some evidence that prefixes influence predictive behavior more strongly than pseudo prefixes do. We suggest that our findings support the hypothesis that speech rhythm (more properly, the metre of speech) provides a fundamental binding principle of speech processing, enabling linguistic structures to be created and matched with similar structures in memory to allow rapid matching of complex sound patterns to meaning.

Supplementary Materials

Stimuli, original data, supplementary figures, R code for the analyses, and additional acoustic analyses of the stimuli not reported in the manuscript are available at https://osf.io/dsyxu. DOI 10.17605/OSF.IO/DSYXU.

Acknowledgements

The authors would like to thank Gerry Altmann for the use of the eye-trackers and lab space; Elaine Tham for data collection; and Marco Piccolino-Boniforti for extensive help with stimulus construction and for many helpful discussions. Portions of this work were presented at the 17th International Congress of Phonetic Sciences, 2011, Hong Kong, the Colloquium of British Association of Academic Phoneticians, Leeds, UK, 2012, and AMLaP 2012, Riva del Garda, Italy. This research was supported by the Marie Curie Research Training Network Sound to Sense, MRTN-CT-2006-035561, including fellowships to Clayards and Piccolino-Boniforti.

Appendix A

The term ‘phonetic detail’ as used in this paper has a very particular meaning . It refers to acoustic-phonetic properties that are systematically distributed and communicatively significant but that are not essential to differentiate phonemes, and hence to distinguish the phonological form of words. This definition hides a number of complexities, discussed for example by Carlson and Hawkins (2007) and Hawkins and Local (2007), but the main points for present purposes are (1) that speech sounds can systematically distinguish meanings and communicative functions without there being a difference in phonemic structure; and (2) that the perceptual significance and hence meaning of any single part of the speech signal depends on the situational and phonetic context in which it is heard. The first claim is widely accepted, although the meanings concerned (e.g. question vs. statement intonation, expressions of doubt) were traditionally designated ‘paralinguistic’ and treated separately both from the types of lexical meaning distinguished in phonemic analysis, and from mainstream psycholinguistic theories of spoken word recognition; the account proposed here makes no such restriction, as exemplified by our focus on grammatical (specifically morphological) linguistic structure. The second claim is likewise widely accepted, but its implications do not always drive the theoretical interpretation of experimental results, nor the design of experiments testing the role of phonetic detail in speech perception.

Appendix B

Sentence pairs used in the experiment. The first sentence of each pair contains the pseudo prefix and the second sentence of each pair contains the true prefix.

I wouldn't be surprised if the boys **discover** them.
I wouldn't be surprised if the boys **discolour** them.

He fell asleep despite all the **discussion.**
He fell asleep despite all the **discomfort.**

The teacher has a very **discursive** style.
The teacher has a very **discourteous** style. (3-syllable pronunciation of target word)

Her paintings are so **distinctive.**
Her paintings are so **distasteful.**

It was difficult because Sam **distracted** him.
It was difficult because Sam **distrusted** him.

A swan **displays** its plumage to its mate.
A swan **displaces** water when it lands.

Alex typically **discards** the fruit.
Alex typically **discounts** the risk.

I'd be surprised if Tess **mistakes** the letters.
I'd be surprised if Tess **mistypes** the letters.

We felt uncomfortable about his **mysterious** demeanour. (3-syllable pronunciation)
We felt uncomfortable about his **mistreatment** of Amina.

The girls were spellbound by tales of Jo's **mystique.**
The girls were spellbound by tales of Jo's **misdeeds.**

We think Jeff **mistook** the tree for a person.
We think Jeff **mistimed** the turning on purpose.

Jo struggled to **recover** her balance.
Joe struggled to **re-cover** the sofa.

They agreed they should **repeal** the verdict.
They agreed they should **re-peel** the carrots.

That's Oscar Wilde. He loved **reposing** quietly.
That's Oscar Wilde. He loved **re-posing** questions.

After the massacre, the armed forces **reformed** their procedures.
After the massacre, the armed forces **re-formed** on the hillside.

After the noisy lawn party, Josh **receded** behind the shed.
After the noisy lawn party, Josh **re-seeded** the trampled lawn.

Everyone was happier after Geoff **restrained** the brute.
Everyone was happier after Geoff **re-strained** the fruit.

He hurried to **relay** the message.
He hurried to **re-lay** the carpet.

They're starting to **redress** the wrong.
They're starting to **re-dress** the wound.

We hoped he'd **release** the catch soon.
We hoped he'd **re-lease** the house soon.

The next job was to **repair** the socks.
They next job was to **re-pair** the socks.

We watched Jess **restore** them.
We watched Jess **re-store** them.

He was punished for **refusing** so rudely.
He was punished for **re-fuelling** so slowly.

The man **rejoiced** as he finished the race.
The man **re-joined** the ends of the rope.

Todd **rebutted** the argument successfully.
Todd **re-baited** the fishing line successfully.

Harry's parents **revoked** his privileges.
Harry's parents **revoiced** his worries again.

We know that Dave **restricts** his arm movements when necessary.
We know that Dave **re-strings** his instrument when necessary.

It's a perfect example of **extravagance** in public spending.
It's a perfect example of **ex-trampoliners'** sense of balance.

We were amused to hear those **expletives** had been censored.
We were amused to hear those **ex-policemen** had been honoured.

There are conflicting views about these **expanders'** roles in orthodontistry.
There are conflicting views about these **expatriates'** roles in this society.
(3-syllable pronunciation of target word)

They didn't understand why these **exponency** terms were important.
They didn't understand why these **ex-pony club** girls were important.

The judge decided he'd **expatiate** at some length.
The judge decided he'd **expatriate** the poor kids.

Appendix C

37 pairs of Filler items specific to this experiment (For the other 30 pairs of filler items, see Heinrich, Flory and Hawkins, 2010.) Words in bold contain either a true (tr) or a pseudo (ps) prefix

We tried to **distinguish** between the twins (ps)
We tried to **distinguish** between the bins

You purify water by **distilling** it (ps)
You purify whisky by **distilling** it

We liked the **description** of the balloons over mountains (ps)
We liked the **description** of the fantastical dragon

The conductor loves his job **despite** being prone to backache (ps)
The conductor likes his job **despite** being prone to motion sickness

We could just **discern** the bridges in the fog (ps)
We could just **discern** the ridges in the fog

Sue **disturbed** the cows (ps)
Sue **disturbed** the sheep

He began to **destroy** the door (ps)
He began to **destroy** the cube

The vandals **distorted** the frame to get revenge (ps)
The vandals **distorted** the wheel to get revenge

They were all impressed with the **disabled** girl's spirit (ambiguously tr)
They were all impressed with the **disabled** boy's spirit

Sugar **dissolves** faster in hot liquids (ps + pronunciation change)
Sugar **dissolves** faster when you stir it

The man had seriously **mistreated** the donkey (tr)
The man had seriously **mistreated** the dog

The drawer was **misaligned** (tr)
The door was **misaligned**

The Lord of **Misrule** as a puppet (tr)
The Lord of **Misrule** as a carving

A **mistrial** is a rollerblader's trick (tr)

A **mistrial** is a lawyer's last resort

Jody had **miscalculated** when to take the toast out (tr)
Jody had **miscalculated** when to jump for the frisbee

The sheep did not **react** to the fox's presence (tr)
The sheep did not **react** to the dog's presence

A **refectory's** where monks eat (ps)
A **refectory's** where monks sleep

We were slow to **repack** because of the baby (tr)
We were slow to **repack** because of the dog

The dog was reluctant to **relinquish** the ball (ps)
The dog was reluctant to **relinquish** the ring

The class **redrew** classical cartoons (tr)
The class **redrew** classical plans

Kate **refused** to buy the coat (ps)
Kate **refused** to buy the dress

We'll **repaint** the dhow (tr)
We'll **repaint** the bow

They did a good job of **re-creating** the original instruments (tr)
They did a good job of **re-creating** the Victorian atmosphere

He **recited** the poem perfectly (ps)
He **recited** the poem passionately

His favorite was this **repeating** rifle (ps)
His favorite was this **repeating** pattern

They planned to **reheat** the risotto (tr)
They planned to **reheat** the lasagna

Cameron **re-sets** the stone (tr)
Cameron **re-sets** the bone

Ali **examined** the book intently (ps)
Ali **examined** the paint intently

Her **ex-husband** is a diver (tr)
Her **ex-husband** is a driver

The men **exchanged** looks (ps)
The men **exchanged** books

She's a really **excellent** musician (ps)
She's a really **excellent** clinician

Nothing like **extorting** promises (ps)
Nothing like **extorting** money

Geoff **extracted** the tooth (ps)
Geoff **extracted** the juice

There was a mountain of plastic **recycling** (tr?)
There was a mountain of plastic for playing in

Sally liked meeting all the **relations** (ps)
Sally like seeing the celebrations

Eddie always takes **revising** seriously (ps)
Eddie always takes his driving seriously

Luke tried hard not to eat the cake
Luke tried hard not to eat the ice cream

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439. https://doi.org/10.1006/jmla.1997.2558

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *TRENDS in Cognitive Sciences, 16*(7), 390-398. doi:10.1016/j.tics.2012.05.003

Baayen, R. H., Levelt, W. M. J., Schreuder, R., & Ernestus, M. (2008). Paradigmatic structure in speech production. *Chicago Linguistics Society, 43*(1), 1-29.

Baker, R. (2008). *The production and perception of morphologically and grammatically conditioned phonetic detail.* University of Cambridge, Cambridge.

Barden, K. J. (2011). *Perceptual learning of context-sensitive phonetic detail*. Doctoral thesis, University of Cambridge). doi:10.17863/CAM.16501

Barden, K. J., & Hawkins, S. (2013). Perceptual learning of phonetic information that indicates morphological structure. *Phonetica, 70*(4), 323-342. https://doi.org/10.1159/000357233

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. doi: 10.18637/jss.v067.i01.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook,* 3, 255-309. doi:10.1017/S095267570000066X

Ben Hedia, S. & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. J*ournal of Phonetics*, 62, 34-49. https://doi.org/10.1016/j.wocn.2017.02.002

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance, 41*(1), 50-68. doi:10.1037/a0038509

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research, 29*(2), 155-168. https://doi.org/10.1023/A:1005140927442

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*, 2670-2676. https://doi.org/10.1121/1.422787

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729. https://doi.org/10.1016/j.cognition.2007.04.005

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience, 29*(9), 1132-1146. doi:10.1080/23273798.2014.894642

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 306-323. doi:10.1037/a0038689

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *TRENDS in Cognitive Sciences, 18*(6), 300-309. https://doi.org/10.1016/j.tics.2014.02.005

Carlson, R., & Hawkins, S. (2007). When is fine phonetic detail a detail? In W. J. Barry & J. Trouvain (Eds.), *16th International Congress of Phonetic Sciences* (pp. 211-214, Paper ID 1721). Saarbrücken: http://www.icphs2007.de/.

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica, 58*, 129-162. https://doi.org/10.1159/000056196

Cho, T. (2016). Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass, 10*(3), 120-141. doi:10.1111/lnc3.12178

Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2016). Social effects of movement synchrony: Increased infant helpfulness only transfers to affiliates of synchronously moving partners. *Infancy, 21*(6), 807-821. doi:10.1111/infa.12140.

Clayards, M., Niebuhr, O., & Gaskell, M. G. (2015). The time course of auditory and language-specific mechanisms in compensation for sibilant assimilation. *Attention, Perception, & Psychophysics*, *77*(1), 311-328. https://doi.org/10.3758/s13414-014-0750-z

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*, 218-236. https://doi.org/10.1016/0749-596X(92)90012-M

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*(5/6), 507-534. https://doi.org/10.1080/01690960143000074

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Learning and Memory, 36*(3), 704-728. https://doi.org/10.1037/a0017449

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803-820. https://doi.org/10.1162/jocn.2009.21059

Davis, M. H., Marslen-Wilson, W., & Gaskell, M. G. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology-Human Perception and Performance, 28*, 218-244. https://doi.org/10.1037/0096-1523.28.1.218

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*(3), 274-294. doi:http://dx.doi.org/10.1016/j.jml.2010.06.003

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua, 142*, 27-41.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1-3), 162-173.

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics, 48*, 60-75. doi:http://dx.doi.org/10.1016/j.wocn.2014.08.001

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic Beta oscillations. *Journal of Neuroscience, 32*(5), 1791-1802.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511-517. doi:10.1038/nn.3063

Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole & J. I. Hualde (Eds.), *Papers in Laboratory Phonology 9: Phonology and Phonetics* (pp. 173-200). Berlin: Mouton de Gruyter.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science, 16*(1), 48-55.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *TRENDS in Cognitive Sciences, 16*(2), 114-121. doi:10.1016/j.tics.2011.12.007

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*, 373-405.

Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. In C. Fougeron, B. Kühnert, M. d'Imperio & N. Vallée (Eds.), *Laboratory Phonology 10: Variability, Phonetic Detail and Phonological Representation* (pp. 479-510). Berlin: Mouton de Gruyter.

Hawkins, S. (2011). Does phonetic detail guide situation-specific speech recognition? Keynote address, *Proceedings of the 17th International*

Congress of Phonetic Sciences. Hong Kong: http://www.icphs2011.hk/ICPHS_CongressProceedings.htm

Hawkins, S. (2014). Situational influences on rhythmicity in speech, music, and their interaction. In R. Smith, T. Rathcke, F. Cummins, K. Overy, & S. K. Scott (Eds.), *Communicative Rhythms in Brain and Behaviour* (Vol. 369, pp. 20130398). London: Philosophical Transactions of the Royal Society B 20130398.

Hawkins, S., Cross, I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech. In M. Orwin, C. Howes, & R. Kempson (Eds.), *Language, Music and Interaction* (Vol. 3, pp. 285-329). London: College Publications.

Hawkins, S., & Local, J. K. (2007). Sound to Sense: Introduction to the special session. In W. J. Barry & J. Trouvain (Eds.), *16th International Congress of Phonetic Sciences* (Vol. , pp. 181-184, Paper ID 1726). Saarbrücken: http://www.icphs2007.de/.

Hawkins, S., & Nguyen, N. (2001). Perception of coda voicing from properties of the onset and nucleus of 'led' and 'let'. In P. Dalsgaard, B. Lindberg & H. Benner (Eds.), *Eurospeech 2001: Proceedings of the 7th International Conference on Speech Communication and Technology,* (Vol. 1, pp. 407-410). Scandinavia.

Hawkins, S., & Smith, R. H. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics-Rivista di Linguistica, 13*, 99-188. http://linguistica.sns.it/RdL/13.1/Hawkins_Smith.pdf

Hay, J. B. (2003). *Causes and consequences of word structure*. New York: Routledge.

Hay, J., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *TRENDS in Cognitive Sciences, 9*(7), 342-348.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865-892.

Hay, J., Hawkins, S., Stuart-Smith, J. Smith, R., Fromont, R. (in preparation) The morphophonetics of word-initial *dis* and *mis* in four varieties of English.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes, 28*(9), 1275-1302. doi:10.1080/01690965.2012.672229

Heinrich, A., Flory, Y., & Hawkins, S. (2010). Influence of English r-resonances on intelligibility of speech in noise for native English and German listeners. *Speech Communication, 52*, 1038-1055.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity. *Memory and Cognition, 33*, 430-446.

Kemps, R., Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes, 20*(1/2), 43-73. doi:10.1080/01690960444000223

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 1969-1988. doi.org/10.1037/xhp0000269

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience, 25*(12), 2179-2188. doi:10.1162/jocn_a_00463

Kuznetsova A, Brockhoff PB and Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software*, **82**(13), pp. 1–26. doi:10.18637/jss.v082.i13.

Lakatos, P., Shah, A. S., Knuth, K., H, Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology, 94*, 1904-1911. doi:10.1152/jn.00263.2005

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron, 53*(2), 279-292. doi:10.1016/j.neuron.2006.12.011.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. *Science, 320*(5872), 110-113. doi:10.1126/science.1154735

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The Leading Sense: Supramodal Control of Neurophysiological Context by Attention. *Neuron, 64*(3), 419-430. doi:10.1016/j.neuron.2009.10.014

Lee-Kim, S. I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic/l. *Laboratory Phonology*, *4*(2), 475-511.

Lehiste, I. (1972). Manner of articulation, parallel processing, and the perception of duration, *Computer and Information Science Research Center Working Papers in Linguistics* (Vol. 12, pp. 33-52): Ohio State University.

Lenth, R.V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1), 1-33. doi:10.18637/jss.v069.i01

London, J. (2012). *Hearing in Time* (2nd ed.). Oxford: Oxford University Press.

Mackenzie, S., Olson, E., Clayards, M., & Wagner, M. (2018). North American /l/ both darkens and lightens depending on morphological constituency and segmental context. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 13. DOI: http://doi.org/10.5334/labphon.104

Mai, G., Minett, J. W., & Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage,* 133, 516-528. doi:10.1016/j.neuroimage.2016.02.064

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology, 38*(4), 465-494. doi:https://doi.org/10.1006/cogp.1999.0721.

Mattys, S., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543-562.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research, 32*(1), 77-97.

Meng, M., Remus, D. A., & Tong, F. (2005). Filling-in of visual phantoms in the human brain. *Nature Neuroscience*, 8(9), 1248-1254. doi:10.1038/nn1518

Morrill, T. H., Heffner, C. C., & Dilley, L. C. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review, 22*, 1451–1457. doi:10.3758/s13423-015-0820-9

Murray, S.O., Kersten, D., Olshausen, B.A., Schrater, P., & Woods, D.L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences,* 99(23), 15164–15169*.* doi:10.1073/pnas.192579399

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*(28), 10234-10240. DOI: 10.1523/JNEUROSCI.0411-11.2011

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*(49), 17572-17581. DOI: 10.1523/JNEUROSCI.3203-12.2012

Oh, G. E., & Redford, M. A. (2012). The production and phonetic representation of fake geminates in English. *Journal of Phonetics, 40*(1), 82-91. doi:http://dx.doi.org/10.1016/j.wocn.2011.08.003

Ogden, R. (2012). The phonetics of talk in interaction--Introduction to the Special Issue. *Language and Speech, 55*(1), 3-11.

Ogden, R., Hawkins, S., House, J., Huckvale, M., Local, J. K., Carter, P., Dankovicová, J., & Heid, S. (2000). ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Computer Speech and Language, 14*, 177-210.

Piccolino-Boniforti, M. A. (2014). *From sound to grammar: Theory, representations and a computational model.* Doctoral thesis, University of Cambridge. https://www.slideshare.net/MarcoPiccolinoBonifo/from-sound-to-grammar-theory-representations-and-a-computational-model

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech, 6*, 151-164.

Pierrehumbert, J. B. and Beckman, M. E. (1988), Japanese Tone Structure, MIT Press.

Philips-Silver, J., & Trainor, L. (2005). Feeling the beat: movement influences infant rhythm perception. *Science, 308*(5727), 1430.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics, 26*(2), 209-228.

Plag, I., & ben Hedia, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and the Role of Discourse-related Factors*. Berlin, New York: de Gruyter Mouton.

Plag, I., Homann, J., Kunter, G. (2015/2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53(1), 181–216. doi:10.1017/S0022226715000183

Post, B., D'Imperio, M., & Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning. In W. J. Barry & J. Trouvain (Eds.), *16th International Congress of Phonetic Sciences* (pp. 191-196, Paper ID 1723). Saarbrücken: http://www.icphs2007.de/

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic word boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51-89.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *TRENDS in Neurosciences, 32*(1), 9-18. doi:10.1016/j.tins.2008.09.012

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *TRENDS in Cognitive Sciences, 12*(3), 106-113. doi:10.1016/j.tics.2008.01.002

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience,* 33(1), 32-49.

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage,* 44, 1133-1143. doi: 10.1016/j.neuroimage.2008.09.045.

Simpson, A. P. (2005). 'From a grammatical angle': Congruence in Eileen Whitley's Phonology of English. *York Papers in Linguistics, 2*(4), 49-90.

Smith, R., Baker, R., & Hawkins, S. (2012). Phonetic detail that distinguishes prefixed from pseudo-prefixed words. *Journal of Phonetics*, 689-705.

Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics, 40*, 213-233.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*, 291-311.

Strycharczuk, P. (2019). Phonetic Detail and Phonetic Gradience in Morphological Processes. In *Oxford Research Encyclopedia of Linguistics*.

Strycharczuk, P., & Scobbie, J. (2016). Gradual or abrupt? The phonetic path to morpholo gisation. Journal of Phonetics, 59, 76–91.

Strycharczuk, P., & Scobbie, J. M. (2017). Whence the fuzziness? Morphological effects in interacting sound changes in Southern British English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8, 7.

Sugahara, M., & Turk, A. (2009). Durational correlates of sublexical constituent structure. *Phonology, 26*, 477-524.

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of/l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology,* 8.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41*, 262-275.

Zuraw, K. & Peperkamp, S. (2015). Aspiration and the gradient structure of English prefixed words. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow. Paper number 0382, 1-5.