HEARING *R*-SANDHI: THE ROLE OF PAST EXPERIENCE

Jennifer Hay
University of Canterbury

Katie Drager
University of Hawai'i at Mānoa

Andy Gibson
University of Canterbury

This article reports on patterns in the production and perception of New Zealand English *r*-sandhi. We report on two phoneme-monitoring experiments that examine whether listeners from three regions are sensitive to the distribution of *r*-presence in linking and intrusive environments. The results provide evidence that sound perception is affected by a listener's experience-driven expectations: greater prior experience with a sound in a given context increases the likelihood of perceiving the sound in that context, regardless of whether the sound is present in the stimulus. For listeners with extremely limited prior exposure to a variant, the variant is especially salient and we also observe an experiment-internal effect of experience. We argue that our results support models that incorporate both word-specific and abstract probabilistic representations.*

Keywords: salience, listener expectations, *r*-sandhi, sociophonetics, phoneme monitoring, speech perception, rhoticity

1. Introduction. Speech perception involves the extraction of words and sounds from a continuous speech signal. This normally occurs with little difficulty for the listener, despite a great deal of variability in the acoustic properties of sounds. Listeners' perception is informed by a large amount of information, including phonological context (Mann & Repp 1981), prior exposure in lexically relevant contexts (Norris et al. 2003, Maye et al. 2008), the lexical status of the word (Magnuson et al. 2003), the speaker who produced the sound (Kraljic & Samuel 2005, 2007), social characteristics attributed to the speaker (Strand & Johnson 1996, Hay et al. 2006), the listener's language background (Ingram & Park 1997), and the acoustic properties of a sound itself. Many of these disparate sources of influence might be understood as stemming from the listener's prior experience: if—based on the listener's experience—there is a high probability of perceiving sound X in a given environment or when produced in a certain way by a given speaker, then this boosts the likelihood that they will interpret an incoming signal as X. (See Davis & Johnsrude 2007 for a review of relevant literature.) If experience influences sound perception in this way, then we would expect variability in prior experience—across contexts and individual listeners—to influence the likelihood of hearing a sound accordingly.

We investigate the influence of language experience on speech perception by looking at a phenomenon that varies within and across speakers of the same language. Specifically, we look at the perception of the sociolinguistic variable *r*-sandhi, produced by a speaker of New Zealand English (NZE). *r*-sandhi has been found to vary in its rate of

* These studies were funded by a Royal Society of NZ Marsden Fund grant, a Rutherford Discovery Grant, a University of Canterbury internal grant to the first author, and a University of Canterbury Erskine Fellowship to the second author. We would like to thank Amalia Arvaniti for allowing us use of the Phonetics Lab at UCSD. The coding of *r* for the corpus work described in §2.2 was completed by Danae McConnel, Amy Wilson, Karen Malcolm, Aaron Nolan, and Abby Walker, with supervision contributed by Margaret Maclagan. The Intermediate Archive recordings were collected by Rosemary Goodyear, Lesley Evans, and members of the Origins of New Zealand English (ONZE) team. The Canterbury Corpus data was collected by members of the NZ English class of the Linguistics Department, University of Canterbury. Work done by the ONZE project team in preparing data, making transcripts, and obtaining background information for the corpus is also acknowledged. The recording and analysis of the small elicitation task reported in §2.2 was conducted by Andrew Lang. We are very grateful to all editors and referees who provided detailed feedback on this article, including Adam Albright, Megan Crowhurst, Greg Carlson, Jim Scobbie, and Natasha Warner. We would also like to thank Jacq Jones for their valuable editorial assistance.

production across regional backgrounds, social factors, and linguistic contexts. Our experiments test the extent to which listeners from three different dialect areas vary in their perception of *r*-sandhi across linguistic contexts, demonstrating that a listener's behavior reflects their previous experience.

To generate predictions for the perception study, we first report on results from a small production study. This is necessary because, while the production of *r*-sandhi has already been investigated for early NZE (Hay & Sudbury 2005), previous work did not analyze variation in the contemporary variety, so we cannot be sure that all of the patterns observed in previous work are indeed an accurate reflection of our listeners' experience with the variable. Therefore, in §2 we review the previous literature on the production of *r*-sandhi, and then present patterns from corpus data from nonrhotic speakers of NZE living in Canterbury (Cantabrians).

After a review of the literature on speech perception in §3, experiment 1 is presented in §4. This involves a phoneme-monitoring task with Cantabrian participants, allowing comparison with the production data reported in §2. The results from experiment 1 demonstrate that participants' perception patterns mirror the production patterns to a considerable degree.

In experiment 2 (§5), we add two new dimensions to the study. First, we vary the acoustic signal in order to investigate how prior experience might drive perception when [ɹ] is not acoustically present. Second, we introduce variability into the listeners' experience by recruiting participants from three regions, each with different *r*-sandhi production patterns. Finally, we summarize the findings from the two experiments (§6) and outline the implications of the results (§7).

2. Production of *r*-sandhi.

2.1. Background on production of *r*-sandhi. *r*-sandhi, which is composed of both 'linking-*r*' and 'intrusive-*r*', occurs in most nonrhotic varieties of English. Linking-*r* refers to the production of *r* across a word or morpheme boundary in contexts in which <r> is present in the spelling, is followed by a vowel, and would be present in rhotic varieties (e.g. *store again*, *storing*). Intrusive-*r* refers to the presence of *r* in analogous contexts that have neither an orthographic <r> nor *r* in rhotic varieties (e.g. *claw again*, *clawing*). The resulting system is a set of alternations between *r*-less forms preconsonantally (e.g. *car door*) and (variably) *r*-ful forms prevocalically (e.g. *car alarm*). The use of intrusive-*r* is phonologically restricted; in most dialects it occurs only after nonhigh monophthongs or after diphthongs with nonhigh offglides.[1]

The appropriate representation of *r*-sandhi has been vigorously debated in the phonological literature. Regardless of theoretical position, however, most formal phonological analyses assume sandhi-*r* to be categorically present across both word and morpheme boundaries, and in both linking and intrusive environments (see e.g. McCarthy 1993, Harris 1994, McMahon 2000, Orgun 2001, Uffman 2007). But quantitative work shows that realization of sandhi-*r* can be variable and socially stratified (e.g. Foulkes 1997, Hay & Maclagan 2010, Mompeán & Gómez 2011, Tan 2011, Pavlík 2016). Empirical data shows that speakers of at least some varieties treat linking- and intrusive-*r* differently (e.g. Mompeán & Gómez 2011), and that the trajectory of evolution differs across word and morpheme boundaries (Hay & Sudbury 2005). Work on British English has shown that sandhi-*r* is durationally shorter than canonical-*r*, and

[1] Note that in order to avoid implications of phonetic or phonemic status, we use the notation *r* throughout this article, rather than [ɹ] or /r/, except when explicitly referring to surface or abstract forms.

that this acoustic difference is used in speech perception by native speakers of that dialect (Tuinman et al. 2011, 2012).

The evolution of the NZE *r*-sandhi system is relatively well documented. Hay and Sudbury (2005) investigated early recordings of NZE, studying how the decline of rhoticity led to the emergence of the *r*-sandhi system for New Zealand speakers born in the late nineteenth century. They show that linking-*r* across morpheme boundaries (e.g. *storing*) remained robustly present during the transition from rhoticity to *r*-sandhi. In contrast, linking-*r* across word boundaries declined with the loss of rhoticity. Intrusive-*r* across word boundaries first appeared while speakers of NZE were still partially rhotic, and then it gradually increased in frequency over time. Intrusive-*r* across morpheme boundaries was a later development.

Hay and Maclagan (2012) investigated corpus data during the subsequent period (speakers born 1900–1935) and show that the system was relatively stable during this time. Across word boundaries, levels of linking-*r* and intrusive-*r* were maintained, with linking-*r* still much more frequent than intrusive-*r*. Although the amount of data available is small, they suggest that intrusive-*r* across morpheme boundaries had increased by this period. Furthermore, they demonstrate lexical differences; words that very commonly occur before vowels (and thus have more opportunity for linking-*r* to surface) were more commonly realized with *r* when they were prevocalic. Taken together, the New Zealand data shows a system in which both types of *r*-sandhi occur with different overall frequencies and with different frequencies across word and morpheme boundaries.

Limited contemporary data on NZE *r*-sandhi has been reported. Hay and Maclagan (2010) conducted a simple reading task with tokens containing potential sites for intrusive-*r*. They found that more *r* was produced by males and by speakers from lower socioeconomic backgrounds, though the extent to which these patterns are found in more natural contexts is unknown. More recently, Gibson (2016) argued that low rates of *r*-sandhi index a youth vernacular style for Pasifika New Zealanders.

2.2. Production of *r*-sandhi in Canterbury. To determine which patterns from early NZE are still observed in NZE today, we provide a small analysis of linking- and intrusive-*r* usage in (nonrhotic) NZE in two data sets: corpus-based spontaneous speech and speech from a reading passage.

For the corpus analysis we combined data from the Canterbury Corpus and the Intermediate Archive of the Origins of New Zealand English (ONZE) project (Gordon et al. 2007). Included in this data set is the data reported by Hay and Maclagan (2012). In total, we analyzed the interviews of 105 speakers born between 1900 and 1978. While we did not explicitly check for rhoticity, the overwhelming majority of speakers in these corpora are nonrhotic.

We limit our analysis to words containing one of the three pre-*r*-sandhi vowels (schwa, THOUGHT, and START[2]) used in our perception experiments. A total of 2,216 such tokens, where *r*-sandhi could occur, were analyzed. Most of the tokens analyzed represent linking environments at a word boundary (n = 2,000). Other tokens are rare: there are 158 tokens of intrusive environments at a word boundary, forty-five tokens of linking environments at a morpheme boundary, and only thirteen tokens of intrusive environments at a morpheme boundary. This distribution is consistent with previous work (see Hay & Sudbury 2005). The environment for intrusion is rare, with a low number of potential types and a relatively low token frequency for these eligible types.

[2] For ease of reference, we use Wells's (1982) lexical set words to refer to particular phonemes.

For each token, two independent coders provided an auditory analysis. For the statistical analysis, a token was treated as containing *r* if both coders independently heard [ɹ]. Tokens coded as containing some other boundary consonant were excluded. By this criterion, 65% of tokens contained *r*.

We fit a mixed-effects logistic regression model to this data set, investigating potential effects of speaker age (birth date—scaled and centered) and gender, as well as the identity of the previous vowel, whether the environment was linking or intrusive (link/intru), and whether it was a word or morpheme boundary (word/morph). Two-way interactions were also explored. Model selection proceeded via ANOVA comparison.[3] The best model of the corpus data is shown in Table 1. Note that preceding vowel does not reach significance in this model.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	0.806	0.431	1.869	0.062
gender = male	0.401	0.095	4.241	0.000
word/morph = word	−1.277	0.404	−3.157	0.002
birthdate	0.563	0.171	3.289	0.001
link/intru = linking	0.924	0.217	4.268	0.000
birthdate : link/intru = linking	−0.607	0.179	−3.395	0.001

TABLE 1. Model A: logistic regression model of corpus data. Included in the model is whether *r*-sandhi occurs across a word or morpheme boundary (word/morph) and whether it is linking- or intrusive-*r* (link/intru).

The model reveals a main effect of gender: males produce more tokens with *r* than females. This supports the findings reported in Hay & Maclagan 2010 in which males produced more intrusive-*r*. Hay and Sudbury (2005) also reported that males used more linking-*r*, even in the very earliest stages of the emergence of *r*-sandhi in NZE.

There is also a significant effect of the morphological boundary type, with morpheme boundaries having more instances of *r* than word boundaries. Finally, linking-*r* is significantly more likely to occur than intrusive-*r*, but this interacts with speaker age: while older speakers are more likely to produce linking-*r* than intrusive-*r*, the difference gradually disappears, with younger speakers equally likely to produce linking- and intrusive-*r*. This interaction is shown in Figure 1.

The intrusive-*r* data available in the corpus is relatively sparse. We therefore double-checked the distribution of intrusive-*r* across word and morpheme boundaries by conducting a small reading study in Christchurch with eighteen nonrhotic participants (ten men, eight women). Ten participants were younger (twenty to thirty-nine years old) and eight were older (forty-five to sixty years old), ages that are comparable to those of the Canterbury Corpus speakers analyzed.

The analysis of this data set reveals trends for gender and age that are consistent with the corpus data. The model also verifies the robustness of the morpheme-boundary ver-

[3] The random-effects structure differs across the models reported in this article due to nonconvergence of some models, and it is described in footnotes along with any other models tested. For the production model, random intercepts were attempted for item and speaker. When both of these intercepts are included, the model does not converge due to the small number of observations per speaker and per word. When either one is included individually, the model converges and is very similar in both cases. The item intercept was included in the reported model because it accounts for more variance than the speaker intercept. Random slopes led to nonconvergent models and were therefore rejected.

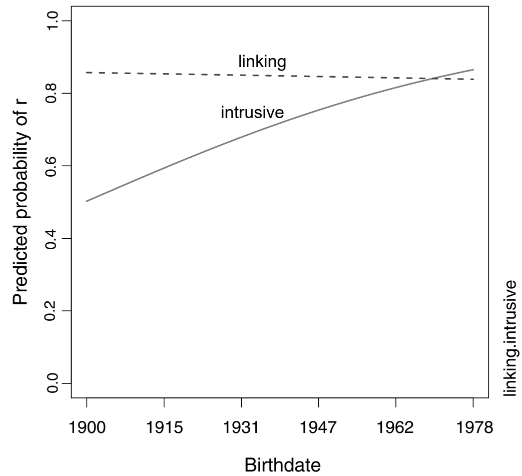

FIGURE 1. The interaction between age and linking/intrusive-*r*, influencing the predicted probability of producing *r* in the corpus data (from model A). Birthdate was scaled and centered in the statistical model, but values have been back-transformed here for interpretation.

sus word-boundary effect: *r*-sandhi is less likely to be realized across word boundaries than across morpheme boundaries (Est. −2.144, $z = -3.365$, $p < 0.001$).[4]

These two small data sets, along with the previously published work outlined above, allow us to make the following statements about the *r*-sandhi system in NZE.

- Rates of linking-*r* were historically higher than intrusive-*r*, but the rates have converged over time.
- Rates of *r*-sandhi are higher at morpheme boundaries than word boundaries.
- Rates of *r*-sandhi appear to be greater for men than women.
- There is no evidence that the preceding vowel robustly predicts the rate of *r*-sandhi.

Historically speaking, rates of linking-*r* were higher than rates of intrusive-*r* (Hay & Sudbury 2005). The difference started to disappear at the end of the nineteenth century, and our data reveals that this was a gradual process that is approaching completion. Phonologically, there is nothing that distinguishes 'linking' from 'intrusive' words. The difference is carried by the words' histories, but—over a 100-year period—it has leveled out. This is a relatively recent phenomenon, so participants in the perception experiments are still likely to have more experience with *r* in linking than intrusive contexts even if they do not have different rates of *r* across the word sets in their own speech.

With these production patterns in mind, we turn now to speech perception, reviewing key issues in the literature in order to provide a background for our perception experiments, which test the degree to which differing experience with *r* across contexts and speakers might lead to different patterns of perception.

3. BACKGROUND ON PERCEPTION OF *r*-SANDHI. Speech perception involves matching an incoming auditory stimulus to cognitive memories and representations in order to identify the nature of the incoming message. While there are many models of speech perception, and many unresolved issues, the literature is broadly converging on a num-

[4] The trends with speaker gender and age are significant only in a model with no speaker random intercept, likely due to small speaker numbers. The difference between morpheme and word boundaries is observed whether or not random intercepts for speaker and word are included.

ber of key points. Specifically, there is a growing consensus that mental representations of linguistic variables include both episodic and abstract representations (though the exact nature of the link between them is still not well understood), and that experience informs perception probabilistically.

3.1. REPRESENTATIONS AND PROBABILISTIC PERCEPTION. Listeners use a vast amount of experience-driven probabilistic information in order to predict the nature of the incoming signal. Speech perception has been shown to be influenced by the probability of encountering a sound given the phonotactics of the language (Vitevitch et al. 1997) and given the following sound across a word boundary (Mitterer & McQueen 2009). Recent work shows how individuals constantly predict the upcoming signal, updating their expectations when their predictions produce errors (see e.g. Kleinschmidt & Jaeger 2015). Further, lexical access by different groups of listeners can be primed to different degrees depending on a combination of their experience with and production of the target sound (Sumner & Samuel 2009).

Because experience with a variable differs across listeners, it follows that listeners should behave differently from one another and this behavior should be somewhat predictable based on their prior experience. Furthermore, in cases where social and linguistic factors interact in production, there should be corresponding effects in perception.

The unit of representation considered to be most important varies across different research traditions. A large body of research illustrates the role of abstracted units, such as phonemes, in production, perception, and grammatical knowledge (see e.g. Labov 1994:Ch. 18, Eulitz & Lahiri 2004, Cutler et al. 2010).

A second body of literature shows that individuals have detailed memories of past speech experiences, and that many aspects of speech production and perception can be well understood as reflecting episodic memory of past experiences with particular words (e.g. Goldinger 1997, Walker & Hay 2011, Hay & Foulkes 2016). In recent work across both literatures, there is an emerging consensus that both episodic and abstract representations are necessary (Pierrehumbert 2002, 2006, 2016, Goldinger 2007, Cutler 2012, Ernestus 2014), although exactly how they work together remains unclear.

We assume that during speech perception, both abstract representations and episodic memories of words are activated. Words (and in some cases frequent phrases) receive activation in proportion with the degree to which the acoustic signal matches past socially and contextually relevant experiences with the word. Phonemes receive activation in proportion to the degree of acoustic match in the signal and the degree of likelihood of occurrence of the phoneme in context. These representations are linked, and feedback flows between them. The role that each level plays in perception varies across different tasks and contexts.

3.2. THE ROLE OF NOVELTY AND ATTENTION. Attention is also known to play a role, conditioning the effects of probabilistic information on phone detection. For example, in a phoneme-monitoring task that compared responses to nonword stimuli with a close neighbor (e.g. *vocabutary*) to nonword stimuli with no close neighbor (e.g. *socubatory*), the effect of probabilistic information from the lexicon on phone identification was found only when attention was drawn away from contexts in which the phone was most likely to occur during the experiment (Wurm & Samuel 1997).

The increased attention that results from a lack of experience with a variable is commonly called 'salience'.[5] In this article, we are concerned with the salience of a sound

[5] The term *salience* is used in various ways and, in the sociolinguistic literature, is often conflated with attention more generally.

that stands out due to its unexpectedness. This concept of salience relates to top-down (expectation-driven) attentional control, which is well studied in visual processing (e.g. Summerfield & Egner 2009, Awh et al. 2012), and contrasts with bottom-up (stimulus-driven) attentional control (e.g. the salience of a loud sound). In work on top-down attentional control, attention is drawn to items that are unexpected given previous exposure and knowledge. For sound perception, top-down influences include words and word sequences and also social information attributed to the speaker and conversational context. Thus, when listeners encounter a phonetic realization that does not match their expectations, they can either (i) dismiss it as a mistake or an idiosyncrasy of the speaker (in which case it may receive little attention) or (ii) attend to it and store it with greater weight so that subsequent encounters with similar realizations result in greater activation.

Work on the processing of nonnative languages has argued for a 'novel pop-out' effect, in which phonotactic sequences that do not occur in one's native language 'pop out' during processing of an L2 (Weber 2001, 2002). The term is borrowed from work on visual processing which shows that novel items in a scene draw visual attention (cf. Johnston & Schwarting 1997). A phonetic analogue of this is reported by Chang (2013), who found that novel voice onset time (VOT) lengths influence productions of an L1 more for early L2 learners than for more advanced learners, and that later learners become accustomed to the novel VOT, after which the L2 VOT patterns have less effect on the L1. Chang argues that the findings are due to 'novelty bias'. The well-attested effect of novelty bias involves stronger attention to, and encoding of, novel items (cf. Tulving et al. 1996). Recent work argues that this type of 'surprisal' upon encountering novel variants is one of the factors leading to sociolinguistic salience (Jaeger & Weatherholtz 2016).

The pop-out effect suggests that, while we expect a general positive correlation between experience and perception, this correlation may break down at the edges, where items that are completely outside of the listener's experience may 'pop out' and become salient. In terms of how the pop-out effect might influence a rhotic listener's perception of *r*-sandhi, Weber would not predict salience of an *r* in something like *clawing* because *r*-sandhi is phonotactically well formed. In contrast (and in line with Chang 2013), we predict that novelty bias will apply because it is a novel form for that lexical item. Furthermore, while the literature on novelty generally focuses on the novel PRESENCE of an unexpected segment, we predict that its novel ABSENCE can also be salient in certain circumstances. Thus, the unexpected absence of *r* in a word like *car* may also emerge as salient for a rhotic speaker who has little experience with nonrhotic speech.

3.3. PHONEME-MONITORING IN THE CONTEXT OF VARIABILITY. The remainder of this article focuses on two phoneme-monitoring experiments in which participants decide whether or not they hear *r* in a series of short spoken phrases, with or without *r*-sandhi contexts. In these experiments, we use a generalized phoneme-monitoring task: participants are asked to listen for a prespecified sound (in this case *r*) that can occur in any position in the utterance. Varying the position of the sound increases the chances that lexical information (and, presumably, other contextual information) will affect prelexical access (Wurm & Samuel 1997). While nonsense-word stimuli are often used in phoneme-monitoring tasks (e.g. Warner & Weber 2001), real-word stimuli can also be used (see e.g. Cho & McQueen 2011). While phoneme-monitoring tasks often focus on an analysis of reaction times (see overview in Connine & Titone 1996), we limit our analysis to the responses.

Our participants complete a short training task in which they are encouraged to respond to sounds rather than spelling, then complete the phoneme-monitoring experiment in which they monitor for *r*. Memory tests occur at various points to encourage processing of word meaning as well as phoneme monitoring. The test stimuli involve positions in which *r*-sandhi may variably surface. In our first experiment, [ɹ] is categorically present in all sandhi contexts where it could appear. In the second experiment, its presence varies. We test the perception of the *r*, across contexts and for listener groups with differing experience with *r*.

Figure 2 provides a sketch of our predictions for phoneme monitoring. On the left are the general predictions, with the probability of hearing a target sound (X) on the y-axis and the listener's prior experience with that sound on the x-axis. We have different predictions for when the sound is present in the stimulus (shown in black) and when it is absent (shown in gray). For both stimulus types, we predict that greater experience with the sound in a given context will result in a higher rate of hearing that sound when encountered in that context. Evidence from phonemic-restoration experiments shows that listeners will report 'hearing' a phoneme that is not present in the acoustics, with the likelihood of restoration related to the listener's expectation of the phoneme occurring in the encountered environment (Samuel 1981). Thus, phonetically deleted-but-expected material can be reported as heard (Cho & McQueen 2011). Because of the novelty bias, however, we also predict that for participants who have little experience with either the sound's presence (shown by the striped box on the left) or the sound's absence (shown by the striped box on the right), its unexpected presence or absence will be salient. Salience of the sound's presence or absence should result in more accurate identification rates.

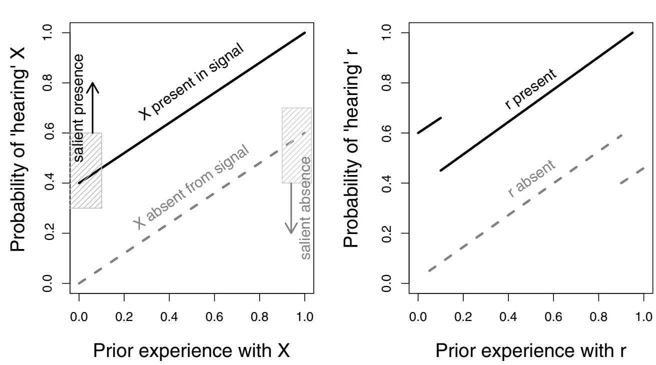

FIGURE 2. Schematic predictions about the relationship between experience and perception, in general (left) and for *r*-sandhi (right). Greater exposure in a given context will increase the likelihood of hearing a variant in that context, regardless of whether it is acoustically present in the signal. However, effects of salience can arise in cases where there is either very little or extensive experience with the variant in that context.

The sketch on the right in Fig. 2 shows these predictions within the context of *r*-sandhi. We expect rates of hearing *r* to be greater for contexts where the *r* is most often experienced. This should differ across participants, depending on the degree to which they have experienced the sound in each of the contexts. Furthermore, for participants having no experience with intrusive-*r*, the presence of [ɹ] in intrusive contexts should be salient and therefore heard at higher rates. Likewise, participants having little experience with the absence of [ɹ] in linking-*r* contexts should be more likely to accurately identify its absence than participants who have experience with its absence.

Our OVERARCHING HYPOTHESIS is as follows:

- OVERARCHING HYPOTHESIS: Listeners' perception of sandhi-*r* will reflect their past experience with *r*. Increased prior experience will lead to increased expectation of *r*, except in the salient boundary zones, as per Fig. 2.

There are two targeted SPECIFIC HYPOTHESES that our experiments were designed to test:

- SPECIFIC HYPOTHESIS 1: *r* will be differently heard in different linguistic contexts, in proportion to the patterns of usage in those contexts.
- SPECIFIC HYPOTHESIS 2: *r* will be differently heard by listeners from different regional backgrounds, reflecting their differing previous experience with *r*.

Experiments 1 (§4) and 2 (§5) were explicitly designed to test specific hypotheses 1 and 2, respectively.

In post-hoc analysis of experiment 2, we also pursue several RELATED PREDICTIONS that follow from the overarching hypothesis:

- RELATED PREDICTION (i): For listeners who are exposed to highly novel variants during the experiment, this new experience may affect perception patterns over the course of the experiment itself (§5.6).
- RELATED PREDICTION (ii): To the extent that speakers from a single dialect region vary in their production of *r*, variation in perception may reflect these individual production patterns (§5.7).
- RELATED PREDICTION (iii): To the extent that lexical items vary in *r*-production, word-specific effects in perception may reflect this variation (§5.8).

4. EXPERIMENT 1. Experiment 1 tests the degree to which listeners perceive both linking- and intrusive-*r* when [ɹ] is invariably present acoustically in the test stimuli. All participants were from nonrhotic areas of New Zealand. Thus, the hypotheses explored in experiment 1 focus on the middle region of the top line in Fig. 2, with [ɹ] always present in the signal and all listeners having some experience with *r*-sandhi. We expect a positive correlation between variable experience (across linguistic contexts) and rate of hearing *r*, and there is no case in which we expect a boost in perception due to salience.

4.1. PARTICIPANTS. Experiment 1 was conducted at the University of Canterbury in Christchurch, NZ. A total of nineteen participants took part, all from nonrhotic parts of New Zealand. They had a median age of twenty-four (range: eighteen to sixty-six), eight male and eleven female.

4.2. STIMULI. All stimuli in both experiments were natural speech produced by the third author, a New Zealand-born male who was twenty-four years old at the time of recording. Stimuli were recorded in a quiet room at the University of Canterbury using Soundforge software, a head-mounted condenser microphone, and a USBPre interface.

First attempts at recording stimuli produced considerably higher F3 for [ɹ] in intrusive than in linking environments. The materials were rerecorded in order to make all instances of [ɹ] comparable. This methodological decision has potential consequences, as the intrusive-*r* tokens may stand out as containing stronger articulations than expected. Since this is contrary to our hypothesis (that intrusive-*r* will be heard less often than linking-*r*), we decided that having comparable realizations of linking- and intrusive-*r* was the preferable option.

All final test stimuli had F3 values under 1900 Hz at their lowest point; the mean F3 was 1798 Hz. Unpaired Wilcoxon tests revealed no significant differences between selected linking versus intrusive tokens, or word- versus morpheme-boundary tokens, in either F3 or the duration of the [ɹ]. To determine whether minor variations in F3 or duration affected results, F3 and duration values were also tested in the statistical models fit to the response data.

The stimuli varied in terms of (i) whether the *r* was linking or intrusive (link/intru), (ii) whether the *r* occurred at a word boundary or word-internally at a morpheme boundary (word/morph), and (iii) the vowel preceding the *r* (vowel). All tokens where *r* occurred at a morpheme boundary were affixed forms of corresponding word-boundary stimuli, followed by *-ing*. The preceding vowels were THOUGHT, START, and schwa.

In experiment 1, half of the tokens that contained schwa were nouns (e.g. *comma*) and half were verbs (e.g. *holler*). As no clear difference in responses was observed (nor was any hypothesized), this distinction was not tested statistically. Only nouns were used for schwa tokens in experiment 2. Example stimuli are shown in 1a–f, labeled according to vowel, word/morph, and link/intru. A full list of stimuli for each experiment is provided in the appendix.

(1) a. Utah and Nevada (START, word, intrusive)
b. starring and leading (START, morpheme, linking)
c. chainsawing and cutting (THOUGHT, morpheme, intrusive)
d. snoring and sleeping (THOUGHT, morpheme, linking)
e. soda and lime (schwa, word, intrusive)
f. clamber and climb (schwa, word, linking)

The stimuli included 330 short phrases: 110 test phrases and 220 filler phrases. Test-phrase types were distributed as shown in Table 2. Some cells, such as intrusive-*r* at a morpheme boundary following schwa, could not be filled because relevant environments are especially rare. However, the distribution of the stimuli allows comparisons of subsets of the data in order to establish the role of vowel, link/intru, and word/morph on *r*-perception. Due to the rarity of appropriate examples, the stimuli are not matched for frequency, and some of the items (e.g. *la-de-da and posh*) may be unlikely to occur in natural speech.

	LINKING		INTRUSIVE	
	WORD	MORPHEME	WORD	MORPHEME
THOUGHT	10	10	10	10
START	10	10	10	0
schwa (noun)	10	0	10	0
schwa (verb)	10	10	0	0
total	*40*	*30*	*30*	*10*

TABLE 2. Distribution of the test-phrase types used in experiment 1 across different preceding vowel environments. The phrase types are linking- and intrusive-*r* across a word and morpheme boundary.

In addition to the test phrases, filler phrases were used to distract participants from focusing on the contexts where linking- or intrusive-*r* could occur and to ensure that—regardless of the participants' responses to the *r*-sandhi stimuli—a reasonable number of 'yes' and 'no' responses would occur throughout the experiment. They were also included so that an effect of spelling (whether or not the word had an orthographic <r>) could be tested; if participants had high rates of misidentifying a (nonrhotic) filler token such as *cavern and cave* as having an *r*, this could indicate reliance on spelling rather than auditory cues.

Half of the fillers included a nonsandhi prevocalic *r* (i.e. an *r* that would be categorically present in most dialects, e.g. *read and listen*), and half did not. The phrases with *r* were balanced so that the *r* occurred in either the first or second word of the phrase and in word-initial or word-medial position. For the other filler phrases, some had no <r> represented in the orthography (e.g. *talk and listen*), and some had an orthographic <r> that would only be realized in a rhotic dialect (e.g. *cart and bus*). The distribution of filler types is shown in Table 3.

FILLER TYPE	
orthographic and realized (e.g. *read and listen*)	110
only orthographic (e.g. *cart and bus*)	40
not orthographic nor realized (e.g. *talk and listen*)	70
total	*220*

TABLE 3. Distribution of the filler-phrase types used in experiment 1.

4.3. PROCEDURE. Participants listened to auditory stimuli over headphones. They were asked to indicate whether or not they heard the sound [ɹ] anywhere in the phrase by pressing keys marked ‘yes’ and ‘no’ on the keyboard. The experiment was conducted in a quiet room with only the experimenter present and was run on a laptop using DirectRT (Jarvis 2006).

Prerecorded instructions were presented using the same voice as was used for the experimental stimuli. The reason for auditory presentation of instructions was to highlight the importance of sounds over spelling in the task, by referring to them phonetically, rather than by their letter names. For example, the instructions state: ‘Remember, it is the SOUND [ɹː] that you are listening for. Whether or not the phrase contains the letter <r> should be irrelevant to your response’. Key instructions, such as which buttons to press, were displayed on the screen as the words were spoken.

The instructions were followed by three short training sessions during which participants received feedback on their responses. In training sessions 1 and 2, participants were asked to monitor for [s] and [f], respectively. The training materials contrasted orthography and phonology to help participants focus on the auditory stimulus and not their knowledge of spelling. When a participant answered a training question incorrectly, the correct answer was given and the sound file was replayed twice. For certain phrases containing a spelling/phonology mismatch, a recorded message explained the mistake. For example, if a participant indicated that the phrase *water and juice* did not contain the sound [s], the recording explained that it did: ‘It is at the end of the word *juice*, even though the spelling contains the letter <c>’. In the third training session, participants were asked to monitor for [ɹ], and they were informed that this was the sound they would be listening for until the end of the experiment. This final training session included initial *r*, as well as specific feedback regarding nonrhotic pronunciations, but no *r*-sandhi examples were included. If a participant responded incorrectly that they heard an *r* in the phrase *north and south*, they received the following feedback: ‘The pronunciation of the phrase *north and south* that you just heard did not contain the sound [ɹː] even though there is an <r> in the spelling. Listen again … ’. Despite our efforts, it is possible that participants were influenced by spelling. We explore this further when presenting the results in §4.5.

The 330 trials of the main experiment were presented in three blocks of ten to fifteen minutes, with a short break between blocks. Responses were speeded: if a participant took longer than one second to respond, a message flashed on the screen asking them to respond more quickly. The stimuli were presented in pseudo-random order, counterbalanced within each block to ensure even distribution of linking- and intrusive-*r* across the different contexts (word/morph) in each part of the experiment. Tokens of the same lexeme in different morphological contexts (e.g. *gnaw and bite* and *gnawing and biting*) always occurred in separate blocks.

To encourage full processing of the phrases, participants were advised during the instructions that the experiment included memory tasks at irregular intervals. During these tasks, participants identified which of three words listed on the screen they had heard during the experiment.

Each participant's own production was tested by recording them reading select phrases from the experiment after completion of the perception task. Unfortunately, due to the list-like nature of the task, almost all participants paused after the target word (e.g. *tuna—and cod*), failing to produce appropriate environments for *r*-sandhi to occur. Therefore, this production data was not analyzed.

4.4. EXPERIMENT 1 PREDICTIONS. Our prediction was that participants' experience-based expectations regarding the presence of an *r* in a given environment would predict accuracy. In this experiment, we focus on participants' experience with differing rates of *r* across different environments, testing specific hypothesis 1: *r* will be differently heard in different linguistic contexts, in proportion to the patterns of usage in those contexts.

Three aspects of the stimuli were manipulated: (i) morpheme boundary versus word boundary, (ii) linking versus intrusive environment, and (iii) previous vowel. Regarding (i), the production data in §2.2 shows that morpheme-boundary environments have higher rates of *r* than word-boundary environments. We therefore predict that listeners from the same region as the production data will more accurately identify *r* at morpheme boundaries than word boundaries.

Regarding (ii), there is a historical difference whereby linking environments attracted more *r* than intrusive environments (Hay & Sudbury 2005), although *r*-rates in these environments have recently converged (see §2.2). We assume that participants' expectations in this task are based not only on their own production but also on exposure to other relevant speakers. Most participants should still have encountered more *r* in linking than intrusive contexts. We therefore have a second, somewhat weaker, prediction that linking environments will be responded to more accurately than intrusive environments.

Regarding (iii), there is no strong evidence for robust differences in the frequency of *r* after different vowels. While this contrasts with Hay and Sudbury's (2005) early data, their data is scarce in some of the relevant cells. Therefore, we do not expect strong vowel effects in perception.

4.5. RESULTS. Trends in the data confirm the predictions outlined above. Participants' mean proportions of 'yes' responses to *r*-sandhi tokens in different environments are shown in a bean plot in Figure 3. Bean plots show data density on the y-axis, providing more information about variability than boxplots. The left panel shows the distribution of participant means by vowel type (including only trials with *r* over word boundaries, since intrusive-*r* across morpheme boundaries occurs only in stimuli with THOUGHT). Two observations are revealed by this plot. First, response patterns vary greatly across participants. Second, there is relatively little variation across vowels. The plot on the right shows responses to trials with the THOUGHT vowel (for which we have data across all boundary types). This plot shows a much higher rate of 'yes' responses to linking-*r* than intrusive-*r* tokens, and a tendency to have higher rates of 'yes' responses to morpheme than word boundaries.

Analysis of individual variation showed that every participant was more likely to indicate hearing linking-*r* than intrusive-*r*, with one participant near-categorically responding 'yes' for linking and 'no' for intrusive trials. As this was likely a spelling

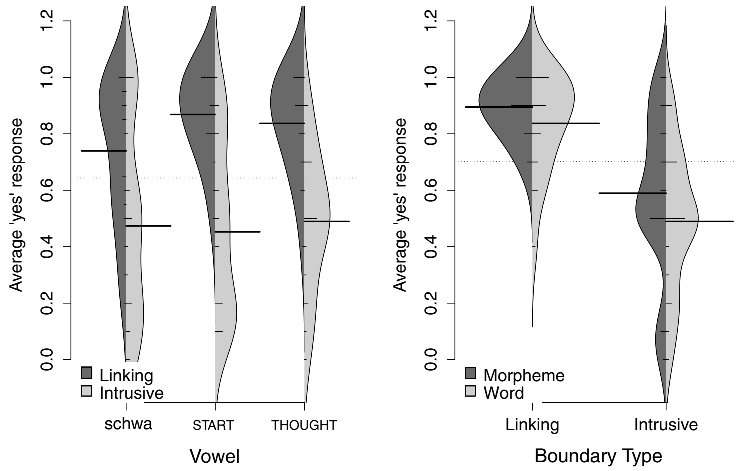

FIGURE 3. Distribution of participant means for the proportion of test items in experiment 1 identified as containing an *r*. Left panel: responses to *r* in linking (dark gray) and intrusive (light gray) environments across vowel contexts. This includes word-boundary data only (morpheme-boundary data is not represented across all vowels). Right panel: responses to *r* across morpheme (dark gray) and word (light gray) boundaries across linking- and intrusive-*r* environments. This includes only responses to words with the THOUGHT vowel (the only vowel that varies in both dimensions). The shape of the bean shows the distribution of the data. Lines represent means: for each participant (short lines), within each factor group (long lines), and overall (dotted lines).

strategy, this participant was removed before statistical testing. Additionally, most participants tended to indicate hearing a greater number of tokens as containing *r* when it occurred across morpheme boundaries than across word boundaries.

A binomial mixed-effects model was fit to the word-boundary data to test the significance of the apparent difference between linking- and intrusive-*r*. It modeled the likelihood of correctly identifying a token as containing an *r* (only test tokens were analyzed, so all tokens contained [ɹ]). Tested as fixed effects were vowel (START, THOUGHT, or schwa), link/intru (linking- versus intrusive-*r*), and the interaction between these factors.[6]

In addition to link/intru and vowel, we also tested the F3 target and the duration of the [ɹ] in order to determine whether acoustic characteristics of the [ɹ] across different stimuli were driving the observed effect. Additionally, we tested participant age and gender, and an interaction between participant age and link/intru.

Because appropriate lexical items for some of our categories are relatively rare, some stimuli are ambiguous with respect to type (e.g. *thaw* could be heard as *Thor*). For these items, the competitor (*Thor*) tends to be less frequent than the intended target (*thaw*), and the target interpretation is always later confirmed by the context (*thaw and melt*). However, the competitor may still have had an effect. We therefore fit this model on two data sets: the complete data set and a more conservative data set that excluded responses to potentially ambiguous stimuli (six intrusive word items, five linking word items, and two linking morpheme items). Vowel surfaced as significant in the model fit to the full data set, but this significance disappeared when potentially ambiguous tokens (which were not equally distributed across vowels) were removed. Based on this differ-

[6] Participant and item were included as random effects, and random by-participant slopes were included for vowel and link/intru.

ence, we opted to report the conservative models in this article, using only the definitively nonambiguous tokens. All significant effects reported are also significant when fit to the full data set.

In the conservative data set of word-boundary tokens, link/intru was the only factor that reached significance. Linking-*r* was heard significantly more often than intrusive-*r* (Est. = 1.9162, $SE = 0.2797$, $z = 6.850$, $p < 0.0001$). Neither the F3 nor duration of the [ɹ] predicted responses significantly. This supports the interpretation that differences in perception are related to participants' previous exposure rather than inherent phonetic cues in the stimuli. No social factors emerged as significant, and neither did vowel.

To explore an effect of word/morph, we next fit a model to the subset of data where *r* was preceded by THOUGHT—the only vocalic environment balanced for whether the *r* occurred at a word or morpheme boundary. The output of the model is shown in Table 4. Again, neither the F3 nor the duration of [ɹ] in the stimuli was predictive. The model shows that linking-*r* is heard more than intrusive-*r*, and that *r* is heard less at word boundaries than morpheme boundaries.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	0.654	0.394	1.659	0.097
word/morph = word	−0.700	0.295	−2.372	0.018
link/intru = linking	1.910	0.331	5.777	0.000

TABLE 4. Model B: regression model of responses in experiment 1, fit only to tokens containing the THOUGHT vowel.

The significant difference between word- and morpheme-boundary *r* is well predicted by the patterns of *r*-sandhi to which participants have been previously exposed. For linking- and intrusive-*r* forms, however, there is no difference in the current usage, nor any phonological differences distinguishing these types. Why, then, do participants have different rates of perceiving *r* across these two classes?

One possibility is that participants have access to individual word histories. The difference between linking- and intrusive-*r* exists at the word level and arises from the path of historical evolution of the *r*-sandhi system in NZE (Hay & Sudbury 2005). In individuals' past experience, words with a linking-*r* context have been experienced more frequently with *r* than words with an intrusive-*r* context (see §2.2). Thus, even without access to the identity of two distinct classes of words (i.e. linking and intrusive), this difference emerges if participants are sensitive to word-level probabilities.

The second possibility is that participants group linking and intrusive words together at a more abstract level, possibly bootstrapped by orthography and/or an observed bimodal distribution in *r*-production rates. Under this interpretation, speakers can classify individual words as either linking or intrusive, and in our experiment, participants may hear a word, classify it as belonging to the linking class, and therefore have a heightened expectation of *r*. For contemporary New Zealanders, this abstract division into two classes seems somewhat unlikely given the results from §2.2, which indicate that differences between the two classes are not being maintained in production.

The final explanation is that the observed difference is merely an artifact of an orthographic response strategy. Indeed, as outlined above, one participant appeared to be extremely reliant on spelling and was therefore excluded from our analysis. We assessed this orthographic-response interpretation by considering response patterns to the fillers. As would be expected, participants rarely responded 'yes' to filler tokens with neither phonological nor orthographic *r*, and almost always responded 'yes' to tokens with nonsandhi prevocalic *r*. However, the rate of response to the (nonrhotic) preconsonantal tokens (e.g. *cart*) varies greatly. Recall that participants were explicitly trained on such tokens and, if they responded 'yes', were given an explanation explicitly contrasting spelling with pronunciation. Despite this, some participants had high rates of 'yes' responses to such tokens during the task.

The high rate of 'yes' responses may indicate a spelling strategy, but it could also reflect participants' previous exposure to such words from a wide range of speakers. The listener's mental representation of *cart*, for example, should contain *r* to the degree that the individual has encountered the word produced by rhotic speakers.

To test whether the significant difference in perception between linking- and intrusive-*r* was an artifact of participants' reliance on spelling, an individual's rate of 'yes' responses to fillers with preconsonantal <r> in spelling was added as an additional factor in the models reported above. This rate did not significantly predict responses on the task when included in either of the models reported above ($p > 0.6$ for both). Nor did it interact significantly with whether the token contained linking- or intrusive-*r*. The lack of this relationship provides evidence that the difference in responses between linking- and intrusive-*r* cannot be explained entirely as an effect of spelling.

Of course, it is possible that spelling contributes to differing behavior across linking and intrusive contexts in some way that is disjunct from the filler items. However, it cannot contribute to the difference in responses across boundary types; within both linking- and intrusive-*r* contexts, most participants accurately identified the *r* across a morpheme boundary more often than across a word boundary, despite there being no orthographic difference between these boundary types. The response patterns overall are therefore most straightforwardly interpreted as related to the different patterns of exposure to *r* in these contexts. This analysis is consistent with our overarching hypothesis and with our specific hypothesis 1: that *r* will be differently heard in different linguistic contexts, in proportion to the patterns of usage in those contexts.

4.6. Experiment 1 summary. In experiment 1 we have shown that for nonrhotic listeners in Canterbury, their phoneme-monitoring patterns broadly reflected the likelihood of *r* occurring across different linguistic contexts. This provides some preliminary evidence in favor of the hypothesis that, in the context of variability, increased contextual exposure should lead to increased expectation. The results thus lend support to our overarching hypothesis by providing evidence for specific hypothesis 1.

In experiment 2, we broaden the domain of inquiry by including stimuli where [ɹ] is not present, and we test specific hypothesis 2 by including listener groups with a wider range of previous experience: Southlanders, who produce less intrusive-*r* than Cantabrians, and San Diegans, who are rhotic and do not have an *r*-sandhi system at all.

5. Experiment 2. Experiment 2 compares participants from three different dialect regions and was conducted in the following locations: Canterbury, using a similar population to experiment 1; Southland, a region of New Zealand with both variable rhoticity (see Bartlett 2002) and *r*-sandhi (confirmed below); and San Diego, California, where nonprevocalic *r* is categorically realized and intrusive-*r* is absent. It was hypothesized that a participant's perception of *r* would vary depending on the extent of their experience with *r* in that context.

In this experiment we also introduce stimuli in which [ɹ] is not acoustically present in *r*-sandhi contexts. We predicted that, for these tokens, participants would 'restore' the *r* in a pattern that can be related to their previous experience of *r* across contexts.

5.1. Participants. There were nineteen participants in Christchurch, fourteen in Southland, and seventeen in San Diego. An overview of participant demographics across the three experiment sites is shown in Table 5.

LOCATION	MEDIAN AGE	AGE RANGE	MALES	TOTAL SPEAKERS
Christchurch	21	18–38	11	19
Southland	54	19–74	3	14
San Diego	20	18–34	7	17

TABLE 5. Overview of participant demographics from experiment 2.

Since rhoticity in Southland is declining, we recruited older participants in this region to maximize both their own rhoticity and their prior exposure to rhotic speech (ages ranged from nineteen to seventy-four). None of the Southland participants had spent any considerable amount of time outside of the area. All Canterbury participants were from a nonrhotic part of New Zealand (thirteen from Christchurch, two from Blenheim, two from Hamilton, one from Napier, and one from Wellington). San Diego participants were restricted to those who had never spent time in an English-speaking country other than the United States and who had not lived outside of California for more than five years. All San Diego participants were rhotic, with minimal exposure to nonrhotic varieties.

For rhotic participants, a response strategy based on spelling would be difficult to distinguish from a strategy based on experience. On the one hand, we would not want to exclude a rhotic participant on the basis of having high rates of 'yes' responses to *cart*-type fillers and linking tokens. On the other hand, we wanted to avoid applying different exclusion criteria to participants from different regions. Data from all participants was therefore included in the statistical analysis, with no exclusion criteria.

5.2. STIMULI AND PROCEDURE. The stimuli were created in the same ways as those used in experiment 1. However, all test phrases in experiment 2 were included in the experiment twice: once with [ɹ] present and once with [ɹ] absent. To reduce strain for the participants, only 200 trials were included; 100 were test items (fifty with *r*, and fifty without). We omitted examples of linking-*r* across a morpheme boundary, since *r* is near-invariant in this position and its absence in this context might be noticeably marked. Because of its high level of variability in production, the test stimuli focused on intrusive-*r* across a word boundary for each of the three vowels. We also included word-boundary linking-*r*, and morpheme-boundary intrusive-*r* for the THOUGHT vowel. The distributions of the test and filler stimuli are shown in Tables 6 and 7. Appendix B gives the full list of stimuli.

	LINKING				INTRUSIVE			
	WORD		MORPHEME		WORD		MORPHEME	
	r	no *r*	*r*	no *r*	*r*	no *r*	*r*	no *r*
THOUGHT	10	10	0	0	10	10	10	10
START	0	0	0	0	10	10	0	0
schwa (noun)	0	0	0	0	10	10	0	0
total	*10*	*10*	*0*	*0*	*30*	*30*	*10*	*10*

TABLE 6. Distribution of the test-phrase types used in experiment 2.

FILLER TYPE	
orthographic and realized	50
only orthographic	25
not orthographic nor realized	25
total	*100*

TABLE 7. Distribution of the filler-phrase types used in experiment 2.

In experiment 2, the presentation order of the [ɹ]-ful and [ɹ]-less versions of each target word was balanced, such that for half of the words, the [ɹ]-ful token occurred first. A minimum of sixty trials separated corresponding target words.

Apart from the change in stimuli, the procedure was otherwise identical to that in experiment 1. The experiment was conducted in quiet rooms at the University of Canterbury (Cantabrians) and University of California at San Diego (San Diegans). Southlanders completed the experiment in their homes.

5.3. Differences in production across region. Following the perception task, participants from New Zealand completed a production task. San Diegans did not complete this task because all participants were assumed to be near-categorical in their realizations.[7]

The production task involved reading aloud twenty full sentences (shown in Appendix C) from index cards, each of which contained an *r*-sandhi environment (e.g. *The ice had already begun to thaw and melt*). There were four sentences for each type of test stimulus included in the perception experiment: word-boundary linking-*r*, word-boundary intrusive-*r* and morpheme-boundary intrusive-*r* following thought, word-boundary intrusive-*r* following start, and word-boundary intrusive-*r* following schwa. In contrast to the attempted production task in experiment 1 where only short phrases (e.g. *thaw and melt*) were elicited, the sentences for experiment 2 were structured to decrease the likelihood of pausing where *r*-sandhi could be realized. An auditory analysis of each token determined whether an *r* was produced.

Both groups produced significantly more intrusive-*r* across morpheme boundaries than word boundaries (paired Wilcoxon, Southland: $p < 0.02$, Canterbury: $p < 0.001$). They also produced more linking-*r* than intrusive-*r* across word boundaries, though this trend is significant only for the Southlanders (unpaired Wilcoxon, $p < 0.02$). Comparing the two populations, Southlanders produced significantly fewer tokens of intrusive-*r* across word boundaries than Cantabrians (unpaired Wilcoxon, $p < 0.02$). This difference is of the same magnitude across morpheme boundaries, but it is not significant, due to having fewer observations.

Two additional sentences, containing a total of six potentially rhotic contexts, were read by the Southland participants. These revealed varying rhoticity levels, from 0 to 83%, with a mean of 36%.

5.4. Experiment 2 predictions. Given the observed production patterns for the Southlanders and Cantabrians, and the assumption that the San Diegans are fully rhotic, we can make the following predictions about the perception patterns. These predictions are based on the different degrees to which participants are likely to have experienced *r*-sandhi, together with (in bold) an expected 'salience boost' in cases of unexpected presence or absence.

(2) Experiment 2 predictions
   a. All groups will hear *r* more in linking positions than intrusive positions, regardless of whether [ɹ] is present.
   b. New Zealanders will hear *r* more at morpheme boundaries than word boundaries, regardless of whether [ɹ] is present.
   c. For intrusive-*r*: We expect the regions to hear *r* to different degrees, in this pattern:
      - [ɹ] present: **San Diego** > Canterbury > Southland
      - [ɹ] absent: Canterbury > Southland > San Diego

[7] For one Southland participant, the production task was not successfully recorded. This production data therefore has one less participant than the perception data.

d. For linking-*r*: We expect the regions to hear *r* to different degrees, in this pattern:
   - [ɹ] present: San Diego > Southland > Canterbury
   - [ɹ] absent: Southland > Canterbury > **San Diego**

As with experiment 1, we expect linking-*r* to be heard more than intrusive-*r*, since all groups are likely to have experienced it more in this position (prediction 2a). We also expect the New Zealanders to hear *r* more at morpheme boundaries than word boundaries, for the same reason (prediction 2b).

With respect to intrusive-*r* (prediction 2c), we know that Cantabrians use it the most, Southlanders use it less, and it is not used in San Diego. When [ɹ] is present, then, we expect it to be very salient to the San Diegans (due to novel pop-out), and we expect the Cantabrians to hear it more than the Southlanders. When [ɹ] is absent, we expect the order of the region-based groups to mirror the degrees of experience, with Cantabrians hearing it most and San Diegans the least.

With respect to linking-*r* (prediction 2d), our production data did not reveal any differences between the Cantabrians and Southlanders. Nonetheless, Southlanders are partially rhotic, with *r* being variably produced in linking-*r* words even when not prevocalic (e.g. in *car door*). This may lead Southlanders to have a slightly higher expectation of *r* in this context, and they may therefore hear linking-*r* more than Cantabrians, both when [ɹ] is present and when it is absent. Presumably, San Diegans categorically produce 'linking-*r*' due to their rhoticity and should expect it more than the other groups. This should lead them to hear it most when [ɹ] is present. But the absence of [ɹ] in linking contexts should be salient, leading San Diegans to notice this absence most and thus 'restore' the missing *r* least.

5.5. RESULTS. Turning now to the perception results, Figure 4 shows the participant means across stimulus type, for the three participant groups. All groups of participants tend to indicate hearing *r* more often for linking-*r* than intrusive-*r* (prediction 2a). Unsurprisingly, all groups were also more likely to hear *r* when [ɹ] was actually in the signal, across all environments. The San Diegans appear to be more sensitive to the presence of [ɹ] than the New Zealand groups (consistent with predictions 2c and 2d).

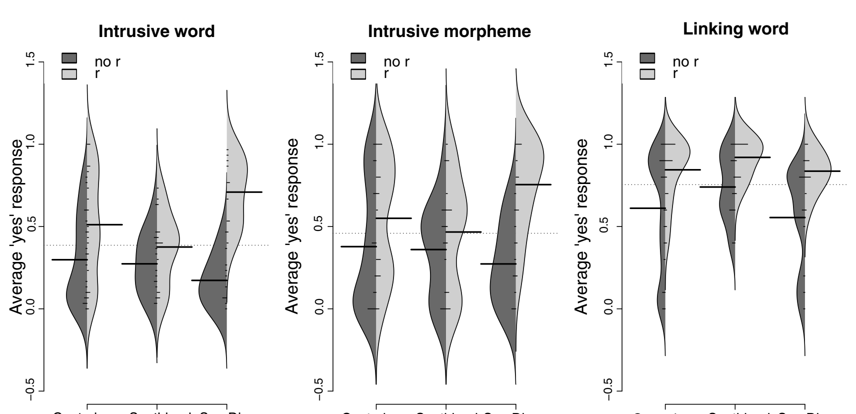

FIGURE 4. Distribution of participants' mean 'yes' responses (i.e. *r* was heard) to test items in experiment 2, shown separately for intrusive-*r* across a word boundary (left) or morpheme boundary (middle), and linking-*r* across a word boundary (right), for stimuli with (light gray) and without (dark gray) [ɹ] in the signal.

As with experiment 1, the three vowels were not distributed evenly across sandhi types. We therefore conducted the analysis using two sets of models. One focused on the THOUGHT data only, testing the difference between linking-*r* and intrusive-*r* across word and morpheme boundaries. The other tested the effect of vowel within the intrusive-*r* word-boundary tokens. As with experiment 1, we excluded potentially ambiguous items from our reported regression models.

All models reported in this section include the interaction between presence/absence of [ɹ] for the given token and the presentation order of [ɹ] presence/absence. This serves as a control on any order effects, and it was left in the models regardless of significance (except where it caused convergence issues).[8]

First, a binomial mixed-effects model explored factors affecting responses to the THOUGHT items, modeling the likelihood of participants responding that they heard *r*. We included the control interaction as outlined above. Word/morph and link/intru were treated as a single factor in the model since the experiment design lacked stimuli with linking-*r* across morpheme boundaries. Factors that were tested included the type (linking word, intrusive word, intrusive morpheme), whether [ɹ] was present, listener region (Christchurch, Southland, or San Diego), and the three-way interaction between these factors. ANOVA comparison justified retention of the three-way interaction in the model. The model is shown in Table 8.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	−0.85	0.42	−2.03	0.04
r = present	1.35	0.49	2.74	0.01
rfirst = y	0.04	0.15	0.30	0.77
type = intru.word	−0.51	0.30	−1.67	0.09
type = linking	1.58	0.30	5.34	0.00
reg = South	0.11	0.60	0.18	0.86
reg = SD	−0.54	0.66	−0.81	0.42
r = present : rfirst = y	−0.28	0.22	−1.28	0.20
r = present : type = intru.word	0.47	0.41	1.14	0.25
r = present : type = linking	0.44	0.43	1.02	0.31
r = present : reg = South	−0.70	0.70	−1.01	0.31
r = present : reg = SD	1.65	0.78	2.12	0.03
type = intru.word : reg = South	0.55	0.41	1.34	0.18
type = linking : reg = South	0.46	0.41	1.13	0.26
type = intru.word : reg = SD	0.28	0.47	0.60	0.55
type = linking : reg = SD	0.03	0.44	0.08	0.94
r = present : type = intru.word : reg = South	−0.79	0.58	−1.37	0.17
r = present : type = linking : reg = South	1.03	0.66	1.55	0.12
r = present : type = intru.word : reg = SD	0.01	0.66	0.02	0.98
r = present : type = linking : reg = SD	−0.43	0.65	−2.19	0.03

TABLE 8. Model C: binomial mixed-effects regression model of experiment 2 participants (Canterbury: default, Southland: South, San Diego: SD) responding to the THOUGHT test items. Intrusive-*r* over a morpheme boundary is treated as the baseline.

Significant in the model is a main effect of type, with participants more likely to hear *r* in linking contexts. Participants are also more likely to indicate hearing an *r* when [ɹ] is acoustically present in the stimulus than when it is absent. Finally, there is a three-way interaction between type, region, and [ɹ]-presence, which is shown in Figure 5. We

[8] The models also include random intercepts for participant and item, nested within lexeme because there was an [ɹ]-ful and [ɹ]-less token for each lexeme. A random slope for [ɹ]-presence by participant was also included. Other slopes were explored but accounted for relatively minimal variance and tended to lead to convergence problems, and so were excluded.

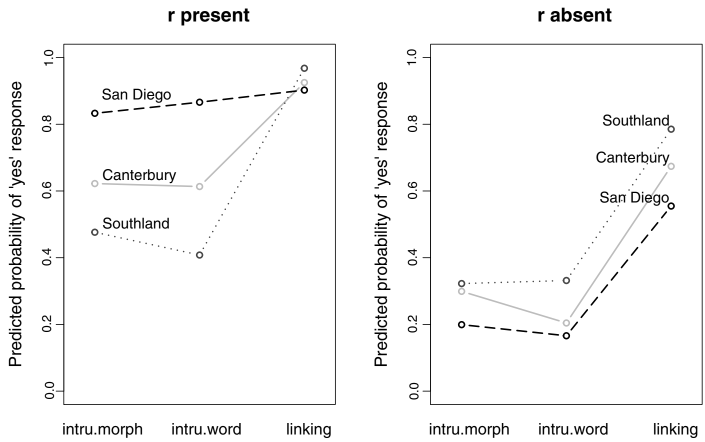

FIGURE 5. Predictions from model C, showing interaction between *r*-sandhi type, the listener's dialect, and whether [ɹ] was present in the stimulus in predicting likelihood of 'yes' response, for tokens where [ɹ] is present in the stimulus (left) and tokens where [ɹ] is absent (right).

also restricted the data to the [ɹ]-ful stimuli and tested for effects of the duration and F3 of the [ɹ] in the stimulus. The model yielded no significant results.

Since San Diegans behaved very differently, we also refit the model to the New Zealand data only and established that the three-way interaction between [ɹ]-presence, type, and region is also justified within the New Zealand data.[9] That is, we confirmed that Southlanders and Cantabrians also behave significantly differently from each other.

Many of the predictions outlined in §5.4 were confirmed by the model. We first predicted that all three groups would have high rates of perceiving linking-*r*, since speakers from all three regions produce *r* in this context. This difference was highly robust.

Second, we predicted that New Zealanders would hear more *r* at morpheme boundaries than word boundaries. This is the direction observed, although not strongly in all parts of the data. The difference is visible for the Southlanders when [ɹ] is present, and for the Cantabrians when [ɹ] is absent. The San Diegans show no strong trend in either case. The difference between word and morpheme boundaries is certainly not as robust in this data set as in experiment 1.

Third, we predicted differences across participant groups for intrusive-*r*, with Cantabrians hearing it more than Southlanders overall, and San Diegans hearing it most in [ɹ]-ful tokens, and least in [ɹ]-less tokens. For the San Diego participants, this was clearly confirmed. They reported hearing intrusive-*r* at the highest rates when it was present, and at the lowest rates when it was absent. Additionally, the Cantabrians heard more intrusive-*r* than the Southlanders when [ɹ] was present. When it was absent, this trend appeared to switch. However, testing this subset of the data reveals no significant difference between the two groups when [ɹ] is absent. Thus, the predicted ranking of the two groups was observed only when [ɹ] was actually in the signal.

Finally, we predicted differences across the regions for linking-*r*, with Southlanders hearing it more than Cantabrians, and San Diegans hearing it most when it was present and least when it was absent. This predicted ranking was observed when the [ɹ] was absent. When [ɹ] was present, all groups tended to hear it very well, creating a ceiling effect that prevents robust discrimination between the groups.

[9] An ANOVA comparison between a model containing the three-way interaction and a model containing all component two-way interactions was significant ($p < 0.03$).

It is noteworthy that, for the New Zealanders, a similar pattern of 'yes' responses occurs for [ɹ]-less tokens as for [ɹ]-ful ones, but at lower rates. Because a similar pattern was observed when there was no acoustic evidence of [ɹ] in the stimuli, the results could not be due to inherent differences in phonetic salience ('degree of standing out') across contexts. Instead, we interpret the similar patterns of responses as evidence that perception is influenced by previous exposure to the sound in these contexts.

The effect of the preceding vowel was tested separately in a model fit only to tokens with intrusive-*r* at a word boundary. This model contained a significant effect of vowel, with THOUGHT tokens eliciting more 'yes' responses than START or schwa. This did not interact with region or stimulus type. Since the response pattern across vowels did not differ for San Diegans and New Zealanders, we assume that this vowel effect does not reflect experience but is instead related to phonetic characteristics of the signal. The model of the intrusive word-boundary data also showed the same interaction between region and [ɹ]-presence that is visible in Fig. 5. San Diegans are more sensitive to [ɹ]-presence than New Zealanders; they are the most likely to indicate that they heard *r* when [ɹ] is present, and that they did not hear it when [ɹ] is absent. In this model, the control interaction between [ɹ]-presence and the order of the [ɹ]-ful and [ɹ]-less tokens also reached significance ($p < 0.02$); participants were more likely to answer 'yes' to an [ɹ]-less stimulus if they had previously heard its [ɹ]-ful counterpart.

In sum, there were some clear differences across regions, providing support for our specific hypothesis 2, that *r*-sandhi would be differently heard by listeners from different regional backgrounds, reflecting their differing previous experiences with *r*-sandhi.

5.6. THE ROLE OF EXPERIMENT-INTERNAL EXPERIENCE. It is interesting to note that, even when [ɹ] is absent in the intrusive contexts, the San Diegans respond 'yes' a nontrivial proportion of the time. This is unexpected since they have no prior expectation of *r*-sandhi in this context. Note, however, that during the experiment, the San Diego participants encountered the absence of [ɹ] where they expected one to be (linking-*r*) as well as the presence of [ɹ] where they did not expect it to be (intrusive-*r*). We hypothesized that this may have influenced their expectations, shifting their behavior as the experiment progressed. In this section, then, we outline a post-hoc analysis exploring related prediction (i): that participants may learn from the novel exposure they gain during the course of the experiment.

If the San Diegan participants shifted their behavior in this way, there should be a significant interaction between trial and stimulus type for the San Diego participants but not for either group of New Zealanders. We therefore fit separate models to each of the regions, testing for this interaction.[10]

For the New Zealanders, the trial number was not significant, in isolation or in interaction. In contrast, the model for the San Diegans (shown in Table 9) revealed a significant interaction between trial and stimulus type. This model also incorporates a significant interaction between [ɹ]-presence and type, showing that the difference between types has a bigger effect when the [ɹ] is absent than when it is present, as discussed in the previous section (refer to Fig. 5).

[10] We first attempted to enter interactions involving trial into the model reported in the previous section. However, the increased complexity resulted in unconverged models and high levels of correlation between predictors. Therefore, we instead fit separate models for each of the three regions, using backward selection, starting from three-way interactions between trial number (centered and scaled), *r*-presence, and stimulus type.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	−1.464	0.588	−2.491	0.013
r = present	3.125	0.922	3.391	0.001
rfirst = y	−0.162	0.420	−0.385	0.700
trial	0.448	0.237	1.889	0.059
type = intru.word	−0.148	0.377	−0.393	0.695
type = linking	1.737	0.353	4.922	0.000
r = present : rfirst = y	0.121	0.719	0.168	0.866
trial : type = intru.word	−0.110	0.270	−0.407	0.684
trial : type = linking	−0.614	0.258	−2.386	0.017
r = present : type = intru.word	0.288	0.541	0.532	0.595
r = present : type = linking	−1.202	0.523	−2.299	0.022

TABLE 9. Model D: San Diego participants only, incorporating significant interaction of trial and stimulus type.

The interaction between trial and stimulus type is plotted in Figure 6. This shows how the San Diegans shifted their behavior during the experiment. The interaction is plotted for the model defaults and therefore shows the predicted probabilities for when [ɹ] is absent. When [ɹ] is present, the predicted probability of hearing it is, of course, much higher. The effect of trial, however, is the same, irrespective of [ɹ]-presence.

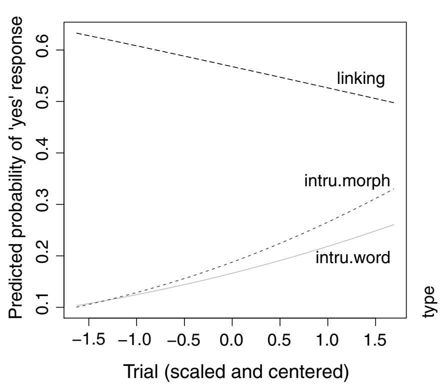

FIGURE 6. Interaction between trial number and stimulus type from model D, predicting San Diego participants' likelihood of 'yes' responses.

As can be seen from Fig. 6, for linking-*r*, the San Diego participants started with a high rate of perceiving *r*. However, the rate lessened as the experiment progressed and they encountered more [ɹ]-less linking tokens. For intrusive tokens, the San Diegans started with very low rates of perceiving *r*, but the rate rose over the course of the experiment after exposure to *r* in this context. These shifts in behavior provide evidence that the San Diego participants found such tokens salient. These tokens were therefore more strongly attended to, influencing subsequent behavior on the task.

Importantly, these changes during the experiment can be seen both when [ɹ] is present in the stimulus and when it is not. As they progress through the experiment, the San Diego participants are increasingly likely to accurately hear an *r* in an intrusive-*r* context when [ɹ] is present. They are also more likely to inaccurately hear an *r* in that context when [ɹ] is actually absent. This provides evidence that it is indeed their experience driving the effect rather than it being (entirely) due to characteristics of the stimulus. We cannot know whether this adaptation effect would generalize to listening to other speakers of NZE. What we are seeing here is likely a local adaptation effect to this particular talker in the context of this experiment.

In contrast to the San Diegans, the New Zealand participants already have extensive exposure to variable production of intrusive-*r*, so hearing *r* in intrusive contexts is not salient to them. Relying on their stored linguistic memories, they had greater variability in perceiving *r* from the beginning of the experiment and no shift in behavior as it progressed.

5.7. Social and individual variation. Given the results of the corpus study outlined in §2.2, we might expect some social predictors to emerge, at least for the Canterbury participants. We wanted to test this and also confirm that the effects of region were not caused by uneven distributions of age and gender across the three regions. We therefore pursued post-hoc analysis as outlined in related prediction (ii): that individual variation in production within dialect area may predict individual variation in the phoneme-monitoring task.

We fit separate models to the thought data, testing for age and gender, for each of the three regions.[11] No social factors proved significant in the Southland or San Diego data. In the Canterbury data, there was a significant effect of gender (gender = male, coef = 1.62, $p < 0.02$).

This result is consistent with the corpus data outlined in §2.2. Cantabrian men are more likely to produce sandhi-*r* than Cantabrian women, and—it transpires—are also more likely to hear it.

To directly test the effect of individual variation in production, we derived a crude measure of production for inclusion in models of the New Zealand data: the proportion of *r*-sandhi a participant realized across all twenty contexts recorded (cf. §5.3).[12] This analysis necessitated the removal of one Southland participant for whom the production data was not successfully recorded. We scaled and centered the proportion of *r*-sandhi produced and then used it as a predictor in the regression models for the thought data, adding it as a fixed effect. A model fit to both data sets together shows a three-way interaction between production rate, sandhi type, and region. This model is shown in Table 10.[13]

[11] To test for age and gender, we first attempted to add these predictors into the model reported in Table 8, adding them both as fixed effects and as random slopes for the item intercept. We also tested interactions with region, but these models did not converge. Examination of the nonconvergent models suggested no effects of age, but a potential significant interaction between gender and region. The final (thought-only) models incorporated random slopes for the social factors, the item intercept, and the sandhi type for the participant intercept. In order to obtain robust converging models for all three data sets, we dropped the interaction between [ɹ]-presence and whether the [ɹ]-ful token came first (rfirst = y). This had been included as a control interaction but did not approach significance in any of these models. We separately tested age and gender in different models.

[12] Over the whole New Zealand data set, this overall individual sandhi-production rate is significantly correlated with the individual production rates for linking-*r*, intrusive-*r* across a word boundary, and intrusive-*r* across a morpheme boundary (Spearman's correlation, $\rho = 0.67$–0.82, $p < 0.00001$). These correlations remain significant when examined within the Canterbury data alone ($\rho = 0.75$–0.88). They are significant within the Southland data set for intrusive-*r* across a word ($\rho = 0.66$) and morpheme ($\rho = 0.75$) boundary, but the overall rate of *r*-sandhi usage does not quite significantly correlate with linking-*r* use for the Southlanders ($\rho = 0.51$, $p = 0.06$). This makes sense since, for the Cantabrians, the *r*-sandhi system is the sole source governing the realization of *r* in a linking position, whereas for Southlanders, the *r* could also surface in this position due to partial rhoticity.

[13] In this model, we tested the control interaction between whether the [ɹ]-ful token was first and whether [ɹ] was present. This did not near significance but caused convergence issues, and so was dropped. The model also converges if this control interaction is retained, but the slope for [ɹ]-presence given participant is omitted. Similarly, a random slope for sandhi rate, on the item intercept, causes convergence issues, but does not affect the significance of the fixed effects. In all models, the reported interaction remains significant and is roughly equivalent.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	−1.115	0.371	−3.006	0.003
sandhi-rate	1.093	0.268	4.080	0.000
type = intru.word	−0.150	0.230	−0.651	0.515
type = linking	1.877	0.231	8.119	0.000
reg = South	0.426	0.499	0.855	0.393
r = present	1.159	0.298	3.888	0.000
rfirst = y	−0.079	0.143	−0.554	0.579
sandhi-rate : type = intru.word	−0.304	0.184	−1.650	0.099
sandhi-rate : type = linking	−0.338	0.194	−1.746	0.081
sandhi-rate : reg = South	0.106	0.585	0.182	0.856
type = intru.word : reg = South	−0.121	0.331	−0.366	0.714
type = linking : reg = South	0.148	0.350	0.423	0.672
sandhi-rate : type = intru.word : reg = South	0.048	0.400	0.120	0.905
sandhi-rate : type = linking: reg = South	−0.944	0.436	−2.163	0.031

TABLE 10. Model E: logistic regression of New Zealand responses to the THOUGHT tokens, incorporating the effect of individual *r*-sandhi production.

The interaction from this model is shown in Figure 7. On the left, we see that the rate at which Southlanders produce *r*-sandhi predicts the degree to which they perceive intrusive-*r*, but not linking-*r*. On the right, however, we see that, for Cantabrians, linking-*r* is fully part of the same system, and the rate of *r*-sandhi production predicts the rate of hearing *r* across all contexts.

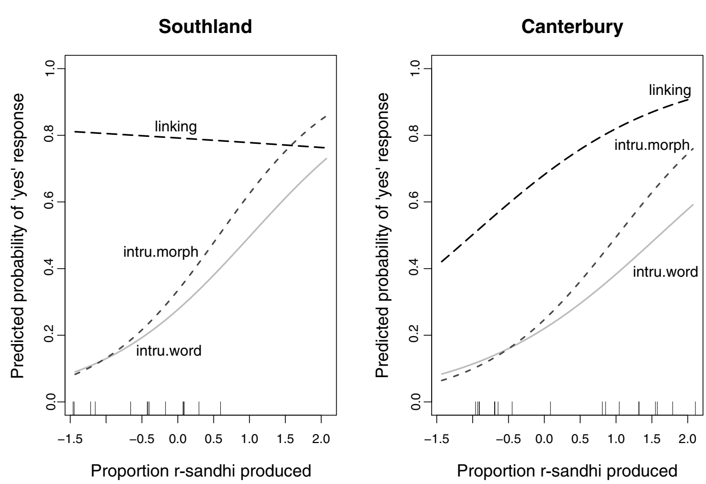

FIGURE 7. Significant interactions from model E, showing the effect of the rate of *r*-sandhi production in the different regions. The rugs along the bottom of the plot show where the actual participants are positioned.

Tested separately, within the Canterbury data, production rate emerges as a significant main effect, with participants who produced more *r*-sandhi also hearing more *r*-sandhi. In the Southland data, it interacts with the sandhi type, showing a positive relationship with the intrusive tokens but not the linking tokens, in the same pattern as shown in Fig. 7. Southlanders' perception of linking-*r* is therefore at least partly driven by rhoticity, rather than *r*-sandhi alone. Indeed, for the Southlanders, we can test for an effect of rhoticity. The degree of rhoticity does not correlate with the degree of *r*-sandhi usage in this data set. However, when scaled and added to the perception model, it emerges as a separate significant predictor of hearing *r*-sandhi in an interaction with sandhi type. The resulting Southland model is shown in Table 11,[14] and the interactions are plotted in Figure 8. On the left, we see the effect of *r*-sandhi production, which increases the probability of hearing intrusive-*r*. On the right, we see that greater rhoticity increases the rate of hearing linking-*r* but not intrusive-*r*.

	EST.	*SE*	*z*-VALUE	Pr(>|*z*|)
(intercept)	−0.714	0.459	−1.557	0.120
sandhi-rate	1.117	0.549	2.034	0.042
type = intru.word	−0.237	0.309	−0.767	0.443
type = linking	2.092	0.337	6.202	0.000
rhotic-rate	−0.175	0.324	−0.540	0.589
r = present	0.965	0.356	2.709	0.007
rfirst = y	0.402	0.324	1.241	0.215
sandhi-rate : type = intru.word	−0.253	0.359	−0.705	0.481
sandhi-rate : type = linking	−1.324	0.394	−3.362	0.001
type = intru.word : rhotic-rate	0.041	0.229	0.178	0.859
type = linking: rhotic-rate	0.656	0.259	2.536	0.011
r = present : rfirst = y	−0.415	0.464	−0.894	0.371

TABLE 11. Model F: Southlanders only, incorporating separate effects of individual *r*-sandhi production and individual rate of rhoticity.

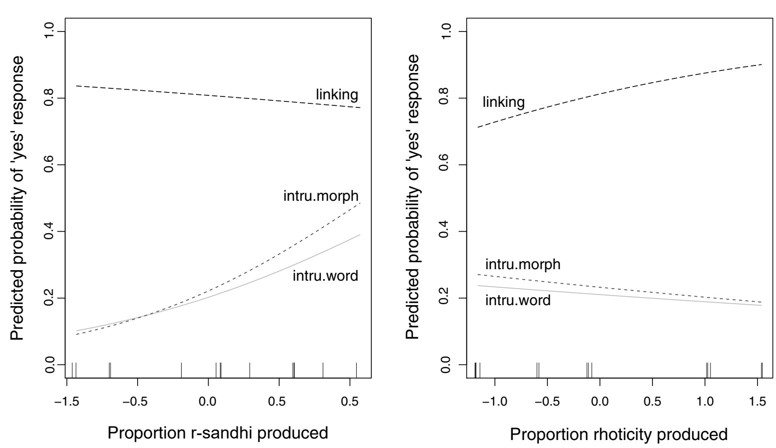

FIGURE 8. Significant interactions from model F, showing the relationship between production and perception in Southland. The rugs along the bottom of the plot show where the actual participants are positioned.

No additional effects of gender emerge in these models when production is included. Therefore, we believe that the previously reported significant effect of gender for the Cantabrians was a proxy for individual degree of *r*-production. The data from the production task accounts for this effect adequately, and in a way that extends to the Southland results where no trends for gender are present.

Overall, we can conclude that the New Zealand data shows a strong relationship between production and perception, in which individual participants who are more likely

[14] The model incorporates random slopes by item for both rhoticity production and *r*-sandhi production, and a random slope by participant for [ɹ]-presence. An interaction between [ɹ]-presence and sandhi type ($p < 0.05$) was dropped from this model in order to obtain convergence. An alternative solution would have been to retain this interaction and remove the random slopes per item. In all converging models, however, interactions between rhoticity and sandhi type, and *r*-sandhi production and sandhi type, are retained as significant and look as pictured in Fig. 8.

to produce *r*-sandhi are also more likely to perceive it. This can be seen in three ways. First, our analysis of the nonrhotic production data in §2.2 revealed that men are more likely to produce *r*-sandhi than women. This is consistent with the findings regarding perception of *r* among participants from the same region: men were more likely to hear *r* than women. In contrast, this was not true of the rhotic (San Diego) and partially rhotic (Southland) participants. Second, if we use a more direct proxy for the degree to which participants use *r*-sandhi themselves, we also observe a link between production and perception: individuals who are more likely to produce *r*-sandhi are also more likely to hear it. This is true in all parts of the New Zealand data, with the exception of Southland linking-*r*, which is partially under the influence of a rhoticity system. Third, we found evidence of a relationship between partial rhoticity and perception of linking-*r* for the Southland participants. Participants who are more rhotic are more likely to hear linking-*r*.

5.8. WORD-BASED EFFECTS. The results from the above analyses support the interpretation that the degree to which participants hear *r* is influenced by their past experience with *r* in the relevant environment. Hay and Maclagan (2012) demonstrated that there was word-specificity in the likelihood of producing linking-*r*, at least in early-twentieth-century NZE. An episodic-memory perspective would therefore predict effects of individual words on the rate of perceiving *r*. This perspective led us to explore evidence for related prediction (iii): that there may be systematic word-level effects.

Certainly, response rates to individual tokens vary, but this could potentially relate to how much the *r* stands out in contrast with its phonological environment, strength of the [ɹ] produced in the stimuli, and/or other random factors. Using an external predictor (such as word frequency or probability of occurring in a sandhi environment) to predict item-based variation is not practical because most of the stimulus items have very low token frequencies, so appropriate statistics cannot be extracted from corpora. To search for evidence of item-based differences in *r*-perception rates, we address two questions, outlined below.

First, we explore whether there is any relationship between the rate of *r* heard in each stem and its corresponding *-ing* form. If there are lexically specific probabilities of encountering *r*-sandhi, then there might be a relationship between probabilities in bases and their associated *-ing* forms. This could occur because such forms are stored in close association and resonance flows between them, or because the latter is derived from the former.

As our experiment was not expressly designed to test item-based variation, we can only investigate it post-hoc and in a relatively limited way. We do this by exploring New Zealanders' responses to post-THOUGHT tokens of intrusive-*r* in experiment 2. We focus on these words because we have a matched set of bases and corresponding inflected forms, both with [ɹ] produced and without. In experiment 1, [ɹ] was invariably present, and in experiment 2, we did not include linking-*r* across a morpheme boundary (see Table 3).

Pooling the data from the Canterbury and Southland participants, we calculated the average number of 'yes' responses to THOUGHT intrusive tokens, across word and morpheme boundaries, when [ɹ] was present and when it was not. We then investigated the relationship between rates of 'yes' responses to particular word-boundary tokens and their corresponding morpheme-boundary (*-ing*) forms, as shown in Figure 9. Tokens in which [ɹ] is present are plotted in bold, and those in which it is absent are plotted in gray. Overall, there is a significant relationship between responses at word boundaries and corresponding morpheme boundaries ($\rho = 0.75$, $p < 0.001$). This correlation approaches, but does not reach, significance when tested separately within the subsets of the data where [ɹ] is present or absent.

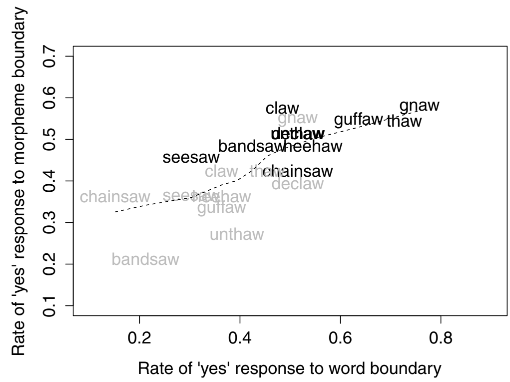

FIGURE 9. The relationship between rates of *r* reported for words and their corresponding affixed forms for participants from New Zealand. Tokens for which the [ɹ] is present are plotted in black, and those for which [ɹ] is absent are plotted in gray.

While there are clearly sparsity issues here, we take the significant correlation to be indicative of variation that exists at the level of the item. Note that in the San Diego data, this relationship does not hold. For the tokens where [ɹ] was present, there is no relationship between 'yes' responses for bases and related inflectional forms ($\rho = 0.17$, $p = 0.64$). In fact, for those tokens where the [ɹ] was absent, there is a negative relationship that reaches significance ($\rho = -0.71$, $p < 0.02$). This negative relationship is driven by responses to *gnaw*, which has the lowest rate of 'yes' responses across a morpheme boundary, but the highest rate of 'yes' responses across a word boundary. The significance disappears when this pair is removed.

The second question we ask in regard to word-based effects is whether there is a relationship between rates of perceiving *r*-sandhi across items when the [ɹ] is and is not present. That is, if an item elicits high rates of 'yes' responses with [ɹ] present, does it also elicit high rates when [ɹ] is absent? We have already shown that such a correspondence is broadly true for the categories of stimuli in our experiment, but does this extend to the patterning of individual items?

Using the same subset of data as described above, we investigated responses to the [ɹ]-ful and [ɹ]-less tokens of each item, as shown in Figure 10, which includes responses from New Zealanders to word-boundary and morpheme-boundary tokens. There is a significant relationship between the [ɹ]-ful and [ɹ]-less forms of a given item (Spearman's $\rho = 0.6$, $p < 0.01$). Items that were frequently perceived as containing an *r* when [ɹ] was present in the signal were also more frequently perceived as having an *r* when [ɹ] was absent. Considered separately within the word-boundary or morpheme-boundary tokens alone, the correlation nears, but does not quite reach, significance. Note, however, that these latter tests are each conducted over only ten pairs of words, so a very strong relationship would be needed in order to observe significance.

In contrast with the New Zealand data, the San Diego data shows no relationship between rates of 'yes' responses for [ɹ]-ful and [ɹ]-less tokens ($\rho = 0.13$, $p = 0.5$). This is expected, since these listeners do not have past experience that leads them to predict high rates of *r* for particular lexical items.

This post-hoc analysis suggests that there may be structure to the item-based variation that we see in the New Zealand data. A base word for which *r* is heard at a particu-

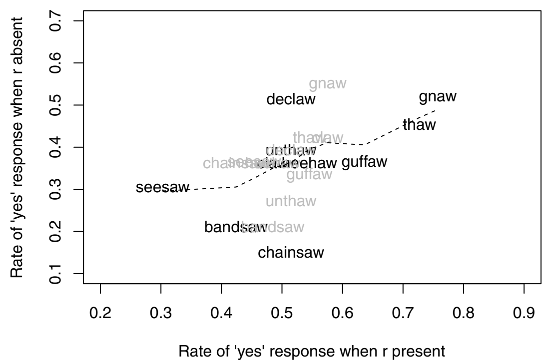

FIGURE 10. The relationship between rates of 'yes' responses to words when [ɹ] is present and [ɹ] is absent for participants from New Zealand. Word-boundary tokens are shown in black, and morpheme-boundary tokens are shown in gray.

larly high rate is likely to attract high rates of *r* responses regardless of whether the [ɹ] is present in the stimulus, as well as across both the word-boundary and corresponding morpheme-boundary tokens. Thus, the data shows evidence of word-specific response patterns, thus lending some support to related prediction (iii).

6. OVERVIEW OF FINDINGS. In production, speakers of nonrhotic NZE appear to maintain a difference between the rates of *r*-sandhi produced at word boundaries and morpheme boundaries, with morpheme boundaries eliciting a higher rate of *r*. There is also a difference between the rates of linking-*r* and intrusive-*r*, although this difference has collapsed in recent years and is not present for the youngest speakers (cf. model A, Table 1).

When the perception results for the nonrhotic New Zealanders (experiments 1 and 2) are taken together, it is clear that these results strongly echo the linguistic-context effects in production. Morpheme-boundary *r* is heard more frequently than word-boundary *r* (model B in Table 4), and this tendency can be observed even when [ɹ] is not present in the stimulus (model C–Cantabrians in Table 8). Additionally, there is a robust effect of sandhi type, with linking-*r* heard more often than intrusive-*r* (models B, C, E in Tables 4, 8, 10).

The results of the phoneme-monitoring experiments involving Cantabrians, then, clearly indicate that there is a relationship between perception patterns and the likelihood of encountering and producing an *r* in a given context, consistent with our specific hypothesis 1.

In order to test specific hypothesis 2, the response patterns of the nonrhotic New Zealanders (the Cantabrians) were compared with the responses of two populations with different production patterns: Southlanders and San Diegans. Southlanders receive exposure to nonrhotic NZE, but they live in an area that is partially rhotic. Because of this partial rhoticity, their experience of linking-*r* is higher than it is for Cantabrians, but they produce less intrusive-*r* (§5.3). In production, San Diegans are rhotic, and intrusive-*r* is entirely absent from their variety. Therefore, the San Diegans should enter the experiment expecting the near-categorical presence of *r* in linking environments, and the absence of *r* in intrusive environments.

When [ɹ] was present in the stimuli, Southlanders heard linking-*r* marginally more often than Cantabrians, and they heard intrusive-*r* significantly less often (model C in Table 8). This is consistent with the observed differences in production. The San Diegan's responses are quite different from those of New Zealanders. Despite the fact that the San Diego participants do not produce intrusive-*r*, they are the most likely group to 'hear' it when it is present (model C in Table 8). We argue that this is due to the high level of salience of the sound, salience that arises because it is directly counter to the listener's previous experience. The results also suggest that salience can arise with the absence of an expected sound: the San Diegans are the least likely group to 'restore' a linking-*r* when it is absent, despite near-categorical production of the *r* in this context. Because it is categorically expected for these words, its absence is likely to draw attention.

The results of experiment 2, then, show differences between the three listener groups, consistent with specific hypothesis 2, that listeners from different regional backgrounds with different *r*-exposure would behave differently in our *r*-monitoring task. The nature of the relationship reflects the overarching hypothesis, that increased exposure will lead to increased expectation (and therefore perception), with the exception of salient boundary zones, occupied here by the San Diegans.

While the San Diego participants are the least likely group to restore an absent intrusive-*r*, they still restore it to some degree, counter to their prior experience. The analysis suggests that this is due to experiment-internal learning; independent of whether [ɹ] was present in the stimulus, the San Diego participants increased their expectation of an *r* in intrusive positions and decreased their expectation in linking positions as the experiment progressed (cf. related prediction (i)). No such trial effect was observed for the New Zealanders.

Additionally, the results provide evidence for individual variation within dialect region. New Zealand participants who were more likely to produce *r*-sandhi were also more likely to indicate perceiving intrusive-*r* in the experiment (model E in Table 10), but only the participants from Canterbury exhibited a similar relationship between *r*-sandhi production and the perception of linking-*r* (model E). For the participants from Southland, there was no such effect. Rather, for the Southlanders, the rate at which they heard linking-*r* was related to their level of rhoticity (model F in Table 11). To the extent that rates in production are linked with exposure, this provides additional support that a listener's experience with a variable affects their perception of that variable (cf. related prediction (ii)).

Finally, we looked for evidence of lexical specificity in the responses. We hypothesized that there may be lexically related variation in the perception of *r*-sandhi because the New Zealand participants have likely encountered *r*-sandhi at higher rates in some items than others (Hay & Maclagan 2012). In the perception data, there are two pieces of evidence for lexically specific response patterns. First, there is a correlation between response rates for intrusive-*r* items when [ɹ] is present and when it is absent. That is, items identified as having an *r* when [ɹ] was acoustically present also tended to have a high rate of *r*-responses when no [ɹ] was present (Fig. 10). Second, there is a correlation between responses to word- and morpheme-boundary tokens that share a stem. If a target (e.g. *gnaw*) had a high rate of 'yes' responses, then its corresponding morpheme-boundary token (e.g. *gnawing*) also did (Fig. 9). Neither of these correlations is significant in the San Diego data, which is consistent with our predictions given that the San Diego participants have no previous experience with intrusive-*r* in these words. Thus, we cautiously interpret these results as revealing a degree of lexical-specificity in the New Zealand participants' responses that is based on their prior experience with the words, as per related prediction (iii). Our caution here is two-fold: first, the data was not explicitly designed to test this hypothesis and the analysis is post-hoc, and second (and not unrelatedly), the analysis is based on correlations across a fairly small number of items.

A schematic summary of the results showing the link between experience and *r*-perception is shown in Figure 11 for the Canterbury (C), Southland (SO), and San Diego (SD) participants. Probabilities of 'hearing' *r* for the three participant groups vary across the different boundary types: linking-*r* across a word boundary (L), intrusive-*r* across a word boundary (W), and intrusive-*r* across a morpheme boundary (M). For the New Zealand participants, the predicted probability of hearing *r* is generated from model C (Table 8). For the San Diego participants, multiple values are extracted from model D (Table 9) in order to capture the effect of trial. This experiment-internal experience includes increased exposure to intrusive-*r* and a decrease in their exposure to *r* in linking positions. These values are distributed across the top and bottom 0.1 of the x-axis.[15] The estimates of experience each participant group has with each *r*-sandhi type come from several different sources. For the Canterbury participants, the values on the x-axis are model predictions from production data (model A, Table 1). We do not have enough data to directly estimate Southlanders' experiences in the same way, but, based on the very small data set described in §5.3, we subtracted 0.175 from the Cantabrian intrusive-*r* values, which was the difference between the mean overall *r*-sandhi rate produced by the Cantabrians and the Southlanders. We also increased the predictions for linking-*r* by an arbitrary 10%, since Southlanders can be partially rhotic and these words are produced with *r* when not prevocalic some proportion of the time. While the values are clearly not precise because they were calculated based on separate models and involved some guesswork, it is the relative placement of the points on the x-axis that is important, not the absolute values.

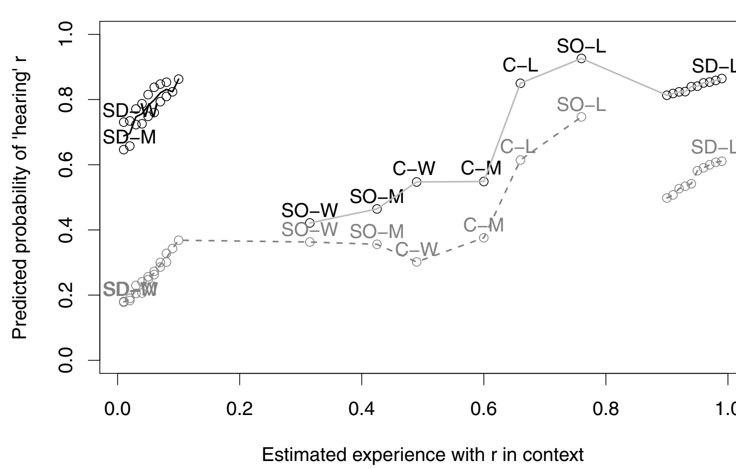

FIGURE 11. The relationship between past experience and probability of hearing *r*. The solid line shows results for when the [ɹ] is present in the stimuli, and the dashed line for when [ɹ] is absent. Points labeled are for participants from Canterbury (C), Southland (SO), and San Diego (SD) responding to linking-*r* (L), intrusive-*r* across a word boundary (W), and intrusive-*r* across a morpheme boundary (M). Changes in experience during the course of the experiment are shown for the San Diegans.

While we show a single point for each of the *r*-categories for the Southland and Canterbury participants, we know—from the models pictured in Figs. 7 and 8—that each of these points actually represents its own small positive correlation. Within each cate-

[15] Note that if a single value for each San Diego type is extracted from model C instead, this value falls near the middle of the range of values shown in the figure.

gory, there is individual variation; greater experience of perceiving and/or producing *r* in the relevant context leads to an increased likelihood of hearing it.

A comparison of the predictions laid out in Fig. 2 with the summary of the results provided in Fig. 11 helps demonstrate the degree to which the results confirm our overarching hypothesis. Overall, greater experience with a sound in a given context results in a higher rate of perception of the sound in that context, regardless of whether the sound is acoustically present in the signal. This can be seen in the middle part of the figure, in which increased experience with *r* across different contexts leads to increased probability of hearing *r*. It can also be seen in the correlations from models E and F, showing that even WITHIN each category, New Zealanders who are more likely to produce *r* in the relevant context are also more likely to perceive it.

An exception to this positive association between experience and perception arises at the edges of the figure, in cases where listeners have near-categorical previous experience. This leads to an effect of salience being applied over and above the effect of experience. If a listener categorically expects a sound to be present, then its absence is highly salient and the listener can more easily recognize its absence. Likewise, if a listener categorically expects a sound to be absent, then its presence is highly salient. Furthermore, we observe the expected effect of experience even within these ‘salient’ regions. Changing people’s experience (here: within the context of the experiment) changes their underlying expectations (at least within the course of the experiment).

The ‘salient’ and ‘nonsalient’ parts of the figure are left unconnected, as we cannot yet know what the transition between these regions looks like. Is there a gradual lessening of the degree to which something is salient as it becomes less unusual? Or does this type of salience act as a binary ‘switch’, in which, after a certain degree of exposure, salience categorically disappears? Weber (2001, 2002) argues that ‘novel pop-out’ requires categorical absence of the pattern from the language. The phenomenon she discusses is somewhat different, though, and her phonotactic account alone would predict that intrusive-*r* should not be salient at all, since rhotic listeners are well acquainted with the presence of *r* intervocalically. Instead, the salience observed in our experiment arises from a mismatch between lexical expectation and phonemic activation. It seems intuitively likely that a small amount of previous exposure would still result in salience in such cases. Our informal experience is that saliently ‘weird’ aspects of other dialects remain weird for some time, until one becomes acclimatized through extended exposure.

In the next section, we step through the implications of these findings for phonological representation and models of speech perception.

7. IMPLICATIONS FOR REPRESENTATION. Taken together, the results of the experiments presented in this article provide support for an experience-based model of speech perception in which the probability of perceiving a sound is influenced by both lexically specific exemplars and probabilistic abstract representations, and that these different levels, together, influence the likelihood that a participant will report hearing a sound.

While a growing number of scholars agree that both lexically specific episodic information and abstract representations play a role in speech production and perception (Pierrehumbert 2002, 2016, Goldinger 2007, Cutler 2012, Ernestus 2014), the link between these levels is not well specified. Here, we summarize the evidence our data provides for the different levels, and we present details of how the two levels might work together to yield our results.

Note that we are not arguing that the levels discussed in this article are the only levels of representation involved in speech perception and production. However, they are the two levels that would appear to be absolutely necessary in order to explain the results presented, so they are the ones we focus on here.

7.1. AN EXPERIENCE-BASED LEXICON. Hay and Maclagan (2012) have shown that the likelihood that *r* will surface prevocalically is related to the overall likelihood of the word occurring prevocalically in general. This suggests that speakers have access to word-specific likelihoods of *r*-realization. Evidence that, like speakers, listeners access word-specific likelihoods of *r* is provided by three sets of results from the current experiments: the difference in responses to linking-*r* versus intrusive-*r*, and, for the New Zealand but not San Diego participants, the correlation between responses to the base and *-ing* forms and the correlation between responses to lexical items when [ɹ] is present and when it is absent. We discuss each of these sources of evidence below.

The history of *r*-sandhi production in New Zealand shows a difference between linking-*r* and intrusive-*r*, with the former eliciting more *r*. Our perception results mirror this difference, with listeners hearing a higher rate of *r* for linking-*r* words. How is this difference maintained? Phonologically, the sets of words have a similar shape, and there is nothing that distinguishes between them at the phonological level. One interpretation of the difference in production and perception is that people refer to the orthography of each word in order to classify it into one of two classes. This classification would then affect production, as well as performance in our task. A second interpretation is that individuals have access to the distribution of each word's production. For historical reasons, the lexical distribution of each linking word has more *r* in it, especially since older speakers maintain some distinction. These word-specific distributions would lead to heightened perception of *r* in linking-*r* relative to intrusive-*r*. The two explanations are not mutually exclusive, and indeed both may contribute to this result. However, the former would simply predict *r*-perception to be uniformly raised in all linking-*r* words relative to intrusive-*r* words. As outlined below, there is also evidence of lexical variation within each class.

The second source of evidence that listeners have access to word-specific likelihoods of *r* is that, for New Zealand but not San Diego participants, post-hoc analysis shows a positive correlation between morpheme- and word-boundary forms for related lexical items. In other words, if a base word is more likely to be perceived with *r*, so is its corresponding *-ing* form. This level of lexical specificity would be difficult to model with an abstract level of representation alone if the model had no role for lexically specific influence. However, if one allows for rich lexical representations, then there are at least three potential mechanisms through which this relationship could come about. One potential mechanism is that many *-ing* forms could be sufficiently frequent that they have their own representation (e.g. *clawing*), a possibility that is well supported by the morphological literature (cf. Stemberger & MacWhinney 1986). Representations of the *-ing* forms would be associated with, but not subsumed by, representations of the words' base forms. Because of their high level of phonetic and semantic similarity, resonance flows between the affixed and base representations, also reinforcing a correlation in the *r*-fulness of each. Indeed, the second author's own experience as a speaker of North American English living in New Zealand reinforces the interpretation that inflected forms can have their own representation. She acquired—as isolated lexical wholes—*r*-ful forms of *scuba-[ɹ]-ing* 'to go scuba diving' and *salsa-[ɹ]-ing* 'to go salsa dancing'. Of course, not all *-ing* forms are stored. For *-ing* forms that are unlikely to have been frequently encountered before (e.g. *heehawing*), we might predict that the expectation is created on the fly via the combination of the base (*heehaw*) and the representation of the productive *-ing* suffix. In this way, the likelihood of *r* in the base form feeds directly into the likelihood for the affixed form, forming a tight correlation. Finally, even for forms where a listener has never encountered the base or the inflected variant, they could generate an expectation of *r* based on analogy to phonologically similar forms. If X is similar to many forms that have a high probability of *r*, it is likely to generate a higher expectation of *r* than if it is similar to many forms that have a lower probability of *r*. Because X and X-*ing* are phonologically similar to each other, they are influenced by an overlapping cohort of analogical forms. This would lead to a correlation in our data. In other words, analogy provides a mechanism through which our observed correlation could arise even if there are no lexically specific representations of *r* for the particular words at hand. Notably, however, the process relies on lexically specific representations of other forms from which the analogy can be extracted. It is likely that all of the above processes operate in parallel, each contributing to the correlations we see. Importantly, all require the positing of lexical specificity in terms of likelihood of *r*-production.

The final source of evidence in our data that listeners are influenced by lexically specific likelihoods of *r* is the correlation between responses for a lexical item when [ɹ] is present and when it is absent. Importantly, we observed this correlation for the (experienced) New Zealand participants but not the (inexperienced) San Diego participants. This follows straightforwardly from a model in which activation of *r* flows from the lexical representation, where greater activation of *r* occurs upon perceiving a word that is frequently encountered with *r*. While we do not have appropriate data to investigate the distribution of *r* across our stimulus words in production, that the correlation is absent for the San Diego participants reinforces the interpretation that it emerges from previous experience with the words. For listeners with exposure to *r*-sandhi, we might also expect a correlation between responses to *r*-ful and *r*-less forms, even for words they have never encountered before (e.g. *seesawing*). This would occur through a process of analogy to existing words (e.g. *seesaw*) for which they have lexically specific representations.

We interpret these lexically specific probabilities as arising from an episodic lexicon, which is consistent with a range of other results indicating the presence of phonetic detail in the lexicon. However, such effects are also consistent with any model where probabilistic information about the link between a sound and a word is stored. Indeed, once a large enough number of probabilities about sounds and contexts are stored, the predictions of such models essentially become indistinguishable from those of exemplar-based models, as pointed out by others (Pierrehumbert 2002:113, Ernestus 2014:38). What is important in accounting for this particular subset of our results is that the degree of experience of *r* in a GIVEN LEXICAL ITEM is encoded.

7.2. AN EXPERIENCE-BASED ABSTRACT /r/. We assume that stored episodic memories are indexed with rich social and linguistic information. When robust patterns exist across such stored associations, more abstract linguistic and social generalizations can emerge. A particularly crucial level of abstract representation is the phoneme, a level that is especially relevant for phoneme-monitoring tasks. The process of phoneme monitoring can be conceived of as the act of checking the activation of the abstract representation of /r/ and answering 'yes' when the activation exceeds some threshold. Thus, the phoneme-level representation may be more involved in this particular task than it is in everyday speech perception, which does not involve the artificial orientation toward this level.

We assume that a phonemic representation emerges as an abstract generalization over the contents of the lexicon, and that these phoneme-level representations are themselves probabilistic. The abstract /r/ is associated with different lexical items to different degrees. For some words (e.g. *sleep*), the association is essentially absent, whereas for others (e.g. *rabbit* and *carry*), it is extremely strong. For words such as *saw* and *soar*, the association is (for some individuals) more variable, and its strength depends on the speaker-hearer's prior experience with the sound.

Activation of an abstract phoneme-level representation occurs when activation of other levels (e.g. those containing acoustic information) flows to it. The presence of acoustic material matching previous experience of /r/ is, of course, a major contributor to whether the /r/ receives sufficient activation. Another is the degree of activation received from any activated lexical representation, as described in §7.1. However, many of our results demonstrate that the presence of [ɹ] in the acoustic signal and the lexical representation are not the only factors influencing activation of the phoneme-level representation.

Activation of the /r/ is also influenced by a variety of contextual factors—contextual generalizations associated with the /r/, representing different probabilities across different social and linguistic contexts. Indeed, the very nature and history of the *r*-sandhi system indicates the involvement of some level of abstraction and generalization. In the latter part of the nineteenth century in NZE, rhoticity declined, with the loss of *r* much more frequent before consonants than vowels. This distribution, like many sociolinguistic variables, can be described as a generalization or statistical 'rule' promoting *r*-loss more in certain contexts than in others. Once the difference in probabilities between preconsonantal and prevocalic environments became relatively marked, it acquired the status of an alternation, setting the stage for the abstract generalization to affect words, such as *clawing*, that did not historically contain *r*. Note that in linking words, the probability of *r* before a vowel remained high and did not dramatically change over this period. Therefore, the high probability of prevocalic *r* alone cannot explain the emergence of *r* in intrusive environments. It was only the evolution of a higher-level generalization, reflecting an alternation, that triggered the emergence of intrusive-*r*. Preconsonantally, *r* now remains categorically absent, again supporting the operation of a statistical generalization prohibiting its emergence in this context.

Recent developments in the *r*-sandhi system also support a relatively abstract analysis. Linking- and intrusive-*r* occur following nonhigh vowels, or vowels with nonhigh offglides. Until recently, this prohibited the emergence of *r* after /au/ (as in *plough*). However, for speakers with a relatively monophthongal production, the offglide is no longer high, and this has triggered the application of intrusive-*r* (Hay & Maclagan 2010). Again, this process is well modeled by assuming that the monophthongization of /au/ makes the vowel subject to a relatively abstract generalization relating to nonhigh vowels.

We assume this 'rule' resides at the phoneme level and consists of a set of probabilities conditioning the realization of *r* in different phonological and morphological contexts. These probabilities are socially indexed: the probability of *r* emerging in a sandhi context is higher for male speakers, for example. The probability of *r* emerging in a nonprevocalic environment is higher for speakers of North American English, and, of course, the probabilities differ across groups and individuals depending on the particular distributions they have been exposed to. It is likely that the relatively abstract contextual and social generalizations linked to the phoneme level most often arise as generalizations over the lexicon. The probabilities are updated on an ongoing basis so that, as experience with a variable increases, the probabilities are more accurately tuned to the distribution of the variable across linguistic environments and society. However, for some phonemic representations, the social indexing may not derive directly from experience, instead coming from stereotypes generated by metalinguistic discussion.

By using the word 'rule', we do not mean to set up a contrast between probabilistic phonological rules and the emergence of online analogical generalizations. Indeed, we believe rules emerge as analogical processes over the lexicon; in regard to *r*-sandhi, there is evidence that the system has been influenced by online analogical pressure between *r*-less and *r*-ful forms in the lexicon (Sóskuthy 2013).

Whether through 'rule' or 'analogy', listeners are highly sensitive to statistical generalizations about phonemic patterns. Furthermore, listeners are able to alter their expectations regarding these patterns relatively quickly, as evidenced by the rapidly changing behavior of the San Diegans in our experiment.

7.3. Salience. As outlined above, an adequate account of our results involves both a rich lexicon that encodes the probability of *r*-production at a lexical level, and an abstract representation of a phoneme, which is associated with probabilities encoding the likelihood of the phoneme surfacing across different linguistic and social contexts. These levels are activated in parallel, and feedback flows between them. In our phoneme-monitoring task, participants answer 'yes' when activation of the /r/ exceeds a certain threshold. The factors contributing to this include evidence from the acoustic signal and prior experience (influencing activation at the lexical level via prior exposure to words, as well as activation at the phoneme level via probabilistic abstractions).

However, listeners with little experience with the relevant variable have not generated a distribution of associated probabilities prior to participating in the task, and their corresponding lexical representations indicate a near-zero association with the phoneme.

After exposure to intrusive-*r* during our phoneme-monitoring task, the acoustic detail in the signal is sufficient to trigger activation of /r/ for the San Diego participants. This phonemic activation runs directly counter to their prior experience with the words, however, resulting in a mismatch. When there is a mismatch between the probabilities that are contributed from different sources, there is surprisal. Surprisal draws attention, and the attention may help to resolve the mismatch. As has been discussed, this is related to a more general bias to attend more heavily to novel items (Tulving et al. 1996).

While the literature on novelty tends to focus on the unexpected presence of a segment, phoneme, or feature, we predicted that the absence of an expected segment could also be salient. Specifically, we predicted that a similar phenomenon would occur when the lexical representation led to strong activation of the /r/ but the acoustic signal did not match this expectation (e.g. the absence of [ɹ] in *store and*), resulting in salience of the sound's absence. The unexpected absence should be somewhat less salient than unexpected presence since listeners regularly encounter the absence of segments in reduced speech (Ernestus 2014); sounds that are frequently reduced are restored by listeners in phoneme-monitoring tasks (Kemps et al. 2004).

Our findings demonstrate that surprisal results in heightened accuracy in the salient zones. As predicted, absence of the sound in a context in which some listeners expect near-categorical presence resulted in the lowest rates of 'hearing' *r* in those contexts for those participants. Conversely, presence of the sound in the context of previous near-categorical absence was also more accurately noticed, resulting in the highest rates of 'hearing' *r*.

Note that we would not want to argue that all mismatches result in surprisal. In order for there to be surprisal (and therefore salience), some minimum threshold of degree of mismatch must presumably be met, and this likely fluctuates depending on the degree to which attention is directed elsewhere. When the signal is degraded (e.g. there is noise such as a cough), expectations from lexical and contextual information play a larger role than the acoustic detail does. When the contextual information is degraded (e.g. no context is provided or it is ambiguous), then the acoustic properties of the signal play a larger role than the context.

For the San Diego participants in our experiment, we assume that—upon encountering the stimulus that caused surprisal—an acoustically rich representation of the word is stored. This lexical representation contributes to the probability of future encounters with the phone in that word and similar words, at least within the course of the experiment. As experience with the sound in a given context increases, the listener updates both the lexicon and the probabilities associated with the representation. This is likely limited to a single talker unless the listener receives exposure to the variable produced by multiple speakers. In the short term (as for the San Diego participants), the effect of experience combines with the effect of attention from the initial mismatch. We suspect that, with ample exposure, a mismatch would eventually no longer occur and the sound would no longer be salient to the listener.

7.4. PHONEMIC RESTORATION. The combined effects of lexical and phonemic representations predict not only that sounds in the signal will be perceived to a degree in line with experience but also that, if expected, sounds will also be perceived when they are not present in the signal. Our results are consistent with this prediction; the New Zealand participants were more likely to hear *r* in linking than intrusive contexts, even when there was no acoustic evidence of [ɹ] in the stimulus. In other words, their likelihood of perceiving a sound when it is absent seems to be driven by their prior experience with the sound in the relevant context. Likewise, the San Diego participants' experiment-internal shift toward hearing the *r* in intrusive contexts even when it was not present in the signal provides evidence that even short-term experience can affect these probabilities.

Thus, our participants 'restored' a phoneme where they were expecting it but where it did not occur in the signal. These results can thus be compared with the wider body of literature on phonemic restoration. Most of this literature concentrates on phonemes that tend to be categorically present but have been artificially removed from the stimulus. Results show that listeners have a tendency to restore the phoneme and that they do so in proportion with their top-down expectation of that phoneme occurring in the relevant context. For example, restoration is more likely in words than nonwords, and the likelihood of restoration increases if the word is primed (Samuel 1981).

Despite the fact that the phonemic-restoration literature deals with phonemes that are expected to be categorically present, the absence of the phoneme does not appear to be salient. This is because, in most of this work, its absence has been masked by noise. The presence of noise lowers the degree of mismatch between the phonetic signal and the probability of the phoneme in the word, biasing perception toward the sound even though it is not acoustically present in the signal. We predict that phonemic-restoration experiments that do not contain noise may well lead to salient nonrestoration, as was the case for our San Diego listeners.

The finding that listeners perceive segments that are not there is also related to the literature on perceptual illusions of vowels (Dupoux et al. 1999, Monahan et al. 2009). This literature shows that listeners perceive vowels that are absent from the acoustic signal in contexts where the absence of the vowel would violate their native phonotactics. Such results have led to proposals that loanword-adaptation patterns stem from restorations that take place in perception (Peperkamp & Dupoux 2003). Our work shows that 'perceptual illusions' may be quite general, extending beyond adaptations across languages or restoration of phonemes that have been noise-masked.

In addition to restoration of an absent segment, the level of accuracy at perceiving *r* when it is present is not particularly high. The influence of expectation and experience is just as important as the acoustic material in driving New Zealanders' responses.

These sizable gaps between what is in the signal and what people report hearing introduce interesting implications for the production/perception loop. Our own account of the phenomena relies on the participants' past experience with *r* heavily driving their perception. But what exactly does this experience look like, and how does it affect representation (and thus future production)? Consider a listener who responds 'no' to an [ɹ]-ful form in our experiment (such as *claw-[ɹ]-ing*). Is this form then encoded as fully *r*-less, with no link to an /r/ phonemic level? Does the remembered acoustic profile of this token contribute to the level of *r* in this word, or is this unattended detail not encoded in the remembered token? Likewise, does a restored *r* contribute to the accumulated experience of *r*-fulness for a particular word to the same degree as an acoustically present *r*? Does a form that is highly salient due to novelty bias have a disproportionately large effect on subsequent production (cf. Chang 2013)? These are open questions; exploring them will shed light on what precisely is heard and stored and will, therefore, help us understand what might affect subsequent production and perception patterns.

8. Conclusion. Our results show that, across words, contexts, and individuals, greater experience with *r* leads to increased expectation of *r*. Increased expectation leads to more accurate perception when [ɹ] is present, and a higher rate of 'restoration' when [ɹ] is absent. These effects emerge from the combined effects of probabilistic/episodic representations in the lexicon and generalized probabilistic information at the phonemic level.

We have also shown that there is not a monotonically linear relationship between experience and perception. In cases where there is an extreme mismatch between the expectation generated from the lexicon and the phonemic activation triggered by the signal, surprisal occurs. Surprisal results in heightened salience of the sound or the absence of the sound, which is manifested as higher perceptual accuracy. Prior experience is central to understanding how and why the sound is salient in certain contexts for certain listeners.

Many questions are raised by this work. What types of variables and conditions lead variants (or their absence) to become particularly salient? To what degree do the patterns observed in phoneme monitoring extend to tasks and contexts where the 'phoneme' is not placed front and center? In what ways do individual patterns of (mis)perception perturb an individual's experience beyond the distributions objectively observed in the acoustic signal? Whatever the answers are to the above questions, it is clear that individuals' past experience, across words and linguistic contexts, is strongly, probabilistically, and nonlinearly connected to their patterns of perception.

Appendix A: Experiment 1 stimuli

Experiment 1 test items

linking word/thought	**linking morpheme/thought**
store and keep	storing and keeping
snore and sleep	snoring and sleeping
score and count	scoring and counting
moor and stay	mooring and staying

bore and excite	boring and exciting
explore and find	exploring and finding
mentor and teach	mentoring and teaching
adore and love	adoring and loving
ignore and avoid	ignoring and avoiding
implore and beg	imploring and begging

linking word/schwa	**linking morpheme/schwa**
smoulder and glow	smouldering and glowing
solder and weld	soldering and welding
answer and ask	answering and asking
teeter and fall	teetering and falling
whisper and shout	whispering and shouting
glitter and shine	glittering and shining
holler and yell	hollering and yelling
enter and exit	entering and exiting
splutter and cough	spluttering and coughing
clamber and climb	clambering and climbing

linking word/START	**linking morpheme/START**
char and blacken	charring and blackening
guitar and sing	guitarring and singing
star and lead	starring and leading
car and bus	carring and bussing
mar and spoil	marring and spoiling
tar and seal	tarring and sealing
bar and lock	barring and locking
par and win	parring and winning
scar and heal	scarring and healing
spar and fight	sparring and fighting

intrusive word/THOUGHT	**intrusive morpheme/THOUGHT**
claw and pick	clawing and picking
gnaw and bite	gnawing and biting
chainsaw and cut	chainsawing and cutting
seesaw and slide	seesawing and sliding
declaw and spay	declawing and spaying
unthaw and cook	unthawing and cooking
guffaw and laugh	guffawing and laughing
bandsaw and file	bandsawing and filing
thaw and melt	thawing and melting
heehaw and neigh	heehawing and neighing

intrusive word/START	**linking word/schwa**	**intrusive word/schwa**
Wichita and Texas	nectar and jam	Anna and Matt
taniwha and ghost	oyster and shell	cola and juice
Pakeha and Asian	pepper and salt	comma and stop
Utah and Nevada	poker and snap	fauna and plants
Omaha and Sandspit	sister and mum	guava and peach
Panama and Suez	sugar and spice	llama and goat
Whangamata and Whakatane	tiger and cat	kava and gin
gaga and senile	badger and mole	soda and lime
blah and sleepy	cedar and oak	sofa and couch
la-de-da and posh	colour and shape	tuna and cod

EXPERIMENT 1 FILLER ITEMS

orthographic and realized

equip and rig	equipping and rigging	boat and raft
sleep and rest	sleeping and resting	cloth and rag

wake and rouse	waking and rousing	Helen and Ray
speak and read	speaking and reading	clouds and rain
clean and rinse	cleaning and rinsing	chest and ribs
talk and react	talking and reacting	record and CD
tell and recite	telling and reciting	reggae and dub
compact and reduce	compacting and reducing	rice and salt
listen and reply	listening and replying	rifle and pistol
think and reflect	thinking and reflecting	rose and tulip
run and jump	running and jumping	rent and cost
ring and call	ringing and calling	ridge and hill
reach and hold	reaching and holding	rind and peel
rise and fall	rising and falling	rock and pebble
risk and win	risking and winning	rope and pulley
rot and decay	rotting and decaying	meeting and forum
rent and lease	renting and leasing	peacock and parrot
row and sail	rowing and sailing	Sue and Mary
ruffle and upset	ruffling and upsetting	Dave and Aaron
regain and awaken	regaining and awakening	apple and carrot
lend and borrow	lending and borrowing	swede and marrow
hold and carry	holding and carrying	tui and sparrow
dig and burrow	digging and burrowing	William and Harry
leave and maroon	leaving and marooning	lamb and curry
tick and correct	ticking and correcting	bush and forest
think and worry	thinking and worrying	carob and chocolate
speak and orate	speaking and orating	Karen and Joanna
sing and carol	singing and carolling	fairy and angel
explode and erupt	exploding and erupting	berries and milk
upset and harrow	upsetting and harrowing	Murray and James
marry and wed	marrying and wedding	
marinate and cook	marinating and cooking	
narrate and talk	narrating and talking	
direct and lead	directing and leading	
orientate and find	orientating and finding	
hurry and pace	hurrying and pacing	
irritate and annoy	irritating and annoying	
forage and hunt	foraging and hunting	
erase and clean	erasing and cleaning	
erode and fall	eroding and falling	

only orthographic

bark and cough	barking and coughing	arch and steeple
cart and tow	carting and towing	carbon and soot
charm and smile	charming and smiling	woodspark and kindling
bargain and exchange	bargaining and exchanging	cork and bottle
unfarm and city	unfarming and citying	hornet and wasp
scorn and hate	scorning and hating	acorn and oak
thwart and avoid	thwarting and avoiding	desert and oasis
snort and chuckle	snorting and chuckling	lantern and candle
cordon and detatch	cordonning and detatching	cavern and cave
divorce and leave	divorcing and leaving	leopard and lion
pertain and decide	pertaining and deciding	
forgive and love	forgiving and loving	
survive and live	surviving and living	
comfort and help	comforting and helping	
pattern and design	patterning and designing	

not orthographic nor realized

blast and explode	blasting and exploding	twit and fool
bleach and dye	bleaching and dying	tube and paste

boost and help
box and post
call and talk
clap and stamp
deal and play
deign and decide
fail and pass
fight and hit
flog and beat
floss and clean
gibe and tack
haul and heave
joust and stab
combine and join
compost and tidy
dissect and analyse
elope and escape
unhuddle and leave
imbibe and excite
jockey and win
lemon and season
needle and sew
vanish and leave

boosting and helping
boxing and posting
calling and talking
clapping and stamping
dealing and playing
deigning and deciding
failing and passing
fighting and hitting
flogging and beating
flossing and cleaning
gibing and tacking
hauling and heaving
jousting and stabbing
combining and joining
composting and tidying
dissecting and analysing
eloping and escaping
unhuddling and leaving
imbibing and exciting
jockeying and winning
lemoning and seasoning
needling and sewing
vanishing and leaving

tweed and cotton
sweat and blood
swab and needle
ode and song
midge and flea
lip and nose
gull and stilt
fizz and wine
beagle and lab
benzine and gas
chaos and quiet
facade and mask
Cletus and Mabel
gossip and talk
hyphen and comma
lemming and mouse
puffin and pop
taffy and sweets

Appendix B: Experiment 2 stimuli

Experiment 2 test items

linking word/THOUGHT
store and keep
snore and sleep
score and count
moor and stay
bore and excite
explore and find
mentor and teach
adore and love
ignore and avoid
implore and beg

intrusive word/START
Wichita and Texas
taniwha and ghost
Pakeha and Asian
Utah and Nevada
Omaha and Sandspit
Panama and Suez
Whangamata and Whakatane
gaga and senile
blah and sleepy
la-de-da and posh

intrusive word/schwa
Anna and Matt
cola and juice
comma and stop
fauna and plants
guava and peach
llama and goat
kava and gin
soda and lime
sofa and couch
tuna and cod

intrusive word/THOUGHT
claw and pick
gnaw and bite
thaw and melt
heehaw and neigh
chainsaw and cut
seesaw and slide
declaw and spay
unthaw and cook
guffaw and laugh
bandsaw and file

intrusive morpheme/THOUGHT
clawing and picking
gnawing and biting
thawing and melting
heehawing and neighing
chainsawing and cutting
seesawing and sliding
declawing and spaying
unthawing and cooking
guffawing and laughing
bandsawing and filing

Experiment 2 filler items

orthographic and realized
rent and cost
rice and salt
rind and peel
rise and fall
rising and falling
risk and win

Murray and James
berries and milk
carob and chocolate
direct and lead
directing and leading
erase and clean

rock and pebble
rope and pulley
row and sail
run and jump
running and jumping
Helen and Ray
boat and raft
chest and ribs
clouds and rain
compact and reduce
equip and rig
listen and reply
sleep and rest
sleeping and resting
speak and read
speaking and reading
think and reflect
thinking and reflecting
fairy and angel
forage and hunt
foraging and hunting
hurry and pace
hurrying and pacing
narrate and talk
Sue and Mary
bush and forest
explode and erupt
exploding and erupting
hold and carry
lamb and curry
lend and borrow
lending and borrowing
meeting and forum
peacock and parrot
sing and carol
swede and marrow
think and worry

only orthographic

acorn and oak
arch and steeple
bargain and exchange
bark and cough
carbon and soot
cavern and cave
charm and smile
charming and smiling
comfort and help
comforting and helping
cork and bottle
desert and oasis
forgive and love
forgiving and loving
hornet and wasp
lantern and candle
leopard and lion
pattern and design
pertain and decide
pertaining and deciding
scorn and hate
scorning and hating
snort and chuckle
survive and live
woodspark and kindling

not orthographic nor realized

beagle and lab
blast and explode
blasting and exploding
bleach and dye
boost and help
boosting and helping
call and talk
chaos and quiet
combine and join
dissect and analyse
dissecting and analysing
elope and escape
fail and pass
fizz and wine
gossip and talk
hyphen and comma
joust and stab
jousting and stabbing
swab and needle
sweat and blood
taffy and sweets
twit and fool
vanish and leave
vanishing and leaving
facade and mask

Appendix C: Experiment 2 production elicitation sentences

Intrusive word

Hamsters gnaw and bite on a daily basis.
The lions tried to claw and bite their way through the tough outer layer.
Horses are boring; all they do is hee-haw and neigh.
The ice had already begun to thaw and melt.
I’m only visiting Utah and Nevada.
Both the Panama and the Suez canals are so large they divide continents.

I've just been feeling so blah and sleepy all day.
Whangamata and Tauranga are great places to go in summer.
The fauna and plants of the island are amazing.
I've never seen so much of Anna and Matt.
She takes her gin with soda and lime.
Guava and peach jam is my favourite.

Intrusive morpheme
The cat was clawing at the edge of the carpet.
The horse was heehawing all night.
I can't imagine chainsawing such a massive piece of wood.
The dog was gnawing on his bone.

Linking word
I absolutely adore and love everything about it.
We explore and find things together.
Please be nice; don't make me implore and beg.
My Mum likes to store and keep EVERYTHING.

Southland rhoticity sentences
The nu**r**se was still in training.
Fi**r**e-fighte**r**s are very ha**r**d wo**r**ke**r**s.

REFERENCES

AWH, EDWARD; ARTEM V. BELOPOLSKY; and JAN THEEUWES. 2012. Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences* 16.437–43. DOI: 10.1016/j.tics.2012.06.010.

BARTLETT, CHRISTOPHER. 2002. *The Southland variety of English: Postvocalic /r/ and the BATH vowel*. Dunedin: University of Otago dissertation.

CHANG, CHARLES B. 2013. A novelty effect in phonetic drift of the native language. *Journal of Phonetics* 41.520–33. DOI: 10.1016/j.wocn.2013.09.006.

CHO, TAEHONG, and JAMES M. MCQUEEN. 2011. Perceptual recovery from consonant-cluster simplification in Korean using language-specific phonological knowledge. *Journal of Psycholinguistic Research* 40.253–74. DOI: 10.1007/s10936-011-9168-0.

CONNINE, CYNTHIA M., and DEBRA TITONE. 1996. Phoneme monitoring. *Language and Cognitive Processes* 11.635–46. DOI: 10.1080/016909696387042.

CUTLER, ANNE. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

CUTLER, ANNE; FRANK EISNER; JAMES M. MCQUEEN; and DENNIS NORRIS. 2010. How abstract phonemic categories are necessary for coping with speaker-related variation. *Laboratory phonology 10*, ed. by Cécile Fougeron, Barbara Kuehnert, Mariapaola D'Imperio, and Nathalie Vallée, 91–111. Berlin: De Gruyter Mouton.

DAVIS, MATTHEW H., and INGRID S. JOHNSRUDE. 2007. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research* 229.132–47. DOI: 10.1016/j.heares.2007.01.014.

DUPOUX, EMMANUEL; KAZUHIKO KAKEHI; YUKI HIROSE; CHRISTOPHE PALLIER; and JACQUES MEHLER. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25.1568–78. DOI: 10.1037/0096-1523.25.6.1568.

ERNESTUS, MIRJAM. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142.27–41. DOI: 10.1016/j.lingua.2012.12.006.

EULITZ, CARSTEN, and ADITI LAHIRI. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16.577–83. DOI: 10.1162/089892904323057308.

FOULKES, PAUL. 1997. English [r]-sandhi—A sociolinguistic perspective. *Histoire Epistemologie Langage* 19.73–96. DOI: 10.3406/hel.1997.2573.

GIBSON, ANDY. 2016. Samoan English in New Zealand: Examples of consonant features from the UC QuakeBox. *New Zealand English Journal* 29&30.25–50.

Goldinger, Stephen D. 1997. Words and voices: Perception and production in an episodic lexicon. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 33–66. San Diego: Academic Press.

Goldinger, Stephen D. 2007. A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, 49–54. Online: http://www.icphs2007.de/conference/Papers/1781/1781.pdf.

Gordon, Elizabeth; Margaret Maclagan; and Jennifer Hay. 2007. The ONZE corpus. *Creating and digitizing language corpora, vol. 2: Diachronic databases*, ed. by Joan C. Beal, Karen P. Corrigan, and Hermann L. Moisl, 82–104. Dordrecht: Springer.

Harris, John. 1994. *English sound structure*. Oxford: Blackwell.

Hay, Jennifer, and Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92.298–330. DOI: 10.1353/lan.2016.0036.

Hay, Jennifer, and Margaret Maclagan. 2010. Social and phonetic conditioners on the frequency and degree of intrusive /r/ in New Zealand English. *A reader in sociophonetics*, ed. by Dennis Preston and Nancy Niedzielski, 41–69. New York: De Gruyter Mouton.

Hay, Jennifer, and Margaret Maclagan. 2012. /r/-sandhi in early 20th century New Zealand English. *Linguistics* 50.745–63. DOI: 10.1515/ling-2012-0023.

Hay, Jennifer, and Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 81.799–823. DOI: 10.1353/lan.2005.0175.

Hay, Jennifer; Paul Warren; and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34.458–84. DOI: 10.1016/j.wocn.2005.10.001.

Ingram, John C. L., and See-Gyoon Park. 1997. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25.343–70. DOI: 10.1006/jpho.1997.0048.

Jaeger, T. Florian, and Kodi Weatherholtz. 2016. What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology* 7:1115. DOI: 10.3389/fpsyg.2016.01115.

Jarvis, B. G. 2006. DirectRT (Version 2006). [Computer software.] New York: Empirisoft Corporation. Online: http://www.empirisoft.com/directrt.aspx.

Johnston, William A., and Irene S. Schwarting. 1997. Novel popout: An enigma for conventional theories of attention. *Journal of Experimental Psychology: Human Perception and Performance* 23.622–31. DOI: 10.1037/0096-1523.23.3.622.

Kemps, Rachèl; Mirjam Ernestus; Robert Schreuder; and Harald Baayen. 2004. Processing reduced word forms: The suffix restoration effect. *Brain and Language* 90.117–27. DOI: 10.1016/S0093-934X(03)00425-5.

Kleinschmidt, Dave F., and T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122.148–203. DOI: 10.1037/a0038695.

Kraljic, Tanya, and Arthur G. Samuel. 2005. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51.141–78. DOI: 10.1016/j.cogpsych.2005.05.001.

Kraljic, Tanya, and Arthur G. Samuel. 2007. Perceptual adjustments to multiple speakers. *Journal of Memory and Language* 56.1–15. DOI: 10.1016/j.jml.2006.07.010.

Labov, William. 1994. *Principles of linguistic change, vol. 1: Internal factors*. Oxford: Blackwell.

Magnuson, James S.; Bob McMurray; Michael K. Tanenhaus; and Richard N. Aslin. 2003. Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science* 27.285–98. DOI: 10.1207/s15516709cog2702_6.

Mann, Virginia A., and Bruno H. Repp. 1981. Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America* 69.548–58. DOI: 10.1121/1.385483.

Maye, Jessica; Richard N. Aslin; and Michael K. Tanenhaus. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science* 32.543–62. DOI: 10.1080/03640210802035357.

McCarthy, John J. 1993. A case of surface constraint violation. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 38.169–95. DOI: 10.1017/S0008413100014730.

McMahon, April. 2000. *Lexical phonology and the history of English*. Cambridge: Cambridge University Press.

Mitterer, Holger, and James M. McQueen. 2009. Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance* 35.244–63. DOI: 10.1037/a0012730.

Mompeán, Jose A., and F. Alberto Gómez. 2011. Hiatus resolution strategies in non-rhotic English: The case of /r/-liaison. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 1414–17. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Mompean/Mompean.pdf.

Monahan, Philip J.; Eri Takahashi; Chizuru Nakao; and William J. Idsardi. 2009. Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. *Japanese/Korean linguistics 17*, ed. by Shoichi Iwasaki, Hajime Hoji, Patricia Clancy, and Sung-Ock Sohn, 391–405. Stanford, CA: CSLI Publications.

Norris, Dennis; James M. McQueen; and Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47.204–38. DOI: 10.1016/S0010-0285(03)00006-9.

Orgun, Cemil Orhan. 2001. English *r*-insertion in optimality theory. *Natural Language and Linguistic Theory* 19.737–49. DOI: 10.1023/A:1013313827798.

Pavlík, Radoslav. 2016. A usage-based account of /r/-liaison in Standard British English. *Journal of Phonetics* 54.109–22. DOI: 10.1016/j.wocn.2015.10.001.

Peperkamp, Sharon, and Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, 367–70. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_0367.html.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101–39. New York: De Gruyter.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34.516–30. DOI: 10.1016/j.wocn.2006.06.003.

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2.33–52. DOI: 10.1146/annurev-linguistics-030514-125050.

Samuel, Arthur G. 1981. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General* 110.474–94. DOI: 10.1037/0096-3445.110.4.474.

Sóskuthy, Márton. 2013. Analogy in the emergence of intrusive-*r* in English. *English Language and Linguistics* 17.55–84. DOI: 10.1017/S1360674312000329.

Stemberger, Joseph Paul, and Brian MacWhinney. 1986. Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition* 14.17–26. DOI: 10.3758/BF03209225.

Strand, Elizabeth A., and Keith Johnson. 1996. Gradient and visual speaker normalization in the perception of fricatives. *Natural language processing and speech technology: Results of the 3rd KONVENS conference*, ed. by Dafydd Gibbon, 14–26. Berlin: Mouton de Gruyter.

Summerfield, Christopher, and Tobias Egner. 2009. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences* 13.403–9. DOI: 10.1016/j.tics.2009.06.003.

Sumner, Meghan, and Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60.487–501. DOI: 10.1016/j.jml.2009.01.001.

Tan, Ying Ying. 2011. To *r* or not to *r*: A sociophonetic analysis of /ɹ/ in Singapore English. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, 1954–57. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Tan/Tan.pdf.

Tuinman, Annelie; Holger Mitterer; and Anne Cutler. 2011. Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *The Journal of the Acoustical Society of America* 130.1643–52. DOI: 10.1121/1.3619793.

Tuinman, Annelie; Holger Mitterer; and Anne Cutler. 2012. Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language* 66.530–44. DOI: 10.1016/j.jml.2012.02.001.

Tulving, Endel; Hans J. Markowitsch; Fergus I. M. Craik; Reza Habib; and Sylvain Houle. 1996. Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex* 6.71–79. DOI: 10.1093/cercor/6.1.71.

Uffman, Christian. 2007. Intrusive [r] and optimal epenthetic consonants. *Language Sciences* 29.451–76. DOI: 10.1016/j.langsci.2006.12.017.

Vitevitch, Michael S.; Paul A. Luce; Jan Charles-Luce; and David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40.47–62. DOI: 10.1177/002383099704000103.

Walker, Abby, and Jennifer Hay. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2.219–37. DOI: 10.1515/labphon.2011.007.

Warner, Natasha, and Andrea Weber. 2001. Perception of epenthetic stops. *Journal of Phonetics* 29.53–87. DOI: 10.1006/jpho.2001.0129.

Weber, Andrea. 2001. Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech* 44.95–118. DOI: 10.1177/00238309010440010401.

Weber, Andrea. 2002. Assimilation violation and spoken-language processing: A supplementary report. *Language and Speech* 45.37–46. DOI: 10.1177/00238309020450010201.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Wurm, Lee H., and Arthur G. Samuel. 1997. Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language* 36.165–87. DOI: 10.1006/jmla.1996.2482.

Hay and Gibson
New Zealand Institute of Language, Brain and Behaviour,
and Department of Linguistics
University of Canterbury
Private Bag 4800
Christchurch, New Zealand
[jen.hay@canterbury.ac.nz]
[andy.gibson@pg.canterbury.ac.nz]

Drager
Department of Linguistics
University of Hawai‘i at Mānoa
561 Moore Hall
1890 East-West Road
Honolulu, HI 96822
[kdrager@hawaii.edu]

[Received 14 August 2011;
revision invited 24 July 2012;
revision received 4 August 2016;
revision invited 30 November 2016;
revision received 9 March 2017;
accepted with revisions 19 June 2017;
accepted 7 August 2017]