Experiences and Expectations in the Perception of Speech

Merten Wiltshire

ABSTRACT. A phoneme-monitoring experiment was conducted in which participants were required to listen for instances of /ɹ/ distributed variably in four different voices. Listeners had different expectations of the occurrence of /ɹ/ in different contexts in each voice, reflecting their experiences with the varieties these voices represented and their beliefs regarding which varieties they were hearing. Listeners' perception of /ɹ/ was strongly conditioned by their expectations, and was also subject to change through experience over the course of the experiment. The study offers tentative evidence that listeners' expectations condition and limit the malleability of their perception, whilst demonstrating listeners' baseline responsiveness to the actual content of language stimuli, independent of the effects of their beliefs and experiences. It is argued that these findings illustrate the importance of bringing elements of predictive processing into usage-based models of language, capturing linguistic knowledge as emerging through the interaction of individuals' experiences with, and beliefs concerning, the language in which they are immersed.

Acknowledgements

I would above all like to thank my supervisors, Professor Jennifer Hay and Dr Daniel Villarreal, for their continuing advice, support, and contributions, in providing the original impetus for this study and many of the ideas developed therein; and would also like to thank the members of the New Zealand Institute of Language, Brain, & Behaviour, for comments offered on earlier presentations of this research; all those who helped in the recruitment of participants, and in particular Hielke Vriesendorp (University of Sheffield), without whom the experiment could not have been conducted; and all participants in the experiment themselves.

Contents

1. Introduction

As knowledge of language has come to be seen as a reflection of real language use in the world, the study of how language is perceived and interpreted has become increasingly integral to theoretical linguistics. Proponents of models of language that assume individuals have rich and detailed memories of language use on which to draw (e.g. Johnson, 1997) have long made the claim that perception does not, as previously assumed, involve an immediate process of normalisation and abstraction away from a stimulus. It is posited instead that speakers are also active listeners who remember different voices and incidents of speech separately and with different associations (Docherty & Foulkes, 2014). As a result, speakers' drawing upon this variable linguistic knowledge emerges in the variation found in all language use.

The implications of such a view of language perception have only more recently begun to be appreciated in their full extent. Following such assumptions, it is permissible and indeed probable that the perception of language is itself as individually variable as its production, likewise conditioned by listeners' exposure to ambient language use. Recent experimental research by Hay, Drager & Gibson (2018) has now shown rather decisively that listeners' past experiences of features of language are highly predictive of the way they perceive these features.

What has been described is still an account of language perception that captures it as a relatively objective process. Listeners' present expectations of language are tied to their past experiences of language use, only insofar as they have objectively, if variably, internalised memories of certain features to different degrees, and match instances of the same or similar features to these memories under new exposure.

It is proposed here that this view of language knowledge can be taken further, by more fully appreciating that our perception of language is profoundly *subjective*. Following earlier suggestions by Grossberg (2003) and Hawkins (2010) regarding the importance of listener expectations in speech perception, and growing empirical support for this position (e.g. van Petten & Luka, 2012; Brunellière & Soto-Faraco, 2013; Lev Ari, 2015; Kroczek & Gunter, 2017; Gnevsheva, 2018), it is argued that all linguistic knowledge, as indeed all knowledge, depends upon the arbitrary categorisation of what we experience in the world around us, in a manner conditioned by expectations that reflect not only what we have objectively heard in the past, but also what we believe we ought to be hearing in the present moment. Expectations are

based on any number of subjective associations we draw in relation to the voice we are hearing, the context in which we are hearing it, or the content of what we are hearing. The endless variation in language thus follows not only from variability in natural constraints, nor even from variability in objective memories of language use, but also from our variable interpretations of these experiences, a reflection of our unique experiences of the world.

This claim follows from experimental research suggesting that listeners' perception of linguistic features heard in different voices is complexly conditioned by the interacting influences of what they have heard in the past, what they believe they ought to be hearing, and what they are indeed hearing. It is in teasing apart these three elements that form our perception – experiences, expectations, and stimuli – that a better understanding of the nature of language perception, and the individual knowledge of language, can be reached.

In order to address this question, a phoneme[1] monitoring experiment was conducted in the present study in which a group of listeners from a New Zealand English language background were played stimuli containing non-prevocalic and intrusive /ɹ/ in voices of four speakers, from Canterbury and Southland in New Zealand and Delaware and Rhode Island in the United States. Listeners proved to have extensive experience with, or at least a strong idea of the Canterbury (General New Zealand) and Delaware (General American) voices, but initially interpreted the Southland voice as an 'ordinary' New Zealand accent, and were challenged in identifying the Rhode Island speaker. Where they had strong, well-founded expectations, listeners displayed no tendency to adjust their responses over time. Where their expectations were weak or proved ill-founded based on evidence from new exposure, their perception of these features proved malleable, and featured either a close estimation or an over-prediction of the actual occurrence of non-prevocalic and intrusive /ɹ/ in these voices. Where there was a significant mismatch between expectations and new experience, listeners' responses changed significantly over time; otherwise, such changes were absent. A general tendency towards greater accuracy and less within-voice learning through exposure to greater variability was also identified, as well as baseline responsiveness to the actual content of the stimuli.

---

[1] The term 'phoneme monitoring' is used throughout because it is the established term for this methodology in the literature (e.g. Connine & Titone, 1996). This should not be taken, however, to suggest that it is any way assumed here that listeners' judgements in experiments of this kind do in fact reflect a knowledge of formal phonological units, as the name would imply.

This study provides preliminary evidence for the importance of understanding of linguistic knowledge as developing out of an interaction between three elements: listeners' expectations, or beliefs, about what they are hearing; their experience with language, which both forms and tests these beliefs; and the features of language that are experienced. While there remains work to be done to better ground this view, it carries the implication that all linguistic knowledge is abstract and subjective, even as it is derived from real-world language use rather than any form of innate knowledge. It is argued that this view of linguistic knowledge is compatible with ideas that have recently attained prominence in other areas of cognitive science, and has meaningful repercussions for the development of usage-based theories of language.

2. Literature

In developing theories of language, the challenge has ever been to reconcile observations of the extraordinary extent and endless variability of individuals' knowledge of language with the intuition that this knowledge is in some way ordered and structured so as not to be prohibitive of communication. The manner in which this issue is approached determines what the perception of language is understood to be. If the emphasis is placed on the primacy of structure over variability, the variable modulation of language through perception is of little import; if structure is understood only to be emergent from variability, then the extent to which perception reflects what is in fact perceived or some subjective interpretation thereof becomes critical to any understanding of language as a whole.

As one path towards the resolution of the variability problem in speech perception, the existence of some form of linguistic knowledge that is abstract and innate has been theorised (e.g. Studdert-Kennedy, 1976). In such a view, there is a tendency to disassociate different forms of language seen at varying levels of granularity, the structural division into phonology, morphology and syntax, for example (e.g. McClelland & Elman, 1986); to disassociate the perception, the knowledge, and the production of language; and to disassociate language itself from its meaning and use, and the functional and social context in which these abide (e.g. Liberman & Mattingly, 1985). Knowledge of language is the knowledge of principles, features, constraints or other distinguishing elements and of the underlying forms of the structures, concepts and sounds they define. Linguistic phenomena follow entirely top-down processes: stimuli are produced and interpreted entirely under the control of pre-existing information, and processing of language always follows a unidirectional, sequential route (e.g. Pisoni &

Sawusch, 1975). Taking this position, the challenge has been to explain away variation, that language might be better understood in isolation from it.

More recently, an entirely different approach to the same issues has been pursued, one that is broadly termed 'usage-based' (e.g. Docherty & Foulkes, 2014). Informed by findings from psychology and sociolinguistics, a theoretical framework has emerged in which the study of variation is vital to the study of language itself. Knowledge of language is conceived of as fundamentally *undifferentiated*, and tied to cognitive processes that of themselves do not privilege language at all (e.g. Langacker, 1987; 1999). Structure emerges from use, acquisition comes through perception (Tomasello, 2000), and functional and social characteristics are as predictive of structure in language as those pressures traditionally described as 'language-internal' that have been a central object of study in theoretical linguistics (Foulkes & Docherty, 2006). In a prominent usage-based approach, that of 'exemplar theory' (e.g. Johnson, 1997; Pierrehumbert, 2001), variation is captured in the form of rich and extensive memories of linguistic experiences. In a usage-based understanding of language, the challenge of variability ceases to be a challenge at all, because variability offers the best evidence of the nature of language itself.

Because of the centrality of use and the primacy of perception, the study of individuals' experiences of language attains a new importance in such a framework. Bottom-up influences of past experiences on linguistic behaviour have been extensively studied in research on how different individuals with diverging long-term experiences of language behave differently (e.g. Evans & Iverson, 2004, 2007; Nycz, 2013), and on how individuals' behaviour changes in the short-term, through 'learning' based on new experiences (e.g. Saffran, Newport, & Aslin, 1996; Clopper & Pisoni, 2004a; Bradlow & Bent, 2008; Kraljic & Samuel, 2011). While some of this literature has focussed on the production of language, it has also been important to understand how the perception of language is itself conditioned by past experiences (Lev-Ari & Peperkamp, 2016; Hay, Drager, & Gibson, 2018; Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018). This provides crucial evidence regarding how variation in perception affects variation in knowledge, and ultimately variation in production.

There is also a growing literature on how the top-down influences of individuals' beliefs concurrently produce variability in perception. Whilst such influences do themselves reflect experiences, findings in this area suggest listeners' experiences are mediated through processes of individual interpretation (e.g. McGowan, 2015; Gnevsheva, 2018; Fiedler, Keller, &

Hanulíková, 2019). Although there has been substantial resistance in the field of psychology to the notion of cognitive influences on perception (e.g. Firestone & Scholl, 2016), an accumulating body of evidence shows that such influences – expectations – are ever-present in some form or another. This literature suggests that the brain above all operates on prediction (e.g. Clark, 2013; Farmer, Brown, & Tanenhaus, 2013; Lupyan, 2015), using interpretations of past experiences to simplify the process of analysing present circumstances.

In light of these findings, it is apparent that there are three principal elements in the perception of language: the objective content of a stimulus, the experiences upon which an individual may draw in response to it, and the expectations they make with regard to its nature. 'Experiences' are considered here to constitute individuals' linguistic knowledge, an assemblage of memories and associations that develop through exposure to language (this corresponds roughly to the distribution of 'exemplars' in an 'exemplar' model). 'Expectations' are defined as individuals' beliefs regarding language they encounter, which are formed through the interpretation of the present context with reference to experiences. Models of language might be better informed by a clearer understanding of how these factors relate.

Relevant to such a question are several issues that will receive attention here: the nature of usage-based approaches to theoretical linguistics, which is discussed in §2.1; the evidence for the role listeners' experiences play in shaping their language, which provides strong support for usage-based approaches, and is dealt with in §2.2; and more recent studies of variability in perception, which suggest that listeners are guided by expectations in processing speech, something detailed in §2.3. The recent study of Hay, Drager, & Gibson (2018), which brings together these issues, is described at length in §2.4, as the present study has been built upon it; sociolinguistic literature relevant to the phenomena and contexts being studied receive attention in §2.5.

2.1. Usage-based models of language

Usage-based models constitute a major departure from previous approaches in theoretical linguistics, offering the potential to explain language with reference primarily to what is known in experimental studies of speech perception and data on real language use, rather than to the assumptions of earlier 'structural' and 'generative' theories of language. The essential principles of any usage-based model are that linguistic knowledge develops from exposure to and use of language, rather than from any innate capacity, and that linguistic knowledge reflects general cognitive processes, undifferentiated from other forms of knowledge. Among usage-

based theories, there is however a considerable diversity of approaches, including 'exemplar' theories in which linguistic knowledge is grounded in specific memories of language, and other models that continue to present linguistic knowledge as entirely abstracted. Whatever the precise approach adopted, following a usage-based model has clear implications for how linguistic knowledge is influenced by the experiences and beliefs of an individual.

An extensive body of evidence, coming largely from experimental research on speech perception, has challenged many of the postulates of previous approaches to theoretical linguistics. This literature has shown that memories of voices influence the processing of other stimuli, in a manner that suggests fine details of auditory stimuli are remembered (e.g. Martin, Mullennix, Pisoni, & Summers, 1989; Mullennix, Pisoni, & Martin, 1989; Goldinger, Pisoni, & Logan, 1991; Nygaard, Burt, & Queen, 2000; Smith & Hawkins, 2012). Knowledge of voice details has been shown to persist long-term (Nygaard, Sommers, & Pisoni, 1994; Goldinger, 1996), and individuals appear to readily learn such details, better comprehending speech in doing so (Nygaard & Pisoni, 1998).

Such evidence suggests that forms of knowledge regarding social information, such as the distinguishing characteristics of a speaker, are not readily dissociable from linguistic knowledge (Pisoni, 1993). It has also been shown that linguistic knowledge may be newly-acquired at any point in the speaker's lifespan, and in the spread of changes, experienced and remembered instances are privileged over others that might be assigned the same abstract category (Goldinger, 2007). As a consequence, there has been a move away from the practice of formal linguistic analysis (e.g. Hawkins, 2003; Port & Leary, 2005) as being ill-supported in experimental evidence (e.g. Ettlinger & Johnson, 2009).

The approach to theoretical linguistics that has emerged in its place is what is commonly termed 'usage-based' linguistics (e.g. Langacker 1987, 1999; Bybee, 1995; Tomasello, 2000; Pierrehumbert, 2001; Docherty & Foulkes, 2014). In 'usage-based' approaches, language is treated as a cognitive phenomenon like any other, developing through experience and use. Usage-based approaches do, however, diverge significantly on many issues. Theoretical approaches within the field range from simple Bayesian models that employ traditional abstract categories such as that of the 'phoneme' (Norris & McQueen, 2008; Kleinschmidt & Jaeger, 2015) to efforts to move entirely beyond analysis of language in terms of discrete, abstract units (Hawkins, 2003, 2010). The approach most often reflected in the speech perception literature is an 'exemplar' or 'episodic' view of linguistic knowledge (e.g. Goldinger, 1997;

Johnson, 1997). This is a model of classification originating in the field of psychology (Medin & Schaffer, 1978; Nosofsky, 1988), in which knowledge is grounded in discrete memories, or 'exemplars', of experienced stimuli, and in which categories only emerge from the similarity of such memories.

Since the inception of 'exemplar theory', it has been applied extensively in investigating phenomena such as the diffusion of language changes and effects of the frequencies of use of stimuli (Pierrehumbert, 2001; Wedel, 2006; Todd, Pierrehumbert, & Hay, 2019), exploring the core assumption of usage-based accounts that the production of language ultimately follows from perception. 'Exemplar' models are also attractive in that they readily accommodate social and contextual information, insofar as such information is remembered together with the speech signal (Foulkes & Docherty, 2006; Docherty, Langstrof & Foulkes, 2013). In a typical 'exemplar' model (e.g. Pierrehumbert, 2001), linguistic knowledge exists in the form of a distribution of memory 'exemplars', where memories that are perceived as more similar are 'closer' together, and memories that are considered more different are further apart. Categories form around 'clouds' of exemplars that are in some way considered to be instances of a category. New exemplars are categorised through comparison with all a category's exemplars (Johnson, 1997). Exemplars are generally treated as having a particular 'strength', reflecting recency and other factors that privilege them in perception and production.

Purely 'exemplar'-based models appear to fall short of being able to account for all generalising phenomena (McQueen, Cutler, & Norris, 2006; Cutler, Eisner, McQueen, & Norris, 2010). In concert with developments in psychology (e.g. Rouder & Ratcliff, 2006) suggesting the necessity of both exemplar-based and more abstract forms of categorisation, there has emerged a consensus in favour of what have been termed 'hybrid models' (Pierrehumbert, 2006; Abbot-Smith & Tomasello, 2006; Goldinger, 2007). These typically contain distinct 'levels' for 'exemplar' and abstract forms of knowledge (Pierrehumbert, 2016; Hay, Drager, & Gibson, 2018). 'Exemplar' knowledge seems to be necessary to account for the complex variation, and abstract knowledge for the often remarkably regular patterning of linguistic phenomena, and in spite of a diversity of approaches, usage-based models of language are increasingly incorporating both.

While a usage-based approach is followed here, agnosticism will be maintained regarding many of the issues surrounding precisely how linguistic knowledge is structured. At the same time, some general points of importance that receive comparatively little attention in the

literature will be emphasised. The first of these points regards the nature of perception. It is assumed here that perception neither follows a sequential process nor operates on distinct 'levels' of information within a stimulus, or with regard to distinct 'levels' of knowledge. This follows the insights of Hawkins (2003, 2010) that perceptual segmentation of a signal is unnecessary for its interpretation, as well as the evidence (van Berkum, 2008; Samuel & Kraljic, 2009) that the human brain operates opportunistically in perceiving evidence, making judgements based on limited cues. If this basic assumption is correct, it is not implausible to posit that individuals rarely listen for or process specific sounds contained in a stimulus.

Following this possibility, a second point is that even as listeners are highly sensitive to variability, abstraction or more precisely generalisation is constantly operating. This reflects a general cognitive impulse towards interpreting the world through the creation of categories (Port & Leary, 2005), and the perceptual parsimony of depending on broader abstractions in repetitive tasks where there is little need for attention to variation (Carr, Brown, & Charalambous, 1989). A third postulate emerges from this, namely that what is attended to and perceived follows from its utility. This is consistent with evidence that variability in stimuli has a comparatively minor effect on perceptual processing where that variability is not considered relevant to the perception of the stimuli (Sommers & Barcroft, 2006). It extends, however, more generally to all aspects of linguistic knowledge, including the knowledge of social information.

A last point regards the locus of linguistic knowledge. It is assumed here that the only form of 'language' that has any meaningful reality of its own is the language of the individual, as alluded to by Docherty & Foulkes (2014). While reference may be made to 'varieties', 'accents', and 'languages', these are all arbitrary, abstracted categories like any other form of linguistic knowledge, and in any analysis, even where data aggregated over multiple individuals are studied, it is important not to lose sight of this essential fact. These are crucial theoretical assumptions that receive less attention within the usage-based paradigm, but which are directly relevant to it.

2.2. Experience and perceptual learning

Central to any usage-based model is the notion that linguistic knowledge is grounded in individuals' experience of the use of language in the world around them. An extensive literature has accumulated demonstrating the importance of learning processes in the development of linguistic knowledge (e.g. Lively, Logan, & Pisoni, 1993; Norris, McQueen, & Cutler, 2003;

Maye, Aslin, & Tanenhaus, 2008), and has shown that learning reflects the statistical properties of ambient stimuli (e.g. Saffran, Newport, & Aslin, 1996). This research has further demonstrated the process by which abstract knowledge might develop from remembered instances of language use (e.g. Clopper & Pisoni, 2004a; Sumner, 2011), and revealed some of the limits to learning (e.g. Kraljic, Samuel, & Brennan, 2008). It is also increasingly clear that everything that is true of the emergence of linguistic structure through experience is also true of the emergence of those aspects of language that have traditionally been treated separately: meaning and social connotations. Much of the variation observed in language use can therefore be accounted for simply as following from individuals' variable exposure to different forms of language.

The discovery of learning phenomena, overturning previous assumptions that knowledge of language was mostly static, has provided the best evidence that this knowledge develops from experience. In an early finding, Lively, Logan, & Pisoni (1993) found that listeners with a Japanese language background, previously tending not to distinguish /l/ and /ɹ/, became consistently more accurate and faster at distinguishing these sounds through active training under exposure to English stimuli both within single experiments and over the course of several weeks of experiments. In other work, it has been shown that in perceptual learning, listeners must continually adjust to changes in the voice being heard (Pisoni, 1993) and that learning varies with attention during exposure (Nygaard & Pisoni, 1998). Saffran, Newport, & Aslin (1996) contributed evidence that learning is based on statistical knowledge. They showed that the presence of variable distributional cues was enough to allow individuals to segment words when exposed to an artificial form of language. More recently, it has been established that statistical learning is indeed a fundamental and independent cognitive function, which is also individually variable (Siegelman & Frost, 2015; Siegelman, Bogaerts, Christiansen, & Frost, 2016). Furthermore, while early studies demonstrating learning effects featured the active training of participants, recent evidence suggests that passive exposure is enough to directly affect language processing in the brain, even among adults (Kurkela, Hämäläinen, Leppänen, Shu, & Astikainen, 2019). The ability to in some manner internalise statistical information regarding ambient stimuli is one of the essential facets of cognition, and underpins all experience-based learning effects.

Another important characteristic of experience-derived effects is that of abstraction, however. It is now well-established that although learning is more difficult in the face of greater variability in experienced stimuli, the same variability facilitates the processing of new stimuli.

Clopper & Pisoni (2004a) found that speaker variability in stimuli during a learning task hindered the learning of new characteristics, but improved perception of these features in stimuli from new speakers. Learning similarly leads to generalisation across variation in sounds, even as improved processing ability is always correlated in a gradient manner with increased similarity between new and previously experienced instances (Lively, Logan, & Pisoni, 1993; Palmeri, Goldinger, & Pisoni, 1993; Kraljic & Samuel, 2006; Dahan & Mead, 2010). In particular, it has been shown that the strength of category learning is associated very directly with absolute variation in stimuli, whether spoken by more or fewer speakers (Sumner, 2011). The implication is that experience with variability increasingly leads to the development of a level of abstract linguistic knowledge, allowing the individual to generalise based on the perceived similarity of what has already been and what is newly experienced (Kleinschmidt & Jaeger, 2015). While the perceptual learning literature provides some of the best evidence for 'exemplar' models of language, it also supports the notion of abstract linguistic knowledge, even if this, too, is in fact learnt.

A separate branch of research has helped to connect findings in sociolinguistics with the perceptual learning literature, supporting the view that there are no qualitative differences between the development of knowledge of language structure and the development of knowledge of meaning, including social meaning. In both cases, linguistic knowledge is entirely founded upon statistical learning through experience, and abstraction of categories from specific instances. Evans & Iverson (2004) studied how listeners variably perceived sounds, depending on whether the listeners were from the south of England, from the north of England but living in the south, or from the north of England and living in the north. Listener behaviour reflected variable experience with southern British features, as well as corresponding to their own variable use. In another study, listeners from southern England and Glasgow were compared, assessing their perception of the speech typical of the other group (Adank, Evans, Stuart-Smith, & Scott, 2009). While the participants from southern England had difficulties processing the Glasgow speech, no such effect was found for the participants from Glasgow when listening to the southern English speech. It is argued that this reflects participants' experience: while the participants from southern England would have had little ambient exposure to Glaswegian speech, the reverse would not be true because of the prevalence of southern English speech both in the media and in certain settings in Glasgow itself.

Effects of experience further interact with the stylistic variation long-studied in the sociolinguistic literature. Nycz (2013) found that participants originating in Canada but

resident in New York produced distinctions between two sounds typically merged in Canadian and unmerged in New York speech, but largely in running speech and not when reading minimal pair words. In the perceptual learning literature, it has been demonstrated that listening to an unfamiliar accent impedes processing (Floccia, Goslin, Girard, & Konopczynski, 2006), but that this improves with exposure, and that greater variability in stimuli again supports the development of a generalised social-indexical category (Bradlow & Bent, 2008). In other research, Lev-Ari (2017, 2018) has argued that social variation influences not only what is perceived, but also the malleability of perception. Specifically, she has presented evidence that perception by individuals with wider social networks is less malleable than perception by individuals with smaller social networks. It is impossible to disassociate the social context in which language is heard from the experience of that language itself.

There are limits to learning effects. As a consequence of abstraction from experience, listeners display significant tolerance for variability, and this reduces the impact of new exposure. Sumner & Samuel (2009) studied the behaviour of listeners from New York who either did or did not tend to have non-prevocalic /ɹ/ in their speech. They found that both groups tended to have perceptual awareness of this feature, in contrast to participants from outside of New York, but that those who did not have /ɹ/-less speech themselves tended to treat /ɹ/-lessness as a variant of an abstract /ɹ/-containing form. Those who tended not to use non-prevocalic /ɹ/ clearly separated the two. Non-/ɹ/ users had apparently learnt two separate abstractions, while the former only perceived one category with reference to the other. Recently, it has been shown that listeners tolerate a remarkable degree of accent variation, absorbing it into existing categories and displaying little learning under new exposure when there is little demand to do so (Shaw, et al., 2018).

Different forms of experience are also more or less likely to produce learning. The experimental evidence seems to indicate that whether or not new experiences produce learning is to some extent determined by the utility of that learning. Kraljic & Samuel (2007) found that where new characteristics were informative of a speaker's identity, they were learnt accordingly, whereas where they were not, any learning that happened occurred independently of the voice being heard. In another study, it was argued that where an unusual feature was experienced as a result of a particular context – in this case, the result of a speaker's having a pen in their mouth (Kraljic, Samuel, & Brennan, 2008) – or where it was predictable based on the linguistic environment, for example always occurring before the same segments (Kraljic, Brennan, & Samuel, 2008), it would not be considered an attribute of a speaker, and would

therefore not be learnt. A further finding was that initial experiences were consequently favoured relative to new experiences, as these would be learnt as characteristic, rather than anomalous, in the absence of any pre-existing information to suggest otherwise (Kraljic, Samuel, & Brennan, 2008). Follow-up work suggested, however, that apparently anomalous features were indeed learnt, but learnt quite separately from characteristic ones (Kraljic & Samuel, 2011). Thus, while learning is to some extent limited by perceived utility, it is more generally constrained by the ways in which new experiences are classified, and there is likely not one uniform pool of 'exemplars' drawn upon in all perception and production.

Learning is an essential characteristic of language. It is the basis for early language acquisition, and it is now quite clear that learning continues in adulthood. Past experiences form the basis for all perception and production of language, and both perception and production may be manipulated through exposure to novel linguistic phenomena. Even so, the literature has revealed apparent limits to the extent to which listeners learn: learning is opportunistic and driven by utility, and because learning leads to abstraction of knowledge, novel linguistic variation is not necessarily learnt, being commonly absorbed into existing, abstract categories. One of the most noteworthy findings is that as a consequence of these limitations, perceptual learning may to some extent be malleable based on the extent of past experiences. This raises the question of whether there is or is not some absolute limit to the extent to which new experiences can influence perception.

2.3. Expectations and speech perception

An improved understanding of language knowledge demands consideration not only of how such knowledge is derived from experiences, but also of how experiences create expectations that in turn influence the ways in which new experiences are interpreted. Although the development of usage-based models has owed much to research in speech perception, there remains significant work to be done in addressing the implications such models have for our understanding of perception. Early findings (e.g. Niedzielski, 1999; Strand, 1999) have been interpreted as suggesting listeners' perception of speech is influenced by their beliefs about a speaker; in other research listeners have been shown to have a remarkable aptitude at predicting speech before encountering it (e.g. Dilley & McAuley, 2008; Hilpert, 2008). Bringing together these two strands – listener beliefs and predictive behaviour – it is possible to conceive of how expectations might operate in a usage-based model, with listeners forming predictions based on their existing linguistic knowledge, which are then tested by exposure to stimuli. It is now

well accepted that variation in language use reflects variable past experiences of language; it is also necessary, however, to consider variability in perception itself.

This issue forms part of an ongoing debate in the field of psychology regarding the extent to which so-called 'top-down' cognitive processes – expectations – influence 'bottom-up' experiences of the world. The possibility of such influences has in some cases been resolutely denied; Firestone & Scholl (2016) contend that experimental evidence for such effects is methodologically unsound. Specifically, they argue that purported cognitive effects on perception are not in fact effects on perception at all, but effects on post-perceptual judgements. Claims regarding top-down influences have been less well-developed with regard to speech perception. A notable exception is the 'adaptive resonance theory' of Grossberg (1980, 2003), in which it is speculated that conscious knowledge, including of language, develops as a 'consensus' between experienced stimuli and expectations thereof. Proposals for top-down effects on language have, however, been rejected due to the paucity of firm experimental evidence for them (Cutler & Norris, 2016), or based on theoretical assumptions: it is claimed that speech perception is optimised for accurate recognition of the content of a signal, to which expectation-derived effects would be prejudicial (Norris, McQueen, & Cutler, 2000). In this view, predictive behaviours, while not denied, can be accounted for as listeners' on-the-spot adjustments to evidence, a purely bottom-up process (Norris, McQueen, & Cutler, 2016). In the face of such challenges, caution is warranted in the framing of claims regarding the role of expectations in the perception of language.

Without taking a strong position regarding cognitive interference with perceptual processes, the notion that expectations are predictive of speech processing can be upheld in some weaker form, for there is substantial recent experimental evidence suggestive of it. Much of this research has focussed on the effects of linguistic knowledge on prediction of language structure and meaning. Hilpert (2008) found that higher-level syntactic knowledge influenced lexical prediction. Expectations regarding prosody, formed significantly in advance, have similarly been shown to influence processing of words (Dilley & McAuley, 2008; Morrill, Dilley, McAuley, & Pitt, 2014). Further evidence indicates the potential malleability of such expectations. Kroczek & Gunter (2017) demonstrated that while individuals came into a syntactic comprehension task with strong expectations regarding particular word orders affecting their processing of sentences, these expectations could be modified through new exposure such that listeners acquired new, long-lasting and speaker-specific expectations regarding syntax.

While most of these studies have depended on evidence from post-exposure tasks involving, for example, comprehension or recall, and would thus be susceptible to the methodological critiques of Firestone & Scholl (2016), there is some evidence of more immediate effects on processing due to such expectations: Fine, Jaeger, Farmer, & Qian (2013) found effects on reaction times of participants' expectations regarding syntactic ambiguities, which are indicative of more fine-grained, direct influences on processing. This body of evidence does of course remain open to the more limited interpretation of Norris, McQueen & Cutler (2016) that while perception is predictive, it only reflects probabilities derived from the incoming signal itself rather than higher-level linguistic knowledge. It does, however, at least raise the possibility that individuals' interpretation of stimuli does draw upon prior knowledge.

As usage-based models offer to break down the traditional separation of social or functional information and what was traditionally conceived of as 'linguistic' knowledge, it is reasonable to assume that where predictive behaviours relating to the latter kind of information have been found, behaviours following expectations regarding the former might be also. In a seminal study, Niedzielski (1999) raised just such a possibility. In this study, residents of Detroit were asked to match resynthesized vowels of varying qualities to those to which they were exposed in audio featuring speech from an individual from Detroit. Their responses varied significantly depending on whether they thought the audio was from a speaker from Detroit or one from Canada, in spite of the invariance of the signal. The signal had features strongly associated with Canada, though also commonly found in speech in Detroit. Niedzielski (1999) argued that listeners' expectations of the speaker were directly influencing what they heard. Because they expected certain features from a speaker from Canada, they heard them in such a context, whereas they did not hear them if they thought the speaker was from Detroit. Other work has shown similar results regarding effects of expectations regarding speaker origin (Hay, Nolan, & Drager, 2006), gender (Strand, 1999), and age (Drager, 2010) on speech perception. The implication is that higher-level knowledge does indeed influence perception in some manner.

Research on the perception of speech believed to be 'non-native' has been similarly telling. In early studies, participants were found to rate a speaker as more accented when perceiving them to be 'Asian' as opposed to 'Euro-American' based upon the presentation of photographs, in spite of invariance in the auditory stimulus, although effects on actual comprehension were less clear (Rubin, 1992; Kang & Rubin, 2009). This is evidence, of course, of 'judgements' rather than 'perception', following the arguments of Firestone & Scholl (2016). Indeed, a further study by Zheng & Samuel (2017) did not find any strong evidence for perceptual

differences. Their conclusions were based on 'selective adaptation' following the method of Eimas & Corbit (1973), in which listeners are played a stimulus repeatedly and their loss of sensitivity to a relevant feature over time is used to gauge perception of the original stimuli. While the finding of Rubin (1992) regarding judgements was replicated, no significant adaptation was found in tests using variable visual stimuli. Other studies featuring relatively direct measures of perception present a different view. Tests featuring audio-visual integration of stimuli combined with assessments of speech intelligibility in noise showed that intelligibility enhancement due to the presence of a visual stimulus was greater with a perceived 'non-native' as opposed to a 'native' speaker (Yi, Phelps, Smiljanic, & Chandrasekaran, 2013). While consistent effects on judgements can be found, the evidence for effects on speech perception in this literature remains mixed.

Unclear as the evidence regarding perceptions of perceived 'non-native' as opposed to 'native' speech is, some of the first attempts to examine the issue of expectations from a usage-based perspective have come in this area. While listener behaviour in these studies has been attributed, following Rubin (1992), simply to bias on the part of the listeners, subsequent studies have offered a somewhat different account. McGowan (2015) found that listeners' ability to perceive Chinese-accented English in noise improved when they perceived the speaker, based on a visual stimulus, to be 'Asian'. This suggests an alternative explanation of previous findings, namely, that the visual stimuli produce expectations, and that the perceived incongruence of expectations and new experience is detrimental to perception. A separate study by Babel & Russell (2015) reached similar conclusions, finding that perceived incongruence was predictive of speech perception in noise where performance in explicit and implicit bias assessment tasks was not. The study further found an effect of experience, whereby listeners with stronger social ties to 'non-native' speakers were more likely to have their perception impaired due to perceived incongruence. More recently, Gnevsheva (2018) found that perceived incongruence even influenced accentedness ratings, such that speakers perceived as 'native' based on visual stimuli produced higher accentedness ratings when the auditory stimuli were perceived to be incongruent. McGowan (2015, pp. 503-504) argues that evidence for the influence of expectations fits with a usage-based model of language in some form; Gnevsheva (2018) specifically favours an exemplar account. It is doubtful that this literature provides substantial evidence for one model over another, but it certainly fits with any understanding of language in which perception, as well as production, varies with individuals' linguistic knowledge.

More recent findings have posed questions for the expectations account. Previously, expectations, while accepted as subjective effects on perception, have been treated as following objectively from past experience. Gnevsheva (2018, p. 585) suggests that effects of expectations follow from 'activation of experience-based representations', which may or may not conflict with new experiences. While this is certainly true, recent work has indicated that expectations, though beginning with experience, may be further modulated by still higher-level mental processes (Hanulíková, 2018; Fiedler, Keller, & Hanulíková, 2019). Fiedler, Keller, & Hanulíková (2019), hypothesise that

> speech intelligibility [in studies of perception of 'non-native' speech] is affected by factors that are not exclusively linked to biases towards non-native speakers […] or to episodic traces and a simple congruency […] It seems that the effect of non-linguistic information on speech processing is based on more complex knowledge about social categories (p. 4)

It also appears that there are limits to the extent to which processing of 'non-native' speech is affected: de Weers (2018) did not find any impediments of listener expectations of 'non-native' speech at perhaps the most fine-grained level of processing, effects on reaction times. Such findings certainly do not go against the notion of expectation-driven effects, but it does suggest that a fuller understanding of the relationship between experiences and expectations is wanting.

Beyond the 'non-native' speech congruity paradigm, there exists other evidence of the operation of expectations relating to social information. It appears that speech processing is affected by beliefs regarding the gender of the speaker (Johnson, 2006), expectations of an impediment to speech (Arnold, Hudson Kam, & Tanenhaus, 2007), the hearing of an unexpected variety (Brunellière & Soto-Faraco, 2013), beliefs regarding the sound characteristics of other forms of language (Lev-Ari & Peperkamp, 2016), or assumptions regarding what form of language is being heard (Gonzales, Byers-Heinlein, & Lotto, 2019).

In a different approach to the question of treatment of 'non-native' speech, Lev-Ari (2015) found significant effects of whether or not 'non-native' speech was being heard on participants' behaviour in a lexical decision task. Listeners appeared more likely to rely on contextual information as opposed to the actual content of the stimuli when listening to 'non-native' speech. Lev-Ari (2015) argued that listeners' expectations that 'non-native' speech would be less readily intelligible produced this behaviour. Importantly, in this study, unlike in those in the paradigm of Rubin (1992), expectations were not overtly primed prior to exposure to

stimuli. Most experiments described here more directly force particular beliefs regarding the stimuli, and thus, it could be argued, suffer from some of the methodological pitfalls described by Firestone & Scholl (2016). A substantial body of evidence has, however, accumulated that at least raises the possibility that expectations regarding a wide variety of social characteristics influence processing of speech, and in some cases, this has been investigated in a more carefully controlled manner.

Given the apparent evidence for the effects of expectations, it becomes increasingly important to consider how expectations condition individuals' behaviour in other contexts. While research on this question remains slim, preliminary evidence of the importance of considering participant expectations in learning studies has come from Siegelman et al. (2018). They found that behaviour in a statistical learning task featuring auditory linguistic stimuli was significantly more variable and inconsistent than behaviour in a statistical learning task using abstract visual stimuli for which the participants could have few prior expectations. This suggests that whatever the degree of new experience an individual acquires with some form of language or feature, their processing is always done with respect to pre-existing beliefs. As such, the effects of experience cannot be considered separately from those of expectations, and there is a need to examine just how malleable individuals' perceptions are when expectations have already been formed.

Recent work suggesting that perception of language is influenced by higher-level knowledge fits in with a significant recent theoretical shift in cognitive science, in which prediction is conceived of as one of the 'fundamental principles of brain functioning' (Bubic, von Cramon, & Schubotz, 2010, p. 11), a domain-general phenomenon in which knowledge develops through adaptation to discrepancies between beliefs and experiences (Clark, 2013). In this model of cognition, the notion of modularity, or the disassociation of different forms of knowledge, ceases to have much relevance (Hilpert, 2008), and perception is not assumed to follow any set sequential process (van Berkum, 2008). Individuals may vary in the extent to which they rely on top-down or bottom-up influences (Tulver, Aru, Rutiku, & Bachmann, 2019), helping to explain otherwise unaccountable variation in individual behaviour. Although there continues to be significant scepticism regarding these claims, evidence from neuroscience is increasingly concordant with them (Davis & Johnsrude, 2007). Neural correlates of context-based predictions reflecting social information are increasingly well understood (e.g. DeLong, Urbach, & Kutas, 2005; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; van Petten & Luka, 2012; Hanulíková & Carreiras, 2015) and have been distinguished from effects of the

better-established phenomenon of priming (Otten & van Berkum, 2008), in which on-the-spot cognitive associations – in an 'exemplar' model, the 'activation' of related episodic memories – affect behaviour. It has further been proposed that predictive behaviour affects processing not only of the speech of others, but also processing of individuals' own speech (Lev-Ari, Ho, & Keysar, 2018), and links the perception of language with the speech production system (Martin, Branzi, & Bar, 2018). The notion of a 'predictive brain' has the potential to unify many different aspects of recent research on cognition, including in linguistics.

In spite of the growing prominence of the notion of expectation-driven processing, little has been done to bring it into usage-based models of language, aside from in Gnevsheva (2018). Findings from the literature regarding prediction based on contextual or linguistic factors have to some extent been taken into account; Morley (2014), for example, offers a computational model featuring expectations related to specifically linguistic knowledge. By contrast, the more complex phenomena of expectations based on social factors are less easy to integrate. The need to do so, however, is increasingly being recognised. Docherty & Foulkes (2014, p. 50), reviewing the present understanding of usage-based models, suggest that 'there are many questions remaining regarding […] how this "bottom-up" channel of learning is influenced by the knowledge […] the individual has already acquired'. Consideration of the relationship of experiences and expectations in forming linguistic knowledge does have significant implications for usage-based models. In 'exemplar'-based models, the importance of concrete memories as opposed to abstract knowledge was originally emphasised, before hybrid models were developed assuming levels featuring both forms of knowledge. Were new experiences indeed being forever filtered through expectations, and were perception, as Samuel & Kraljic (2009, p. 1207) suggest, never 'an objective translation of reality', it would suggest all linguistic knowledge is in fact highly abstracted, although generalised to varying degrees. Insofar as linguistic knowledge depends on some combination of the objective content of stimuli and one's interpretation of it reflecting experiences and expectations, the relative importance of these different elements needs to be better understood.

2.4. Recent findings on experience-driven perception

Many of the issues relating to the nature of linguistic knowledge in a usage-based model, the effects of past and new experiences and the influences of expectations have recently been examined by Hay, Drager, & Gibson (2018). As this study is in many ways the precursor to the current study, its methods, findings, and implications will be explained in more detail here.

This study investigated whether or not listeners' past experiences of a particular linguistic feature influenced their ability to identify that feature in stimuli due to the expectations such experiences would create. Specifically, Hay, Drager, & Gibson (2018) studied the perception of what are known in the phonological literature as '/ɹ/ sandhi' phenomena: 'linking /ɹ/' in which final /ɹ/, represented in orthography as <r>, occurs as the onset of the following syllable, regardless of whether non-prevocalic /ɹ/ is typically used; and 'intrusive /ɹ/', in which this linking behaviour is extended to some intervocalic contexts at morpheme and word boundaries in spite of the absence of orthographic <r>. In two experiments, the authors looked to see whether variable experience with linking and intrusive /ɹ/ across speakers, words, and sound environments translated into variable perception of these features. Specifically, it was argued, following a usage-based account, that where listeners had greater past experience with these features, they would be more likely to hear their presence, regardless of whether they actually occurred in the stimuli.

In the first experiment, the effect of variable experience of /ɹ/ in different linguistic contexts was investigated, and it was found that perception did indeed follow clearly from past experience. Listeners from a New Zealand language background listened to stimuli spoken by an individual with a similar variety of speech, in which /ɹ/ consistently occurred in linking and intrusive contexts. A phoneme monitoring method was adopted in which listeners listened for the relevant sound and indicated whether or not they heard it in each stimulus. Stimuli included instances of intrusive and linking /ɹ/ at morpheme and word boundaries. Findings of earlier research and new examination of corpus data showed that /ɹ/ is more likely to occur at morpheme than at word boundaries and in linking than in intrusive contexts, at least historically, among speakers in New Zealand. Based on this, Hay, Drager, & Gibson (2018) predicted that listeners would more often hear /ɹ/ in linking contexts and at morpheme boundaries than in intrusive contexts and at word boundaries, reflecting their own probable ambient exposure. Both hypotheses were well-supported by the experimental evidence. Furthermore, the results regarding linking versus intrusive /ɹ/ were not merely explicable in terms of orthography (a tendency to indicate the presence of /ɹ/ more often when the corresponding symbol occurred), as indicated by participant behaviour in filler items featuring a transparent orthography. It appears then that what is and is not perceived does indeed follow in some manner from past experiences in a quantifiable way. Hay, Drager, & Gibson (2018) interpret these results to suggest that experiences create expectations that affect perception during listening.

Variation in experience of /ɹ/ across listeners was also examined. In the second experiment conducted by Hay, Drager, & Gibson (2018), listeners from Canterbury, New Zealand, Southland, New Zealand, and San Diego, California heard stimuli spoken in the same voice used in the first experiment, but with variable linking and intrusive /ɹ/. Listeners from these separate areas were assumed to have variable experience with intrusive /ɹ/. Based on production findings in these three areas, it was anticipated that listeners from Canterbury would have substantial experience with intrusive /ɹ/, while listeners from Southland would have less experience, and listeners from San Diego would have almost no experience with the feature.

It was hypothesised that greater experience would on the whole be associated with greater identification of the feature, but also that where listeners had virtually no experience of a feature, it would be particularly salient and would therefore be *more* often identified. Evidence for these suggestions was indeed found. The San Diego participants, for whom intrusive /ɹ/ was most novel, were most likely to hear it when it was present, but the Canterbury participants were more likely to hear it than the Southland participants. Trends in behaviour towards intrusive /ɹ/ absence and towards linking /ɹ/, present or absent, appeared to go as predicted, but were less robust. Overall, however, the results of this experiment supported the view that perception follows from experience, complicated by the effect of salience of highly novel features.

The study was able also to show more directly that individuals' own use of linking and intrusive /ɹ/ was predictive of their perception of these features. Using production data from listeners, it was shown that participants from Canterbury who had more /ɹ/ sandhi in their speech were more likely to hear both linking and intrusive /ɹ/. Participants from Southland who had more /ɹ/ sandhi in their speech were also more likely to hear intrusive /ɹ/, but their perception of linking /ɹ/ was not affected in this same way, as their perception of this feature was to some extent predicted instead by the presence of non-prevocalic /ɹ/ in their speech; people from this language background maintain a clearer distinction between these phenomena. Although particular behaviours may emerge as characteristic of a category of individuals with similar experiences, these effects may be localised to the level of an individual and their unique past experiences of language.

Post-hoc analysis by Hay, Drager, & Gibson (2018) also revealed an effect of new exposure local to the second experiment. The participants from San Diego, listening to stimuli with variable linking and intrusive /ɹ/, started from a position of hearing /ɹ/ very often in linking

contexts, regardless of its actual presence, and scarcely hearing the sound in intrusive contexts, though it was present equally often. Over the course of the experiment, however, the San Diego participants increasingly heard intrusive /ɹ/, even just in stimuli where it was in fact absent, and conversely ceased to hear linking /ɹ/ as much where it was absent. No equivalent perceptual learning behaviour was found among the Canterbury and Southland participants.

These results carry two implications. Firstly, they corroborate findings in the perceptual learning literature (e.g. Lively, Logan, & Pisoni, 1993) in demonstrating that listeners will readily adjust to hearing a particular feature in spite of having little experience with it, or at least to a particular speaker in a particular context. Presumably exposure to other, similar speakers in other contexts is necessary for the development of more generalised knowledge (Clopper & Pisoni, 2004a; Kraljic & Samuel, 2006). Secondly, these results suggest that listeners' past experiences influence not only *what* they perceive but also *how likely* their perception is to change in such a setting. A plausible explanation of the finding is that the San Diego listeners had little prior experience of intrusive /ɹ/ and little prior experience with New Zealand voices, and as such their perception of these stimuli was highly malleable. Participants from Southland and Canterbury, by contrast, regardless of the extent of their own use of intrusive /ɹ/, would be highly familiar with typical New Zealand voices featuring extensive /ɹ/ intrusion, and as a result would have much more fixed expectations of these stimuli. The study of Hay, Drager, & Gibson (2018) demonstrates how perception of language changes through new experiences, and it raises the issue of how much change is possible given listeners' varying expectations.

Hay, Drager, & Gibson (2018) have amply demonstrated, as predicted by a usage-based model of language, that an individual's perception of language follows from their past experiences of language. They also offer a useful methodology for the further study of the relationship between experiences, expectations, and language. While Hay, Drager, & Gibson (2018) studied how participant behaviour in experiments monitoring linking and intrusive /ɹ/ in stimuli from a single speaker varied with their varying experiences of these features, this form of experiment might be inverted to examine how listeners with similar experiences – sharing a New Zealand background, in this case – variably perceive these features where their expectations of different stimuli are variable. The present study pursues this avenue, in order to better study how the perception of language is influenced by listeners' subjective interpretations of what they are hearing.

2.5. Intrusive and non-prevocalic /ɹ/

As the present work follows the model of Hay, Drager, & Gibson (2018), but introduces variation in voices into the stimuli, the assumptions that have been made regarding the voices used and the literature on the forms of speech they reflect will be detailed at some length. While reference is made throughout to specific 'varieties' for convenience, it is understood that these are purely social constructs formed through the generalisation and stereotyping of certain characteristics, which are important insofar as they create expectations in listeners, but which do not have any reality of their own. Likewise, although reference will be made to certain linguistic features as they have commonly been classified and understood in the phonological literature, this should not be taken to imply that any assumptions are being held regarding the status of such concepts in real linguistic knowledge.

Phenomena relating to English rhotic sounds – hereafter '/ɹ/' – such as those studied by Hay, Drager, & Gibson (2018) continue to be particularly useful in the investigation of the effects of expectations insofar as they are highly variable across speakers, and this variation is often strongly socially coded, being a marker of commonly distinguished varieties associated with particular geographic regions. As a consequence, if voices featuring variable /ɹ/ phenomena are connected with different varieties and regions, it may be anticipated that this will create different expectations regarding the sounds in the voice, the influence of which would be detectable in a phoneme monitoring task.

Notably, the occurrence of /ɹ/ in syllable codas – hereafter 'non-prevocalic /ɹ/' – is more or less common in different forms of English, and is variably marked as either a 'standard' or a 'non-standard' feature depending on the region. In North America, Ireland and Scotland, non-prevocalic /ɹ/ typically occurs; in England and the Antipodes, it is much less often encountered. Intrusive /ɹ/ generally follows the reverse distribution as a result of its gradual emergence through generalisation of linking /ɹ/ amidst the decline of non-prevocalic /ɹ/ use (Hay & Sudbury, 2005; Sóskuthy, 2013). Four commonly-identified varieties of English were of interest to the present study, because they display a series of oppositions ideal for investigating how experiences and expectations influence the perception of speech. These forms of English are that which is most typically found throughout New Zealand, hereafter termed 'General New Zealand English'; that which is found in the Southland region of New Zealand, hereafter referred to as 'Southland English'; American speech that lacks strong regional marking, identified as 'General American English'; and a form of speech traditionally characteristic of

much of the New England cultural region in the United States, hereafter referred to as ‘Eastern New England English’.

Because two of these varieties (General New Zealand English and General American English) are considered to be standards of language throughout a geographic area that encompasses one or other of the remaining varieties, while the others are regional varieties with which many listeners will be unfamiliar, it was anticipated that there would be significant variation in the strength of expectations created by stimuli heard in voices corresponding to these varieties. In particular, when a voice is associated with a variety with which the listener has considerable experience or which they strongly stereotype, this is likely to create very strong, and relatively fixed, expectations. With familiar varieties, expectations are likely to be weaker. In such a situation, perception will either represent the stimuli very accurately, or it will initially be guided by expectations of a standard variety, but will change more easily through new exposure. The distribution of non-prevocalic and intrusive /ɹ/ in each of these varieties is detailed here, as being relevant to how such voices would be interpreted and what experience listeners from a New Zealand background would be likely to have with these phenomena across these different varieties.

2.5.1. Non-prevocalic and intrusive /ɹ/ in General New Zealand English

Although intrusive /ɹ/ has long been typically present, and non-prevocalic /ɹ/ typically absent, in most speech in New Zealand as anticipated in this study, both characteristics, while still predominant, have undergone something of a decline. Most accounts treat ‘New Zealand English’ – defined as New Zealand speech generally, excluding that of the Southland region, and the occasional occurrence of rhotic vowels as an ethnic marker of Māori or Pasifika speech (Starks & Reffell, 2005; Kennedy, 2006; Gibson, 2016) – as ‘non-rhotic’, lacking non-prevocalic /ɹ/ (e.g. Hay, Maclagan, & Gordon, 2008). There is, however, recent evidence for the emergence of rhoticity in younger speakers in New Zealand, particularly in the form of a rhotic vowel in the NURSE environment (Marsden, 2017).

The current status of intrusive /ɹ/ is unclear; while the corpus analysis of Hay, Drager, & Gibson (2018) showed increases in intrusive /ɹ/ over time until it became as frequently used as linking /ɹ/, recent evidence suggests a potential decline in /ɹ/ sandhi (Marsden, 2017). As a consequence of growing variability in the use of these features, it is likely that listeners from a New Zealand language background will have more experience with non-prevocalic /ɹ/ and less experience with intrusive /ɹ/ in New Zealand voices than might otherwise be thought.

Presumably, however, non-rhoticity and /ɹ/ intrusion remain the dominant forms that would be heard and most likely expected in New Zealand voices.

2.5.2. Non-prevocalic and intrusive /ɹ/ in Southland, New Zealand

A single localised form of English in New Zealand is generally treated as a distinct variety (Hay, Maclagan, & Gordon, 2008). This is a variety associated most strongly with Southland, an area in the far south of New Zealand's South Island. Historically distinguished by other features such as greater /ʍ/ retention (Schreier, Gordon, Hay, & Maclagan, 2003), the variety is now more or less indistinguishable from the speech of the rest of the country save for its retention of non-prevocalic /ɹ/, commonly attributed to the influence of the Scottish English spoken by the majority of its first European settlers (E. Gordon, et. al., 2004). This feature is somewhat more widely spread beyond Southland proper, being also prevalent in the adjacent Otago region (Kennedy, 2006), but does not predominate anywhere else in the country.

Early references to the variety noted the more frequent occurrence of rhotic sounds in the southern South Island and suggested the sound was 'not trilled or different' from that found elsewhere in New Zealand (Turner, 1966). It appears the Southland rhotic consonant has always or has long been comparable in articulation to that of General New Zealand English, contrary to popular assertions that Southland English is characterised by a 'rolled "R"' (Hay, Maclagan, & Gordon, 2008); the predominant Scottish articulations [r], [ɾ] and [ɻ], which are believed to have been present in the speech of early Scottish settlers (E. Gordon, et. al., 2004), have left no trace. Importantly, this carries the benefit that listeners are unlikely to perceive it as a markedly different sound to that heard in the other voices included in this study.

Sociolinguistic studies of Southland English since 1990 have made clear that non-prevocalic /ɹ/ has declined significantly in most contexts, but persists as a rhotic vowel in NURSE environments, where its use has in fact increased among younger speakers. Bartlett (2002), in data drawn from sociolinguistic interviews of sixty-nine speakers from the Southland region, conducted in 1991-1992, found that younger participants were more likely to have rhoticity in NURSE environments than members of older age groups, irrespective of gender or socioeconomic background. While the variety used to be characterised by general rhoticity undergoing a gradual decline, it is now strongly associated with the presence of a stressed rhotic vowel, and no non-prevocalic /ɹ/ elsewhere. As a consequence, it is not reasonable to assume that speakers would be able to produce the sound in non-prevocalic contexts other than NURSE, and nor is it clear that such rhoticity would be associated with Southland.

Perceptual studies offer information on the extent to which Southland English is known about in other parts of the country (e.g. Bayard & Bartlett, 1996; P. Gordon, 1997), but are not indicative of real experience with the variety. A perceptual dialectological study conducted on a broader sample of New Zealand residents at four different universities throughout the country also found Southland English to be the most-commonly identified regional variety in the country, however, with descriptions often relating to a distinct pronunciation of /ɹ/ (Nielsen & Hay, 2005). There is clear evidence for high levels of awareness of a Southland variety, but this is not indicative of the extent to which people from outside the region will actually have heard Southland rhotic speech.

The occurrence or otherwise of intrusive /ɹ/ in Southland English does not form a part of popular ideas of the variety, and nor has it received much attention in the literature. As the occurrence of the feature tends to be inversely correlated with the occurrence of non-prevocalic /ɹ/, we would expect intrusive /ɹ/ to be less common in Southland than elsewhere in New Zealand. Indeed, Hay, Drager, & Gibson (2018) found that a group of Southland participants did produce significantly fewer tokens of intrusive /ɹ/ at word boundaries than a group of participants from elsewhere in New Zealand, while more limited data for morpheme boundary environments suggests the same trend is valid here as well. It is therefore reasonable to assume that intrusive /ɹ/ is a more recent, and less widespread, feature in Southland English. There is nothing in any of this to suggest that participants from a General New Zealand English language background would have any expectations regarding intrusive /ɹ/ in Southland English. Overall, they would likely hear a Southland voice as an ordinary New Zealand voice, given the overwhelming similarities between the varieties, but might have some higher-level awareness of its being a Southland variety based on the occurrence of non-prevocalic /ɹ/, even if this not reflected in any subconscious expectations.

2.5.3. Non-prevocalic and intrusive /ɹ/ in General American English

In the United States, in contrast to New Zealand, non-prevocalic /ɹ/ is a recognisably standard feature. With noteworthy exceptions correlated with ethnicity (African-American) and region (New England, New York, and parts of the Southern United States), non-prevocalic /ɹ/ is used categorically throughout the United States, and its occurrence is a necessary feature of any ‘unmarked’ or ‘General American’ English (Kretzschmar, 2004); this usage is furthermore spreading into those communities where it was scarce historically. Intrusive /ɹ/ in the United States now appears to be an exclusively New England usage, being absent in the Southern

United States (Thomas, 2004) and ill-attested in New York.[2] As such, Eastern New England English is the most appropriate regional North American variety for the purposes of this study.

Individuals from a New Zealand English language background have been shown to readily identify a speaker voice as American, whether on the basis of the presence of non-prevocalic /ɹ/ or other features, while they also tend to identify Canadian voices as American (Bayard, 1990), suggesting they have a deeply entrenched stereotype of 'American' speech connected with general characteristics of North American varieties. Listeners from New Zealand will readily identify American voices, and this reflects above all their extensive exposure to creative and news media produced in the United States (Bayard, Weatherall, Gallois, & Pittam, 2001), whatever their actual experience of General American voices in conversation. As such, it is highly likely that listeners from this background will readily be clued to a speaker's being American, and will have strong expectations of the occurrence of non-prevocalic /ɹ/; potentially, they may also not expect intrusive /ɹ/, at least at a more subconscious level, although this is less likely given the lower salience of this feature.

2.5.4. Non-prevocalic and intrusive /ɹ/ in Eastern New England English

A form of speech once prevalent throughout the New England region of the United States, and most strongly associated with the city of Boston, this variety is noteworthy in that, at least in its canonical description, it shares the same mixed distribution of non-prevocalic /ɹ/ as that described for Southland English (§2.5.2), but is under the influence of a rhotic standard, where Southland English is under the influence of a standard lacking non-prevocalic /ɹ/. Typical accounts of Eastern New England English do, however, suggest that the variety features intrusive /ɹ/ (e.g. McCarthy, 1991; Halle & Idsardi, 1997), which is expected to be less common in Southland English.

Non-prevocalic /ɹ/ has historically been rare in some regions of the United States, namely Eastern New England (Van Riper, 1957), New York City (Labov, 1966), and areas of the Southern United States (Feagin, 1990; Labov, Ash, & Boberg, 2006). In all of these areas, however, non-prevocalic /ɹ/ has more recently spread to the point of becoming overwhelmingly predominant (Becker, 2009, 2014; Mather, 2012; Feagin, 1990; Stanford, 2018).The most extensive use of non-rhotic speech is likely found in New England, but here, there is no

---

[2] M. Gordon (2004) offers a rare reference to the occurrence of intrusive /ɹ/ in New York City; the phenomenon is not mentioned in major recent accounts of /ɹ/ in New York, including Becker (2009, 2014) and Mather (2012).

evidence of historic non-rhoticity in western areas of the region (Boberg, 2001), and more recently this feature has begun to rapidly disappear from eastern New England (Nagy, 2001; Villard, 2009; Stanford, Leddy-Cecere, & Baclawski, 2012; Stanford, Severance, & Baclawski, 2014). Non-rhoticity is now largely confined to parts of Maine, eastern Massachusetts, and Rhode Island (Stanford, 2018). Even here, however, it is increasingly confined to older and less-educated speakers, and NURSE contexts predominantly feature a rhotic vowel (Irwin & Nagy, 2007; Nagy & Irwin, 2010).

Intrusive /ɹ/ is regularly identified as a feature of New England English in the phonological literature, along with hypercorrective /ɹ/ in non-prevocalic positions (e.g. Kahn, 1976; McCarthy, 1991; 1993). While there is some historical evidence of its actual occurrence (Whorf, 1943), evidence for its present use is limited (Stanford, Severance & Baclawski, 2014; Stanford, 2018); it appears to have declined significantly with the spread of non-prevocalic /ɹ/.

In spite of its decline, non-prevocalic /ɹ/ remains the most characteristic, typifying feature of New England, and particularly Boston, speech (Irwin & Nagy, 2007); intrusive /ɹ/ is presumably much less salient. However strongly non-rhoticity is connected with New England in the United States, it is an open question whether listeners from a New Zealand language background will make the same connection. The ability to discriminate different varieties is coupled with listeners' experience with variation (Clopper & Pisoni, 2004b), and it is doubtful that New Zealand listeners would have the requisite experience of variation to have the kind of sensitivity to non-rhoticity as a marker of New England speakers that is found in the United States. As such, if they are exposed to stimuli from such a voice, lacking non-prevocalic /ɹ/ and including intrusive /ɹ/, it is reasonable to assume they will have few expectations regarding the occurrence of these features, or will expect a distribution similar to that found in the General American voice, mirroring the Southland-New Zealand voice relationship.

3. Direction of study

Hay, Drager, & Gibson (2018) showed how listeners' past experiences created expectations that influenced their perception of features, and further demonstrated the effect of new experiences – perceptual learning – of a less-familiar linguistic feature on the perception of the presence or absence of that feature. San Diego listeners, from an American English speech environment, became more likely to perceive an unfamiliar feature, intrusive /ɹ/, as present, and less likely to perceive a familiar feature, linking /ɹ/, as absent, over the course of the experiment, regardless of these features' actual presence or absence. In that experiment, the

language variety whence the stimuli were derived was kept constant for all stimuli; what was variable was the extent of listeners' experiences with /ɹ/ sandhi, depending on its occurrence in the ambient environment in different areas – Canterbury, Southland, or San Diego – and in different – linking and intrusive – contexts. The effects of expectations being measured followed directly from listeners' experiences.

The literature on expectations, however, shows that expectations reflect not only listeners' past experiences, but also the manner in which these are interpreted and drawn upon in a particular context, based on other information derived from the stimulus source, including beliefs regarding a particular speaker, their social background, and other relevant information (e.g. Niedzielski, 1999; Strand, 1999; McGowan, 2015; Gnevsheva, 2018; Fiedler, Keller, & Hanulíková, 2019). In inverting the experimental method of Hay, Drager, & Gibson (2018), aspects of this more complex relationship between experiences, expectations, and stimuli might be established.

In particular, it remains to be seen whether variability in the social characteristics with which stimuli are contextualised contributes to the formation of expectations that influence the perception of speech, and the phenomenon of perceptual learning. Reflecting the evidence that listeners form expectations based on the perceived language variety of a speaker (Brunellière & Soto-Faraco, 2013), it is anticipated that variation in the source language variety for linguistic features will influence both how these features are perceived and the degree to which perception changes under increasing exposure to these features.

In this view, if a listener has more extensive experiences and a better-established 'idea' of one language variety than another, they might be more likely to expect, and therefore hear, a particular feature that is common to both in the former, rather than the latter. Their perception of features in the less familiar variety would also potentially be governed by expectations, but in a more complex way, not following from any (limited) experiences of this variety. They might interpret the voice as being from a speaker of another, more familiar variety, and expect the characteristics of that variety, or else their expectations might be limited, such that what they would perceive would closely reflect the actual content of the stimuli. In either case, their perception would be more malleable, depending less on a pre-existing idea of the variety, and more substantial perceptual learning would occur. The present study investigates whether there is indeed such an effect of different varieties on listener perception and its relative malleability.

3.1. Research questions

The central issue with which this study is concerned is the relationship between experience, expectations, and stimuli in speech perception. Two main questions present themselves. Firstly, does variation in listeners' expectations of different language varieties affect their perception of stimuli they associate with these varieties? Secondly, does variation in listeners' expectations of different language varieties affect the manner in which their perception changes through new experiences of stimuli?

It is hypothesised that where listeners have weaker existing expectations of a language variety, they will tend to adjust more over time to the presence of unfamiliar features in that variety, but where there is a discrepancy between listener expectations and novel experience of a language variety, it will take longer for listeners to adjust. Where listeners have stronger existing expectations of a variety, they will adjust very quickly, but less significantly. As a result, strong expectations will limit the role of new experiences in perception.

3.2. Expected outcomes

Changes over time in participants' 'hearing' of instances of non-prevocalic and intrusive /ɹ/ in four voices representing different language varieties – General New Zealand, Southland, General American and Eastern New England English – are being studied. It was anticipated that the voice in which stimuli occurred, the /ɹ/ phenomenon they contained (non-prevocalic versus intrusive /ɹ/) and the environment of /ɹ/ (the preceding vowel, and, in the case of intrusive /ɹ/ the kind of boundary preceding it) would influence listeners' responses.

It was thought that listeners would initially identify the Southland voice as a General New Zealand voice, and the Eastern New England voice as a General American voice, as a result of their limited experiences with comparable voices. From this followed several hypotheses:

1. It was hypothesised that perceptual responses to the less familiar varieties (Southland and Eastern New England) would change more quickly and significantly than perceptual responses to the more familiar varieties (New Zealand and General American)
    a. Listeners would increasingly hear the presence of /ɹ/ in non-prevocalic contexts in the Southland voice after originally under-predicting it whilst continuing not to hear it to any significant degree in the General New Zealand voice.

b. Conversely, they would hear less /ɹ/ in intrusive contexts in the Southland voice over time, whilst hearing it consistently to a significant degree in the General New Zealand voice.
c. They would increasingly cease to hear /ɹ/ in non-prevocalic contexts in the Eastern New England voice after originally over-predicting it, whereas they would consistently hear it to a significant degree in the General American voice with little change over time.
d. Listeners would increasingly hear /ɹ/ in intrusive contexts in the Eastern New England voice, whereas they would consistently hear very little intrusive /ɹ/ in the General American voice.

Were these hypotheses confirmed, it would convey an impression of speech perception as an interaction of experiences and expectations, in which incorrect expectations would lead to perception that did not mirror the actual content of stimuli, but where perception would change rapidly through new exposure due to the weakness of expectations. Were this indeed the case, a further possibility would be that listeners' perception of the more familiar voices, with which they had much more extensive experience, would be less malleable. Evidence for this would come from the following:

2. Listeners would come to more strongly predict the presence or absence of /ɹ/ in non-prevocalic and intrusive contexts in the unfamiliar than in the corresponding familiar voices, limited only by their responsiveness to the actual distribution of these features in each voice. Specifically:
    a. They would come to hear more non-prevocalic /ɹ/ in the Southland NURSE context than in the General American NURSE context, but not in other contexts.
    b. Listeners would come to hear less non-prevocalic /ɹ/ in the Eastern New England voice than in the General New Zealand voice, save for in NURSE.

Thus, while listeners would have stronger, and on the whole rather accurate expectations of those varieties with which they had more extensive experience, they would be less able to overcome whatever inaccuracies they retained under the influence of new experiences due to the strength of their expectations.

3. It was further hypothesised that there would be an additional effect of salience. Insofar as intrusive /ɹ/ was anticipated to be less salient than non-prevocalic /ɹ/, it was predicted that the similarity between General New Zealand and Southland English would inhibit

a significant change over time in perception of the feature in the latter variety, while a more substantial effect was expected in perception of the feature in Eastern New England English, where the difference was believed to be starker. This would prevent the extension of those effects described in hypotheses 2a and 2b to intrusive /ɹ/.

4. A final prediction was that listeners' perceptions would be mediated by the preceding vowel context. Since /ɹ/ after NURSE is strongly associated with Southland English, and contradicts the typical lack of non-prevocalic /ɹ/ in Eastern English, perception of the presence of this feature was expected to pattern differently to perception of the presence of non-prevocalic /ɹ/ in START and NORTH environments. There would be rapid adjustment to hearing /ɹ/ presence after NURSE for Southland English, and less of a decline after NURSE in Eastern New England English, although there would still be some change under the influence of the growing experience of non-prevocalic /ɹ/ in this voice.

These outcomes, as they would appear in experimental evidence, are represented in Figures 1 and 2, which display the extent to which listeners indicate they believe /ɹ/ to be present, represented on the *y*-axis, over time, represented on the *x*-axis, as hypothesised.

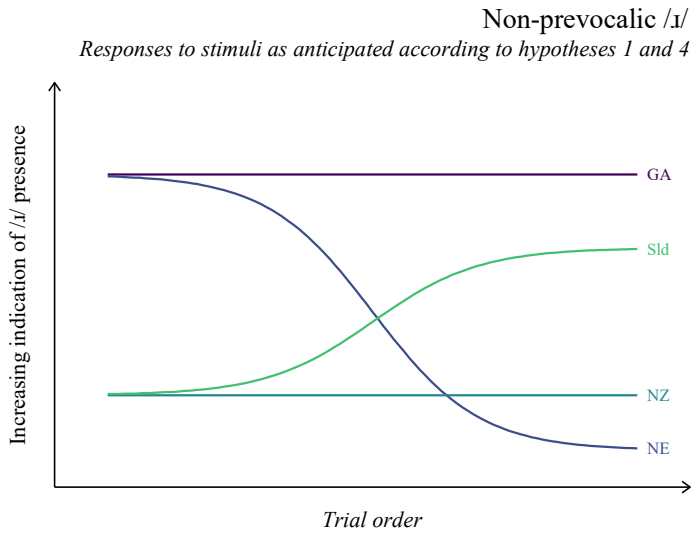

Figure 1: Expected outcomes for identification of /ɹ/ presence in non-prevocalic contexts over time, in stimuli from voices representing four varieties. 'GA' = General American voice; 'Sld' = Southland voice; 'NZ' = General New Zealand voice; 'NE' = Eastern New England voice.

Figure 1 represents hypotheses relating to non-prevocalic /ɹ/ (hypotheses 1a, 1c, 2a, 2b, and 4). According to these hypotheses, it was proposed that while listeners would consistently predict

high rates of /ɹ/ in the General American, and low rates of /ɹ/ in the General New Zealand voices, they would increasingly hear /ɹ/ in the Southland, and increasingly cease to hear /ɹ/ in the Eastern New England voice, according to a typical change-over-time trajectory. The change would be greater, however, with the Eastern New England voice than with the Southland voice, because non-rhoticity is more widespread in the former than rhoticity in the latter (hypothesis 4). As such, although listeners' perception of both voices would be very malleable, would come to expect less rhoticity in the Eastern New England voice than in the General New Zealand voice (hypothesis 2b), they would not come to hear more rhoticity in the Southland than in the General American voice (following from hypothesis 2a), save for in NURSE.

Equivalent phenomena relating to intrusive /ɹ/ are represented in Figure 2, following from hypotheses 1b, 1d, and 3. It was anticipated that listeners would consistently hear high levels of intrusive /ɹ/ in the General New Zealand, and low levels in the General American voices. By contrast, they would increasingly hear intrusive /ɹ/ in the Eastern New England, and decreasingly hear it in the Southland voice. These changes over time would, however, be more limited than the effects on perception of non-prevocalic /ɹ/, due to the relative lack of salience of the feature. The effect on perception of intrusive /ɹ/ in the Southland voice, in particular, would be very limited, whereas that on perception of intrusive /ɹ/ in the Eastern New England voice would be greater.

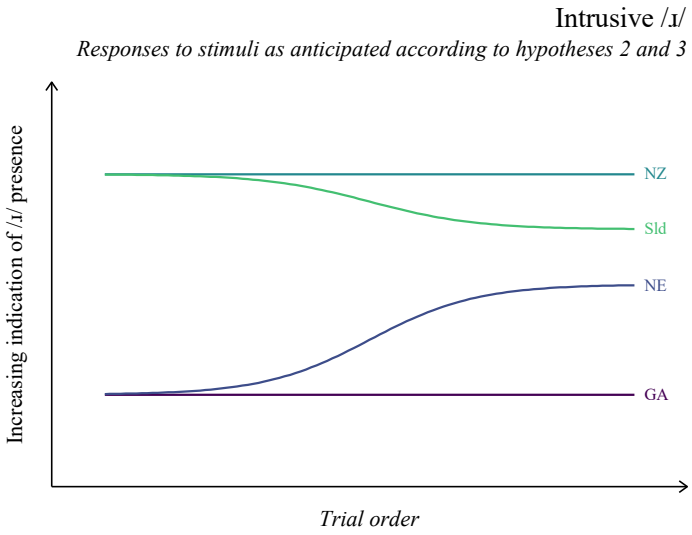

Figure 2: Expected outcomes for identification of /ɹ/ presence in intrusive contexts over time. 'NZ' = General New Zealand voice, 'Sld' = Southland voice, 'NE' = Eastern New England voice, 'GA' = General American voice.

4. Methodology

The basis of the study was a phoneme monitoring experiment adapted from that conducted by Hay, Drager, & Gibson (2018). Phoneme monitoring tasks, following the methodology established by Foss & Lynch (1969), require a participant to identify whether a sound is present or absent in a stimulus. Of particular interest is the phenomenon of phoneme restoration, where listeners 'restore' a sound not present in the stimulus, providing evidence of a mismatch between a stimulus and its perception (Warren, 1970; Samuel, 1981).

In an attempt to reduce the role of orthography on participant behaviour in phoneme monitoring as attested by Dijkstra, Roelofs, & Fieuws (1995), which of itself can lead listeners to 'hear' sounds other than those present in the stimuli (Hallé, Chéreau, & Segui, 2000), a training task was included before the main phoneme monitoring task.

Because the present experiment was geared towards assessing the relationship between listener expectations of /ɹ/ usage in voices associated with four different varieties, it followed a within-subjects design with counterbalancing, reflecting the intention to measure individual listeners' responses to different voices, rather than different listeners' responses to individual voices, whilst compensating for potential order effects. All listeners were presented all stimuli, but half heard the two New Zealand voices before the two American voices, and half heard them in the reverse order. The more familiar voices always preceded the less familiar voices as an influence of the former on expectations of the latter was anticipated regardless, whilst it was imperative that exposure to the less familiar varieties in no way complicated listeners' hearing of the more familiar varieties.

Participants were further required to complete a production task and a questionnaire. There existed a danger that the assumptions made regarding listeners' experiences with and expectations regarding different voices were incorrect, and these sources of information were intended to test them. Participants were required to read aloud sentences containing environments where intrusive /ɹ/ and non-prevocalic /ɹ/ occurred. The questionnaire was directed towards drawing information from participants on their relative exposure towards the language varieties used in the experiment, and their identification of each voice.

The research hypotheses and methods, including statistical models, were preregistered with the 'AsPredicted' online service (aspredicted.org), on 7 November 2018. The preregistration form is included in the appendix (§8.5).

4.1. Variables

In all tests, participants' responses as to whether or not /ɹ/ was present were measured. The primary independent variables were the voice and the trial order.

Additional variables expected to influence the data included the quality of the vowel preceding /ɹ/, whether it occurred at a morpheme or word boundary in the case of intrusive /ɹ/, and whether it was intrusive or non-prevocalic /ɹ/. Potential sources of random error included participant-specific effects, and effects relating to unanticipated characteristics of specific stimuli.

4.2. Stimuli

Stimuli for the experiment were elicited from four individuals, one from each of the language backgrounds of interest. To ensure some consistency across stimuli with regard to the fundamental frequency of speech, two female speakers were chosen for the New Zealand voices, and two male speakers for the American voices.

Phrases for the General New Zealand, Southland and General American speakers were recorded in a quiet environment at the University of Canterbury in Christchurch, New Zealand. The Eastern New England speaker was recorded separately in a sound-proofed recording studio at the University of Sheffield in the United Kingdom. All audio was recorded at a sample rate of 44.1 kHz and 16-bit depth, operated through a Praat (Boersma & Weenink, 2018) script (§8.1.2). This displayed each phrase orthographically on the screen, and played a low-pass filtered audio model of the same phrase spoken in a General New Zealand voice and recorded by the same means. The audio model, filtered with a Praat script (§8.1.1) retaining sound in the 0-300 Hz range, with smoothing of 100 Hz, was provided to achieve greater consistency in the speech rate, intonation and stress patterns used across stimuli.

Consistent inclusion or exclusion of non-prevocalic or intrusive /ɹ/ was required of all speakers, in a manner unreflective of real, variable use but necessary in the study to provide a consistent model for learning. Successive recordings of each stimulus phrase for each speaker were preserved, allowing for an assessment of the differences between their initial use of non-prevocalic or intrusive /ɹ/, without the overt influence of the author, and the ultimate, consistent pronunciations. Speakers' behaviour with regard to non-prevocalic /ɹ/ was largely as anticipated based on the findings on the typical use of this feature among speakers from the relevant areas (§2.5). Use of intrusive /ɹ/ had, by contrast, been overestimated. The use of

intrusive /ɹ/ by the Eastern New England speaker was entirely forced, and had to be extensively forced in the General New Zealand voice as well.

Stimuli consisted of audio recordings of ninety short English phrases in which two semantically-related words were conjoined, for example <boarding and leaving>, as in Hay, Drager, & Gibson (2018). This semantic connection was intended to reduce the possibility of participants misinterpreting the words they heard as other, homophonous words.

Of the ninety phrases, sixty contained a single environment in which non-prevocalic or intrusive /ɹ/ might occur, and thirty were control phrases lacking such environments. In those phrases containing the relevant environments, half occurred in the word preceding, and half in the word succeeding the conjunction, removing any element of predictability with regard to the location of /ɹ/ in the phrases. Half the control phrases included a single instance of prevocalic /ɹ/, and half lacked any environment in which /ɹ/could occur. Prevocalic /ɹ/was included to give participants cases where /ɹ/was clearly present, and the other control words to supply cases where /ɹ/ was clearly absent, reducing the tendency of participants to guess. The breakdown of stimuli by class is illustrated in Table 1.

Table 1: Breakdown of perception experiment stimuli, according to class and vowel context. A table annotated to display which stimuli contained /ɹ/ in each voice appears in §8.4.2.

	Variables		*Control stimuli*		
Vowel context	Non-prevocalic /ɹ/	Intrusive /ɹ/	Prevocalic /ɹ/	/ɹ/-less	*Total*
BATH/PALM/START	10	10	7	8	*35*
THOUGHT/NORTH/FORCE	10	20	8	7	*45*
NURSE	10				*10*
Total	*30*	*30*	*15*	*15*	*90*

All phrases included at least one stressed vowel, limited to the lexical sets BATH/PALM/START, THOUGHT/NORTH/FORCE, and NURSE.[3] These restrictions were imposed due to the demands of the linguistic phenomena of interest, which occur in relatively few environments, and in order to keep some consistency across phrases. Intrusive /ɹ/ occurs only after mid- or open vowels and diphthongs with a schwa offglide (Gimson, 1962, pp. 204-205). Non-prevocalic /ɹ/ occurs

[3] 'Lexical sets', and corresponding 'keywords', as defined by Wells (1982), are followed throughout. THOUGHT, NORTH, and THOUGHT are considered to have comparable vowels in all voices. BATH, PALM, and START are considered to have the same vowel quality save for where BATH in American voices is potentially distinct from PALM/START.

in START, NORTH, NURSE, LETTER, NEAR, SQUARE and CURE. With diphthongs and reduced vowels excluded, there remain the environments BATH/START and THOUGHT/NORTH/FORCE for intrusive and non-prevocalic /ɹ/, and NURSE specifically for the latter. NURSE was included due to its irregular occurrence across the four varieties. All phrases contained instances of these vowels, either preceding or succeeding /ɹ/.

The inclusion of the same vowels in the control phrases lacking rhotic consonants was intended to further serve to disassociate sounds and orthography, as listeners would hear environments in which /ɹ/ was phonotactically possible but where it clearly never occurred. All stimuli are listed in the appendix (§8.4.1).

Although effects of the occurrence of intrusive /ɹ/ at either a morpheme[4] or a word boundary such as those identified by Hay, Drager, & Gibson (2018) were not a subject of study, phrases featuring both of these contexts were included to balance the set of stimuli, and boundary effects were factored into statistical modelling.

4.3. Participants

Listeners in the experiment were thirty-nine individuals from a New Zealand English language background, eighteen female and twenty-one male, born between 1944 and 2000, with a median birth year of 1993 due to the presence of a greater number of younger participants. All listeners identified as having normal speech and hearing.

All listeners grew up in New Zealand, and identified having English as a native language. Participants potentially had an additional native language, and thirteen had parents from a different language background. Information on these factors was collected in the questionnaire (§4.4.4), but did not lead to their exclusion from the sample.

Given the variability of non-prevocalic /ɹ/ in New Zealand today (§2.5.1), it was entirely possible that some participants would produce /ɹ/ in some non-prevocalic environments, and particularly in words of the NURSE lexical set. Six participants pronounced non-prevocalic /ɹ/ in more than one of twenty-five stimuli in the production task (§4.4.3), while a further five had a single case of non-prevocalic /ɹ/. Rates of rhoticity among the former group varied between

---

[4] All instances of potential intrusive /ɹ/ at morpheme boundaries were in the gerundive forms of words, for example <gnawing and chewing>. To avoid a predictable association of gerundive forms with intrusive /ɹ/ presence, gerundive forms were included in equal numbers in non-prevocalic /ɹ/ and /ɹ/-less stimuli.

six and fifteen in twenty-five, and were readily explicable in terms of unusual characteristics of participants' backgrounds, including connections with the southern South Island in three cases. Occurrences of non-prevocalic /ɹ/ in other participants appeared comparatively anomalous, and these listeners were not excluded from the sample, while the six participants with more frequent use of non-prevocalic /ɹ/ were.

4.4. Procedure

The experiment was conducted in a quiet laboratory environment at the University of Canterbury. Listeners were seated in front of a laptop computer screen displaying the experiment, and listened to stimuli through headphones. Interaction between participants and the experimenter was kept to a minimum beyond the requirement of obtaining informed consent, to limit experimenter effects such as those found by Hay, Drager, & Warren (2010).

The experiment program was built using the Python experiment design library PsychoPy 3 (Peirce, 2007). The program included all tasks listeners were required to complete to prevent interruption by the experimenter. Instructions for the experiment prior to the commencement of the perception task, excepting messages asking for a key response, were conveyed through pre-recorded audio, intended to attune participants to spoken rather than written language from the beginning of the experiment.

Audio instructions were recorded from a different female speaker of between 20 and 30 years from a New Zealand English language background, which given its familiarity to participants would not convey any novel linguistic information prior to the commencement of the perception experiment. To further prevent any influence on participant performance, words in which intrusive or non-prevocalic /ɹ/ could occur were excluded.

Listeners were required to press the LEFT arrow key to indicate they believed the sound indicated in the preceding instructions was *present*, and the RIGHT arrow key to indicate they believed it was *absent* (§8.4.1). The computer displays for both the training task and the main perception task included arrow symbols pointing left and right with <Yes> and <No> displayed in text above them to help establish a relationship between key presses and answers, reducing any potential effect resulting from initial difficulties in adjusting to the setup of the experiment. In both the training task and the perception task, stimulus phrases were played for two seconds. All stimuli appeared in a random order within each block, negating stimulus order effects.

4.4.1. Training task

In the training task, unlike in the main perception task (§4.4.2), listeners were not required to provide responses within a certain time, allowing them to adjust to the setup of the experiment and receive any necessary feedback on performance without being penalised for taking too long to answer.

The training task included two sets of four phrases each, where a different sound was being monitored. This followed the model of Hay, Drager, & Gibson (2018), where participants were asked to listen for /f/ and /s/. These are particularly suited given the substantial variation found in these sounds' orthographic representation. Training task phrases were recorded from the same speaker providing the audio instructions, and avoided any potential /ɪ/ instances (§8.4.3).

Listeners received feedback on their responses through audio messages (§8.2.3) and were notified of whether their answer was correct or incorrect. Where there was a mismatch between orthography and sound and the participant provided a wrong answer, an explanation of the mistake was given in audio feedback and the recording of the phrase was replayed at three-quarters of the tempo.

4.4.2. Perception task

In the main task, listeners were required to respond within the two seconds for which each phrase ran; if they did not, a message appeared on the screen, asking the listener to respond more quickly. Listeners were notified of this in the instructions preceding the task (§8.2.4), and key-answer correspondences were further reinforced.

At the beginning of the experiment, all listeners were assigned a number and group. The group assigned, A or B, was used in counterbalancing for the perception task. Participants in Group A received the New Zealand stimuli prior to the American stimuli, and participants in Group B received the reverse. Listeners were able to take breaks between the four blocks. Excluding breaks, the perception task ran for no more than 12 minutes to limit fatigue, though varying with listeners' response rate.

4.4.3. Production task

In the production task, participants were recorded reading out sentences displayed on the screen. Recording was conducted in Audacity (Audacity Team, 2019) on a separate computer, with a sample rate of 44.1 kHz and 16-bit depth. Stimulus sentences were twenty simple, single-clause sentences (§8.4.4). They contained fifteen environments in which intrusive /ɪ/ could occur, including five tokens where it could occur at word boundaries after BATH, five

where it could occur at word boundaries after THOUGHT, and five where it could occur at morpheme boundaries after THOUGHT. Some perception task stimuli were repeated for want of alternatives, though this was avoided where at all possible to limit the effect of participants' experience of the perception task on their performance in the production task.

The production sentences further contained twenty-five environments in which non-prevocalic /ɹ/ could occur, including eight for each of the lexical sets START and NORTH/FORCE, and nine of NURSE. Lastly, the production sentences include sixteen instances of prevocalic /ɹ/, eight of which preceded a THOUGHT vowel and eight of which preceded a BATH vowel, to deemphasise the features of interest.

4.4.4. Questionnaire

Initial questions asked for basic demographic information. Subsequently, participants were asked to identify as many 'accents' as they believed they heard in the perception task, in the order that they heard them.

Following this, participants were asked to indicate, on four on-screen clickable rating scales of five points each, how often they typically heard each of these varieties used in different social contexts: among family, among friends, at work, and through media. Each point on a scale represented a time period, either 'Daily', 'Weekly', 'Monthly', 'Yearly' or 'Never'. Participants were told that 'Daily' represented a rate of more than once a week, that 'Weekly' represented a rate of more than once a month, and so on. A transcript of the questionnaire is included in the appendix (§8.3).

4.5. Treatment of data

Data resulting from the experiment would be potentially incomplete, were participants to take too long to respond to a phrase in the perception task, or fail to answer all questions in the questionnaire. Participants displaying regular use of non-prevocalic /ɹ/ were excluded as detailed in §4.4.3. Participants whose mean accuracy in control items was 2.5 standard deviations below the mean accuracy for the sample as a whole were also removed from the sample.

4.6. Method of analysis

Results were analysed in R (R Core Team, 2018), using logistic mixed-effects regression modelling based on the package 'lme4' (Bates, Mächler, Bolker, & Walker, 2015), assessing

the relationship between participants' responses with regard to /ɹ/ presence or absence and trial order in the experiment.

Separate models were planned for non-prevocalic and intrusive /ɹ/, each featuring a three-way interaction between trial order, voice, and some criteria relating to the environment of /ɹ/. For non-prevocalic /ɹ/ stimuli, this was the lexical set of the preceding vowel; for intrusive /ɹ/, it was both the lexical set and the nature of the boundary preceding the /ɹ/ token, either a word or morpheme boundary. Random intercepts would be included for both the listener and the unique stimulus (out of 360), allowing for baseline variation in /ɹ/ presence identification across listeners and for different stimuli, and a random by-listener slope would be included for trial order, allowing for variation in individuals' rate of perceptual learning over time, consistent with individual variability in statistical learning (Siegelman et al., 2016). The details of both models were preregistered (§8.5).

Model 1: non-prevocalic /ɹ/ (preregistered)

$$\text{keyPressed} \sim \text{orderTrialCentred} * \text{variety} * \text{lexicalSet} + (1 + \text{orderTrialCentred} \mid \text{participant}) + (1 \mid \text{stimUnique})$$

The model preregistered for non-prevocalic /ɹ/ stimuli is displayed here (Model 1), where `keyPressed` represents the binary participant response, `orderTrialCentred` represents trial order for stimuli of each voice, centred to standardise across varying numbers of trials due to variable rates of completion, `variety` may be one of four voices played in the experiment, `lexicalSet` represents the category of the vowel – NORTH/FORCE, START or NURSE – preceding the /ɹ/ token, `participant` represents each unique listener, and `stimUnique` represents the number of each unique recording of a stimulus presented in the experiment.

Model 2: intrusive /ɹ/ (preregistered)

$$\text{keyPressed} \sim \text{orderTrialCentred} * \text{variety} * \text{vowelBoundary} + (1 + \text{orderTrialCentred} \mid \text{participant}) + (1 \mid \text{stimUnique})$$

For the preregistered model for intrusive /ɹ/ (Model 2), the terms are the same as those for Model 1, save for the replacement of the `lexicalSet` predictor with `vowelBoundary`, a factor with three levels: BATH at a word boundary, THOUGHT at a word boundary, and THOUGHT at a morpheme boundary.

The practical implementation of these models necessitated the introduction of some modifications. Trial order was scaled as well as centred (`orderTrialSC` in models) and the optimiser BOBYQA was added to resolve non-convergence problems. Significantly, the

random slope for trial order had to be discarded in models for non-prevocalic, though not intrusive /ɹ/, in that the slopes' inclusion led to singular fit. Models were pared-down through stepwise analysis of variance (ANOVA) comparison; where a more complex model did not perform significantly better than any of a number of possible simpler models, the Akaike Information Criterion (AIC) was employed in comparing these reduced models and whichever model performed best according to this criterion was selected.

Separate tests were conducted in order to ascertain whether the order in which a participant heard each voice, reflecting which group they were in for counterbalancing purposes, had a significant effect on their perception of /ɹ/ presence or absence. As there did appear to be such an effect, at least for intrusive /ɹ/, counterbalancing group was introduced into modified forms of the preregistered models as a new predictor, and stepwise reduction of these models proceeded again.

Attempts were made at modelling potential non-linear effects with restricted cubic spline regression modelling, using the function rcs() from the 'rms' package in R (Harrell, 2018), and at using participants' familiarity with each voice as indicated in questionnaire responses as a predictor in linear modelling. Both avenues of analysis proved unviable however, due to convergence issues and the lack of a suitable means of capturing varying and opposite effects of familiarity based on voice: a less familiar voice type might be associated with either increasing or decreasing indication of /ɹ/ presence, and there was no suitable measure for the degree of change irrespective of direction that could feasibly be used as a dependent variable.

4.6.1. Analysis of questionnaire responses

Two kinds of data from the questionnaire – five-point scale responses regarding listeners' familiarity with different varieties in different contexts, and written responses providing their guesses regarding which accents they had heard in the experiment – received attention in the study.

Listeners' responses to how often they tended to hear a variety in a particular context, on a scale from 1 to 5 representing increasingly more frequent time intervals, were added together to provide a rough estimate of their experiences with each kind of voice. No attempt was made to give variable weight to the different contexts in which a voice might be heard due to the complexities this would introduce, though doubtlessly there would be such variability.

Out of one-hundred and twenty answers (for four voices, for thirty participants) to the voice identification questions, one-hundred were received. The answers 'Southland' and

‘Invercargill’ were taken to indicate a Southland voice. No ‘Otago’ responses remained in the sample after exclusions, although these were also included in the latter category. All other references to New Zealand were taken to indicate a ‘General New Zealand’ voice. The responses ‘Texas’, ‘Southern US’, and ‘Tennessee’ contributed to a ‘Southern United States’ category. ‘New York’ and ‘Chicago’ were categorised separately, and responses mentioning ‘New England’ or ‘Boston’ were categorised as ‘Eastern New England’ responses. All other references to the United States were put in a ‘General American’ category.

5. Results

Inclusion of the order of voices produced noteworthy changes in both the model for non-prevocalic /ɹ/ and that for intrusive /ɹ/. While only the modified intrusive model performed significantly better than the corresponding simpler model in ANOVA comparison, it remains worthwhile to retain the voice order predictor in both models because it is theoretically well-founded, and because its inclusion reveals important patterns that would otherwise be obscured. These changes to the models do not amount to a significant deviation from the method originally proposed insofar as they follow from the original theoretical assumptions, but do ameliorate problems of interpretation found in the original models proposed. Predictions of both the modified and the original models are described and compared here.

Results for participant responses to non-prevocalic /ɹ/ stimuli are described in §5.1; these are relevant to hypotheses 1a, 1c, 2a, 2b, and 4. Results for responses to intrusive /ɹ/ are detailed in §5.2; these are relevant to hypotheses 1b, 1d, and 3. Conclusions that hold generally are discussed in §5.3. Of interest in all models are interactions between trial order and voice – relevant to hypotheses 1a, 1b, 1c, 1d, and 3 – the absolute differences in response rates for different voices – relevant to hypotheses 2a and 2b – and differences with respect to NURSE as opposed to other vowel contexts, relevant to hypothesis 4.

Model predictions are displayed in tabulated data (Tables 1-4), listing estimates for fixed effects, and in graphs (Figures 3-11). Each graph is organised such that the *x*-axis represents the time course of each block of stimuli for a particular voice, measured by trial order within each block. The *y*-axis in each case measures the extent to which listeners indicated that they believed /ɹ/ was *present* in a stimulus. As such, all graphs display listeners’ perception of /ɹ/ presence *over the course of a block*. In different figures, this is broken up according to the relevant voice, voice order (according to counterbalancing group), and vowel context. In all graphs, ‘NZ’ (light blue) denotes the General New Zealand voice, ‘Sld’ (green) denotes the

Southland voice, ‘GA’ (purple) denotes the General American voice, and ‘NE’ (deep blue) denotes the Eastern New England voice. The translucent shaded bands behind each trend line represent the corresponding ninety-five per cent confidence intervals.

5.1. Models for non-prevocalic /ɹ/

5.1.1. Model 1a (preregistered model for non-prevocalic /ɹ/)

Model 1a, the optimal model following reduction (§4.6) of the preregistered model for non-prevocalic /ɹ/ (§8.5), took the form

Model 1a: non-prevocalic /ɹ/ (preregistered model, simplified)

$$\text{keyPressed} \sim \text{variety} * (\text{orderTrialSC} + \text{lexicalSet}) + (1 \mid \text{participant}) + (1 \mid \text{stimUnique})$$

Predictions of Model 1a, in which the order of voices is not accounted for, are displayed in Table 2, and graphed in Figure 3, where estimates are not divided according to vowel context, mirroring the hypotheses represented in Figure 1. Figure 4 displays the same estimates, broken up according to vowel context.

As suggested in Table 2, according to this model, listeners clearly differentiated each voice (the main effects for each voice), and tended to hear less rhoticity in the Eastern New England voice than in the General New Zealand voice, and to some extent heard more rhoticity in the Southland than in the General American voice, at least in some contexts. Listeners under-predicted /ɹ/ in the Eastern New England voice to some extent relative to the General New Zealand voice, and over-predicted /ɹ/ in both NURSE and NORTH/FORCE contexts in the Southland versus the General American voice.

Table 2: Summary for Model 1a, the original model for non-prevocalic /ɹ/, featuring interactions between voice and preceding vowel quality, and voice and trial order.

Fixed Effects

	Estimate	**Std. Error**	***z***	***p***	
Main effects					
(Intercept)	1.8422	0.2623	7.024	<0.001	***
General New Zealand voice	-1.6278	0.2574	-6.324	<0.001	***
Southland voice	-1.4866	0.2562	-5.802	<0.001	***
Eastern New England voice	-1.9939	0.2576	-7.74	<0.001	***
NORTH/FORCE contexts	-0.783	0.2608	-3.002	0.0027	**

NURSE contexts	-0.3098	0.267	-1.16	0.2461	
Trial order	-0.2588	0.0883	-2.931	0.0034	**
Interactions					
NORTH/FORCE: General NZ	1.0266	0.3521	2.915	0.0036	**
NORTH/FORCE: Southland	1.0895	0.3512	3.103	0.0019	**
NORTH/FORCE: ENE	1.389	0.3513	3.955	<0.001	***
NURSE: General NZ	0.5204	0.3562	1.461	0.1441	
NURSE: Southland	1.5679	0.3653	4.292	<0.001	***
NURSE: ENE	1.7474	0.3618	4.83	<0.001	***
General NZ voice by trial	0.3111	0.1174	2.65	0.0081	**
Southland voice by trial	0.3307	0.1183	2.796	0.0052	**
ENE voice by trial	0.2934	0.1179	2.488	0.0128	*

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

There also appear perceptual learning trends of varying significance in Table 2: contrary to what was anticipated (hypotheses 1a and 1c), listeners appear to have increasingly heard /ɪ/ in the General New Zealand and the Eastern New England voices as well as the Southland voice over time, although the latter effect was most significant.

The nature of these learning trends is more clearly displayed in Figure 3.

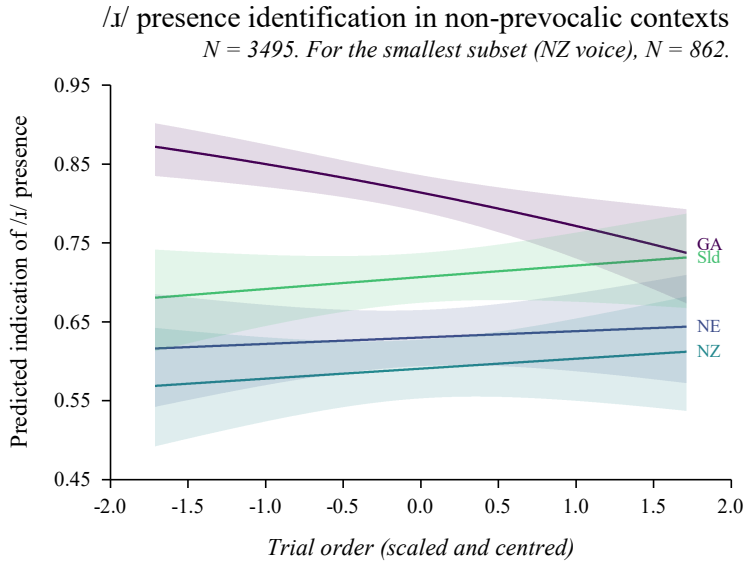

Figure 3: Model 1a predictions for hearing of /ɪ/ over time, by voice.

As anticipated (hypotheses 1a and 1c), listeners initially expected very high levels of /ɹ/ in the General American voice, and relatively low levels in the General New Zealand voice (but note that listeners were still more likely than not to hear /ɹ/ in the latter voice, significantly over-predicting its presence). Against the hypotheses, listeners appeared to expect substantially more /ɹ/ in the Southland compared with the General New Zealand voice from the beginning, and expected relatively little /ɹ/ in the Eastern New England voice from the beginning.

The most significant learning trend appears to be the decreasing hearing of /ɹ/ in the General American voice (the intercept in Table 2), something that was not expected. While listeners did indeed increasingly hear /ɹ/ in the Southland voice according to this model, this phenomenon was comparatively weak. Figure 3 also appears to show that listeners did not in fact hear less non-prevocalic /ɹ/ in the Eastern New England than in the General New Zealand voice, contrary to what was anticipated (hypothesis 2b).

There are complexities relating to the vowel contexts of stimuli that are obscured in this representation of the data, however. The same predictions are broken down by environment in Figure 4. The learning effect for the General American voice occurs quite consistently regardless of vowel context, always appearing as a decrease in hearing of /ɹ/ over time from near-ceiling levels. Listeners appeared significantly less likely to hear /ɹ/ in START environments in the Eastern New England voice than in the General New Zealand voice, but about as likely to hear it in NORTH/FORCE environments, and significantly more likely to hear it in NURSE environments. Listeners were always more likely to hear rhoticity in the Southland voice than in all other voices bar the General American voice, though to varying degrees: they were most likely to hear it in NURSE, and least likely to hear it in START. Listeners appeared noticeably more likely to hear NURSE in a Southland voice after 90 trials than in a General American voice after 90 trials, as hypothesised (hypothesis 2a), though this over-prediction does not occur in START contexts, and is insignificant in NORTH/FORCE contexts.

Based on Model 1a predictions, we would assume that none of the anticipated perceptual learning trends occurred, while there was a perceptual learning trend relating to the General American voice that was not anticipated. The hypothesis that listeners would overall find the Southland voice more rhotic than the General American, at least in NURSE, is by contrast well-supported, although there is no evidence for the related hypothesis that listeners would find the Eastern New England voice less rhotic than the General New Zealand voice.

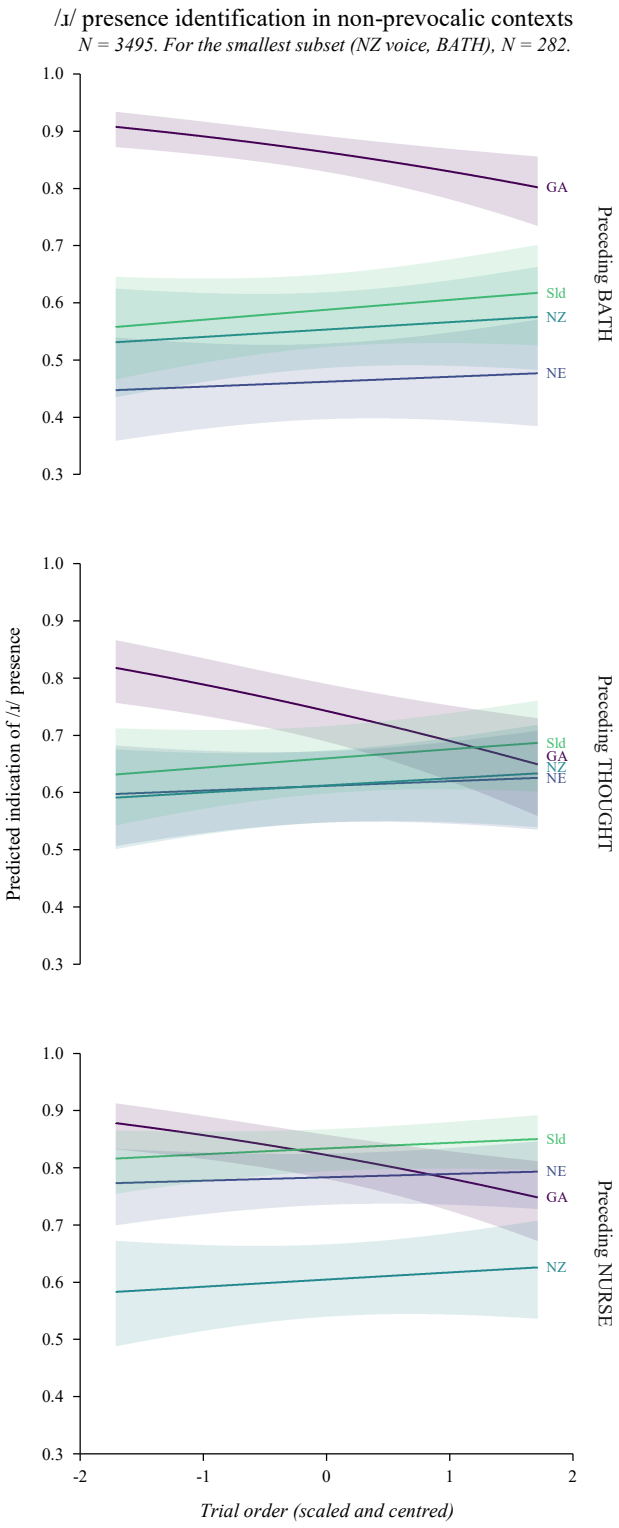

Figure 4: Model 1a predictions for hearing of /ɹ/ over time, by vowel context and voice.

5.1.2. Model 1b (modified model for non-prevocalic /ɹ/)

With the inclusion of the order of voices as a variable, the modified model for non-prevocalic /ɹ/, Model 1b, took the form

Model 1b: non-prevocalic /ɹ/ (simplified, and modified to include counterbalancing group)

$$\text{keyPressed} \sim (\text{set} * \text{variety}) * (\text{orderTrialSC} + \text{lexicalSet}) + (1 \mid \text{participant}) + (1 \mid \text{stimUnique})$$

where `set` represents the counterbalancing group or order of voices (whether the listener heard the New Zealand or the American voices first).

Significant perceptual learning trends that occurred are obscured in the predictions of Model 1a due to the omission of voice order as a predictor. Model 1b rectifies this, and reveals a clearer picture relevant to hypotheses 1a and 1c in particular. Table 3 displays the predictions of this model, with comparatively fewer, but more relevant, significant effects. The only significant perceptual learning trend listed is that for the Southland voice, a tendency for listeners to increasingly hear /ɹ/. While there are no significant interactions featuring the order of voices as a variable, factoring voice order in clarifies the issues raised by Model 1a.

Table 3: Summary of Model 1b, the new model for non-prevocalic /ɹ/, featuring interactions between voice order, voice and preceding vowel quality, and voice order, voice and trial order.

Fixed Effects

	Estimate	**Std. Error**	***z***	***p***	
Main effects					
(Intercept)	1.510	0.345	4.382	<0.001	***
Set B (American voices first)	0.714	0.485	1.472	0.141	
New Zealand voice	-1.412	0.324	-4.357	<0.001	***
Southland voice	-1.160	0.322	-3.608	<0.001	***
New England voice	-1.740	0.323	-5.389	<0.001	***
THOUGHT contexts	-0.440	0.329	-1.337	0.181	
NURSE contexts	-0.253	0.332	-0.761	0.446	
Trial order	-0.256	0.122	-2.091	0.037	*
Two-way interactions					
General New Zealand voice, for Set B	-0.490	0.429	-1.141	0.254	
Southland voice, for Set B	-0.693	0.426	-1.626	0.104	
Eastern New England voice, for Set B	-0.570	0.429	-1.329	0.184	
NORTH/FORCE contexts, for Set B	-0.738	0.438	-1.685	0.092	

NURSE contexts, for Set B	-0.138	0.454	-0.303	0.762	
Trial order, for Set B	0.011	0.178	0.064	0.949	
NORTH/FORCE: General NZ	0.772	0.448	1.723	0.085	
THOUGHT: Southland	0.693	0.445	1.557	0.120	
NORTH/FORCE: ENE	0.906	0.445	2.038	0.042	*
NURSE: General NZ	0.462	0.448	1.030	0.303	
NURSE: Southland	1.305	0.458	2.850	0.004	**
NURSE: ENE	1.905	0.460	4.140	<0.001	***
General NZ voice by trial	0.238	0.164	1.447	0.148	
Southland voice by trial	0.491	0.165	2.975	0.003	**
ENE voice by trial	0.191	0.167	1.139	0.255	
Three-way interactions					
Set B: NORTH/FORCE: General NZ	0.583	0.576	1.013	0.311	
Set B: NORTH/FORCE: Southland	0.834	0.573	1.455	0.146	
Set B: NORTH/FORCE: New England	1.036	0.574	1.807	0.071	
SET B: NURSE: General NZ	0.146	0.587	0.249	0.803	
Set B: NURSE: Southland	0.602	0.611	0.986	0.324	
Set B: NURSE: ENE	-0.260	0.600	-0.433	0.665	
Set B: General NZ voice by trial	0.125	0.236	0.530	0.596	
Set B: Southland voice by trial	-0.353	0.239	-1.477	0.140	
Set B: ENE voice by trial	0.180	0.238	0.755	0.450	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Model 1b predictions are displayed in Figure 5, where estimates are broken up by voice order. The upper panel in this figure shows predictions for responses of listeners in Group A, those who first heard the General New Zealand voice, followed by the Southland voice, the General American voice, and lastly the Eastern New England voice. The lower panel displays predictions for Group B, where the General American voice was played first, followed by the Eastern New England voice and the two New Zealand voices. A very pronounced difference is apparent. While the hypothesised learning trend did not occur with stimuli from the Southland voice when all other voices had already been heard (if anything, there was a reverse trend), there was a highly significant trend towards increased hearing of non-prevocalic /ɪ/, where the only reference point was a General New Zealand voice, to the point that listeners over-predicted rhoticity in the Southland versus the General American voice in this condition by the end of each block.

The trend towards decreased hearing of /ɹ/ in the latter voice occurred regardless of voice order. There were no clear perceptual learning trends with the other voices, though there was an overall tendency for listeners to increasingly hear non-prevocalic /ɹ/ in these voices over time. Figure 5 does not reveal any tendency for listeners to under-predict non-prevocalic /ɹ/ in the Eastern New England relative to the General New Zealand voice.

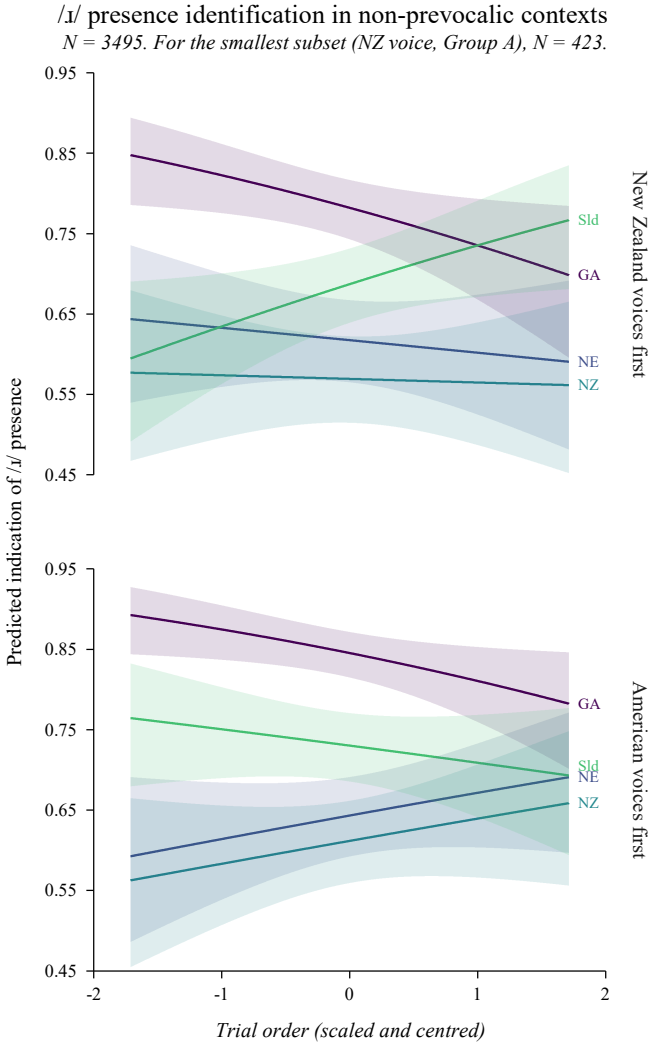

Figure 5: Model 1b predictions for hearing of /ɹ/ over time, by voice order and voice.

Further complexities are revealed in Figures 6-8. Each of these figures displays predictions for each vowel context separately, with estimates for each voice separated and ordered according to voice order, such that the overall trajectory of listener responses over the course of the experiment is visible.

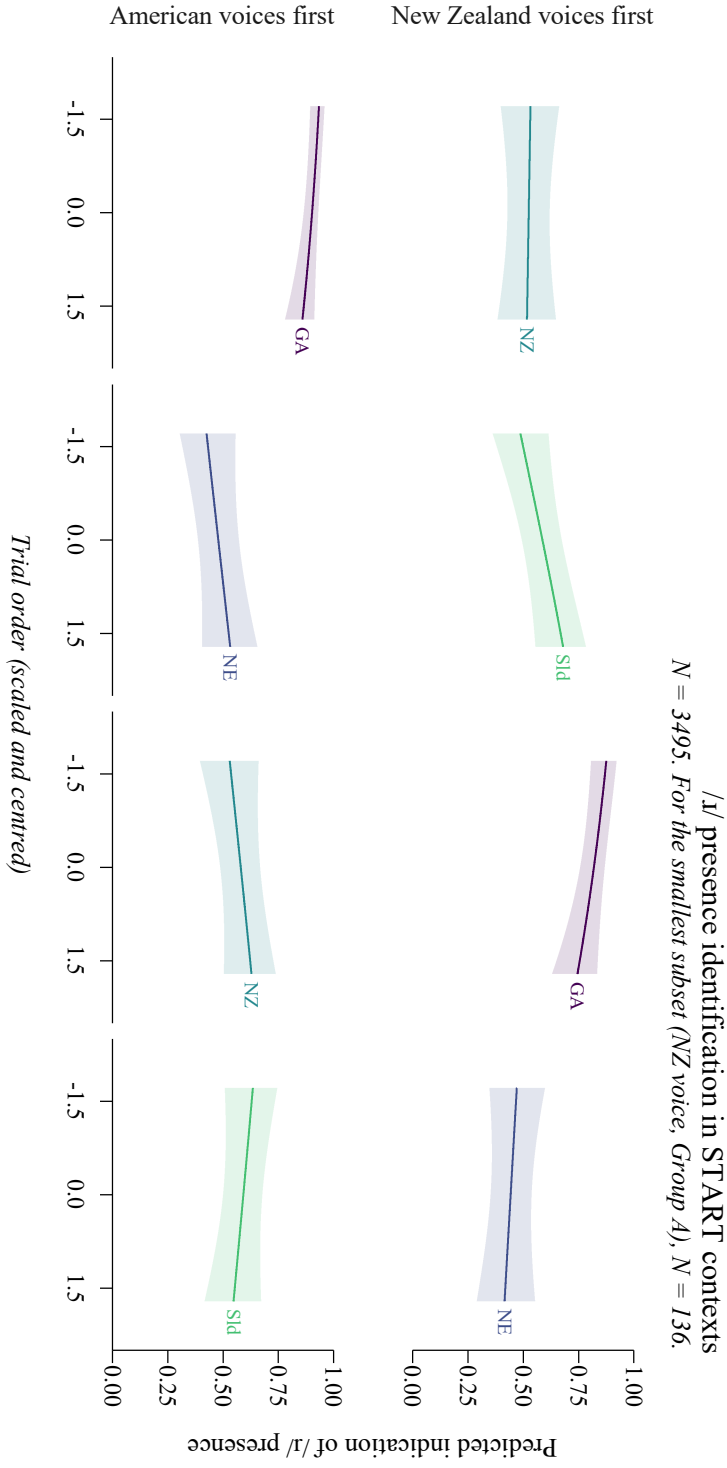

Figure 6: Model 1b predictions for hearing of /ɹ/ over time in START contexts, showing the trajectory of responses over the course of the whole experiment for each voice order.

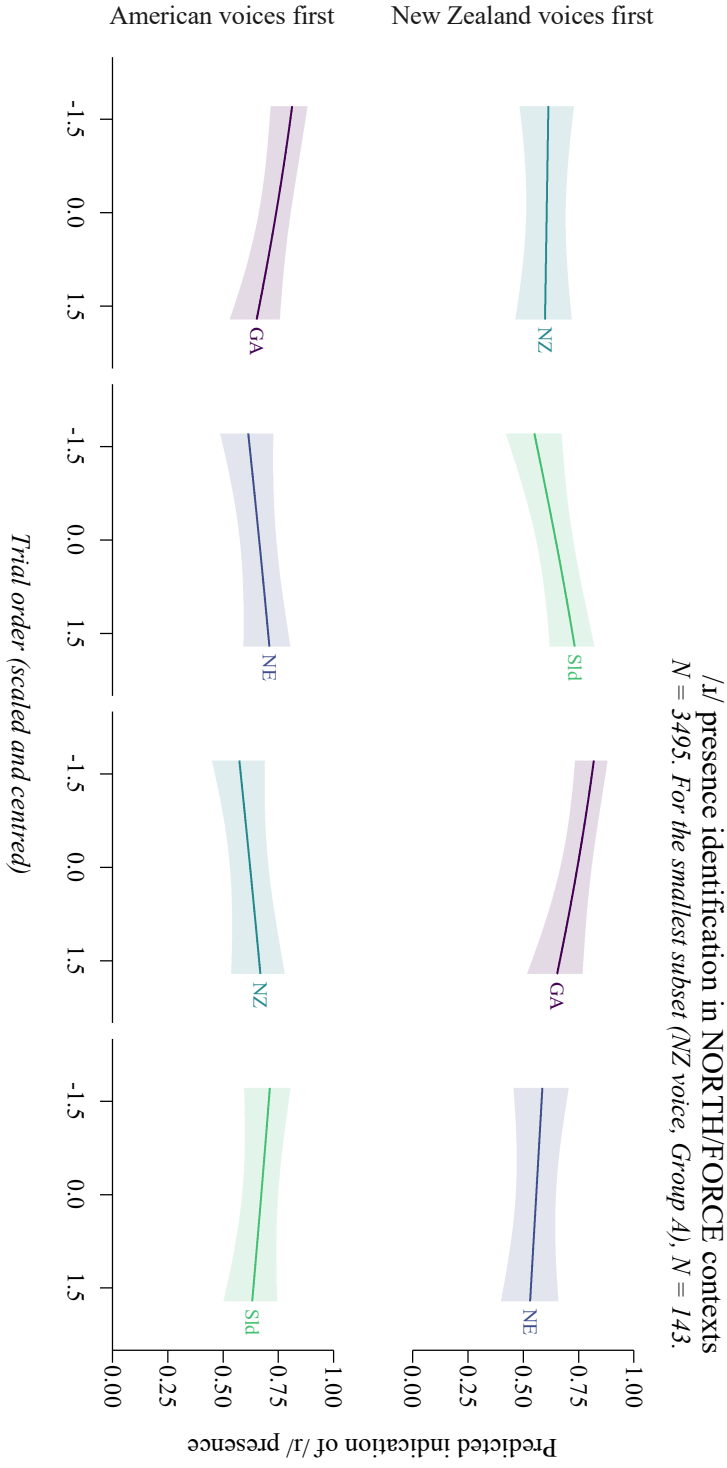

Figure 7: Model 1bpredictions for hearing of /ɪ/ over time in NORTH/FORCE contexts, showing the trajectory of responses over the course of the whole experiment for each voice order.

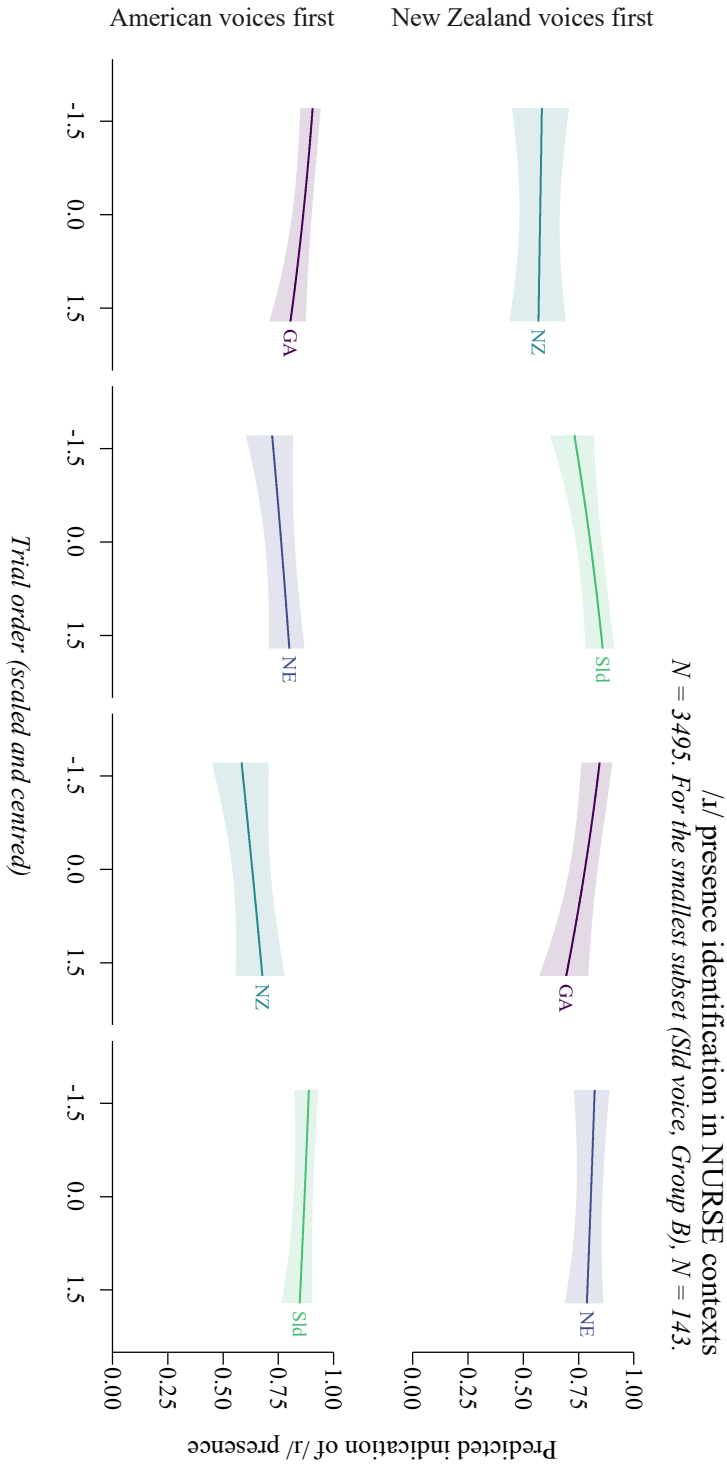

Figure 8: Model 1b predictions for hearing of /ɪ/ over time in NURSE contexts, showing the trajectory of responses over the course of the whole experiment for each voice order.

Of particular interest is the evidence of both continuities and discontinuities in hearing of /ɹ/ at when the voice being heard changed. It is known that listeners must always to some extent adjust their perception to a new voice (Pisoni, 1993). As such, discontinuities – significant differences in initial estimates of hearing of /ɹ/ in one voice relative to final estimates of hearing of /ɹ/ in the preceding voice – are evidence of significant responses to characteristics of a new voice, in which new expectations are formed based on limited auditory cues.

Continuities appear whenever the General New Zealand voice follows the Eastern New England voice, save for in NURSE contexts (Figure 8), and typically when the Eastern New England voice follows the General American voice, but not in START contexts (Figure 6), or NURSE contexts when all other voices have already been heard (Figure 8). Continuities also typically occur whenever the Southland voice follows the General New Zealand voice, except for with NURSE stimuli. Major discontinuities, by contrast, occur with NURSE stimuli when the Southland voice is heard after the General New Zealand voice or the Eastern New England voice is heard last after the General American voice (Figure 8), and with START stimuli when the Eastern New England voice is heard after the General American voice, and the General American voice is heard after the Southland voice. Listeners were highly sensitive to Eastern New England START, distinguishing it significantly from General American START as a non-rhotic context, even though they did not draw the same distinction with NORTH/FORCE. They were also more likely to hear START as non-rhotic in an Eastern New England than in a General New Zealand voice. Listeners were sensitive to Southland NURSE stimuli, from the outset finding them more rhotic than the General New Zealand NURSE stimuli. They were also generally more likely to be sensitive to rhotic NURSE in the last block of the experiment than in earlier blocks.

Importantly, there are perceptual learning trends that hold regardless of these continuities or discontinuities. The trajectories of perception of the General New Zealand and Eastern New England stimuli, given the continuities across voices and the variability in trends for each voice, are not noteworthy. While the perceptual learning trend towards hearing decreased /ɹ/ in the General American voice is noteworthy for its consistency, it is always from near-ceiling level. There is, however, always a perceptual learning trend towards increased hearing of /ɹ/ in the Southland voice, in all vowel contexts though to varying degrees, when the Southland voice is heard second, in accordance with hypothesis 1a. With START and NORTH/FORCE, listeners initially expected levels of rhoticity comparable to that which they predicted in the General

New Zealand voice, but by the end of the block heard significantly more /ɹ/, in spite of the absence of any /ɹ/ in any of these stimuli.

5.2. Models for intrusive /ɹ/

5.2.1. Model 2a (preregistered model for intrusive /ɹ/)

The optimal model for intrusive /ɹ/ based on that which was preregistered (§8.5), Model 2a, took the form

Model 2a: intrusive /ɹ/ (adjusted)

$$\text{keyPressed} \sim \text{variety} + \text{orderTrialSC} + \text{vowelBoundary} + (1 + \text{orderTrialSC} \mid \text{participant}) + (1 \mid \text{stimUnique})$$

Predictions for listeners' hearing of /ɹ/ intrusive contexts, when voice order is not accounted for, are displayed in Table 4 and Figure 9. No significant perceptual learning trends emerge in this model, neither those that were predicted (hypotheses 1b and 1d) nor any others.

Listeners were overall more likely to expect intrusive /ɹ/ in the Eastern New England than in the General New Zealand voice, which would be in accordance with the overall malleability argument, though it was not specifically hypothesised. Going against this, they were less likely to expect intrusive /ɹ/ in the General American than in the Southland voice.

Table 4: Summary of Model 2a, the original model for intrusive /ɹ/, featuring main effects of voice, morphophonemic context and trial order.

Fixed Effects

	Estimate	**Std. Error**	***z***	***p***	
(Intercept)	-1.566	0.284	-5.505	<0.001	***
General New Zealand voice	1.898	0.169	11.206	<0.001	***
Southland voice	0.521	0.166	3.143	0.002	**
Eastern New England voice	2.432	0.174	14.008	<0.001	***
THOUGHT (word)[5]	0.572	0.144	3.974	<0.001	***

[5] The study successfully replicated the effect of the context of intrusive /ɹ/ on the hearing of this feature found by Hay, Drager, & Gibson (2018). Listeners were significantly more likely to hear intrusive /ɹ/ after THOUGHT than after BATH/PALM, and more likely to hear it at morpheme than at word boundaries. It was found that this effect operated irrespective of the voice in which intrusive /ɹ/ was heard. As there was no such interaction with the other variables of interest to this study, this phenomenon was not explored further. The finding does, however, further

THOUGHT (morpheme)	0.831	0.146	5.703	<0.001	***
Trial order	-0.039	0.055	-0.701	0.483	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

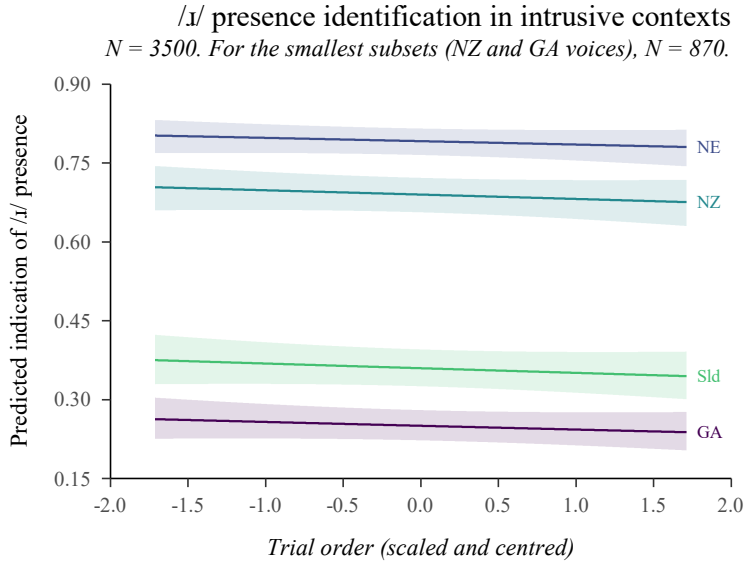

Figure 9: Model 2a predictions for hearing of /ɹ/ over time in intrusive contexts, by voice, showing no significant interactions with trial order.

5.2.2. Model 2b (modified model for intrusive /ɹ/)

The optimal model for intrusive /ɹ/, including voice order (Model 2b), took the form

Model 2b: intrusive /ɹ/ (simplified, and modified to include counterbalancing group)

$$\text{keyPressed} \sim (\text{set} * \text{variety} * \text{orderTrialSC} + \text{vowelBoundary} + (1 + \text{orderTrialSC} \mid \text{participant}) + (1 \mid \text{stimUnique})$$

The addition of the order of voices as a model predictor has significant consequences for the interpretation of results. In particular, the introduction of voice order as a fixed effect leads to significantly greater complexity in the model ultimately selected, and the emergence of some perceptual learning effects.

The predictions of this model are displayed in Table 5. In addition to apparent significant interactions between the voice order, voice, and trial order, an interesting distinction that

support the notion that listeners' perception of sounds is heavily influenced by the perception of these sounds, with sensitivity to the fine detail of particular sounds in particular contexts.

emerges is that listeners were significantly more likely to hear /ɹ/ in the Eastern New England voice relative to other voices when they had not yet heard either of the New Zealand voices ('Eastern New England voice, for Set B' in Table 5).

Table 5: Summary of Model 2b, new model for perception of intrusive /ɹ/, featuring interactions of voice order, voice and trial order, with a main effect of morphophonemic context.

Fixed Effects

	Estimate	**Std. Error**	***z***	***p***	
Main effects					
(Intercept)	-1.572	0.391	-4.017	<0.001	***
Set B (American voices first)	-0.028	0.530	-0.053	0.958	
General New Zealand voice	1.828	0.204	8.955	<0.001	***
Southland voice	0.715	0.197	3.626	0.000	***
Eastern New England voice	2.056	0.205	10.022	<0.001	***
THOUGHT (word)	0.578	0.145	3.976	<0.001	***
THOUGHT (morpheme)	0.843	0.147	5.730	<0.001	***
Trial order	-0.284	0.127	-2.245	0.025	*
Two-way interactions					
General New Zealand voice, for Set B	0.196	0.245	0.799	0.425	
Southland voice, for Set B	-0.451	0.241	-1.868	0.062	
Eastern New England voice, for Set B	0.862	0.257	3.355	<0.001	***
General New Zealand voice by trial	0.303	0.164	1.851	0.064	
Southland voice by trial	0.128	0.162	0.785	0.432	
Eastern New England voice by trial	0.302	0.164	1.842	0.065	
Trial order, for Set B	0.719	0.193	3.723	<0.001	***
Three-way interactions					
Set B: General NZ voice by trial	-0.698	0.244	-2.862	0.004	**
Set B: Southland voice by trial	-0.786	0.248	-3.168	0.002	**
Set B: ENE voice by trial	-0.835	0.255	-3.281	0.001	**

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The nature of the apparent perceptual learning effects is displayed in Figure 10. There appears to be a strong trend towards increased hearing of /ɹ/ in the General American voice, though from near-floor levels, when this is the first voice being heard, contrary to what was anticipated

(hypothesis 1b). In the other voice order, there is a somewhat weaker trend towards *decreased* hearing of /ɹ/ in this voice. A similar trend is found with stimuli from the Southland voice.

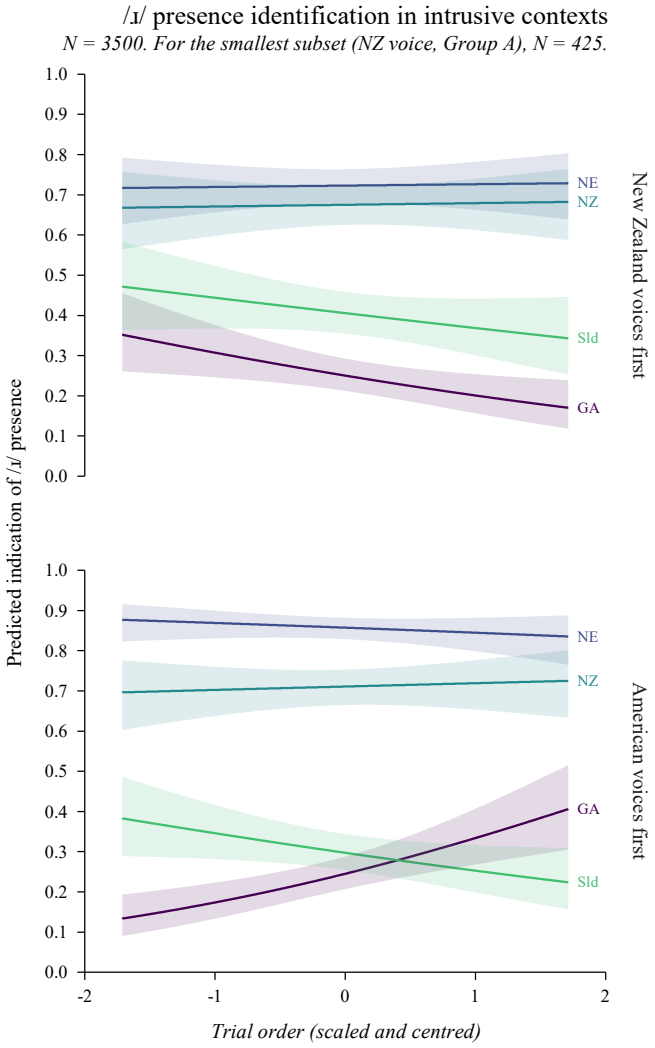

Figure 10: Model 2b predictions for hearing of /ɹ/ in intrusive contexts, by voice order and voice.

Listeners in Group B were less likely to hear intrusive /ɹ/ in the Southland than in the General American voice by the final trial, though primarily due to their increased hearing of intrusive /ɹ/ in the General American voice. It is also apparent that although listeners were always more likely to hear intrusive /ɹ/ in the Eastern New England than in the General New Zealand voice, this effect is only significant when they had not yet heard the latter voice.

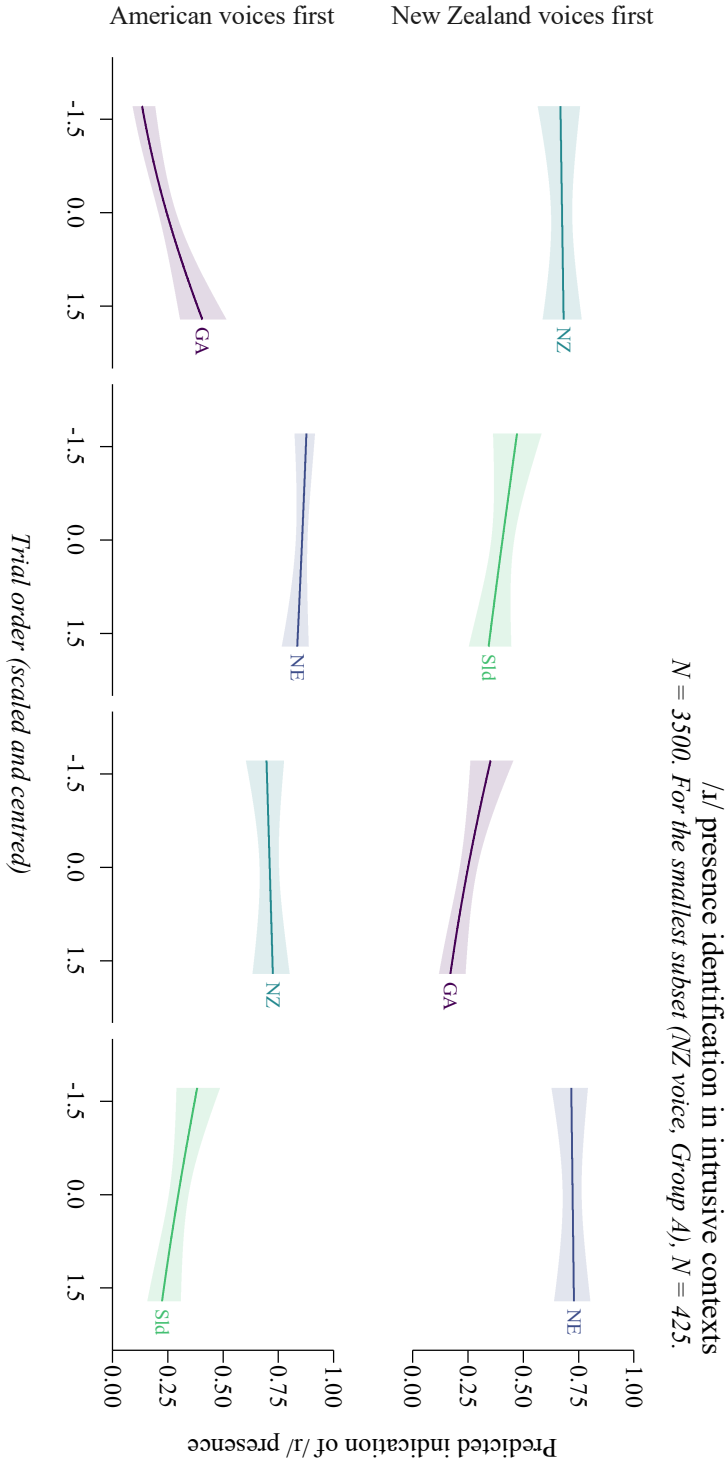

Figure 11: Model 2b predictions for hearing of /ɹ/ over time in intrusive contexts, showing the trajectory of responses over the course of the whole experiment for each voice order.

Also of import is the overall trajectory of hearing of intrusive /ɹ/ over the course of the experiment, as displayed in Figure 11, following the layout of Figures 6-8. The decreasing hearing of /ɹ/ in the General American voice among listeners in Group A, seen in this light, is merely a continuation of a trend beginning when listeners first adjusted to the Southland voice. The effect of voice order in the hearing of intrusive /ɹ/ in the General New Zealand and Eastern New England voices follows from listeners' tendency to hear less intrusive /ɹ/ in the former relative to the latter voice, rather than a tendency to hear lower intrusive /ɹ/ in the former voice overall.

5.3. Summary

Listener behaviour in the perception task as predicted by the voice in which they heard a phrase and the time at which they heard it displays both highly consistent perceptual learning trends and other phenomena less conducive to explanation. In all analyses, there emerges a general trend distinguishing participants' behaviour towards stimuli in the Southland voice from their responses to stimuli in other voices, whilst their treatment of the remaining three voices is much more variable and complex. Some of this complexity is explicable through effects of the order in which participants heard each voice, and the inclusion of the counterbalancing order variable in Models 3 and 4 helps to give an indication of the significance of these effects. Additional variability in the effects of voices and the context in which /ɹ/ does or does not occur appears to be the result of variability in the stimuli for which the experiment did not control. The results also reveal considerable variation across listeners in baseline performance in the perception task. There persists a level of complexity in individuals' perceptual learning behaviour for which the models presented here cannot account.

In accordance with hypothesis 1a, there is a consistent, positive trend in Group A listeners' identification of non-prevocalic /ɹ/ presence in stimuli in the Southland voice, independent of vowel context. Providing some support for hypothesis 1b, listeners were also less likely to hear intrusive /ɹ/ in the Southland voice over time, although they were from the beginning very responsive to the absence of intrusive /ɹ/ in this voice, contrary to the salience hypothesis 3. Baseline responsiveness to the actual content of stimuli is revealed in listeners' immediate hearing of greater rhoticity in Southland, as well as Eastern New England NURSE, relative to other vowel contexts, in accordance with hypothesis 4.

Surprisingly, there appears to be a consistent trend towards decreasing identification of non-prevocalic /ɹ/ in stimuli in the General American voice, though from near-ceiling levels,

and among Group B listeners a comparable trend towards increased identification of intrusive /ɹ/, though from near-floor levels. Against hypotheses 1c and 1d, no perceptual learning trends were found with the Eastern New England voice, either for non-prevocalic or for intrusive /ɹ/.

In addition to trial order effects, there are some noteworthy effects of the context in which /ɹ/ occurs that operate independently. Participants generally treated the Eastern New England voice as less rhotic than the General New Zealand voice, though this may largely be a consequence of their hearing of START in this voice as particularly non-rhotic; they also heard more intrusive /ɹ/ in this voice than in the General New Zealand voice, however. More clearly, listeners somewhat over-predicted rhoticity in the Southland versus the General American voice, though it is less clear whether they heard less intrusive /ɹ/ in the Southland than in the General American voice. Taken all in all, these results offer some evidence in support of hypotheses 2a and 2b.

There are also apparent, unanticipated, effects of listener experience. When the Southland and Eastern New England voices were heard last, perceptual learning trends were reduced, and listeners consistently rated NURSE stimuli as more highly rhotic.

Overall, these findings provide mixed evidence in support of the hypotheses originally laid out (§3.2). In particular, while those regarding perception of the Southland voice largely held up, those relating to the Eastern New England voice did not pattern as anticipated. This may not be a consequence, however, of incorrect assumptions regarding the role of expectations versus experiences in the perception of stimuli, but rather of incorrect assumptions regarding these particular listeners' expectations of the voices selected. Examining this issue further, results from the questionnaire are provided in §5.4.

5.4. Questionnaire results

Of relevance are listeners' responses regarding how often they hear each variety, which gives some information on their experiences, and their guesses regarding which voices they heard, which are informative of their expectations during the experiment. Given a highest possible score of 20 and a lowest possible score of 4, listeners' mean familiarity with the General New Zealand voice was 19.025, their mean familiarity with the Southland voice was 8.912, their mean familiarity with the General American voice was 11.545, and their mean familiarity with the Eastern New England voice was 5.4. This follows what was anticipated: listeners had little experience with the regional varieties, and considerable experience with the General New Zealand and General American varieties, although arguably their experience with the latter

variety was somewhat lower than anticipated, being more comparable to their experience with the Southland voice than to their experience with the General New Zealand voice.

Of the one-hundred accent identifications received, the General New Zealand, General American, and Eastern New England voices had twenty-seven each, and just nineteen were obtained for the Southland voice. Equal, high numbers of listeners provided 'General New Zealand' and 'General American' answers, twenty-four of twenty-seven responses in both cases. By contrast, just five listeners provided a 'Southland' response for the Southland voice; eleven provided a 'General New Zealand' answer, and three provided another response ('Southern British' or 'Australian'). This is as anticipated: listeners had strong, accurate expectations regarding the General New Zealand and General American voices, and more weakly expected a General New Zealand accent when they heard the Southland voice.

Identification of the Eastern New England voice did not go as expected, however. While indeed only two listeners provided an 'Eastern New England'-related response, only four offered a 'General American' answer. Ten identified the voice with the Southern United States, four with New York City, and seven offered other answers (these included instances of 'Irish-American', 'Scottish', 'Canadian', 'Southern British' and 'Chicago'). It is evident that while almost no listeners believed they were hearing an Eastern New England voice, most had expectations of a non-standard American accent, rather than a General American voice. All of these responses were received after the experiment, of course, and are therefore not directly indicative of what listeners initially heard. They do, however, provide strong evidence of listeners' higher-level thoughts from the experiment, and as such are relevant to the interpretation of the perception experiment results.

6. Discussion

At its inception, this study was intended to explore the relationship between individuals' experiences of language, their expectations regarding language, and the language they hear. Specifically, it was hypothesised firstly, that listeners' varying expectations regarding the occurrence of linguistic features in different voices would influence their perception of these features, and secondly, that varying expectations would influence their ability to adjust their perceptions under new exposure. Where they had stronger expectations of a variety, they would show less of a tendency to adjust.

In the context of the perception of non-prevocalic and intrusive /ɹ/ in General New Zealand, Southland, General American, and Eastern New England voices, it was anticipated that participants would initially hear no non-prevocalic /ɹ/ in the Southland voice, expecting a typical New Zealand voice, but that over time they would increasingly hear non-prevocalic /ɹ/, whereas their behaviour towards the General New Zealand voice would change little. Conversely, it was thought that they would initially hear non-prevocalic /ɹ/ in the Eastern New England voice, expecting a typical American voice, and that over time they would adjust to its absence, where there would be no observable change in their behaviour towards the General American voice. Furthermore, it was hypothesised that behaviour towards intrusive /ɹ/ would follow a similar pattern, but that these changes would be less pronounced, due to the lower salience of the feature. Lastly, it was anticipated that rates of /ɹ/ identification in NURSE contexts in the Southland and Eastern New England voices would be higher overall due to listener responsiveness to the actual content of the stimuli.

The extent to which the experimental results presented in §5 support these hypotheses is discussed firstly, before their broader implications for the research questions regarding expectations and experience in speech perception are considered.

6.1. Hypothesis 1: effects of listener expectations on perception

In this study, it was anticipated that listeners would at the outset of the experiment believe the regional varieties they heard (the Southland and Eastern New England varieties) to be examples of more familiar varieties (General New Zealand and General American, respectively) and expect distributions of non-prevocalic and intrusive /ɹ/ characteristic of these latter varieties. It was also anticipated, however, that their expectations regarding these voices would be relatively weak, and change with the effects of new experience over the course of the experiment. As Floccia et al. (2006) demonstrated, although listeners are initially challenged in processing unfamiliar varieties of speech, over time they learn their characteristics. As such, the effects of listeners' original expectations would be overcome through new experiences. This would be consistent with models such as those proposed by Clark (2013) and Lupyan (2015), in which perception develops through the interaction of the bottom-up experience of stimuli and top-down predictions, where the former 'correct' when the latter impede perception in some way.

Specifically, it was hypothesised that listeners would initially expect low levels of non-prevocalic /ɹ/ (hypothesis 1a) and high levels of intrusive /ɹ/ (hypothesis 1b) in the General

New Zealand and Southland voices, whilst they would expect high levels of non-prevocalic /ɹ/ (hypothesis 1c) and low levels of intrusive /ɹ/ (hypothesis 1d) in the General American and Eastern New England voices. Over time, however, they would increasingly hear non-prevocalic /ɹ/ (hypothesis 1a) in the Southland voice, and cease to hear intrusive /ɹ/ (hypothesis 1b), whilst their perception of these features in the General New Zealand voice would not change. Conversely, they would increasingly cease to hear non-prevocalic /ɹ/ (hypothesis 1c) and increasingly hear intrusive /ɹ/ (hypothesis 1d) in the Eastern New England voice, whilst their perception of the General American voice would not change.

While the predictions regarding behaviour towards the General New Zealand and Southland voices are largely borne out in the data, the same is not true for the Eastern New England and General American voices. Listeners did indeed initially expect low levels of non-prevocalic /ɹ/ (§5.1.2), and to some extent higher levels of intrusive /ɹ/ not only in the General New Zealand but also in the Southland voice (§5.2.2). Furthermore, while their perception of the General New Zealand voice remained quite constant, they increasingly ceased to hear intrusive /ɹ/, and increasingly heard non-prevocalic /ɹ/ in the Southland voice, although only when they had not heard the American voices. Importantly, this occurred independently of vowel context. Although there was not a single instance of /ɹ/ in START and NORTH/FORCE in this voice, listeners increasingly heard it in these contexts under the influence of their growing experience of Southland NURSE rhoticity, suggesting some sensitivity to a broader 'rhoticity' category as well as to less-generalised levels of categorisation.

There were no such perceptual learning shifts with the Eastern New England voice however, whilst there was a regular perceptual shift towards the General American voice, although always from near-floor or near-ceiling levels. From the beginning, listeners expected little non-prevocalic /ɹ/ (even in NURSE when they had yet to hear the New Zealand voices) in the Eastern New England voice (§5.1), and considerable intrusive /ɹ/ (§5.2), contradicting hypotheses 1c and 1d, even as their expectations regarding the General American voice did pan out as anticipated. There were no significant interactions between the trial order variable and the Eastern New England voice type when the order of voices was accounted for.

Whilst hypotheses 1c and 1d are clearly not supported by the results, when viewed in the light of participants' responses in the questionnaire, these findings potentially offer further insights regarding the role of expectations in perception. The questionnaire responses diverged from what was anticipated in two ways. Firstly, although listeners clearly had had the most

experience with the General New Zealand and General American voices (§5.4), their experience with the latter voice was not as considerable as believed, which would render perception of it more malleable under the influence of new experiences according to the theoretical assumptions of this study, something indeed borne out in the experiment.

Secondly, and most importantly, although both the Southland and the Eastern New England voices were rarely accurately identified, there were important differences in listeners' expectations of these voices. Fewer participants had ideas regarding the former voice than had ideas regarding the latter, and of fourteen participants who incorrectly identified the Southland voice, eleven indicated that they believed it was a General New Zealand voice, as anticipated (§5.4). Of the twenty-five listeners who incorrectly identified the Eastern New England voice, just four believed they had heard a General American voice, however. The largest number thought they had heard a Southern United States accent; a smaller group identified the speaker as from New York City.

While these are both regions that have historically had non-rhoticity (Labov, 1966; Labov, Ash, & Boberg, 2006), it is doubtful that listeners made this association, given the extent to which the feature has declined in these locations (Feagin, 1990; Becker, 2009; 2014; Mather, 2012). More likely, they identified the voice with the United States, but found it to be sufficiently unusual that they tended to associate it with the non-standard American accent best known to them, that of the American South. Anecdotal information from participants at the conclusion of the experiment suggests that many reacted very strongly and very negatively to this voice based on such beliefs. This fits with the well-described tendency of individuals to have varying degrees of detail in 'folk-linguistic' awareness, where the most salient examples of 'non-standard' speech (notably, Southern United States English) are those which are most readily identified (Preston, 1996). It is therefore not the case that listeners initially believed the Eastern New England voice to be a General American voice, and as such, there were no perceptual learning effects. While their expectations regarding this voice were relatively weak, listeners were cued to its being a different kind of accent.

6.2. Hypothesis 2: malleability of perception through experience

It was predicted in this study that listeners' perception of the Southland and Eastern New England voices would be more malleable than their perception of the other voices under the influence of new experiences (in the context of the experiment itself), following from the relatively weak expectations of these voices originally hypothesised and largely confirmed in

the results, as discussed in §5.1. The implication would be that listeners' expectations regarding speakers not only guide perception, as has been attested (McGowan, 2015; Lev-Ari, 2015; Gnevsheva, 2018), but also interact with learning effects coming from the bottom-up perception of stimuli that have also already been extensively studied (e.g. Lively, Logan, & Pisoni, 1993; Pisoni, 1993; Clopper & Pisoni, 2004a; Kraljic & Samuel, 2006, 2011).

The perceptual learning trends discussed in §5.1 do not provide any complete evidence either in support of or against this hypothesis. It could be that listeners' perception of the General New Zealand and General American varieties is not very malleable simply because it already very closely approximates the actual content of the stimuli (although note that this is not true in the case of the General New Zealand voice, where rhoticity is significantly over-expected). Based on this, it could be that perceptual learning based on experience is simply a process of correcting for expectations that lead perception astray. The possibility that listeners' expectations could to some extent limit perceptual learning would however be evidenced in a tendency to over-predict newly-learnt features in the regional varieties relative to the other varieties. That is, while listeners would continue to hear a feature to a particular, not entirely accurate, degree in the General New Zealand and General American voices, in accordance with their expectations, they would come to expect the presence of a feature newly-learnt as being present, and to expect the absence of a feature newly-learnt as being absent, to a greater degree than they would expect these features where learning was limited by existing expectations.

Specifically, it was predicted that listeners would hear more non-prevocalic /ɹ/ in Southland NURSE contexts than in General American NURSE contexts by the final trial of each block (hypothesis 2a), and that listeners would hear less non-prevocalic /ɹ/ in the Eastern New England voice than in the General New Zealand voice, save for in NURSE contexts, by the final trial (hypothesis 2b). No specific hypotheses were made regarding the hearing of intrusive /ɹ/, because the latter feature was assumed to be less salient than non-prevocalic /ɹ/ (hypothesis 3).

There is some evidence in support of these hypotheses, but there are also many complexities. On the whole, in hearing the less familiar voices, listeners' responses more closely mirrored the actual content of the stimuli. In fact, this was true of intrusive /ɹ/ as well as non-prevocalic /ɹ/, contrary to what was anticipated but in line with the broader thesis.

Listeners tended to hear less non-prevocalic /ɹ/ in an Eastern New England than in a General New Zealand voice, save for with NURSE contexts, and secondly, they consistently heard more intrusive /ɹ/, although both voices had precisely the same distribution of the latter feature. This

was particularly true when they had not yet heard either of the New Zealand voices (§5.2.2.). Although the tendency to under-predict rhoticity in the Eastern New England relative to the General New Zealand voice could be argued to be a consequence of listeners' particular sensitivity to non-rhotic START in the former voice, an issue discussed in §5.3, this finding regarding intrusive /ɹ/ cannot be dismissed in the same way.

Group A listeners did over-predict rhoticity in the Southland voice relative to the General American voice by the final trial (§5.1.2, Figure 5), and not only in NURSE but overall. This is in fact a stronger effect than was expected, although it is partly a consequence of the changing perception of the General American voice. The same is even truer of the tendency of Group B listeners to hear less intrusive /ɹ/ in the Southland than in the General American voice by the final trial.

The results relating to the second hypothesis are therefore mixed. While listeners tended overall to hear more non-prevocalic /ɹ/ and less intrusive /ɹ/ in the Southland voice than in the General American voice, and less non-prevocalic /ɹ/ and more intrusive /ɹ/ in the Eastern New England voice than in the General New Zealand voice by the final trial of each block, these patterns may be artefacts of the particular experiment setup, given the complexities that appear when specific vowel contexts and voice orders are considered.

Another possible argument is that whatever effects of expectations were found merely follow from beliefs regarding orthography: listeners from a New Zealand language background will have a stronger belief that people who sound like them are more likely to speak in a way that mirrors the spelling of words, and therefore persist in hearing non-prevocalic /ɹ/ in this voice even as they hear less in the Eastern New England voice. This cannot however explain their tendency towards hearing greater non-prevocalic and less intrusive /ɹ/ over time in the Southland relative to the General American voice. While there is no firm evidence that listeners' stronger expectations of the General New Zealand and General American voices did indeed limit the malleability of their perception of these voices, these results do raise this possibility, and it is an issue that ought to be explored further in a modified experimental setup, an issue addressed in §7.2.

6.3. Hypotheses 3 and 4: responsiveness to stimuli

It was assumed that the effects of expectations and experiences on listener behaviour would to some extent be limited by responsiveness to the actual properties of stimuli. While it was hypothesised that listeners would adjust more to non-prevocalic than to intrusive /ɹ/ (hypothesis

3) due to the former feature's greater salience, limiting their responsiveness to stimuli, it was also thought that they would overall hear more non-prevocalic /ɹ/ in NURSE contexts in the Southland and Eastern New England voices than in other vowel contexts in these voices (hypothesis 4), a reflection of the actual content of the stimuli. No effect of salience was in fact found, but there was clear evidence in support of hypothesis 4 and indeed other evidence of listeners' responsiveness to stimuli.

While there were stronger perceptual learning effects when listeners heard non-prevocalic /ɹ/ in the Southland voice (§5.1.2) than when they heard intrusive /ɹ/ in this voice (§5.2.2) this was in fact a consequence of their greater responsiveness to intrusive than to non-prevocalic /ɹ/ tokens, rather than of the lower salience of the former feature. Listeners overall more accurately perceived intrusive /ɹ/ than non-prevocalic /ɹ/ from the first trial, likely reflecting for the most part their tendency to overestimate non-prevocalic /ɹ/, presumably from orthography-derived expectations or other, related beliefs. Because of this, listeners had to adjust more to non-prevocalic than to intrusive /ɹ/. Hypothesis 3 was therefore clearly unsubstantiated.

It is apparent that listeners displayed baseline accuracy in identifying the presence or absence of /ɹ/, regardless of their experience or expectations. In confirmation of hypothesis 4, there were substantial, immediate increases in listeners' estimation of /ɹ/ presence in NURSE contexts when hearing a Southland after hearing a General New Zealand voice, and similar immediate decreases in hearing intrusive /ɹ/ in the same transition, leading to discontinuities in the perception of /ɹ/ at the changes of voices rather than the gradual adjustment to a new voice that would normally be expected (Pisoni, 1993).

Listeners were similarly responsive to the presence of non-prevocalic /ɹ/ in Eastern New England NURSE, and the absence of intrusive /ɹ/ in this voice, from the beginning. Interestingly, however, they were also highly sensitive to the absence of non-prevocalic /ɹ/ in START contexts in the same voice (§5.1.2, Figure 6) whilst showing comparatively little sensitivity to Eastern New England NORTH/FORCE (§5.1.2, Figure 7). This may be a consequence of the peculiar, more fronted quality of START in this voice, often found in Eastern New England accents, although primarily in the Boston area (e.g. Stanford, 2018).While listeners found START to be less rhotic than other vowels overall (§5.1.1, Figure 4) with the other voices, the vowel quality is such that it may be easier for listeners to internalise the vowel and rhotic consonant separately, based on their experience with low, unrounded vowels that are variably followed by rhotic consonants. With the noticeably fronted START in the Eastern New England voice,

listeners may not be able to form the same segmental generalisations, lacking experience with fronted START vowel-rhotic consonant combinations. This would be consistent with the evidence that experience with variability is necessary for the formation of such generalisations (e.g. Clopper & Pisoni, 2004a; Sumner, 2011). As such, fronted START would be perceived as an environment in which /ɹ/ could not occur. Alternatively, fronted START may simply be surprising to listeners, and cause them to believe that they are listening to some variation in /ɹ/ use given the context of a task that requires them to listen for this sound.

It is of particular concern to models that propose the existence of top-down influences on perception that they have the potential to produce hallucinatory behaviour (Norris, McQueen, & Cutler, 2016). It appears that in this experiment, listeners were quite able to be guided by their expectations regarding what they would hear and adjust over the course of the experiment whilst retaining some amount of absolute sensitivity to the actual content of the stimuli. The influences of expectations are broadly reconcilable with some essential deference on the part of listeners towards sense perception, even as the relative importance of expectations versus experience presumably varies with the context; listeners likely rely on prediction to a greater or lesser extent based on its varying utility in simplifying the effort of processing versus hindering perception through apparent inaccuracies.

There are nevertheless clear limits to individuals' abilities to respond based on hearing alone in a phoneme monitoring task; participants tended to significantly overestimate the presence of non-prevocalic /ɹ/ where it was absent, and particularly in the General New Zealand voice, without changing their behaviour over time. This is presumably an effect of orthography or some other source of beliefs regarding the sounds in words. Anecdotally, more than one listener, in discussing this issue after completing the experiment, said that they struggled to distinguish /ɹ/ from some larger unit, for example, /aɪ/, or from the word in which it occurred.

Although not the focus of the current study, this problem serves as a reminder that processing speech bottom-up through the segmentation of a signal into discrete sounds is *not* the normal path in speech perception as claimed by, for example, Norris, McQueen, & Cutler (2000, 2016), and demonstrates the value of moving beyond structural analyses of language, as proposed by Hawkins (2010). Listeners are ordinarily able to get by on minimal auditory cues, augmented by significant contextual information, and even where they are forced to focus on sounds in decontextualized phrases, they struggle to do so, even with extensive training and cuing to the relevant sounds.

6.4. Effects of experience of variation on voices on perception

Whatever the influences, and indeed the potential blocking effects of listener expectations, it is also apparent that experience, and in particular variable experience, over time can bring listeners' expectations more into line with the actual content of stimuli. The perception experiment results revealed a trend relating to the order of voices that had not been hypothesised, and the order of voices also limited the effects of expectations discussed in §5.1. Notably, the trend from hearing low to hearing high rates of non-prevocalic /ɹ/ in the Southland voice only occurred where listeners had not yet heard either of the American voices; where they had already heard these voices, their perception more closely mirrored the content of the stimuli. Responses to the Eastern New England voice were also more closely-aligned with the stimuli when this voice was heard last as opposed to second. In both cases, listeners heard much higher levels of rhotic NURSE from the first trial of the block when they already had experience with three other voices, including voices where NURSE was rhotic and voices where it was not.

It has already been extensively shown (Clopper & Pisoni, 2004a; Kraljic & Samuel, 2006; Dahan & Mead, 2010; Sumner, 2011) that greater experience with variability leads to greater generalisation, or, in the view of many, produces abstractions. The present findings suggest that greater experience with variability also facilitates listeners' ability to distinguish between different variants, and potentially their ability to 'hear' segments within a signal more acutely, rather than depending too heavily on expectations of a particular voice. Over the course of the experiment, listeners appear to have become increasingly sensitive to /ɹ/ in different vowel contexts, disassociated from the voices in which it was being heard, such that by the end of the experiment expectation-driven processing had been significantly reduced.

6.5. Instability in perception of the General American voice

Contrary to what was predicted, perception of non-prevocalic and intrusive /ɹ/ in the General American voice was not entirely stable. Regardless of vowel context and voice order, hearing of non-prevocalic /ɹ/ declined over time in the General American voice; hearing of intrusive /ɹ/ increased over time when the General American voice was heard first, and decreased over time when heard after the New Zealand voices. This behaviour is not readily explicable in terms of learning under new exposure, because perception is initially very accurate and becomes increasingly less so.

One possibility is that this is a response strategy, following from the fact that listeners are hearing non-prevocalic /ɹ/ at such high levels, and intrusive /ɹ/ at such low levels that over time

they increasingly guess the other way, expecting a more regular distribution. This account is difficult to sustain, however, not only because the pattern is not universal (there is an exception in listeners' hearing of intrusive /ɹ/ in the General American after having heard the New Zealand voices) but also because it is unlikely listeners would have had such a specific response strategy for these kinds of stimuli. They heard instances of non-prevocalic and intrusive /ɹ/ in the General American voice amidst thirty filler items, half of which included and half of which did not include instances of /ɹ/. For these effects to be a response strategy, listeners would have to have an acute awareness of non-prevocalic and intrusive /ɹ/ as phenomena separate from other instances of /ɹ/, in order to isolate out these particular kinds of stimuli from the rest.

Alternatively, it could be that over time listeners are increasingly ceasing to hear the distribution of /ɹ/ sounds characteristic of this voice, instead falling back on word-level knowledge reflecting their own use or that of the language with which they are most familiar with around them, a General New Zealand variety. Such an account would also require an explanation of why there are no comparable effects with other voices (the Southland and Eastern New England voices), however. While it does not undermine the main conclusions of this study, this issue remains difficult to resolve.

6.6. Implications for usage-based theories of language

While many of these issues require much deeper investigation, the experimental results presented here suggest that linguistic knowledge emerges from the interaction of established expectations with new experiences of linguistic signals (§6.1), in which the strength of expectations has the potential to limit learning effects from new experiences (§6.2), but where perception continues to always be responsive to the actual content of stimuli (§6.3-6.4).

This carries a significant implication for usage-based models of language. The debate over the episodic versus abstract representation of language has now largely been resolved through the postulation of hybrid models featuring levels of both 'episodic' and 'abstract' information (e.g. Goldinger, 2007; Pierrehumbert, 2016), but what has been termed 'abstract' in this sense should more properly be termed 'generalised': it is the process by which larger categories are built up around categories that have already formed, through learning across variability in experience, as described in Clopper & Pisoni (2004a), Dahan & Mead (2010), Sumner (2011), and the present study (§5.4). In usage-based 'exemplar' models of language (e.g. Johnson, 1997; Pierrehumbert, 2001), perception is typically represented as a straightforward process in which stimuli are compared with and added to the distribution of existing 'exemplars', even as

what is perceived may be limited by past experiences (Hay, Drager, & Gibson, 2018). In other models (e.g. Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2016), perception is more restricted still, being conceived as an entirely bottom-up process of signal decoding.

Insofar as perception is influenced by complex expectations derived from listeners' beliefs regarding a speaker, as implied by findings such as those of Gnevsheva (2018), Fiedler, Keller, & Hanulíková (2019), and the present work, all linguistic knowledge is in fact abstracted away from what is actually perceived, and there is no objective translation of stimuli into 'exemplars' in linguistic knowledge. While all linguistic knowledge is derived from experience, this experience is always being modulated due to predictive behaviour. Similar phenomena are found with regard to social meanings in linguistic knowledge: individuals' interpretation of potential speech cues for particular social categories is modulated by their own beliefs regarding these categories (e.g. Campbell-Kibler, 2009, 2010; Levon, 2014), and social judgements become similarly established in real-time exposure based on minimal, salient cues (Watson & Clark, 2015). In the future, work in usage-based models of language will increasingly have to consider how these forms of modulation occur. It is probable that newly-heard stimuli are unable to attain anything like the status of pre-existing knowledge, because they are forever being filtered through the lens of existing expectations. At the extreme, this would even lead to a 'blocking' effect in which perceptual learning is prevented because of the strength of expectations. Many of these issues are currently being studied in the broader cognitive science literature (e.g. Clark, 2013; Lupyan, 2015), and usage-based linguistics generally would benefit from engagement with it.

7. Conclusion

With the development of usage-based models of language (e.g. Johnson, 1997; Pierrehumbert, 2001, 2006, 2016; Todd, Hay, & Pierrehumbert, 2019) and the now well-supported notion that variable language use proceeds from individuals' variable experiences of language (e.g. Evans & Iverson, 2004, 2009; Adank et al., 2009; Lev-Ari & Peperkamp, 2016; Lev-Ari, 2017, 2018), it is increasingly important to consider the extent to which individuals' perception of language reflects similar influences. The present study, supporting earlier findings that suggested that listener behaviour in perceptual task was influenced by their expectations regarding a speaker or context (Niedzielski, 1999; Strand, 1999; McGowan, 2015; Gnevsheva, 2018), demonstrates how listeners' beliefs regarding a voice they are hearing influences their processing of speech, even in the absence of any explicit cue intended to generate specific expectations. Even so, it

also illustrates how this behaviour may change under the influence of new experiences, and establishes clear limits to experience- and expectation-driven processing of language. Importantly, it provides some early indications of how expectations may limit perceptual learning, an area that needs considerable further attention. Based on these findings, it may be worthwhile in usage-based theoretical linguistics to be cognisant of emerging ideas in other fields regarding predictive behaviour, and to reconsider the framing of 'episodic versus abstract levels' of information.

Taken altogether, these results do strongly indicate that listeners' expectations, based on whatever cues they receive, do indeed influence what they hear. Importantly, in this experiment, as opposed to many earlier studies on predictive behaviour (Rubin, 1992; Niedzielski, 1999; McGowan, 2015; Gnevsheva, 2018) nothing was done to actively promote particular expectations in participants before their exposure to the stimuli, and their responses came entirely in the form of instantaneous decisions regarding /ɹ/ presence rather than judgements after the fact. Much of the criticism levelled at the notion of top-down effects, whether in speech perception specifically (Cutler & Norris, 2016) or in cognitive science more generally (Firestone & Scholl, 2016) has been directed at experimental methodologies that have measured the latter kinds of responses.

In the present study, listeners had no time to reflect on their thinking, having to make snap decisions with each new stimulus. They were compelled to make decisions based on limited auditory cues as to the kind of voice they were hearing and the likelihood of /ɹ/ occurring in different contexts within it. Hearing a Southland voice with features such as particular vowel qualities that were familiar to them as those of a General New Zealand English speaker, listeners initially expected similar levels of non-prevocalic and intrusive /ɹ/ to that which they expected in a General New Zealand English speaker, but over time learned to perceive the voice as rhotic, to the point of significantly over-predicting non-prevocalic /ɹ/ (§6.1).

Such a phenomenon may still be amenable to interpretation in a model based on prediction through entirely bottom-up processes (e.g. Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2016), though it does not fit so easily with the teleological claims that have been made on behalf of such models, that 'the primary function of perception is to construct the best possible model of the world' (Norris, McQueen, & Cutler, 2016). There is nothing in this behaviour to suggest perception very closely mirrors what is actually heard in the world; listeners veer between a non-rhotic and a rhotic model of the stimuli, neither of which is totally

accurate. The precise flow of information remains unclear, but it does appear that listeners' processing of speech is heavily guided by their beliefs about what they *ought* to hear.

This seems compatible with a cognitive model such as that proposed by Clark (2013). In such a model, the brain pre-emptively forms hypotheses of what will be perceived based on pre-existing knowledge (expectations), which is tested against evidence from sense perception (the stimulus). Where these two forms of information appear poorly-aligned, 'error signals' lead to the adjustment of knowledge (the effect of new experiences) which in turn can influence future expectations. In such a model, this interaction is further modulated by 'precision-weighting' (Clark, 2016), in which the corrective influence of 'error signals' varies with listeners' beliefs regarding the usefulness or reliability of the stimuli.

It is conceivable that strong expectations limit the extent to which linguistic knowledge is influenced by new experiences through such a channel. The present study offers some tentative evidence that listeners' expectations regarding the distribution of non-prevocalic and intrusive /ɹ/ in different voices influenced not only their perception of these features, but also the malleability of their perception. On the whole, where listeners had stronger expectations of particular voices, they displayed less of a tendency to 'error correct' (§6.2). While much more evidence is needed to confirm this, it is an interesting possibility and one that would be explicable through a tendency of listeners to give less weight to new information when expectations are very strong. It would therefore be likely that as listeners' expectations grow increasingly strong, the stimulus must more sharply contradict their predictions in order to be the source of new experiences forming part of their linguistic knowledge. Lev-Ari (2017, 2018) argued that individuals with larger social networks and greater experience with variable language use would have less malleable perception because they would place less weight on newly-perceived stimuli; here, it is suggested that listeners' perception may be similarly constrained by the strength of their beliefs regarding the speakers to which they are listening.

Clearly, however, individuals are able to maintain some essential sensitivity to the content of stimuli. While listeners initial underestimated the rhoticity of Southland NURSE, they still immediately heard greater rhoticity in such stimuli than in the General New Zealand NURSE stimuli, for example. As such, they are not prone to 'hallucination', which according to Norris, McQueen, & Cutler (2016) follows from models including expectation-driven influences.

Given these competing sources of information, it is plausible that there is no one path in speech perception that all listeners generally follow. Rather, they may rely variably on what

they believe they should hear or on what is contained in the stimulus they are hearing, just as they variably do or do not absorb what they hear as new experiences in their linguistic knowledge. This fits with a view of speech perception where listeners behave opportunistically (van Berkum, 2008), relying on as little information as possible to develop an interpretation that is useful to them. It is also consistent with models in which linguistic knowledge does not feature consistently-defined, formal units (Hawkins, 2010). It is a view of linguistic knowledge as deeply subjective: whilst all knowledge of language begins with the experience of language in the external world, the knowledge of language increasingly comes to influence how language in the external world is interpreted.

7.1. Problems and limitations

Some outcomes of the experiment raised issues that do not readily fit into the theoretical framework followed here, particularly in relation to the findings of Hay, Drager, & Gibson (2018), and are discussed in §7.1.1. The extent to which this study is informative of the role of listeners' expectations and experiences of stimuli in speech perception is further limited by factors relating to the format of the experiment conducted, something addressed in §7.1.2.

7.1.1. Issues of interpretation

As discussed in §6.5, a major unexpected result occurred with listeners' perception of the General American voice. Rather than perceiving constant levels of non-prevocalic and intrusive /ɹ/ over the course of the experiment, listeners tended to increasingly hear less of the former, and more of the latter, at least under some conditions. This does not refute the view that where listeners have strong expectations, the influence of new experiences is reduced, because here the perceptual shift veers away from, rather than towards, the actual content of the stimuli: it is not, in fact, a perceptual learning trend.

Possibilities that this might be a response strategy, or alternatively a shift in listeners' attention to sound versus word-level perception, were considered in §6.5, and it was suggested that these were unsatisfactory, as they cannot account, in the former case, for why there was no such response strategy with other classes of stimuli, and in the latter case, for why the perception of other voices was not similarly affected. It is difficult to say what the basis for this trend is, which is all the more remarkable given its consistency across conditions. It can, however, be said that it clearly does *not* show processing of language through the regular segmentation and analysis of a speech signal, directed at improving accuracy in perception over time.

Also problematic were the effects of listeners' beliefs regarding the occurrence of non-prevocalic /ɹ/ on their responses. Although much was done to make listeners focus on sounds, rather than orthography (§4.4.1), most listeners found it nonetheless difficult not to respond based on beliefs about the occurrence of /ɹ/ in non-prevocalic contexts where the sound is in fact absent. Although this phenomenon offers an interesting insight regarding individuals' lack of any acute knowledge of language in the form of segmental phonology, it renders it more difficult to make judgements regarding what listeners were in fact perceiving in the non-prevocalic /ɹ/ stimuli in this experiment. It should be noted, however, that listeners did not tend to have the reverse problem, that of failing to hear intrusive /ɹ/ due to the absence of orthographic cues (§5.2); this defies ready explanation.

Another issue relates to the stimuli used. As has been described (§4.2), although much work was put into trying to find speakers whose actual use of non-prevocalic and intrusive /ɹ/ was comparable to that which was required in the experiment stimuli, and although attempts were made to achieve uniformity across stimuli through the use of an audio model, inevitably, problems arose. In particular, intrusive /ɹ/ in the Eastern New England stimuli had to be entirely forced, and there were also problems in matching the speaking rates in these stimuli to those of the other speakers. It is possible that listeners responded to peculiarities of these stimuli as a consequence, for example in being particularly sensitive to the Eastern New England intrusive /ɹ/ stimuli from the outset (§5.2). Although varieties with canonically identical alveolar approximant consonants were selected, it is also possible that listeners were attuned to subtler variations in speakers' articulation of these sounds.

A separate problem concerns the interpretation of the results presented by Hay, Drager, & Gibson (2018) in the light of the arguments presented here (§6.1). It was argued that the experimental results closely followed the hypotheses for listeners' perception of the Southland voice, but not for their perception of the Eastern New England voice, but that this latter outcome reflected problems with the assumptions that were made regarding listeners' expectations of this voice. Notably, while listeners did tend to expect a General New Zealand accent upon hearing the Southland voice, they rarely expected a General American voice upon hearing the Eastern New England voice, even though their expectations were weak in both cases. As such, a perceptual learning trend was only observed with the Southland voice, because listeners were compelled to adjust to unexpected information, a case of responsiveness to 'error signals' in the Clark (2013) model.

Hay, Drager, & Gibson (2018) found a perceptual learning trend when listeners from San Diego heard a New Zealand voice, a context that would more closely mirror the Eastern New England condition of the present experiment than the Southland condition. Presumably, listeners from San Diego would have little idea of what to expect with a New Zealand voice, though perhaps associating it with Australia, as in Bayard (2001), and would not need to 'correct' in the same way. Lacking any information on how listeners interpreted the voice in the experiment conducted by Hay, Drager, & Gibson (2018), it is difficult to explain these differences. It therefore remains an interesting question exactly which conditions produce perceptual learning and which do not, and why.

7.1.2. Methodological limitations

While the experiment conducted in this study offers insights into the effects of listeners' expectations and experiences on their perception of speech, various possible interpretations of the data have not been exhaustively tested. The method followed in this study has potentially led to some issues of interest being obscured, whilst others have not been examined at all.

Although the use of a within-subjects experimental design ensured that the same individuals' expectations of different voices could be tested, it rendered it much harder to isolate the effects of each voice on listeners' responses. In particular, although the latter blocks demonstrate interesting effects of exposure to variability that would not otherwise have been found (§6.4), it is more difficult to interpret what listeners were basing their predictions on by this point.

Another inevitable limitation concerns the level of analysis: as has been stressed (§2.1), the object of investigation is individual linguistic knowledge, the only form of 'language' which has a meaningful reality (Docherty & Foulkes, 2014), but conclusions have been made based on data aggregated over a sample of individuals. Although the mixed-effects modelling method employed to some extent accounts for variability between individual listeners, such as individual variation in learning competency (Siegelman & Frost, 2015), it would still be worthwhile to examine how individuals behaved over the course of the experiment, how their behaviour varied, and how variation might be accounted for in terms of their individual expectations and experiences of the stimuli. The conditions off this experiment, featuring relatively few listeners (thirty in the final sample) and few trials per voice (thirty non-prevocalic and thirty intrusive /ɹ/ stimuli per voice), are unfortunately prohibitive of any such fine-grained analysis.

A possible line of reasoning against the suggestions presented in this study is that phoneme monitoring responses still constitute post-perceptual judgements, rather than being indicative of actual perception, the criticism that has been levelled against much of the evidence put forward for strong arguments in favour of 'top-down' influences on perception (Firestone & Scholl, 2016), arguments that are not specifically being made here. At this point, it must be asked what exactly would in fact be evidence of perception. It would, nevertheless, be useful to look at subtler measures of processing such as reaction times to explore possible impacts on listeners' perception, isolated from the formation of judgements after the fact.

Similarly, there are limits to what can be said about listeners' expectations regarding each voice. While the questionnaire responses have been taken as evidence of such expectations, these were of course given only after listeners had extensive exposure to each voice, such that it is impossible to know what precisely their beliefs about these voices were at the beginning of the experiment. This setup was necessary, of course, so that listeners were not primed for each voice by exposure to stimulus examples, but it does mean that what is suggested here regarding listeners' expectations (§6.1) must carry the proviso that this does not reflect evidence of their original beliefs.

There are also limitations to the claims that can be made regarding one of the original research questions, that regarding the effects of expectations on the malleability of perception. While evidence was presented (§6.2) suggesting listeners were more able to adjust their knowledge of voices based on new experiences when their expectations were weaker, it has also been noted that there are complexities that have potentially been obscured in this analysis and which would not support this thesis. The method that has been followed here, whilst able to provide insights relating to both research questions (§3.1), is not able to address the latter question in the depth that it demands, because although it presents cases where the unexpected features of a voice are being learned, and others where listeners' perception of familiar features associated with voices for which they have strong expectations is not changing, it lacks cases where listeners have to respond to anomalous features in voices for which they have strong expectations. Possible approaches to addressing this issue are discussed in §7.2.

7.2. Future directions

The experiment conducted here demonstrated how listeners' expectations regarding different voices, derived only from in-the-moment interpretation of these voices' characteristics, influenced their perception, and also how their perception increasingly changed as they

absorbed new experiences of these voices. It has also been suggested that where listeners had stronger expectations of particular voices, their perception was less likely to change. The evidence for this was, however, more limited than that for the direct effects on perception. In order to take this further, it would be useful to introduce some modifications to the experimental design.

The introduction of low-level variability in stimuli for each voice would provide a means of gauging how listeners deal with such unexpected variability given their varying expectations of different voices. Are listeners more likely to ignore such variability when they have stronger expectations regarding the nature of the voice? This would provide support for the argument that the strength of expectations limits the impact of newly-acquired experiences. More difficult to achieve, but also informative, would be an experimental setup in which the feature being monitored would be one the distribution of which in a particular variety would not be well-known to listeners in spite of their being very familiar with the variety. Although the occurrence of non-prevocalic /ɹ/ in New Zealand English in the present study is in some ways a case of this, such a situation disassociated from orthography would provide a better context for observing listeners' perceptual learning ability.

To better understand the relationship between listeners' experiences, and expectations which ultimately depend on them, it would also be worthwhile to investigate what beliefs individuals hold regarding language, associated with different contexts and speakers, and what the basis for such beliefs is. Furthermore, it would be useful to study precisely what expectations develop in-the-moment when listeners hear a voice, and what awareness listeners have of these beliefs.

With regard to theoretical issues, it has been argued here that it may be worthwhile to give more attention to the concept of predictive processing (e.g. Clark, 2013, 2016; Lupyan, 2015) in usage-based models of language. In particular, it is important to consider how the concepts of prediction, prediction error, and precision weightings might or might not be compatible with well-established features of usage-based models such as the exemplar distribution, the perception-production loop, and parameters that influence perception. Finally, the necessity of moving beyond the notion of linguistic knowledge as formal, as argued by Hawkins (2010), has been reiterated here, with the further suggestion with regard to the issue of 'episodic versus abstract' knowledge that linguistic knowledge may be more properly said to vary in terms of specificity or generality, even as it is all, in a sense, abstract, being modulated through listeners'

expectations. This would allow usage-based models to account both for individuals' sensitivity to variation, and to their predilection for category formation, without necessitating the introduction of discrete forms of linguistic knowledge. These suggestions are, however, largely speculative. Regardless of whether or not it would in fact be worthwhile to implement them in models, usage-based theories should increasingly pay heed to how listeners interpret stimuli based on expectations, rather than simply matching them to particular past experiences of similar stimuli.

7.3. Summary

Hay, Drager, & Gibson (2018) demonstrated that individuals' perception of speech sounds reflected their own past experience with these sounds, and that they were sensitive not only to the occurrence or otherwise of these sounds, but also their distribution in different linguistic contexts. These effects on perception were such that listeners even heard sounds where they were absent from the stimuli, when they thought they would likely be there. Their behaviour in a phoneme monitoring experiment demonstrated that they were guided by expectations of the occurrence of sounds that followed from their own variable experiences of these sounds. Such findings strongly support a usage-based model of language, in which linguistic knowledge develops through the language individuals experience, and in which this knowledge influences not only individuals' use of language, but also their perception of it.

This latter conclusion carries a profound implication for the usage-based paradigm: if perception is influenced by expectations, to what extent does linguistic knowledge develop directly from new experiences of stimuli, and in what way can that knowledge be said to be abstract or otherwise? To what extent are individuals capable of learning, if their expectations are so strong as to lead them to ignore the evidence of stimuli to which they are newly exposed?

The present study, drawing upon the growing body of literature demonstrating how individuals' beliefs regarding what they are hearing influence what they actually hear (e.g. McGowan, 2015; Gnevsheva, 2018; Fiedler, Keller, & Hanulíková, 2019), and the concept of predictive processing being explored in other areas of cognitive science (e.g. Clark, 2013; Lupyan, 2015), attempted to extend the study of Hay, Drager, & Gibson (2018) in examining the relationship between listeners' beliefs about what they will hear – their expectations – and their knowledge derived from what they do hear – their experiences – with respect to the stimuli to which they are exposed.

A phoneme monitoring experiment such as that carried out by Hay, Drager, & Gibson (2018) was adopted, introducing variation into the associations of the stimuli and the expectations listeners would hold regarding them rather than the experiences listeners would have with the features of interest.

It was shown that listeners' perception of non-prevocalic and intrusive /ɹ/ in different voices for which they had varying expectations did indeed vary significantly with these expectations. Listeners had strong, largely accurate expectations of the occurrence of these features in some voices, reflecting either their extensive experience with or their extensive stereotyping of similar voices. When this was the case, their perception remained largely constant over the course of the experiment, and never displayed a tendency to converge on the actual content of the stimuli, even when there were inaccuracies in their responses. For other voices, listeners had weak or incorrect expectations. When this was so, they were either very responsive to the actual content of the stimuli from the beginning, or their perception changed significantly through the absorption of new experiences. Listeners' varying expectations regarding different speakers had a significant influence on their perception of speech.

It was also argued that there was some evidence that listeners' perception of sounds might be less malleable – that is, less likely to change based on new experiences – when they had strong expectations regarding what they were hearing. On the whole, listeners by the end of the experiment tended to find a partially-rhotic voice of which they had weak expectations to be more rhotic than a fully-rhotic voice of which they had strong expectations, for example. There were, however, complexities in these data that hindered interpretation. The study does at least present this as an interesting possibility, and one that might be further investigated.

Against older approaches to theoretical linguistics in which the knowledge of language was conceived as an abstract formal system isolated from the contexts in which language is used, the functions its serves, and the memories it leaves, usage-based theories of language have presented a compelling argument that knowledge of language is entirely explicable in terms of individuals' experience of language. As it becomes increasingly apparent that the perception of language varies with listeners' expectations, and that these expectations affect how listeners acquire experiences of language, the challenge for usage-based models of language is that of taking this relationship into account. Individuals' knowledge of language is on some fundamental level subjective, emerging out of interaction of their beliefs regarding, and their experiences of, the world.

8. Appendix

8.1. Audio-editing scripts

8.1.1. Filtering script

```
form Low pass filter files
   comment Directory of source sound files
   text sound_directory
   sentence Sound_file_extension .wav
   comment Directory of created filtered files
   text end_directory
   comment Filter above this frequency
   positive Low_pass 300
   positive Smoothing 100
   comment Scale intensity to
   positive intensity 80.0
endform
Create Strings as file list... list 'sound_directory$'*'sound_file_extension$'
numberOfFiles = Get number of strings
for ifile to numberOfFiles
   filename$ = Get string... ifile
   Read from file... 'sound_directory$''filename$'
   Filter (pass Hann band)... 0 low_pass smoothing
   Scale intensity... intensity
   Write to WAV file... 'end_directory$''filename$'
   select Strings list
endfor
select all
Remove
```

8.1.2. Recording script

```
form Stimulus Recording Control
   comment Speaker
   text speakerCode A
   comment Directory for audio model files
   text indir phrases_filtered
   comment Output directory
   text outdir phrases_completed
   positive Recording_time 4
endform
recordTime = recording_time
strings = Create Strings as file list: "list", indir$ + "/*.wav"
numberOfFiles = Get number of strings
for ifile to numberOfFiles
    selectObject: strings
    fileName$[ifile] = Get string: ifile
    sound[ifile] = Read from file: indir$ + "/" + fileName$[ifile]
endfor
beginPause: "Instructions"
   comment: "You will be asked to read a series of phrases into the
microphone."
   comment: "You will hear a filtered audio model of the phrase, which will be
displayed in writing on the screen."
   comment: "When the audio model has finished playing, please repeat the
phrase."
   comment: "IMPORTANT: Please follow the speaking rate and stress pattern of
the model."
   comment: "You will be asked to repeat, or go onto the next word."
   comment: "This will continue until all words have been recorded."
   comment: "You may take a break whenever you choose."
endPause: "Continue", 1
pauseScript: "Click continue to begin"
for file from 1 to numberOfFiles
   phrase$ = fileName$[file] - ".wav"
   phraseWords$ = replace$(phrase$, "_", " ", 0)
   echo Repeat "'phraseWords$'"
   rep = 0
   repeat
```

```
            sleep(0.25)

            rep += 1

            selectObject: sound[file]

            Play

            recSound[rep] = Record Sound (fixed time): "Microphone", 0.99, 0.5,
"44100", recordTime

            Save as WAV file: outdir$ + "/" + speakerCode$ + "__" + phrase$ +
"__" + string$(rep) + ".wav"

            beginPause: "Re-record?"

                   comment: "What would you like to do?"

            clicked = endPause: "Redo", "Next Phrase", 2

   until clicked = 2

   for rec from 1 to rep

            removeObject: recSound[rec]

   endfor

endfor

selectObject: strings

for file from 1 to numberOfFiles

   plusObject: sound[file]

endfor

Remove

pauseScript: "This is the end of the recording task.  Thank you very much for
your time."
```

8.2. Experiment instructions

8.2.1. At the commencement of the experiment

You will be played successive phrases and asked to identify if a sound is present within them. Try to focus on the sound, instead of the spelling, which does not always reflect the actual sounds in these phrases.

Press the LEFT arrow key if you think the indicated sound is present in the phrase.

Press the RIGHT arrow key if you think it absent from the phrase.

You will be asked to complete a brief training task in this way, and then the main task, which will go at some length.

[press any key to continue]

8.2.2. At the commencement of the training task, varying randomly between /f/ and /s/

Now please identify if the sound {/f/~/s/} is present in the following few phrases. You will be notified if a response is correct. In some cases, if you make a mistake, the phrase will be replayed and the mistake explained.

[press any key to begin]

8.2.3. Training task feedback

Correct.

Incorrect. (For §8.4.3, Nos. 1, 4, 6, and 8)

Incorrect. Please listen attentively as the sound is replayed. Although no <f> is present in the spelling, the sound /f/ is present in the middle of <coughing>. (For §8.4.3, No. 5)

Incorrect. Please listen attentively as the sound is replayed. Although no <f> is present in the spelling, the sound /f/ is present at the beginning of <phase>. (For §8.4.3, No. 7)

Incorrect. Please listen attentively as the sound is replayed. Although no <s> is present in the spelling, the sound /s/ is present at the beginning of <city>. (For §8.4.3, No. 2)

Incorrect. Please listen attentively as the sound is replayed. Although <s> is present in the spelling, it represents the sound /z/. (For §8.4.3, No. 3)

8.2.4. *At the commencement of the perception task:*

From now on, in the main task, you will not receive any feedback. Please identify if the sound /ɹ/, as in <tomo*rr*ow>, is present in all following phrases. You may take a break at the end of each block of phrases.

If you take too long, a message will be displayed asking you to increase the speed at which you respond, before the program skips to the next phrase.

[press any key to begin]

8.3. Experiment questionnaire

1. Gender:
2. Place of birth:
3. Year of birth:
4. Please list any places other than your place of birth where you have lived for a year or longer, and indicated how long you have lived in these places, below:
5. Where did your parents grow up, and what were their native languages?
6. Have you learnt any languages other than English? How old were you when you began learning?
7. You have been listening to various different accents. Which accents do you think these were?
8. What experience do you have with the following accent?

   {‘General American’, ‘New Zealand’, ‘Southland’, ’Boston’}

   Click on the sliders {indicating ‘Daily’, ‘Weekly’, ‘Monthly’, ‘Yearly’, ‘Never’} to indicate how often you would expect to hear it in the identified contexts {‘Among family’, ‘Among friends’, ‘At work’, ‘From media’}.

8.4. Experiment stimuli

8.4.1. Perception task stimuli

/ɹ/ type	*Phrase*	No.	*Phrase*	No.
Intrusive	claw and bite	1	Utah and Idaho	21
	gnaw and chew	2	shah and king	22
	seesaw and slide	3	la-de-da and posh	23
	heehaw and neigh	4	Panama and Suez	24
	jigsaw and game	5	gaga and senile	25
	thaw and melt	6	blah and nonsense	26
	jaw and teeth	7	ma and pa	27
	coleslaw and salad	8	spa and sauna	28
	law and statute	9	Omaha and Utah	29
	guffaw and cackle	10	ha-ha and chuckling	30
	clawing and biting	11		
	cutting and sawing	12		
	gnawing and chewing	13		
	pitching and yawing	14		
	seesawing and sliding	15		
	cheeping and cawing	16		
	guffawing and cackling	17		
	chainsawing and cutting	18		
	thawing and melting	19		
	heehawing and neighing	20		
/ɹ/-less	wattle and daub	31	façade and disguise	38
	dawn and dusk	32	avocado and guacamole	39
	eagle and hawk	33	donations and alms	40
	thoughts and ideas	34	balm and ointment	41
	taunting and intimidating	35	embalming and mummifying	42
	sighing and yawning	36	soothing and calming	43
	walking and ambling	37	palming and stealing	44
			singing and psalming	45
Prevocalic	fraught and difficult	46	branch and twig	54
	loud and raucous	47	file and rasp	55
	prawn and fish	48	rascal and nuisance	56
	sketch and draw	49	draught and gust	57

	aural and visual	50	grasping and holding	58		
	meeting and forum	51	distinguishing and contrasting	59		
	overwrought and emotional	52	rafting and canoeing	60		
	plaguing and marauding	53				
Non-prevocalic	port and haven	61	sharpen and blunt	71	herbs and spices	81
	smooth and coarse	62	joke and farce	72	beech and birch	82
	doors and windows	63	marsh and swamp	73	irk and vex	83
	knife and fork	64	boat and barge	74	second and third	84
	hornets and wasps	65	farm and livestock	75	infirm and sick	85
	corking and bottling	66	carding and spinning	76	working and toiling	86
	compelling and forcing	67	dealing and bargaining	77	tending and nursing	87
	boarding and leaving	68	parting and leaving	78	earning and spending	88
	hating and scorning	69	assessing and marking	79	sipping and slurping	89
	hoarding and keeping	70	barking and shouting	80	hurting and aching	90

8.4.2. Rhoticity chart: listed varieties

('NZ' = General New Zealand, 'Sld' = Southland, 'GA' = General American, 'NE' = Eastern New England) indicate which stimuli contain /ɹ/ in each corresponding voice.

	Variables		*Control stimuli*	
Vowel context	Non-prevocalic /ɹ/	Intrusive /ɹ/	Prevocalic /ɹ/	/ɹ/-less
BATH/PALM/START	GA	NZ, NE	NE, Sld, GA, NE	None
THOUGHT/NORTH/FORCE	GA	NZ, NE	NZ, Sld, GA, NE	None
NURSE	Sld, GA, NE			
Total	*30*	*30*	*15*	*15*

8.4.3. Training task stimuli

	Phrase (/s/)	No.	*Phrase (/f/)*	No.
	looking and seeking	1	coughing and sneezing	5
	city and capital	2	twisting and folding	6
	picking and choosing	3	cycle and phase	7
	passage and lane	4	aviation and flight	8

8.4.4. Production task stimuli

	Sentence	No.
	The Panama and Suez Canals serve much of the world's shipping.	1
	The last Shah of Iran was overthrown thirty-nine years ago.	2
	The resort has a spa and a swimming pool.	3
	I do not plan on visiting any part of Utah or Colorado.	4
	Omaha and Juno were codenames for two of the Normandy beaches.	5
	The aeroplane began to yaw as it turned in to land.	6
	A whole raft of law and order issues remained fraught with controversy.	7
	The flaw in the plan was hardly obvious.	8
	I am still in awe of their efforts.	9
	We saw another trawler turning far from the shore.	10
	The children had been drawing on the whiteboard with markers.	11
	The ice will be thawing away before long.	12
	The crafty cat was clawing at the carpet to get attention.	13
	I hope it will not be too overawing.	14
	They heard the dog pawing at the door.	15
	We grow strawberries and raspberries in the garden.	16
	Those rascals had a raucous party last night after work.	17
	I have already drafted a response to the card they brought us.	18
	There was no end to the drama.	19
	They sell a broad range of things there.	20

8.5. Preregistration form

Hypothesis. What's the main question being asked or hypothesis being tested in this study?

The central hypothesis of this research is that individuals' adjustment to novel linguistic features will be conditioned by their expectations relating to the language varieties in which they hear these features. In the context of this research, the linguistic features will be non-prevocalic and intrusive /r/, and the language varieties will be General New Zealand English, Southland New Zealand English, General American English and Boston English. Listeners will listen to stimuli from these varieties in a phoneme-monitoring task. The following predictions with regard to listeners' responses have been made:

For General New Zealand English, non-prevocalic /r/ is absent, and listeners will expect that it is absent. It is therefore predicted that they will consistently hear non-prevocalic /r/ as

absent over the course of the task. Intrusive /r/ is present, and listeners will expect its presence, so they will consistently hear intrusive /r/ as present. Both of these predictions are expected to hold regardless of the vowel preceding /r/.

Non-prevocalic /r/ in Southland New Zealand English is absent following START and NORTH and present following NURSE, but listeners will expect that it is absent following all vowels. Listeners will increasingly hear non-prevocalic /r/ as present after NURSE, and may also hear it somewhat more in other non-prevocalic contexts in spite of its consistent absence, under the influence of NURSE. Intrusive /r/ is absent, but listeners will expect that it is present, so listeners will increasingly hear intrusive /r/ as present over the course of the task, though this change is not expected to be as substantial as that with non-prevocalic /r/, given the feature is not as salient.

General American English has non-prevocalic /r/ in all environments, and listeners will consistently expect its presence. The reverse is true for intrusive /r/. Both of these predictions are expected to hold regardless of the vowel preceding /r/.

Non-prevocalic /r/ Boston English is absent following START and NORTH and present following NURSE, but listeners will expect that it is present in all contexts. As a result, listeners will increasingly hear non-prevocalic /r/ as absent after START and NORTH, and may also hear it somewhat less after NURSE in spite of its consistent presence, under influence of the other environments. Intrusive /r/ is present, though listeners will expect its absence: consequently, listeners will increasingly hear intrusive /r/ as present over the course of the task. Because listeners will have less experience with this variety than with Southland English, the change in the perception of it is expected to be greater.

Dependent variable. Describe the key dependent variable(s) specifying how they will be measured.

The dependent variable being measured is listeners' responses as to whether or not the sound /r/ is present in the stimulus.

Conditions. How many and which conditions will listeners be assigned to?

The experiment will use a within-listeners design with counterbalancing, reflecting the intention to measure individual listeners' responses to different varieties, rather than different listeners' responses to individual varieties, whilst compensating for potential order effects. There will be two conditions: all listeners will listen to stimuli from all four varieties, but half

will hear the two New Zealand varieties before the two American varieties, and half will hear them in the reverse order; the Southland English block will always follow the General New Zealand English block, and the Boston English block will always follow the General American block.

Analyses. Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Results will be analysed with logistic mixed-effects regression modelling. Intrusive /r/ and non-prevocalic /r/ will be modelled in separate models. For non-prevocalic /r/, fixed effects will be a three-way interaction between trial order (centred), preceding vowel, and language variety. For intrusive /r/, fixed effects will be a three way interaction between trial order, vowel and boundary category (word vs. morpheme), and language variety. Random intercepts for both models will be listener and stimulus, and random slopes for trial order by listener will be included.

Potential non-linear effects will be modelled with restricted cubic spline regressions, with the same dependent and independent variables included in linear modelling. Models will be pruned using backwards stepwise comparisons by means of ANOVA comparison.

If questionnaire data reveals adequate variability across listeners in exposure to the target varieties, degree of exposure may also be explored in post-hoc analyses.

Outliers and Exclusions. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Where listeners display non-prevocalic rhoticity in the accompanying production task, they will be excluded from any analysis dealing with the group of listeners as a whole, though their results may be analysed separately. If listeners hesitate in reading the production task sentences in environments where intrusive /r/ might occur, these tokens will be excluded from any analysis of their use of the feature.

Any listener who displays 2.5 standard deviations below mean accuracy on control trials will be excluded from analysis.

Sample Size. How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

There will be thirty listeners, each providing 90 responses to four varieties each, giving a total of 10,800 points of data.

Other. Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

The main phoneme-monitoring task will be accompanied by a production task assessing listeners' own use of non-prevocalic and intrusive /r/, and a questionnaire assessing their linguistic experience. Variables collected in the questionnaire include listeners' parents' language backgrounds, the language varieties they believe they were hearing in the perception task, and the rate at which they hear these varieties spoken in different contexts. Because these additional tasks will occur after the main task so as not to influence them, it is likely that results will be affected by listener experience of the main task, but results from them may be used in post-hoc interpretation of results.

9. References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-Learning and Schematization in a Usage-Based Account of Syntactic Acquisition. *The Linguistic Review, 23*(3), 275-290.

Adank, P., Evans, B., Stuart-Smith, J., & Scott, S. (2009). Comprehension of Familiar and Unfamiliar Native Accents under Adverse Listening Conditions. *Journal of Experimental Psychology, 35*(2), 520-529.

Arnold, J., Hudson Kam, C., & Tanenhaus, M. (2007). If You Say Thee Uh You Are Describing Something Hard: The On-Line Attribution of Disfluency during Reference Comprehension. *Journal of Experimental Psychology, 33*(5), 914-930.

Audacity Team. (2019). Audacity(R): Free Audio Editor and Recorder, Version 2.3.2 [Computer software]. Retrieved May 20, 2019, from https://audacityteam.org/

Babel, M., & Russell, J. (2015). Expectations and Speech Intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823-2833.

Bartlett, C. (2002). The Southland Variety of New Zealand English: Postvocalic /r/ and the BATH Vowel (Unpublished doctoral dissertation). University of Otago, Dunedin, New Zealand.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bayard, D. (1990). 'God Help Us if We All Sound Like This': Attitudes to New Zealand and Other English Accents. In A. Bell, & J. Holmes (Eds.), *New Zealand Ways of Speaking English* (pp. 67-96). Wellington: Victoria University Press.

Bayard, D., & Bartlett, C. (1996). "You Must Be From Gorre": Attitudinal Effects of Southland Rhotic Accents and Speaker Gender on NZE Listeners and the Question of NZE Regional Variation. *Te Reo, 39*, 25-46.

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. (2001). Pax Americana? Accent Attitudinal Evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics, 5*(1), 22-49.

Becker, K. (2009). /r/ and the Construction of Place Identity on New York City's Lower East Side. *Journal of Sociolinguistics, 13*(5), 634-658.

Becker, K. (2014). (r) We There Yet? The Change to Rhoticity in New York City English. *Language Variation and Change, 26*(2), 141-168.

van Berkum, J. (2008). Understanding Sentences in Context: What Brain Waves can Tell Us. *Current Directions in Psychological Science, 17*(6), 376-380.

Boberg, C. (2001). The Phonological Status of Western New England. *American Speech, 76*(1), 3-29.

Boersma, P., & Weenink, D. (2018). Praat: Doing Phonetics by Computer, Version 6.0.42 [Computer software]. Retrieved August 15, 2018, from http://www.praat.org/

Bradlow, A., & Bent, T. (2008). Perceptual Adaptation to Non-Native Speech. *Cognition, 106*(2), 707-729.

Brunellière, A., & Soto-Faraco, S. (2013). The Speakers' Accent Shapes the Listeners' Phonological Predictions during Speech Perception. *Brain & Language, 125*(1), 82-93.

Bubic, A., von Cramon, Y., & Schubotz, R. (2010). Prediction, Cognition and the Brain. *Frontiers in Human Neuroscience, 4*.

Bybee, J. (1995). Regular Morphology and the Lexicon. *Language and Cognitive Processes, 10*(5), 425-455.

Campbell-Kibler, K. (2009). The Nature of Sociolinguistic Perception. *Language Variation and Change, 21*(1), 135-156.

Campbell-Kibler, K. (2010). The Effect of Speaker Information on Attitudes toward (ING). *Journal of Language and Social Psychology, 29*(2), 214-223.

Carr, T., Brown, J., & Charalambous, A. (1989). Repetition and Reading: Perceptual Encoding Mechanisms Are Very Abstract but Not Very Interactive. *Journal of Experimental Psychology, 15*(5), 763-778.

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences, 36*, 181-204.

Clark, A. (2016). Attention Alters Predictive Processing. *Behavioral and Brain Sciences, 39*, 23-24.

Clopper, C., & Pisoni, D. (2004a). Effects of Talker Variability on Perceptual Learning of Dialects. *Language and Speech, 47*(3), 207-239.

Clopper, C., & Pisoni, D. (2004b). Homebodies and Army Brats: Some Effects of Early Linguistic Experience and Residential History on Dialect Categorization. *Language Variation and Change, 16*(1), 31-48.

Connine, C., & Titone, D. (1996). Phoneme Monitoring. *Language and Cognitive Processes, 11*(6), 635-646.

Cutler, A., & Norris, D. (2016). Bottoms Up! How Top-Down Pitfalls Ensnare Speech Perception Researchers, Too. *Behavioural and Brain Sciences, 39*, 25-27.

Cutler, A., Eisner, F., McQueen, J., & Norris, D. (2010). How Abstract Phonemic Categories are Necessary for Coping with Speaker-Related Variation. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 91-111). Berlin: De Gruyter Mouton.

Dahan, D., & Mead, R. (2010). Context-Conditioned Generalization in Adaptation to Distorted Speech. *Journal of Experimental Psychology, 36*(3), 704-728.

Davis, M., & Johnsrude, I. (2007). Hearing Speech Sounds: Top-Down Influences on the Interface between Audition and Speech Perception. *Hearing Research, 229*, 132-147.

DeLong, K., Urbach, T., & Kutas, M. (2005). Probabilistic Word Pre-Activation during Language Comprehension Inferred from Electrical Brain Activity. *Nature Neuroscience, 8*(8), 1117-1121.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic Effects on Phoneme Monitoring. *Canadian Journal of Experimental Psychology, 49*(2), 264-271.

Dilley, L., & McAuley, D. (2008). Distal Prosodic Context Affects Word Segmentation and Lexical Processing. *Journal of Memory and Language, 59*(3), 294-311.

Docherty, G., & Foulkes, P. (2014). An Evaluation of Usage-Based Approaches to the Modelling of Sociophonetic Variability. *Lingua, 142*, 42-56.

Docherty, G., Langstrof, C., & Foulkes, P. (2013). Listener Evaluation of Sociophonetic Variability: Probing Constraints and Capabilities. *Linguistics, 51*(2), 355-380.

Drager, K. (2010). Speaker Age and Vowel Perception. *Language and Speech, 54*(1), 99-121.

Eimas, P., & Corbit, J. (1973). Selective Adaptation of Linguistic Feature Detectors. *Cognitive Psychology, 4*(1), 99-109.

Ettlinger, M., & Johnson, K. (2009). Vowel Discrimination by English, French and Turkish Speakers: Evidence for an Exemplar-Based Approach to Speech Perception. *Phonetica, 66*(4), 222-242.

Evans, B., & Iverson, P. (2004). Vowel Normalization for Accent: An Investigation of Best Exemplar Locations in Northern and Southern British English Sentences. *The Journal of the Acoustical Society of America, 115*(1), 352-361.

Evans, B., & Iverson, P. (2007). Plasticity in Vowel Perception and Production: A Study of Accent Change in Young Adults. *The Journal of the Acoustical Society of America, 121*(6), 3814-3826.

Farmer, T., Brown, M., & Tanenhaus, M. (2013). Prediction, Explanation, and the Role of Generative Models in Language Processing. *Behavioral and Brain Sciences, 36*(3), 211-212.

Feagin, C. (1990). The Dynamics of a Sound Change in Southern States English: From r-less to r-ful in Three Generations. In J. Edmondson, C. Feagin, & P. Mülhäusler (Eds.), *Development and Diversity: Language Variation in Time and Space* (pp. 129-146). Arlington, Texas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Fiedler, S., Keller, C., & Hanulíková, A. (2019, March). Social Expectations and Intelligibility of Arabic Speech in Noise. Paper presented at the International Conference of Phonetic Sciences, Melbourne, Australia.

Fine, A., Jaeger, F., Farmer, T., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *Public Library of Science One, 8*(10).

Firestone, C., & Scholl, B. (2016). Cognition Does Not Effect Perception: Evaluating the Evidence for "Top-Down" Effects. *Behavioral and Brain Sciences, 39*, 1-19.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a Regional Accent Perturb Speech Processing? *Journal of Experimental Psychology, 32*(5), 1276-1293.

Foss, D., & Lynch, R. (1969). Decision Processes during Sentence Comprehension: Effects of Surface Structure on Decision Times. *Perception & Psychophysics, 5*(3), 145-148.

Foulkes, P., & Docherty, G. (2006). The Social Life of Phonetics and Phonology. *Journal of Phonetics, 34*(4), 409-438.

Gibson, A. (2016). Samoan English in New Zelaand: Examples of Consonant Features from the UC QuakeBox. *New Zealand English Journal, 29-30*, 25-50.

Gimson, A. (1962). *An Introduction to the Pronunciation of English.* London: Edward Arnold.

Gnevsheva, K. (2018). The Expectation Mismatch Effect in Accentedness Perception of Asian and Caucasian Non-Native Speakers of English. *Linguistics, 56*(3), 581-598.

Goldinger, S. (1996). Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology, 22*(5), 1166-1183.

Goldinger, S. (1997). Words and Voices: Perception and Production in an Episodic Lexicon. In K. Johnson, & J. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 33-66). San Diego, California: Academic Press.

Goldinger, S. (2007, August). A Complementary-Systems Approach to Abstract and Episodic Speech Perception: Paper presented at the International Conference of Phonetic Sciences, Saarbrücken, Germany.

Goldinger, S., Pisoni, D., & Logan, J. (1991). On the Nature of Talker Variability Effects on the Recall of Spoken Word Lists. *Journal of Experimental Psychology, 17*(1), 152-162.

Gonzales, K., Byers-Heinlein, K., & Lotto, A. (2019). How Bilinguals Perceive Speech Depends on Which Language They Think They're Hearing. *Cognition, 182*, 318-330.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its Origins and Evolution.* Cambridge: Cambridge University Press.

Gordon, M. (2004). New York, Philadelphia, and Other Northern Cities: Phonology. In E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 282-299). Berlin: Mouton de Gruyter.

Gordon, P. (1997). What New Zealanders Believe About Regional Variation in New Zealand English: A Folklinguistic Investigation. *New Zealand English Journal, 11*, 14-25.

Grossberg, S. (1980). How Does a Brain Build a Cognitive Code? *Psychological Review, 87*(1), 1-51.

Grossberg, S. (2003). Resonant Neural Dynamics of Speech Perception. *Journal of Phonetics, 31*(3), 423-445.

Halle, M., & Idsardi, W. (1997). R, Hypercorrection, and the Elsewhere Condition. In I. Roca (Ed.), *Derivations and Constraints in Phonology* (pp. 331-348). Oxford: Clarendon Press.

Hallé, P., Chéreau, C., & Segui, J. (2000). Where Is the /b/ in "absurde" [apsyrd]? It Is in French Listeners' Minds. *Journal of Memory and Language, 43*(4), 618-639.

Hanulíková, A. (2018). The Effect of Perceived Ethnicity on Spoken Text Comprehension under Clear and Adverse Listening Conditions. *Linguistics Vanguard, 4*(1).

Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of Subject-Verb Agreement Mediated by Speakers' Gender. *Frontiers in Psychology, 6*.

Harrell, F. (2018, January 7). rms: Regression Modeling Strategies. *5.1-2*. Retrieved from http://biostat.mc.vanderbilt.edu/rms

Hawkins, S. (2003). Roles and Representations of Systematic Fine Phonetic Detail in Speech Understanding. *Journal of Phonetics, 31*(3), 373-405.

Hawkins, S. (2010). Phonological Features, Auditory Objects, and Illusions. *Journal of Phonetics, 38*(1), 60-89.

Hay, J., & Sudbury, A. (2005). How Rhoticity Became /r/-Sandhi. *Language, 81*(4), 799-823.

Hay, J., Drager, K., & Gibson, A. (2018, June). Hearing R-Sandhi: The Role of Past Experience. *Language, 94*(2), 360-404.

Hay, J., Drager, K., & Warren, P. (2010). Short-Term Exposure to One Dialect Affects Processing of Another. *Language and Speech, 53*(4), 447-471.

Hay, J., Maclagan, M., & Gordon, E. (2008). *New Zealand English.* Edinburgh: Edinburgh University Press.

Hay, J., Nolan, A., & Drager, K. (2006). From Fush to Feesh: Exemplar Priming in Speech Perception. *The Linguistic Review, 23*(3), 351-379.

Hilpert, M. (2008). New Evidence against the Modularity of Grammar: Constructions, Collocations, and Speech Perception. *Cognitive Linguistics, 19*(3), 491-511.

Irwin, P., & Nagy, N. (2007). Bostonians /r/ Speaking: A Quantitative Look at (R) in Boston. *University of Pennsylvania Working Papers in Linguistics, 13*(2), 135-147.

Johnson, K. (1997). Speech Perception without Speaker Normalization. In K. Johnson, & J. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-165). San Diego, California: Academic Press.

Johnson, K. (2006). Resonance in an Exemplar-Based Lexicon: The Emergence of Social Identity and Phonology. *Journal of Phonetics, 34*(4), 485-499.

Kahn, D. (1976). Syllable-Based Generalizations in English Phonology (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, Massachusetts.

Kang, O., & Rubin, D. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology, 28*(4), 441-456.

Kennedy, M. (2006). Variation in the Pronunciation of English by New Zealand School Children. Wellington: Victoria University of Wellington.

Kleinschmidt, D., & Jaeger, F. (2015). Robust Speech Perception: Recognize the Familiar, Generalize to the Similar, and Adapt to the Novel. *Psychological Review, 122*(2), 148-203.

Kraljic, T., & Samuel, A. (2006). Generalization in Perceptual Learning for Speech. *Psychonomic Bulletin & Review, 13*(2), 262-268.

Kraljic, T., & Samuel, A. (2007). Perceptual Adjustments to Multiple Speakers. *Journal of Memory and Language, 56*(1), 1-15.

Kraljic, T., & Samuel, A. (2011). Perceptual Learning Evidence for Contextually-Specific Representations. *Cognition, 121*(3), 459-465.

Kraljic, T., Brennan, S., & Samuel, A. (2008). Accommodating Variation: Dialects, Idiolects, and Speech Processing. *Cognition, 107*(1), 54-81.

Kraljic, T., Samuel, A., & Brennan, S. (2008). First Impressions and Last Resorts: How Listeners Adjust to Speaker Variability. *Psychological Science, 19*(4), 332-338.

Kretzschmar, W. (2004). Standard American English Pronunciation. In E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 257-269). Berlin: Mouton de Gruyter.

Kroczek, L., & Gunter, T. (2017). Communicative Prediction can Overrule Linguistic Priors. *Scientific Reports, 7*(1).

Kurkela, J., Hämäläinen, J., Leppänen, P., Shu, H., & Astikainen, P. (2019). Passive Exposure to Speech Sounds Modifies Change Detection Brain Responses in Adults. *NeuroImage, 188*, 208-216.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington, District of Columbia: Center for Applied Linguistics.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change.* Berlin: Mouton de Gruyter.

Langacker, R. (1987). *Foundations of Cognitive Grammar.* Stanford, California: Stanford University Press.

Langacker, R. (1999). *Grammar and Conceptualization.* Berlin: Mouton de Gruyter.

Lev-Ari, S. (2015). Comprehending Non-Native Speakers: Theory and Evidence for Adjustment in Manner of Processing. *Frontiers in Psychology, 5*.

Lev-Ari, S. (2017). Talking to Fewer People Leads to Having More Malleable Linguistic Representations. *Public Library of Science One, 12*(8).

Lev-Ari, S. (2018). Social Network Size Can Influence Linguistic Malleability and the Propagation of Linguistic Change. *Cognition, 176*, 31-39.

Lev-Ari, S., & Peperkamp, S. (2016). How the Demographic Makeup of Our Community Influences Speech Perception. *The Journal of the Acoustical Society of America, 139*(6), 3076-3087.

Lev-Ari, S., Ho, E., & Keysar, B. (2018). The Unforeseen Consequences of Interacting with Non-Native Speakers. *Topics in Cognitive Science, 10*(4), 835-849.

Levon, E. (2014). Categories, Stereotypes, and the Linguistic Perception of Sexuality. *Language in Society, 43*(5), 539-566.

Liberman, A., & Mattingly, I. (1985). The Motor Theory of Speech Perception Revised. *Cognition, 21*(1), 1-36.

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese Listeners to Identify English /r/ and /l/. II: The Role of Phonetic Environment and Talker Variability in Learning New

Perceptual Categories. *Journal of the Acoustical Society of North America, 94*(3), 1242-1255.

Lupyan, G. (2015). Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *Review of Philosophy and Psychology, 6*(4), 547-569.

Marsden, S. (2017). Are New Zealanders "Rhotic"? The Dynamics of Rhoticity in New Zealand's Small Towns. *English World-Wide, 38*(3), 275-304.

Martin, C., Branzi, F., & Bar, M. (2018). Prediction is Production: The Missing Link between Language Production and Comprehension. *Scientific Reports, 8*(1), 1079-1087.

Martin, C., Mullennix, J., Pisoni, D., & Summers, V. (1989). Effects of Talker Variability on Recall of Spoken Word Lists. *Journal of Experimental Psychology, 15*(4), 676-684.

Mather, P.-A. (2012). The Social Stratification of /r/ in New York City: Labov's Department Store Study Revisited. *Journal of English Linguistics, 40*(4), 338-356.

Maye, J., Aslin, R., & Tanenhaus, M. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science, 32*(3), 543-562.

McCarthy, J. (1991). Synchronic Rule Inversion. In L. Sutton, C. Johnson, & R. Shields (Eds.), *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society.* Berkeley, California: Berkeley Linguistics Society.

McCarthy, J. (1993). A Case of Surface Constraint Violation. *Canadian Journal of Linguistics, 38*, 169-195.

McClelland, J., & Elman, J. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology, 18*(1), 1-86.

McGowan, K. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech, 58*(4), 502-521.

McQueen, J., Cutler, A., & Norris, D. (2006). Phonological Abstraction in the Mental Lexicon. *Cognitive Science, 30*(6), 1113-1126.

Medin, D., & Schaffer, M. (1978). Context Theory of Classification Learning. *Psychological Review, 85*(3), 207-238.

Morley, R. (2014). Implications of an Exemplar-Theoretic Model of Phoneme Genesis: A Velar Palatalization Case Study. *Language and Speech, 57*(1), 3-41.

Morrill, T., Dilley, L., McAuley, D., & Pitt, M. (2014). Distal Rhythm Influences Whether or Not Listeners Hear a Word in Continuous Speech: Support for a Perceptual Grouping Hypothesis. *Cognition, 131*, 69-74.

Mullennix, J., Pisoni, D., & Martin, C. (1989). Some Effects of Talker Variability on Spoken Word Recognition. *The Journal of the Acoustical Society of America, 85*(1), 365-378.

Nagy, N. (2001). "Live Free or Die" as a Linguistic Principle. *American Speech, 76*(1), 30-41.

Nagy, N., & Irwin, P. (2010). Boston (r): Neighbo(r)s Nea(r) and Fa(r). *Language Variation and Change, 22*(2), 241-278.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology, 18*(1), 62-85.

Nielsen, D., & Hay, J. (2005). Perceptions of Regional Dialects in New Zealand. *Te Reo, 48*, 95-110.

Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115*(2), 357-395.

Norris, D., McQueen, J., & Cutler, A. (2000). Merging Information in Speech Recognition: Feedback is Never Necessary. *Behavioural and Brain Sciences, 23*(3), 299-370.

Norris, D., McQueen, J., & Cutler, A. (2003). Perceptual Learning in Speech. *Cognitive Psychology, 47*(2), 204-238.

Norris, D., McQueen, J., & Cutler, A. (2016). Prediction, Bayesian Inference and Feedback in Speech Recognition. *Language, Cognition and Neuroscience, 31*(1), 4-18.

Nosofsky, R. (1988). Exemplar-Based Accounts of Relations between Classification, Recognition, and Typicality. *Journal of Experimental Psychology, 14*(4), 700-708.

Nycz, J. (2013). New Contrast Acquisition: Methodological Issues and Theoretical Implications. *English Language and Linguistics, 17*(2), 325-357.

Nygaard, L., & Pisoni, D. (1998). Talker-Specific Learning in Speech Perception. *Perception & Psychophysics, 60*(3), 355-376.

Nygaard, L., Burt, A., & Queen, J. (2000). Surface Form Typicality and Asymmetric Transfer in Episodic Memory for Spoken Words. *Journal of Experimental Psychology, 26*(5), 1228-1244.

Nygaard, L., Sommers, M., & Pisoni, D. (1994). Speech Perception as a Talker-Contingent Process. *Psychological Science, 5*(1), 42-46.

Otten, M., & van Berkum, J. (2008). Discourse-Based Word Anticipation during Language Processing: Prediction or Priming? *Discourse Processes, 45*(6), 464-496.

Palmeri, T., Goldinger, S., & Pisoni, D. (1993). Episodic Encoding of Voice Attributes and Recognition for Spoken Words. *Journal of Experimental Psychology, 19*(2), 309-328.

Pavlík, R. (2016). A Usage-Based Account of /r/-Liaison in Standard British English. *Journal of Phonetics, 54*, 109-122.

Peirce, J. (2007). PsychoPy: Psychophysics Software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

van Petten, C., & Luka, B. (2012). Prediction during Language Comprehension: Benefits, Costs, and ERP Components. *International Journal of Psychophysiology, 83*(2), 176-190.

Pierrehumbert, J. (2001). Exemplar Dynamics: Word Frequency, Lenition and Contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137-157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2006). The Next Toolkit. *Journal of Phonetics, 34*(4), 516-530.

Pierrehumbert, J. (2016). Phonological Representation: Beyond Abstract versus Episodic. *Annual Review of Linguistics, 2*(1), 33-52.

Pisoni, D. (1993). Long-Term Memory in Speech Perception: Some New Findings on Talker Variability, Speaking Rate and Perceptual Learning. *Speech Communication, 13*(1), 109-125.

Pisoni, D. (1997). Some Thoughts on "Normalization" in Speech Perception. In *Talker Variability in Speech Processing* (pp. 9-32). San Diego, California: Academic Press.

Pisoni, D., & Sawusch, J. (1975). Some Stages of Processing in Speech Perception. In A. Cohen, & S. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 16-35). Berlin: Springer-Verlag.

Port, R., & Leary, A. (2005). Against Formal Phonology. *Language, 81*(4), 927-964.

Preston, D. (1996). Where the Worst English is Spoken. In E. Schneider (Ed.), *Focus on the USA* (pp. 297-360). Amsterdam: John Benjamins.

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rouder, J., & Ratcliff, R. (2006). Comparing Exemplar- and Rule-Based Theories of Categorization. *Current Directions in Psychological Science, 15*(1), 9-13.

Rubin, D. (1992). Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking Teaching Assistants. *Research in Higher Education, 33*(4), 511-531.

Saffran, J., Newport, E., & Aslin, R. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*(4), 606-621.

Samuel, A. (1981). Phonemic Restoration: Insights from a New Methodology. *Journal of Experimental Psychology, 110*(4), 474-494.

Samuel, A., & Kraljic, T. (2009). Perceptual Learning for Speech. *Attention, Perception, & Psychophysics, 71*(6), 1207-1218.

Schreier, D., Gordon, E., Hay, J., & Maclagan, M. (2003). The Regional and Sociolinguistic Dimension of /hw/ Maintenance and Loss in Early 20th Century New Zealand English. *English World-Wide, 24*(2), 245-269.

Shaw, J., Best, C., Docherty, G., Evans, B., Foulkes, P., Hay, J., & Mulak, K. (2018). Resilience of English Vowel Perception across Regional Accent Variation. *Laboratory Phonology, 9*(1), 1-36.

Siegelman, N., & Frost, R. (2015). Statistical Learning as an Individual Ability: Theoretical Perspectives and Empirical Evidence. *Journal of Memory and Language, 81*, 105-120.

Siegelman, N., Bogaerts, L., Christiansen, M., & Frost, R. (2016). Towards a Theory of Individual Differences in Statistical Learning. *Philosophical Transactions of the Royal Society B, 372*.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic Entrenchment: Prior Knowledge Impacts Statistical Learning Performance. *Cognition, 177*, 198-213.

Smith, R., & Hawkins, S. (2012). Production and Perception of Speaker-Specific Phonetic Detail at Word Boundaries. *Journal of Phonetics, 40*(2), 213-233.

Sommers, M., & Barcroft, J. (2006). Stimulus Variability and the Phonetic Relevance Hypothesis: Effects of Variability in Speaking Style, Fundamental Frequency, and Speaking Rate on Spoken Word Identification. *The Journal of the Acoustical Society of America, 119*(4), 2406-2416.

Sóskuthy, M. (2013). Analogy in the Emergence of Intrusive-r in English. *English Language and Linguistics, 17*(1), 55-84.

Stanford, J. (2018). *New England English: Large-Scale Acoustic Sociophonetics and Dialectology.* Manuscript submitted for publication.

Stanford, J., Leddy-Cecere, T., & Baclawski, K. (2012). Farewell to the Founders: Major Dialect Changes along the East-West New England Border. *American Speech, 87*(2), 126-169.

Stanford, J., Severance, N., & Baclawski, K. (2014). Multiple Vectors of Unidirectional Dialect Change in Eastern New England. *Language Variation and Change, 26*(1), 103-140.

Starks, D., & Reffell, H. (2005). Pronouncing Your Rs in New Zealand English? A Study of Pasifika and Maori Students. *New Zealand English Journal, 19*, 36-48.

Strand, E. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology, 18*(1), 86-99.

Studdert-Kennedy, M. (1976). Speech Perception. In *Contemporary Issues in Experimental Phonetics* (pp. 243-293). New York: Academic Press.

Sumner, M. (2011). The Role of Variation in the Perception of Accented Speech. *Cognition, 119*(1), 131-136.

Sumner, M., & Samuel, A. (2009). The Effect of Experience on the Perception and Representation of Dialect Variants. *Journal of Memory and Language, 60*(4), 487-501.

Thomas, E. (2004). Rural Southern White Accents. In E. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (pp. 300-324). Berlin: Mouton de Gruyter.

Todd, S., Pierrehumbert, J., & Hay, J. (2019). Word Frequency Effects in Sound Change as a Consequence of Perceptual Asymmetries: An Exemplar-Based Model. *Cognition, 185*, 1-20.

Tomasello, M. (2000). First Steps toward a Usage-Based Theory of Language Acquisition. *Cognitive Linguistics, 11*(1-2), 61-82.

Tulver, K., Aru, J., Rutiku, R., & Bachmann, T. (2019). Individual Differences in the Effects of Priors on Perception: A Multi-Paradigm Approach. *Cognition, 187*, 167-177.

Turner, G. (1966). *The English Language in Australia and New Zealand.* London: Longmans.

Van Riper, W. (1957). The Loss of Post-Vocalic R in the Eastern United States (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, Michigan.

Villard, S. (2009). Postvocalic /r/ in the Upper Valley of Vermont and New Hampshire. Paper presented at New Ways of Analyzing Variation 38, Ottawa, Canada.

Warren, R. (1970). Perceptual Restoration of Missing Speech Sounds. *Science, 167*(3917).

Watson, K., & Clark, L. (2015). Exploring Listeners' Real-Time Reactions to Regional Accents. *Language Awareness, 24*(1), 38-59.

Wedel, A. (2006). Exemplar Models, Evolution and Language Change. *The Linguistic Review, 23*(3), 247-274.

de Weers, N. (2018). Measuring the Effect of Speaker Ethnicity on Online Perception: Evidence from a Response Time Study. *The Journal of the Acoustical Society of America, 144*(3), 1802.

Wells, J. (1982). *Accents of English* (Vol. I). Cambridge: Cambridge University Press.

Whorf, B. (1943). Phonemic Analysis of the English of Eastern Massachusetts. *Studies in Linguistics, 2*(2), 21-40.

Yi, H.-G., Phelps, J., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced Efficiency of Audiovisual Intergration for Nonnative Speech. *The Journal of the Acoustical Society of America, 134*(5), 387-393.

Zheng, Y., & Samuel, A. (2017). Does Seeing an Asian Face Make Speech Sound More Accented? *Attention, Perception, & Psychophysics, 79*(6), 1841-1859.