THE EVOLUTION OF MEDIAL /t/ OVER REAL AND REMEMBERED TIME

JENNIFER HAY
University of Canterbury

PAUL FOULKES
University of York

This article follows a change in pronunciation of word-medial intervocalic /t/ in New Zealand English, as it unfolds over 120 years. Data are analyzed in the context of questions about the role of experience-based lexical representations and their potential impact on the time course of sound change in progress. Three major results are reported. First, frequent words lead the change. Second, the distributions of individual words affect their participation in the change: words favored by younger speakers are produced with newer variants. Finally, the topic of conversation affects which variant is favored: older topics elicit older variants. Together, these findings provide evidence that phonetic distributions of word-level representations are implicated in the course of sound change.*

Keywords: sound change, New Zealand English, lexical frequency, topic, lexical representation, exemplar theory, episodic memory

1. INTRODUCTION. There is now abundant evidence that people possess extensive episodic memories about their linguistic experiences. These memories may reflect, for example, aspects of speaker identity (whether a word was heard from a man or woman, a specific individual or an unfamiliar talker) and stylistic and situational variation (whether a word was heard in a careful or casual manner, spoken face to face, or filtered by telephone transmission). Numerous experimental studies have shown that such details, in certain circumstances, affect speech production and perception processes (e.g. Goldinger 1997, Strand 1999, Hawkins & Smith 2001). In recent years a consensus has emerged that linguistic knowledge is likely to include both abstract representations (e.g. phonemes) and, at the lexical level, episodic memories that include noncontrastive ‘fine phonetic detail’ associated with indexical properties of speakers and the contexts in which speech occurs (e.g. Pierrehumbert 2006). Debate is now centered not on whether representations include fine phonetic detail, but on the nature of how this detail is represented and how abstraction might emerge from the repository of episodic memories, and on the relationship between the abstract and detailed levels of representation in speaking and listening (Foulkes & Hay 2015).

The present study is concerned with the role of experience-based lexical representations in sound change. There remains considerable controversy over the role of lexical representations in language variation and change (cf. Labov 1994, Phillips 2006). We investigate lexical factors with reference to an ongoing change affecting word-medial intervocalic /t/ in New Zealand English. In words such as city the dominant pronuncia-
The evolution of medial /ɹ/ over real and remembered time has shifted over time from [t] to [d/ɾ]. Voicing or tapping processes such as this have often been described as lenition changes.

In many ways, our overall analysis is similar to many standard and well-known studies of sound change. We investigate the production of this common variable in speakers with birth dates spanning a considerable time range and determine the factors conditioning its variation. However, the types of potential factors we consider are much broader than is typical in studies of language variation and change. In particular, we focus on potential effects that would be predicted to be present if lexically specific episodic memory is involved in the production (and, presumably, perception) of lexical items. There is no denying that the existence of abstract phonemic categories plays an important role in governing the progress of sound change. Sound changes seldom completely abandon particular lexical items, for example: all tokens belonging to a particular category tend to participate eventually in a change. However, our focus is on the additional role of episodic memory and the potential consequences of ongoing activation of fine phonetic detail at the lexical level. We hypothesize that such representations are involved in speech production and perception, and are thus unavoidably central to mechanisms governing the spread of sound change across words and speakers.

In particular we test three predictions regarding the role of episodic memory in the progress of sound change. Predictions 2 and 3 are interwoven, both concerned with the ways that individuals’ speech productions may reflect change in progress. The three predictions are as follows.

• **Prediction 1: Frequent words lead leniting change.** It is a common prediction among proponents of lexical effects in sound change that frequent words lead lenition changes (e.g. Phillips 2006). Indeed, such a prediction falls out from implemented models of speech production containing phonetically rich memories (Pierrehumbert 2001). Despite the fact that this frequency effect is commonly claimed, however, most literature on the topic does not directly test the prediction in a rigorous way. In order to demonstrate definitively that frequency leads change, we predict an interaction between word frequency and speaker age, with younger speakers producing higher rates of innovative variants in more frequent items.

• **Prediction 2: Words used more by younger speakers are produced with more new variants.** If fine phonetic detail is contained in lexical representations and influences subsequent productions, then it follows that words that are more commonly used by younger speakers (‘young words’) are predicted to be more advanced in the change. By virtue of the fact they are used more by younger people—the speakers who lead the change—these words should be produced more often with the innovative variant than words used more by older speakers. This prediction is independent of the age of the speaker: ‘young words’ should attract more innovative variants no matter who is speaking.

• **Prediction 3: Speech about older events elicits older variants.** If some episodic memory is retained in memory over a speaker’s life, then it is possible that older episodic memories are triggered by particular contexts or topics. Specifically, we explore the idea that older phonetic variants may be used more often when speakers discuss events in the distant past, since they may draw on distributions in their episodic memory that have a long time depth. Newer variants, by contrast, are predicted to be found in greater frequency in discussions of more recent events.

Before describing the empirical part of the study we first outline the motivations for the study more fully.
1.1. Abstract and episodic representations. Structuralist-generative linguistics is built on insights surrounding abstract linguistic representations such as phonemes. There is ample experimental evidence that such categories exist and are important in governing speech production and perception. However, evidence for a concurrent role for episodic memory is also reasonably robust (e.g. Goldinger 1997, Lachs et al. 2003, Smith & Hawkins 2012). People are sensitive to fine phonetic details and access them in the course of listening and speaking. To some extent these details permit speaker-listeners to signal and monitor indexical, conversational, or discourse-level information (Thomas 2002, Local 2003, Ogden 2012). The details may also contribute to more ‘basic’ processes of speech perception and production such as phoneme or word recognition, at least in some circumstances (e.g. Strand & Johnson 1996, Gahl 2008, Hay & Drager 2010).

Evidence thus suggests that linguistic representations are, at some level, constantly updated with experience and are rich in phonetic and indexical detail. Models that make this assumption central differ to some extent in the levels of representation upon which they concentrate, but they are commonly referred to as exemplar or episodic models of representation (Nosofsky 1988, Goldinger 1997, Johnson 1997, Barlow & Kemmer 2000, Pierrehumbert 2001, 2002, 2003, Lachs et al. 2003).

Importantly, the presence of experience-based representations does not preclude a role for abstraction (a point commonly misunderstood or misrepresented in the literature; Foulkes & Hay 2015). Indeed, episodic representations are hypothesized to provide rich distributions that are necessary for abstractions to be formed. It is now widely accepted that some aspects of language structure can be learned through stochastic or probabilistic processes (Saffran et al. 1996, Maye et al. 2002, Chater et al. 2006, Maye et al. 2008). Abstract categories such as phonemes can be derived from the statistical distributions inherent to the speech samples constituting the input for acquisition (Pierrehumbert 2003). While many researchers would now agree that both episodic memory and abstraction play a role in representation, production, and perception, what is less clear is the nature of the balance of the ‘specific’ and the ‘abstract’, their roles across differently sized units, and the degree and manner in which each is involved in different types of linguistic task (Cutler 2010, Foulkes 2010, Foulkes & Hay 2015).

With respect to studies of language variation and change, the notion of experience-based representations is particularly relevant, due to common assumptions that episodic learning involves memory not only of detailed phonetic form, but also of associations between the word and contextually relevant nonlinguistic factors such as the identity and demographic profile of the interlocutor, or the general topic of discussion in which the word was situated. Episodic memories are relevant for language variation and change because they can capture processes of stochastic learning relevant for the development of both linguistic and nonlinguistic knowledge and provide a base over which associations can emerge between the two (Docherty & Foulkes 2014). For example, in British English, exemplars of words ending in /-t/ may vary in phonetic form, with possible variants including both [-tʰ] and [-ʔ]. For sociolinguistic or dialectological reasons these variants may be produced statistically more frequently by particular groups of speakers. In most dialects [-tʰ] is produced more often by women than men, for instance, with the reverse true for [-ʔ]. With a learning process based on detailed exemplars, knowledge will accrue that both variants are possible alternatives in the same set of words. Through association with particular speakers or speaker groups, the association between variants and speaker sex or gender may also be learned.

Regular sound change has traditionally been investigated and modeled in terms of abstract structures such as phonemes, and it is beyond question that change eventually
affects abstract categories. However, questions relating to word-specificity remain controversial, especially in predictions about language change. Early literature on the trajectory of sound change predicted word-specific effects and generated polarized debate about the degree to which there can be lexical diffusion of change (e.g. Schuchardt 1885, Zipf 1929, Trubetzkoy 1969 [1939], Chen & Wang 1975, Labov 1994; for a thorough review see Phillips 2006). Labov (1994), for example, acknowledges lexical diffusion only in the case of phoneme substitution changes (i.e. lexical rules) and rejects diffusion as a factor in phonetically conditioned postlexical rules (which would presumably include /t/-voicing/tapping).

Our aim here is not to enter the polarized debate. Instead, our starting point is to accept levels of representation that include both abstract phoneme-like elements and phonetically detailed, word-specific elements. Our particular interest in this study is the nature of episodic memories for words and the degree to which detailed word-specific memories might impact the course of regular sound change. There is good evidence for word-specific learning in speech production and perception. How, then, might the mechanics of such learning influence the trajectory of change? Our three specific predictions inform the empirical study we present here, as we now explain in more detail.

### 1.2. Prediction 1: Frequent Words Lead Leniting Change

Episodic/exemplar models assume that word memories are phonetically detailed and constantly updated with experience. If we couple this assumption with the existence of sound change in progress, then word frequency is predicted to be relevant to that change. This is because frequent words are encountered more often, and so their representations are (all else being equal) updated at a faster rate. The distribution for a more frequent word is therefore predicted to shift in the direction of an incoming variant more rapidly than the distribution for an infrequent word. Evidence in support of this stance is provided by Pierrehumbert (2001), who implemented a simple model of sound change driven by a leniting bias, that is, a tendency toward undershoot inherent to the speech production system. Speakers sample from their distribution of memories to create a production target. This target is subject to the leniting bias, and so the produced token may be more lenited than the target. The produced token is heard by the speaker as well as the addressee, so a version of it is then placed back into the exemplar store, illustrating what Pierrehumbert calls the ‘production-perception loop’. The bias in the loop drives the sound change forward via more frequent words, as these are produced, experienced, and thus fed back into the exemplar store more often. An increasing number of lenited exemplars in the exemplar store shifts the overall distribution of remembered forms, thus gradually increasing the likelihood of selecting a lenited form as the target for speech production. In Pierrehumbert’s simple model, the more iterations a word has through the loop (i.e. the more frequent it is), the more advanced it becomes in the sound change. The same prediction should hold for any change where some ongoing bias is involved. The bias need not be articulatory. For example, a social bias could be operative, motivating young speakers to assert a particular identity or stance by favoring one extreme of a particular phonetic distribution.

Pierrehumbert’s model, then, predicts that frequent words should lead sound change that is driven by a constant bias, such as a leniting bias. Several scholars working in usage-based models have searched for word-frequency effects of this kind (e.g. Bybee 2001, 2002, Phillips 2006). Bybee, for example, studied /-/t/, d/-deletion in coda clusters in Chicano English. She showed that deletion rates in high-frequency words were statistically higher than in low-frequency words (54% against 34%; Bybee 2002:264). Her interpretation is that high-frequency words therefore appear to be driving this change.
However, Bybee does not examine frequency effects at multiple time points in the change. What her data show is that, at a particular time point, there is a significant effect of frequency, with greater rates of deletion in higher-frequency items. This is problematic because frequency also has strong effects on patterns of speech production that are independent of language change. In particular, the more predictable a word or phrase is, the more likely it is to be reduced phonetically. Frequent words are inherently more predictable than infrequent ones and thus more likely to be subject to reduction. There are many demonstrations of this fact (e.g. Phillips 1984, 1999, 2006, Fowler & Housum 1987, Bybee 2000, Lavoie 2001, Jurafsky et al. 2002, Bell et al. 2003, Aylett & Turk 2004, Wright 2004, Baker & Bradlow 2009, Bell et al. 2009). If frequency can have a major effect on production at any time—particularly when lenition is involved—then showing an effect of frequency on the production of a given variable at a given time is insufficient to warrant claims about its involvement in sound change. If word frequency is important for sound change, then what is expected is an increasing effect of word frequency as the sound change begins to accelerate. A stable effect of frequency—which is present throughout the change or at a given point in the change—does not constitute sufficient evidence. A stable effect of frequency could be due to factors unrelated to change.

Whether word frequency is relevant to sound change is, as already noted, controversial. Yet proponents on both sides of the debate use the same types of data to support their case. Dinkin (2008) and Labov (2010), for example, test lexical frequency as a predictor in several different phonetically gradual changes. They report small or nonsignificant effects and conclude that there is no evidence for the involvement of lexical frequency in change. What they test, however, is a stable main effect of lexical frequency.

To establish that frequency drives leniting change, the correct prediction is not simply that frequent words behave differently from infrequent words. Instead, what is expected is an interaction between frequency and time. Hay and colleagues (2015) demonstrate an interaction between frequency and time for a very different type of sound change—vowels undergoing change due to a push-chain mechanism (specifically, the short front vowels in New Zealand English). They show that such changes operate quite differently. In their vowel-shift change, low-frequency words move the fastest, which they attribute to psycholinguistic mechanisms relating to the different treatment of high- and low-frequency words in the context of phonetic overlap. As far as we know, however, no one has tested the prediction that more canonical regular sound changes, such as leniting changes, should show high-frequency words moving faster than low-frequency words. Or, put differently, no one has tested for an interaction between frequency and time in predicting the progress of such a change.

1.3. Prediction 2: words used more by younger speakers are produced with more new variants. As explained above, we expect that words that are produced and experienced more often might lead the course of a change. However, not all words are produced and experienced equally often by every speaker, or by particular types of speakers. Some words, for example, might be used more frequently by students, or women, or rugby players. Because different speakers produce words differently, experience-based representations of words that are disproportionately experienced from one group may be dominated by phonetic characteristics associated with that group.

This hypothesis has been pursued by Walker and Hay (2011). In a study based on the same New Zealand English corpus that we analyze here, they identified words used more often by older speakers and words used more often by younger speakers. By conducting a lexical-decision task in two voices (an older voice and a younger voice), they
showed that word recognition was faster and more accurate when the age of the voice matched the ‘age of the word’. That is, if a word tended to be used more by older speakers, then listeners could more easily recognize that word when it was produced by an older voice.

A parallel prediction follows for production, in the context of change in progress. Younger speakers tend to produce innovative variants more often than older speakers do (indeed, the existence of such a distribution is the very diagnostic to identify change in the apparent-time framework). Remembered distributions for words that have been encountered more from younger speakers are therefore more likely to contain innovative variants. If word production involves sampling from the remembered word distribution in order to create a production target, then ‘young words’ should be further ahead in an ongoing sound change than ‘old words’. This should hold true regardless of the identity of the speaker: ‘young words’ should be more likely to contain an innovative variant even when being produced by an older speaker. There is some evidence to support this claim. For instance, Gordon and colleagues (2004) report a higher use of conservative variants of /r/ in New Zealand English in words with overt ‘old-time’ associations or describing activities related to early settler lifestyle such as farming and mining (e.g. cart, quartz). We test the claim with a much larger set of data here, and also explore the entire lexicon rather than solely words with overt associations.

The idea that word-level representations reflect the phonetics of past experience with that word receives some support from work showing that homophones are disambiguated differently depending on tone of voice (Nygaard & Lunders 2002). For example, when spoken in a happy voice, ‘bridal’ is more likely to be guessed by listeners than ‘bridle’, which presumably reflects greater experience with the former in a happy voice than the latter. While this does not relate to the distribution across speakers, it is likely to be a result of the same general mechanism.

It is important to note that we are not claiming that such an effect should be observable for every sound change. Production does not completely reflect perception. Even if it is the case that a remembered distribution for an older word contains more conservative variants, it does not follow that individuals will produce more conservative variants for that word. Younger speakers, for example, may not have the more conservative variant in their active repertoire, or may resist using it. For example, in British English /θ, ð/-fronting is well established and may be perceptually salient to listeners, but not all speakers use the innovative nonstandard variants [f, v] (Foulkes & Docherty 2007). It is thus in cases where most individuals use more than one variant that we expect to be able to observe this predicted effect in production. Medial /u/ in New Zealand English is one such case, and our data set provides an ideal test for this prediction.

1.4. Prediction 3: speech about older events elicits older variants. If knowledge of linguistic structures is updated throughout life in response to experience, the overall shape of the exemplar store may change noticeably over time if there are significant changes in the individual’s experience. For example, new dialects or radically different speakers may be encountered in situations of prolonged contact. Ongoing systematic change in experience over the life course might therefore yield systematic change in speech production as the statistical balance of the exemplar store shifts in response to that experience. Some previous work has revealed changes in adults’ speech production. This is particularly clear in cases where individuals move to new dialect regions, shifting production patterns in line with newly experienced local forms (Chambers 1992). A further implication of this position is that we should be able to find evidence for ongoing episodic learning if we compare individuals’ speech at different time points in
their lives. There is indeed evidence that individuals reflect sound changes in their community. Sankoff and Blondeau (2007), for instance, discuss the relationship between individual and community change in /r/ in Montreal French. A widely cited case concerns the speech of Queen Elizabeth II, which has been analyzed from recordings over a thirty-year period (Harrington et al. 2000). Some aspects of the Queen’s speech have changed in line with community change. For example, her /u:/ vowel was distinctly back in the 1950s, but had shifted forward by the 1980s, just as it had more generally in standard British English.

One open question with respect to this prediction is how much time depth is associated with episodic representations. Are exemplars experienced earlier in life forgotten and irrelevant for speech production later in life? Or does the remembered distribution contain layers of learning, reflecting some time depth in the stored experiences? At least some literature indicates that early memories may be particularly important and do not decay completely. Speakers who shift to another dialect region during adulthood revert back to their original accent when interacting with someone from ‘home’ or when under degraded auditory feedback (Howell et al. 2006). It has also been claimed that some speakers revert to the accent of their youth in extreme old age (Wright 1905:vii).

In the present study we explore the evidence for ongoing learning by accessing what we refer to as ‘remembered time’: we attempt to access layers of learning by examining speakers’ discussions of different times in their lives. If individuals store phonetically detailed memories over a long time period, they may access older variants when talking about older events. In other words, if there is a linguistic change in progress, the nature of that change should be replicated by speakers, visible across their speech about distant versus recent events.

A number of caveats are offered with respect to this prediction. First, we do not assume that ‘time depth’ is represented in episodic representations in a simplistic way, akin to a time stamp on a digital photograph. We consider it an open question how time depth of exemplars might be represented. Second, we do not assume that time depth is necessarily the focus of any particular level of indexical tagging in an exemplar representation. The time depth of an event is likely to be related to several other dimensions of experience, including topic and addressee. That is, discussions in or about the past may well be dominated by particular topics with particular audiences. Any episodic information related to time depth could be strongly linked to these other indexical properties. Finally, we also expect any effect of time depth on contemporary speech production to be subtle, subordinate in its influence on speech production relative to major sociolinguistic constraints connected to complex issues of style, attitude, or identity. We discuss these issues further in §4.3.

We now turn to our study of New Zealand English (-t-), outlining the data set used to test the three predictions.

2. New Zealand English (-t-). The variable chosen to test these predictions is (-t-) in New Zealand English (NZE). The variable is circumscribed as word-medial intervocalic /t/, in all positions except the onset of a stressed syllable. Examples from our data set, in decreasing frequency, include better, pretty, amalgamated, Waikato (a region and river of New Zealand), and pittosporum (a native plant). The heritage variant is generally assumed to be [t], imported by settlers to New Zealand from the 1840s onward. Although historical records of nineteenth-century migration are lacking, the available evidence shows that the largest numbers of early settlers came from London and the south of England (Gordon et al. 2004). In these areas, [t] seems to have been the default
variant in /-VtV/- contexts, although taps and glottal stops are also reported for Cockney by Sivertsen (1960). (Wells 1982 notes that [d] was characteristic of the west of England in the Survey of English dialects materials, and also of Ulster. Neither area contributed large numbers of migrants to New Zealand.)

The variable has been undergoing change for some time (Bell 1977, Bauer 1986, Holmes 1994, Fiasson 2013). In current NZE, voiced variants (ranging from [d] to lenited stops and taps) are increasingly common, similar to the variants used in Australia or North America (Wells 1982:250). We annotate this variant as [d], reflecting the most frequent phonetic type observed in the data (see further §2.2). Voiceless variants are also used, including plosive [t] (similar to standard British English) and a range of fricatives similar to those reported for Australian English (Tollfree 2001, Jones & McDougall 2009). In the most complete quantitative study of the variable in NZE, Holmes (1994) identifies change in progress toward [d], with strong sex and class correlations. Young working-class males are leading the change, and their usage of [d] is described as ‘semi-categorical’ (Holmes 1994:215). The change can therefore be characterized as a ‘change from below’ (Labov 1994:78), driven forward by men and members of lower socioeconomic groups. Wells (1982) suggests that the [d] variant might be an import from American usage, but it might also result from a supraregionalization process whereby NZE is converging with Australian English and diverging from its largely British roots. However, we leave aside the social explanations for the change here (see further Gordon et al. 2004).

Several writers treat voicing/tapping as a lenition process, at least historically if not necessarily as a synchronic phonological process (e.g. Wells 1982:94, Wolfram & Schilling-Estes 2006:55). As befits a lenition process, previous studies have revealed a stable lexical-frequency effect for this variable. In American English, for example, Patterson and Connine (2001) report 95% [d/r] on high-frequency words in the Switchboard corpus compared with 76% in low-frequency words.

2.1. Data set: ONZE corpus. The variable (-t-) was analyzed in the ONZE (Origins of New Zealand English) corpus (Gordon et al. 2007). ONZE contains recordings of around 600 speakers. Most material is spontaneous speech, consisting largely of personal narratives. ONZE combines a number of subcorpora, collected at different times and for different purposes. These include the Mobile Unit corpus (henceforth MU), a collection made in the 1940s of speakers who were born in the second half of the nineteenth century, and the Canterbury Corpus (CC), recordings made in the 1990s of speakers born from the 1930s to 1980s. The CC speakers include a well-balanced sample of men and women of different ages, categorized for social class (‘professional’ versus ‘nonprofessional’). The MU speakers are mainly male and not differentiated by social class, which in the nineteenth century was especially difficult to define (Gordon et al. 2004:58ff.). The ONZE corpus is tagged and searchable for numerous properties of phonology, grammatical category, speaking rate, and various types of social and discourse-level information. These are stored in, and were analyzed via, the LaBB-CAT interactive software system (Fromont & Hay 2012; formerly ONZE Miner—Fromont & Hay 2008).

For the analysis of (-t-) we selected a subset of speakers from the MU and CC, in order to sample the entire age range available in the ONZE corpus. Since ONZE remains a work in progress, we selected all and only those speakers for whom topic tagging was complete at the time of analysis. All speakers, as far as we know from the ONZE records, were born in New Zealand. The data set discussed in the present article is summarized in Table 1.
The final data set includes ninety-eight speakers born from 1862 to 1982. The initial analysis generated a larger data set, but tokens (and a small number of speakers) were eliminated for a variety of reasons, including poor recording quality, overlap between speakers, mispronunciations, and lexicalized phrases that had been transcribed as single items in ONZE (gotta, sorta). Māori words, other than place names and similar items that have entered common usage in NZE, were removed from the analysis, following Holmes (1994). A small number of unusual variants, including glottal stops and a zero form, were also removed prior to the statistical analysis (see further §2.2 below). Of the final set of tokens, 67% were two-syllable words. The /t/ formed the onset of the final syllable in 87% of tokens.  

2.2. Data analysis and coding. Each token of (-t-) was examined both auditorily and acoustically. The analysis was conducted by the second author using the LaBB-CAT software system. Tokens identified by LaBB-CAT’s search function were opened in Praat such that the relevant word could be heard within the whole utterance in which it was produced. Initial auditory impression was verified by inspection of both spectrogram and waveform. No quantification of detailed acoustic features was attempted, in part because of differences in recording procedures across the corpus as a whole and the suboptimal quality of some of the oldest recordings.

Given previous discussions of the variable in NZE and other dialects, we anticipated phonetic variation based on voicing, and we also expected a range of possible variants (Bell (1977:350), for example, identified six variants in an auditory analysis of the same variable in NZE). Complexity was indeed observed in the phonetic forms produced in the ONZE data. Variants found included [tʰ tɾ t d θ h] and a zero form (deletion), although four variants dominated: voiceless plosives (with or without aspiration or preaspiration), voiceless fricatives, voiced stops, and taps.

A subset of tokens was checked independently by the first author fairly early in the analysis process in order to establish criteria for classification. Agreement was extremely close, especially with respect to the broad voiceless/voiced distinction. Problematic tokens were classified after discussion, and criteria for further analysis were established. Formal comparison of interrater agreement was not attempted as we did not consider it necessary given the eventual reduction of the data to a coarse-grained binary distinction. There were two main sources of difficulty, both restricted in the main to certain individ-

<table>
<thead>
<tr>
<th>GROUP</th>
<th>CORPUS</th>
<th>SEX</th>
<th>CLASS</th>
<th>BIRTH YEARS</th>
<th>N SPEAKERS</th>
<th>N TOKENS</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM</td>
<td>MU</td>
<td>m</td>
<td>—</td>
<td>1862–1900</td>
<td>16</td>
<td>703</td>
</tr>
<tr>
<td>OF</td>
<td>MU</td>
<td>f</td>
<td>—</td>
<td>1863–1886</td>
<td>6</td>
<td>130</td>
</tr>
<tr>
<td>MMp</td>
<td>CC old</td>
<td>m</td>
<td>p</td>
<td>1936–1950</td>
<td>12</td>
<td>352</td>
</tr>
<tr>
<td>MMn</td>
<td>CC old</td>
<td>m</td>
<td>np</td>
<td>1932–1951</td>
<td>14</td>
<td>301</td>
</tr>
<tr>
<td>MFp</td>
<td>CC old</td>
<td>f</td>
<td>p</td>
<td>1933–1950</td>
<td>10</td>
<td>278</td>
</tr>
<tr>
<td>MFn</td>
<td>CC old</td>
<td>f</td>
<td>np</td>
<td>1933–1953</td>
<td>13</td>
<td>230</td>
</tr>
<tr>
<td>YMp</td>
<td>CC young</td>
<td>m</td>
<td>p</td>
<td>1970–1980</td>
<td>8</td>
<td>229</td>
</tr>
<tr>
<td>YMn</td>
<td>CC young</td>
<td>m</td>
<td>np</td>
<td>1975–1982</td>
<td>9</td>
<td>178</td>
</tr>
<tr>
<td>YFp</td>
<td>CC young</td>
<td>f</td>
<td>p</td>
<td>1968–1970</td>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>YFn</td>
<td>CC young</td>
<td>f</td>
<td>np</td>
<td>1965–1970</td>
<td>5</td>
<td>190</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>98</td>
<td>2,679</td>
</tr>
</tbody>
</table>

Table 1. Summary of data set (group codes: OM: old males; MMp: middle-aged males, professional; and so forth; class codes: p: professional, np: nonprofessional).

\[1\] A referee questioned whether syllable count played a role. We tried adding syllable count into the models reported in this article. It did not statistically improve those models and so was discarded.
ual speakers in the corpus and limited to a relatively small number of tokens. First, some
tokens gave an auditory impression of being voiced, but presented acoustic evidence of
a short break in phonation. These tokens also generally had short voice onset time and
were thus classified as (unaspirated) \([t]\). Second, there were tokens that appeared not to
have a break in voicing but that also contained distinct high-frequency fricative energy
more typically associated with voiceless plosives. Such tokens were generally heard as
\([t]\) rather than \([d]\), and thus were classified as \([t]\). The apparent continuity of voicing in
such cases was often difficult to resolve via acoustic analysis because of suboptimal
recording quality, often accompanied by a degree of low-frequency hum.

The initial coding took the phonetic detail into account. However, the initial cate-
gories were collapsed into a binary opposition based on voicing for the purposes of sta-
tistical analysis. This decision was made in order to follow the practice adopted by
previous studies, for example, Bell 1977 and Holmes 1994, and also because variation
centering on other factors appeared idiosyncratic. For example, the voiceless fricative
variant accounted for just under 20\% of the data overall, but for five speakers it ac-
counted for more than 70\% of their data, while another thirty-two speakers did not pro-
duce it at all. We label the collapsed variants \([t]\) and \([d]\), in order to capture voicing as
the main distinction. The former combines voiceless stops and voiceless coronal fri-
catives, while the latter includes voiced stops and fricatives as well as taps. As noted ear-
lier, a small number of unusual variants such as \([h]\) and \([ʔ]\) were removed prior to the
analysis reported here.

Several social and linguistic factors were also coded, as summarized in Table 2. Age
was categorized by year of birth. This was centered and scaled for the purposes of model
fitting (i.e. we subtracted the mean and divided by the standard deviation, so the variable
is centered around zero). Sex and class were included since previous studies have all
shown the variable (-t-) to be sensitive to these factors in NZE. Class was coded only for
the CC speakers, as already noted. The phonological codes considered whether the word
contained another /t/ or /d/ (to explore potential assimilation/dissimilation effects),
while the morphological factor considered whether the variable was located at an affix
juncture, for example, batting versus bottom (following Holmes 1994). Speech rate was
included because the voicing or tapping process is widely considered a lenition phe-
nomenon, at least historically (e.g. Wells 1982:94). It is thus likely to be sensitive to ar-
ticulation rate, with faster speech promoting \([d]\). In ONZE, speech rate is calculated as
the number of phonological syllables per second on the relevant line of the transcript. In
the ONZE protocols, transcribers introduce time-stamped breakpoints into the tran-
scripts, dividing turns into ‘lines’, which would typically represent one or two clauses.
Guidance to transcribers on the relevant scope is somewhat loose. The particular scope
over which speech rate is calculated is thus very local, but varies in a way that is inde-
pendent of (and unlikely to affect) the current analysis. Whether the word was a number
term was considered in light of Gordon and Maclagan (2008:74) and Horvath (2008:
100), who report higher rates of voiced variants in numbers (thirteen, fourteen, eighteen,
thirty, forty, eighty).

Word frequency was quantified within the ONZE corpus itself rather than with refer-
ence to independent resources such as CELEX (Baayen et al. 1995). This was because
many of the relevant words do not appear in other corpora that are based largely on
modern written texts, being particular to New Zealand or NZE (Waikato, pittosporum),
or because they are now relatively archaic (perambulator, petticoat). Frequency was
classified according to word form rather than lemma, based on previous studies indicat-
ing that word-form frequency offers a more explanatory insight into frequency effects
on lenition than lemma frequency (e.g. Jescheniak & Levelt 1994, Jurafsky et al. 2002, Gahl 2008, etc.), and because morphological structure was itself initially a point of interest. Word-form frequency was quantified on a logarithmic scale, following Baker and Bradlow (2009) and many others, and was centered and scaled using the same technique as for year of birth.

To calculate ‘word age’ we took the words in our data set and extracted their frequencies in both the CC and ONZE’s Intermediate Archive (IA). The latter is another subcorpus, comprising a set of speakers born between approximately 1900 and 1930 (i.e. intermediate between the MU and CC speakers) and thus representative of ‘older’ speakers that our CC speakers may have been exposed to. To align these frequency counts with each other we then centered (subtracted the mean) and scaled (divided by standard deviation) the two sets, and added one to each (to ensure that the full range of both estimates was positive). We then took the ratio of these values: CC/IA. This ‘word age’ value would be 1 if the words were equally represented in the two corpora. High values indicate overrepresentation in the CC (defined as ‘young words’). Low values indicate overrepresentation in the IA (‘old words’). Examples of young words include computer, metres, and university. Examples of old words include visitors, quarter, and beautiful. The resulting ratio values range from 0.35 to 3.37, with a median of 1.18 and a mean of 1.15.

The factor labeled ‘repetition’ indicates whether the word was mentioned for the first time in the discourse, or whether it was a repetition of a word already produced. This factor was again included because the variable is widely considered a lenition change, and it is well known that first productions tend to be more citation-like, with greater lenition occurring in repetitions (e.g. Fowler & Housum 1987, Bell et al. 2003, Aylett & Turk 2004, Baker & Bradlow 2009, Bell et al. 2009). This difference reflects the relative predictability of the word in context: repeated words are more predictable because they are primed by the first production. Previous research on repetition has not delimited a clear time frame in which to consider the priming effect, especially with respect to speech production. For pragmatic reasons we therefore elected to classify words as repetitions if they had already been produced by the same speaker within the preceding sixty seconds. The decision to limit to the same speaker was because many of the

<table>
<thead>
<tr>
<th>FACTOR</th>
<th>RANGE OR VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOCIAL</td>
<td></td>
</tr>
<tr>
<td>year of birth</td>
<td>1862–1982</td>
</tr>
<tr>
<td>sex</td>
<td>male/female</td>
</tr>
<tr>
<td>social class (CC only)</td>
<td>professional/nonprofessional</td>
</tr>
<tr>
<td>PHONOCOLOGICAL</td>
<td></td>
</tr>
<tr>
<td>preceding /t/ in word</td>
<td>yes/no</td>
</tr>
<tr>
<td>following /t/ in word</td>
<td>yes/no</td>
</tr>
<tr>
<td>preceding /d/ in word</td>
<td>yes/no</td>
</tr>
<tr>
<td>following /d/ in word</td>
<td>yes/no</td>
</tr>
<tr>
<td>MORPHOLOGICAL</td>
<td></td>
</tr>
<tr>
<td>variable at affix juncture</td>
<td>yes/no</td>
</tr>
<tr>
<td>PHONETIC</td>
<td></td>
</tr>
<tr>
<td>speech rate</td>
<td>0.433–16.0 sylls/second</td>
</tr>
<tr>
<td>LEXICAL</td>
<td></td>
</tr>
<tr>
<td>number term</td>
<td>yes/no</td>
</tr>
<tr>
<td>log word frequency</td>
<td>0–9</td>
</tr>
<tr>
<td>word age (ratio CC/IA)</td>
<td>0.35–3.37</td>
</tr>
<tr>
<td>(CC only)</td>
<td></td>
</tr>
<tr>
<td>DISCOURSE</td>
<td></td>
</tr>
<tr>
<td>repetition</td>
<td>not repeated (first mention)/repeated (within 60 secs)</td>
</tr>
<tr>
<td>topic</td>
<td>distant past/recent</td>
</tr>
</tbody>
</table>

Table 2. Factors coded in analysis of data set.
recordings were largely monologues, with limited involvement of other speakers. The sixty-seconds constraint was in order to retain a sufficiently large number of tokens in each category. (We trialed statistical models with longer and shorter time frames. Results were similar, but levels of significance weakened with smaller numbers of tokens in one or the other category.)

Topic codes were distilled from the wide range of codes available. The topic coding was jointly developed by Abby Walker and Jen Hay as part of the ONZE project. A number of topic-based codes were added to the data manually by trained research assistants. We examine here the effect of one particular topic: the time period the speaker is talking about. The original coding for this topic had a number of fine-grained time-based categories (‘last week, last month, last year’, etc.). Coding was done on general topic chunks, and coders were instructed not to be too specific or to switch the topic for every utterance or clause, but rather to try to capture the overall topic being discussed. Comparisons between ‘the old days’ and ‘now’, for example, were usually given a code that reflected the distant past, since this was usually the most informative part of the comparison, and—in this context—discussion of the present was often only relevant in that it was able to shed light on the past. The coding was done with reference specifically to the time of the speaker’s experience of the topic being discussed, not the experience of any third party being discussed. For example, if a speaker was relating a story about their friend’s childhood that they clearly learned about in the previous week, this would be coded as ‘last week’ (because it is recent for the speaker). The codes referring to more distant times were reserved for cases in which speakers appeared to be relating a memory or event that was associated with their own distant experience. Given the concentration of personal narrative in the corpus it is not surprising that topic codes mostly referred to the speaker’s own experience (64%) or that of friends and family (17%). Only 7.8% was classified as ‘other’, that is, referring to the experience of unknown or unfamiliar people.

In our data set many of the original topic codes had very few observations. For the purposes of our analysis we therefore collapsed topic into a binary category for which there are a large number of observations in each category and which also enabled us to contrast the speakers’ experience over different time points in their lives: speech referring to a decade or more ago (‘distant past’) versus speech referring to a more recent time or, in a few cases, the future (‘recent’).

2.3. Summary of predictions. To summarize our predictions for the data set, we expected to find more of the innovative variant, [d], for younger speakers, men, nonprofessionals, in faster speech rates, in repetitions, recent events, high-frequency words, and ‘young words’. To establish that frequency is involved in the progression of the change toward [d] we also predicted an interaction between year of birth and frequency, with effects of frequency stronger for younger speakers.

2.4. Statistical analysis. Binomial mixed-effects models were hand-fitted to the entire data set with the lme4 library in the software package R (Bates et al. 2011, R Core Team 2012), using the bobyqa optimizer. The dependent variable was a binomial factor distinguishing between voiced and voiceless variants. Voiceless variants were set to the default, meaning that presented models are always modeling the log odds of a [d] production. Levels of collinearity between the continuous predictors in the overall data set were checked using the collin.fnc function from the languageR library and were found to be nonproblematic (condition number < 7). In each model, word and speaker were treated as random intercepts. The large number of factors explored precluded testing of all pos-
sible interactions, due to data sparseness and convergence issues. Our model fitting was therefore hypothesis-driven, as follows. We identified our ‘test’ predictors, relating to our main hypotheses, as those appearing in the ‘social’, ‘lexical’, and ‘discourse’ categories in Table 2. The ‘control’ predictors, not relating to our primary hypotheses, fall in the ‘phonetic’, ‘phonological’, and ‘morphological’ categories. We started with a model incorporating all independent factors listed in Table 2 and all two-way interactions between test predictors—that is, we tested for interactions between all factors listed in the ‘social’, ‘lexical’, and ‘discourse’ categories in Table 2. We then pruned the model, removing non-significant factors. Factors involved in significant two-way interactions were retained. We then tested for the significance of three-way interactions involving all factors that had been retained in two-way interactions. Model selection was guided by $\chi^2$ likelihood tests, Akaike information criterion (Akaike 1974), and Bayesian information criterion (Schwarz 1978). A final model was settled upon for the entire data set, as discussed in §3 below. This model was subsequently tested on subsets of the data containing the MU and CC speakers, respectively, for reasons explained in §3.2. These subset models were further pruned using the same procedure. Note that all figures presented below retain the same scale on the y-axis for ease of comparison.

3. Results.

3.1. Full data set. Figure 1 provides an overview of variant usage by individual speaker across the entire corpus. Trend lines by group (subcorpus, sex, and social class) are also shown. An element of care should be exercised in interpreting Fig. 1, as the number of tokens per speaker is highly variable (mean = 27.3, median = 18.5, range = 4–122). Furthermore, it does not take into account the various other factors tested in the statistical modeling. It does, however, provide a useful snapshot of the broad patterns in the data. First impressions suggest clear differences between the MU and CC subcorpora. The MU speakers show a marked separation by sex, with the men using far more [d] than the women. There is no clear evidence of patterning by age within this subcorpus. The male trend line is flat, and the apparently rising female one appears to be driven largely by the one woman who uses a high rate of [d] (over 60%, although note that this reflects 8/13 tokens). There is also considerable variation within the male group, with speakers varying across the full range from 0% to 100% [d]. The CC speakers, like the MU speakers, show clear differences by sex and also by class. In contrast to the MU group, they also show evidence of change: in each of the four sex × class groups the trend is rising, with younger speakers using more [d]. The distribution of individuals within groups indicates that the trend lines are more meaningful than the one for the MU women. Fiasson (2013) investigated patterns for the same variable in a different subset of ONZE, including speakers born between 1900 and 1930. His results are consistent with ours for the MU and CC speakers. Those born 1900–1930 use around 25% [d] overall, varying in line with sex and social class.

The binomial mixed-effects model fit over the entire data set is shown in Table 3. This table of coefficients shows (in the ‘estimate’ column) the effect of each independent variable on the log odds of [d]. It contains noninteracting main effects of speaker sex and speech rate. The model confirms the impression from Fig. 1 that males produce significantly more [d]. This can be seen by inspecting the coefficients for sex, which show that if the speaker is male the log odds of [d] increase (i.e. the coefficient is positive) by 2.335. The significance of this effect is shown in the rightmost column. With respect to the control predictors, there is a significant dissimilatory effect of a previous /d/ in the word (e.g. daughter, dating). That is, a previous /d/ reduces the likelihood of
The evolution of medial /t/ over real and remembered time

A [d] realization of the medial /t/. The coefficient for speech rate shows the rate at which the log odds of [d] increases as speech rate increases. The faster the speech, the more [d] is predicted.

All other factors in the model enter into significant interactions. Figure 2 plots the significant interaction between whether the word form is repeated and its lexical frequency (i.e. the effect shown in the bottom row of Table 3). The top panel shows the probability of [d] as predicted by the statistical model, holding other factors constant. For ease of interpretation, we convert log odds into probabilities in all of the figures showing model predictions. The bottom panel shows the trend as it exists in the raw data. The same overall pattern is visible in both plots. Note that the absolute scale cannot be compared directly across the two plots. This is because the model predictions show predicted values assuming a particular speaker and linguistic context. The raw data show actual proportions as averaged across all speakers and contexts. Thus in this pair of plots (and also in Figs. 3 and 4 below) what should be compared are the relative patterns in the plots, rather than the absolute scales.
What we see here is that there are effects of repetition and of frequency, but that these are not independent of one another. Rather, only nonrepeated words are affected by frequency; that is, the frequency effect is manifested only in nonrepeated words. Repeated words are not affected by frequency and are predicted by the model to act like frequent words. So, the lowest rates of predicted [d] are seen only in the case where the word is relatively low frequency and where it has not been used previously within the last minute. Put another way, words that are relatively predictable in context have higher rates of [d]. This predictability may come from the local context (i.e. repetition in discourse) and/or high lexical frequency. Words that are not predictable—because they are low frequency and have not been mentioned previously—have the lowest rates of [d]. This shows that while word frequency is important, it cannot be considered independently of other factors that affect predictability (such as repetition). Post-hoc regression modeling confirms that the effect of word frequency remains significant if tested within the set of nonrepeated words alone.

![Figure 2](image)

**Figure 2.** The interaction between log lexical frequency (which has been centered) and whether the word form is being repeated. The top panel shows the model predictions (holding other factors in the model constant). The bottom panel shows the trend as it exists in the raw data, by dividing the frequencies into bins 0.5 wide and showing the proportion of [d] productions within each bin for repeated and nonrepeated words. Lowess lines are fitted through these points. Note that the raw data are inherently noisy, because they average over uneven distributions of many things that influence the variable (e.g. different speech rates, sexes, ages, etc.). The overall trend is nonetheless still visible.
Figure 3 shows the significant interaction between year of birth and lexical frequency. If, as has often been claimed, frequent words tend to lead leniting changes, then we expected to see frequent words progressing faster in the change than infrequent words. This is indeed exactly what the model shows. The top panel of Fig. 3 shows the model predictions for three different word frequencies—the upper quartile, the median, and the lower quartile—holding other factors constant.

The factor ‘repetition’ is fixed to nonrepetition—that is, predictions are shown for words that are not being repeated. However, the year of birth × frequency coefficient is constant, meaning that—regardless of repetition—low-frequency words do not change with year of birth and high-frequency words change fastest.

The bottom panel shows raw data for low-frequency (below median) and high-frequency (above median) words. It can be seen clearly that the high-frequency words are leading the way in the change. The lowest-frequency words, in fact, do not show any increase in [d] over the time course of the change. Thus, for increasingly younger speakers we have a strong difference between high- and low-frequency words. This type of difference has often been reported for leniting changes (§1.2). The important aspect of the (-t-) data set is that it shows that the frequency effect was not present for the oldest speakers. Thus we can show that, not only are frequent words ahead in the change, but they are also leading the change.

As can be seen from Fig. 3, the change has progressed over the 120 years covered by our data. It is important to keep in mind, however, that the year-of-birth effect, although modeled as continuous, comes from two disjoint sets of speakers. The lines in the top panel of Fig. 3 thus traverse the middle period that links our two data sets, and no data are in fact included for the years 1900–1932.

Repetition also interacts with topic. As was seen in Fig. 2, there is a significant interaction between lexical frequency and repetition. But we also have an added interaction between repetition and topic. These two effects are shown together in Figure 4. As has already been seen, there is a frequency effect, which exists for nonrepeated forms only (top left panel). If a form is being mentioned for the first time, this frequency effect dominates, and there is no extra visible effect of topic. The top right panel of Fig. 4 shows repeated forms. Repeated forms that relate to recent events act just like frequent words: they have a high rate of [d]. However, the predicted probability of [d] for discussion of distant events is markedly lower. For repeated forms, distant events are associated with significantly lower rates of [d] than recent events. Post-hoc regression modeling confirms that the effect of topic remains significant if tested within the set of repeated words alone.

In addition to the full analysis, we also conducted separate analyses of the two subcorpora. There were a number of reasons for this. First, year of birth is modeled as continuous in the overall data set, but as already noted we are in fact dealing with two groups of speakers with some gap between them, and the cursory inspection afforded by Fig. 1 suggests the two subcorpora display different patterns. It therefore seemed desirable to better understand what was happening within each subcorpus, representing the earlier and later periods of the change. Second, the CC is stratified by social class, but this information is unavailable for the MU speakers. In order to test for an effect of social class, we needed to run a model on the CC only. Finally, we predicted an effect of ‘word age’, but this can only be tested for the CC. This is for two reasons. It is clear from the overall model that the oldest speakers have not experienced as robust an association between age and variant use as the speakers born after 1930, so it is for the more recent speakers that we expect any effect of word age to be observable. Furthermore, it
is also for the CC speakers that we have the best means of assessing the likely distribution of particular words over younger and older speakers. We do not have any recordings of speakers older than those in the MU corpus, so cannot assess the statistical distribution of word age for the MU speakers.

3.2. Mobile unit and Canterbury corpus analyses. We fitted a model of the MU speakers by starting with the final model from the full data set and then pruning out nonsignificant predictors. Table 4 shows this model. Only one factor emerged as significant: speaker sex. Even the effect of speech rate, which is very robust over the whole data set, does not reach significance within the MU alone. In this subcorpus, then, we are apparently dealing with an emergent change. The variation between [t] and [d] is already associated with some social meaning—males produce significantly more [d] than females. However, the remaining variation (and there is a lot of variation, shown clearly in Fig. 1) is at this stage apparently rather unstructured, even chaotic. It does not show the classical hallmarks of change in progress, which is usually characterized by complex conditioning by both linguistic and social factors.
The evolution of medial /u/ over real and remembered time

The CC model stands in marked contrast, as shown in Table 5. The CC analysis tests two new factors that could not be assessed in the overall data set: social class and word age. We added these as independent predictors in the model and also tested interactions between word age and the other test predictors. A wide array of factors emerged as significant for this group of speakers. Nearly all of the factors that were significant in the overall data set (Table 3) remain significant. One exception is log frequency × year of birth. Since this effect is clearly driven by the large increase in the effect of frequency between our MU and CC speakers, the omission of the earlier speakers eliminates this
interaction. The other exception is the effect of a previous /d/, which did not reach significance. The interaction between repetition and frequency remains robust and shows the same patterning as displayed in Fig. 2. The interaction between repetition and time also remains robust and shows the same patterning as displayed in Fig. 4.

The three new significant main effects are as follows. First, social class, as expected, is robustly significant: professionals use significantly fewer [d] variants than nonprofessionals, although the effect size is considerably smaller than that seen for sex. Second, year of birth is significant within the CC, younger speakers having higher rates of [d]. In the overall model this main effect was not significant, as the MU speakers showed a high degree of variation and no clear direction of change. It is only for speakers born after 1930 that the change toward [d] has progressed. Third, we also observe the predicted effect of word age: ‘younger words’ are more likely to be produced with [d] than ‘older words’ (Figure 5). This effect is separate from that of topic, and it does not interact with other factors. It is important to note that word age is not problematically confounded with topic. While there is a slight tendency for talk about the past to use older words (distant past: mean word age 1.13; recent: mean word age: 1.17), this difference is not significant.

![Figure 5](image)

**Figure 5.** Effect of word age on predicted probability of [d]. ‘Young words’ (i.e. words overrepresented in younger speakers) are associated with higher rates of [d].

It should be noted that this effect of word age is not apparent in simple plots of the raw data, which are therefore not shown here. It is a minor effect, especially in compar-
The evolution of medial /t/ over real and remembered time

3.3. Meaningful differences or sample-size differences? The MU data, representing a stage close to the beginning of the change, seems to show remarkably little structure. Even effects commonly associated with stable leniting variables, such as speech rate and lexical frequency, do not reach significance. The CC data, by contrast, show very fine conditioning, with variation based on speaker, topic, word, and context. So, are we seeing complex patterning emerging over the course of the change? This seems possible. However, there is an important difference between the MU and CC data sets: the number of observations. The total of 833 tokens in the MU equates to 45% of the size of the CC data set (1,846 tokens). Is /t/-realization in the CC subject to more complex conditioning? Or is it simply that we have a larger data set, in which significance can emerge more robustly?

In order to answer this question we further investigated the CC data set in order to determine how sensitive our observed effects are to sample size. From the CC we randomly selected 1,000 subsets of the data, each of which was equivalent in size to the MU data set. We then ran the exact model reported in Table 5 on each subset of the data.

In Table 6 we report, for each factor in our reported CC model (Table 5), the proportion of these 1,000 reruns in which that factor reached significance, as assessed by the simple criterion of \(|z| > 2\). This approach enables us to establish a degree of confidence in terms of which factors pattern in a meaningfully different way in the CC compared with the MU. For the factors that are not significant in most reruns on the smaller data sets we cannot state with confidence that they were also not relevant during the early period. Their lack of significance in the MU data set could simply be due to the smaller token count in that data set.

While Table 6 gives the proportion of models in which each factor reaches significance, it is also instructive to consider the distribution of \(z\)-scores returned for each factor. This is shown in the form of boxplots in Figure 6. From this we can see, for example, that although both sex and class reach significance 100% of the time, sex has a uniformly stronger effect than class (i.e. it yields a higher \(z\)-score in the statistical models). Furthermore, while word age is significant in only 38% of the models, the center of the distribution, while not above 2 (generally taken as the threshold for significance), is nonetheless close to it.

For factors that are significant in a minority of runs of the smaller models (e.g. word age), there are two possible interpretations about why this could be so. One is that the effect is not real, is present by chance in the full CC data set, and does not survive removal of crucial tokens that appeared to be supporting it. The other is that it is a real effect, but it is not as strong as other effects in the model and needs a sufficiently large data set in order for there to be enough power for it to emerge as significant.

The reduction in sample size is likely to be particularly harmful for the interactions. Consider, for example, the interaction between time and repetition. Figure 3 shows that words that are repeated, and that are being used to discuss the distant past, have an overall lower level of [d]. Tokens that fall into this category constitute about 8% of our total data. In a data set of 1,846 tokens this is still a reasonable number of tokens. If we cut the sample back radically by randomly sampling a subset, however, then there is a good chance we might end up with an insufficient distribution of tokens in the subset to show
the interaction (for example, there might be very few repeated tokens occurring in talk about the distant past). In the case of the two significant interactions reported in our overall model, each reaches significance in over 500 of the 1,000 reruns. We therefore interpret this outcome as a real effect, but an effect that is sensitive to the size of the data sets.

Certainly what this validation exercise does confirm is that the absence of expected effects such as lexical frequency and speech rate in the MU data is not simply due to its smaller sample size. If effects similar to those found in the CC had been present in the MU, we can be confident that our sample of 833 tokens should have been able to detect them.

3.4. Summary of findings. Our predictions for the data set (listed in §2.3) were all supported. We found higher usage of [d] for younger speakers, men, and nonprofessionals. As befits a lenition process, [d] was also used more in faster speech rates and repetitions. The main hypotheses were borne out, with more [d] in words used more by younger speakers, in discussions of recent events, and in high-frequency words. The predicted interaction between year of birth and frequency was also robust in the analysis, with effects of frequency stronger for younger speakers.

4. Discussion. Overall, the data set shows that there were high rates of [d] in NZE from the very earliest stages. That the voiced variant was used by speakers born in the
The evolution of medial /t/ over real and remembered time

nineteenth century provides counterevidence to previous assumptions that this variant was imported wholesale from the United States (Holmes 1994:218, 1995:66, Wells 1982:250). It does not, of course, rule out the possibility that increasing prestige associated with American accents may have triggered or accelerated the move from the relatively chaotic early variation to the later pattern showing increased rates of voicing and the highly structured variation typical of change in progress. There has certainly, over this time period, been an orientation away from Britain in terms of a prestige norm, and increasing orientation toward American English (Holmes 1995, Vine 1999). While this re-orientation is not responsible for importing the variation in medial /t/ to New Zealand, it may nonetheless have been responsible for influencing social evaluation of the variant, and in turn the way in which the change has unfolded. It is also important to keep in mind, however, that there are other potential external influences that could be at play, most notably Australian English, which also shows high rates of /t/-voicing (Horvath 2008).

We cannot know the exact mechanisms involved in the shift from the early largely unstructured patterning to the highly structured later variation. However, there are a number of surprising elements about the trajectory. One is the apparent absence of typical lenition-based effects in the early data. There is no effect of word frequency, for example, nor is there a speech-rate effect. The analysis in §3.3 shows that the size of the data set should have been big enough to pick these up, at least if the size of the effect were comparable to that observed in the contemporary data. This pattern rules out a trajectory in which a latent lenition-based process gradually acquires momentum and social meaning. Socially correlated patterning appears to be there from the very beginning, in that we observe a strong sex effect. The lenition-based factors only emerge later. It was, of course, one of our central predictions that the frequency effect should increase over time. It is nonetheless surprising to observe that it was actually completely absent in the early speakers.

The time depth of the change thus provides new evidence about the history of medial /t/ in NZE, showing that it is not a variable adopted from American English, as previously thought. More important for the current article, however, is the very subtle patterning underpinning the subsequent variation, which shows the active involvement of word-specific distributions in the course of the sound change and the important influence of contextual factors. We thus now turn to a discussion of what we believe are our most important results, organized according to our three initial predictions.

4.1. Prediction 1: frequent words lead leniting change. We predicted that word frequency should contribute to sound change in progress. Specifically, in the context of a lenition change, we expect frequent words to lead the change. As outlined in §1.2, in order to demonstrate definitively that frequency contributes to change, we predicted an interaction between word frequency and speaker age, with increasing effects of frequency as the change accelerates toward its peak. In this data set we do indeed see the predicted interaction. We have a robust frequency effect that emerges over the course of the change. From this we can infer that frequency is integral to the transmission of this leniting change, in support of claims made by Bybee (2000), inter alia. To our knowledge this is the first time that such an interaction has been shown.

We should note that in predicting an interaction between frequency and speaker age we do not mean to imply that frequent words should be moving uniformly faster at all stages of a change. While frequency effects are predicted to accelerate as a change progresses, this must naturally also reverse if a change starts to reach completion. For a change nearing completion, it may be possible to observe a period where high-frequency words show little movement (because they are at the end of the ‘S curve’), whereas low-
frequency words show more (because they are lagging behind, and so still have scope to continue to advance before reaching the end of the change). In the time period we are studying, our change does not near completion for the community as a whole, although the young nonprofessional males are approaching categorical usage of [d], as noted by Holmes (1994). We therefore see a steeper slope for frequent words than infrequent words. However, at some stage in the future, should the change approach completion, we expect to see a period where frequent words are uniformly produced with [d] (and so have a zero slope), whereas infrequent words are still changing from [t] to [d] (and so have a positive slope). If frequent words ‘lead sound change’, then we expect to see interactions between year of birth and lexical frequency, but the nature of the expected interactions depends on the time course of the change.

A second caveat regarding our frequency effect is that it exists in interaction with a repetition effect. In our data, repeated words tend to act like frequent words. The effect of lexical frequency is therefore not uniformly present in our data, but is most strongly observable in words that are being used for the first time in the local discourse context. Other work has also shown interactions between second-mention reduction and frequency, albeit on a different type of data and phenomenon, and patterning somewhat differently (e.g. Baker & Bradlow 2009). What is clear is that frequency cannot be considered in isolation, and that it patterns in complex ways with other factors. Frequency, then, is related to, but not the same as, predictability in context. In cases where it is of theoretical importance to establish whether frequency effects do or do not exist in a data set, it is therefore important to consider frequency as part of a wider set of constraints that can affect production patterns (cf. Labov 2010). Finding that frequency effects are small relative to other constraints, or that they occur only in interaction, should not be interpreted as indicating that frequency is irrelevant in explaining change.

The result that lexical frequency is leading our lenition-based change is compatible with the model outlined by Pierrehumbert (2001). In her model there is a slight articulatory bias toward lenition applied to every production. Frequent words undergo this bias more often than infrequent words, and so run ahead in the change. Sóskuthy (2013) rightly argues that such a model, if left unchecked, would predict that all phonemes should progress rapidly toward lenited variants, yet such rampant universal change is not observed. The model thus has to be considered within the context of a wider system in which various constraints can keep articulatory biases in check. These include constraints on contrast maintenance and perceptibility as well as social constraints. Such biases do not always lead to change, then, but when they do, a frequency effect is predicted.

Finally, we note that while the observed frequency effect is compatible with Pierrehumbert’s (2001) model, there may well be other mechanisms through which such an effect could arise or be strengthened. Seyfarth (2014), for example, argues that word-specific effects of predictability are observable. Words that are predictable in context are more reduced. In exemplar theory, argues Seyfarth, we should expect to see word-specific reflexes of this effect that are independent of the local context: words that are more often predictable will more often be reduced. This will affect the content of the exemplar store, and so we should observe words that are very often predictable to be more reduced even in low-predictability contexts. This reasoning follows exactly the same structure as the reasoning used to predict the word-age effect in our own data set (see §4.2 below). Indeed, Seyfarth finds that words that are commonly in high-predictability contexts are in general more reduced across all contexts. While Seyfarth does not explicitly discuss the consequences for frequency effects, it is reasonable to infer that high-frequency words will more often be predictable than low-frequency
words and so would—through this mechanism—come to have more reduced representations. Thus, some of the observed frequency effects may come from the different contextual pattern of high-frequency and low-frequency words, rather than increased numbers of cycles through an articulatory bias. These two hypothesized mechanisms for frequency-based effects in leniting changes need not be mutually exclusive and, indeed, may mutually reinforce each other.

Finally, we note that although our results show a robust effect in which frequent words lead, we do not predict that such an effect should be present for all types of sound changes (as discussed in detail by Phillips 2006). There may be cases where other predictors are so overwhelmingly strong that no frequency effect is observed. Alternatively, the change may be driven by factors that do not involve any systematic articulatory bias at all (cf. the very different frequency effects reported for a chain shift in Hay et al. 2015). Frequency effects are predicted in usage-based models, but they are not predicted to look the same across all types of change or all stages of change.

4.2. Prediction 2: words used more by younger speakers are produced with more new variants. Our second prediction was that ‘word age’ would contribute to sound change in progress. We predicted that words that are more commonly used by younger speakers will be more advanced with respect to the change, containing higher incidence of the innovative variant. This trend was indeed present in the data. The prediction follows from work by Walker and Hay (2011), which shows that words used more often by younger speakers tend to be more easily recognized when produced in a younger voice. Walker and Hay interpret their results as showing word-level representation of the distribution of voices in which a word is encountered. Here we show that this effect can be observed in production as well as perception.

It is important to make clear that what we are seeing here is an effect of the distribution of encountered variants at the word level. It is not a more general effect of topic. It is possible that older words are more often used to refer to older events, but word age and topic are accounted for separately in our model (see discussion in the next section). We can therefore be confident that the word-level effect is not an artifact of the topic effect.

This effect of the population distribution of individual words has been foreshadowed by Gordon and colleagues (2004) and Yaeger-Dror (1996). For example, Yaeger-Dror (1996:282) argues that in a sound change in progress in Montreal French, more conservative pronunciations were apparent in words that had connotations of ‘the olden days’. She argues that the diffusion of sound change might involve ‘a culturally determined cognitive semantic/associative network’. Similarly, Gordon and colleagues (2004) observe that the loss of rhoticity in early NZE appeared to proceed more slowly in words that were associated with early settler life. In both of these cases, ‘old-timey’ words were subjectively identified by the researchers.

Our data set builds on this work in two ways. First, it incorporates this kind of effect in a larger, more statistically complex model, in order to assess its robustness in the face of the many other factors that are affecting variation. Second, and more importantly, it moves beyond consideration of words that may overtly and/or subjectively have old-timey associations, to consider instead the overall population distribution of words. Walker and Hay (2011) showed in their perceptual work that the population distribution of words itself affected people’s production, and that their effect was not driven by overt semantic associations. It is the same type of effect that we are showing here. While some words that are more frequently used by older speakers may also have old-timey associations, most do not. The population distribution of words affects represen-
tation, which in turn affects perception and—we have now shown—production. Overt knowledge of the population distribution is not necessary for such effects to occur.

We reiterate the caveat given in §§1.3 and 4.1, however, that such effects are not necessarily predicted for all types of sound change. In particular, there are some types of sound change (and/or stages within sound changes) where some individuals do not use the full repertoire of potential variants in their own production. We would therefore predict that this kind of effect should be more robustly observed for perception than for production. However, this data set shows that it is certainly possible for it to surface in production as well. It is not a strong effect, compared to some of the other predictors in our model, and so may not be present or observable in all contexts. It is not easily visible in the raw data, but it can be seen in our statistical model, once other—larger—effects are being held constant. It is also clear that a reasonably sized data set is needed for the effect to emerge as significant, as the analysis of subsets of the data in §3.3 reveals that the result is not robust once the sample size is considerably reduced. Exactly how widespread such effects are, and the types of variables and contexts in which they can surface, remains an open question for future work.

4.3. Prediction 3: Speech about older events elicits older variants. Our final prediction was that speech production, if driven by a changing exemplar store, may reflect sound change in progress. We addressed this prediction by examining discussions of different time points in speakers’ lives. We found evidence to support the prediction, in that the new [d] variant was used more frequently in discussions of recent events, with more [t] in discussion of the distant past. However, the effect of time depth was visible only in interaction with repetition, the difference between recent and distant events emerging in subsequent mentions. When a word was first mentioned no such effect was found. The interaction was robust, being significant at p < 0.005 in the model of the whole CC data set and replicated in 53% of the 1,000 applications of the model to a smaller subset.

The effect is clearly consistent with the idea that ongoing learning is reflected in lexical representations. Explaining it, however, is not straightforward. At least three explanations are possible to account for the time-depth effect, singly or in combination.

First, the simplest account is that the time-depth effect reflects remembered time. The store of exemplars has a broadly chronological history, and its structure is therefore archaeological, offering direct information about the changing distribution of variants over time. This structure is undoubtedly mediated by the effects of attrition and other factors that affect memory, and thus the exemplar store should not be conceived of as a set of frames taken from a long time-lapse movie. It will, however, broadly reflect the changing composition of linguistic experience through the distribution of exemplars: exemplars committed to memory early in life contain more [t], while more recent exemplars contain more [d], both reflecting the changing distribution of variant usage by speakers in the community as a whole.

A possible mechanism underpinning the activation of older variants for older stories relates to the coactivation of linguistic and nonlinguistic memories. Accounts of autobiographical memory distinguish between episodic memories for past events and more abstracted, semanticized memories—broadly what is ‘remembered’ versus what is ‘known’ about the past (Tulving 2002). Accessing episodic memories ‘entails a vivid sensory-perceptual reexperiencing of the event, including first-person perceptions, thoughts and emotions that accompanied the original experience’ (Prebble et al. 2013: 818). Linguistic episodic memory is generally believed to be linked or indexed to social and contextual information (see review in Foulkes & Hay 2015). It is thus reasonable to
assume that when past nonlinguistic memories are accessed—memories relating to the relevant event, the context, and the identity of the speakers—these may in turn increase activation of the linguistic representations or memories with which they are associated. This would lead to an increased proportion of older variants being activated during discussion of older events. Such an account would predict that talk about the past indicating increased episodic recall would show the linguistic effect most strongly.

A second possibility is that the effect of time depth in our data set might reflect variable issues of speaker identity. Some recent work emphasizes shifts in identity that may be associated with particular topics, showing effects of shifting to a topic associated with particular identities (e.g. more ‘local’ neighborhood talk, more talk associated with particular cultural behaviors such as fighting; Eckert 2000, Becker 2009, Lawson 2009, Love & Walker 2013). In discussing older events speakers may portray themselves as historical versions of themselves. In doing so they may draw on linguistic resources to convey that identity. Experience of sound change in progress may yield an abstract understanding of the relationship between speaker age and variant type, similar in kind to the associations people learn about variant distributions across dialects of a language (e.g. that postvocalic [r] is characteristic of North American varieties of English). The development of such an understanding at an abstract level is presumably derived from analysis of specific experiences and could coexist with detailed information stored in episodic form. Based on a general awareness that variants are associated with particular age groups, speakers might invoke a greater use of [t] to index a person born some time ago, while more [d] signals a more contemporary identity. Labov (2010:344) presents a similar proposal to account for how younger speakers manage to participate in ongoing change, pushing indexical patterns further along a statistical trajectory. Younger speakers come to recognize the statistical correlations between variants and speaker age and position themselves accordingly in line with this trajectory. In our /t/ data we may be seeing the inverse of this process, with speakers aligning themselves with historically earlier speakers in order to project discussion of the distant past.

A third possibility is that our time-depth effect is explained as a factor of audience design. Bell (1984) offers a number of testable hypotheses about the effect of audience on speech style. One such hypothesis is summarized as follows:

speakers associate classes of topics or settings with classes of persons. They therefore shift style when talking on those topics or in those settings as if they were talking to addressees whom they associate with the topic or setting. Topics such as occupation or education, and settings such as office or school, cause shifts to a style suitable to address an employer or teacher. Similarly, intimate topics or a home setting elicit speech appropriate for intimate addressees—family or friends. (Bell 1984:181).

The examples Bell gives are associated with shifts in formality across different topics, and indeed this is the type of topic-based style shifting that has been reported most often (see e.g. Rickford & McNair-Knox 1994). His hypothesized effect in fact follows directly from the notion of experience-based lexical representations. We assume that episodic memories are indexed to social factors and contextual factors, potentially including speaker identity and topic. These socioindexical associations will be active in speech production and influence the selection of target forms. There is rich evidence for effects of speaker convergence in interaction (e.g. Giles et al. 1991). Thus, if particular topics tend to be discussed with particular addressees, then forms associated with that topic will have a phonetic profile that is biased toward those addressees. In future interactions, when the same topic is discussed (and so activated), the topic’s phonetic profile could then be one of the many factors that influence production. We might explain the time-depth effect in our data set in similar terms, if older events are more frequently dis-
cussed with a subset of people for whom a particular distribution of phonetic forms is found. In our case, for example, if older events are more typically discussed with older people, the speaker may accommodate to the older audience by using variants characteristic of that audience, in this case using more [t] in line with the observable distribution of variant usage by age. Of course, we have no direct evidence that this is the case here, since our recordings do not involve speakers with different interlocutors. If supported, however, this explanation would invoke some form of tagging of exemplars without necessitating time depth and without assuming direct access to knowledge reflecting lifelong learning at the level of exemplars.

Also to be explained is why the effect of time depth should emerge only in interaction with repetition. Such an effect is consistent with claims that frequent words can act as the ‘locus of style’. Hay and colleagues (1999:1391), for example, claim that it is in ‘highly frequent words that a speaker finds the crucial combination of a) ease of processing … and b) the repeated opportunity of presentation that would be needed to layer new grammatical or social meaning’. Both of these factors certainly apply also to repeated words—possibly even to a greater extent than for frequency. There are also previous reports of ‘easier’ words displaying more indexical information (Munson 2007, Clopper & Pierrehumbert 2008, Mendoza-Denton 2008, Scarborough 2010). Repetition is not investigated by these authors, but these studies do seem to show that words that are in some sense easier (via high lexical frequency, low neighborhood density, or semantic predictability) are more prone to display indexical variability. Munson (2007) and Clopper and Pierrehumbert (2008) both identify two possible sources of such an effect. One relates to the need for successful communication. In easy words, the talker can ‘afford to index social information to the listener without significantly affecting intelligibility’ (Clopper & Pierrehumbert 2008:1687). This explanation would certainly extend to our results, if we assume that the topic-based effect relates to some form of socioindexical meaning.

A second possibility is that topic effects may relate to the time course of processing and access. There are documented time-course effects in lexical processing, where indexical effects emerge in tasks where the time course of processing is slow, but they are not apparent in contexts where processing occurs rapidly (McLennan & Luce 2005). In other words, in perception, speeded or fast tasks do not seem to show indexical effects to the same degree as tasks that proceed more slowly. It is possible that related time-course effects might also be observed in speech production, although here the prediction would go in the opposite direction. In perception, fast access of lexical items does not allow time for activation to spread to indexical features. In production, however, words that are easily accessible may allow more time for subsequent activation and influence of indexical characteristics (this possibility is also discussed in Munson 2007 and Clopper & Pierrehumbert 2008). A possible interpretation of such effects in terms of episodic representations would be as follows. Speech production involves selecting a word form to produce and then choosing a contextually relevant subset of exemplars over which to generalize in order to create a production target. On the first production of the word, the selection of the word form is slower, and selection of the contextually relevant part of the distribution occurs quickly and is somewhat crude. On subsequent productions, the word form itself is primed and quickly accessed, and thus contextual factors may play a greater role. This would be because there is more time available for activation to spread to the most contextually relevant part of the distribution, and also because the relevant exemplars have been somewhat boosted from the previous repetition. Under such an interpretation each subsequent repetition, within a given interac-
tion, may show greater specificity and indexicality (assuming lexical selection is further facilitated by further repetitions). It should be noted that such an explanation is highly speculative, because it is based on several layers of assumptions beyond documented time-course effects. We cannot know for certain that such effects extend to production in this way, nor that they are relevant to the type of socioindexical information we are dealing with, which is much more richly contextual than the factors manipulated in the experimental literature.

In sum, our investigation into a possible source of time depth of discussion has revealed a very interesting effect, which warrants further investigation. The interaction with repetition serves to make it even more intriguing. Our current data set does not enable us to disentangle the exact cause or causes of the observed effects. The mechanisms we discuss are not, in fact, mutually exclusive, and we suspect more than one may be at play at once. For example, the effect of topic could be driven by both an element of identity work and a more automatic effect of accessing ‘layers of memory’. Both of these could be working together to facilitate the emergence of the effect in repeated words only. Our finding does suggest, however, that future experimental work designed to investigate these issues could bear considerable fruit.

5. Conclusion. This study has provided a comprehensive analysis of sound change in progress, using data from a 120-year span. Over this time period, New Zealand English medial /t/ evolves from relatively chaotic and unstructured variation, patterning only with respect to speaker sex, into a highly structured and intricately conditioned sociolinguistic variable. The course of this trajectory is shaped by a wide range of social, linguistic, lexical, and discourse-level factors. Three of these effects are of particular importance. First, there is an effect of word frequency, which gradually strengthens throughout the course of the change. In showing a robust interaction between word frequency and speaker age, this study is the first to show unequivocally that word frequency is involved in the progression of a leniting change in progress. Second, the population distributions of individual words affect the nature of their participation in the change. An individual who encounters a word very frequently with innovative variants is more likely to produce that word subsequently with an innovative variant. This finding clearly shows that the phonetic distributions of word-level representations are implicated in the course of sound change. Taken together, these two results provide strong evidence in favor of models that contain experience-based representations at the word level. Finally, we show that the topic of conversation affects which variant is favored. Older topics elicit older variants. Thus, in addition to demonstrating the history of a variable as it unfolds over 120 years, we have also demonstrated that this history is at least partly represented within individual speakers and can be observed in their discourse as they traverse apparent trajectories of ‘remembered time’.

REFERENCES


Gordon, Elizabeth; Lyle Campbell; Jennifer Hay; Margaret Maclagan; Andrea Sudbury; and Peter Trudgill. 2004. New Zealand English: Its origins and evolution. Cambridge: Cambridge University Press.


Harrington, Jonathan; Sallyanne Palethorpe; and Catherine Watson. 2000. Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen’s Christmas broadcasts. Journal of the International Phonetic Association 30.63–78. DOI: 10.1017/S0025100300006666.


Hay, Jennifer; Janet Pierrehumbert; Abby Walker; and Pat LaShell. 2015. Tracking word frequency effects through 130 years of sound change. Cognition 139.83–91. DOI: 10.1016/j.cognition.2015.02.012.


Howell, Peter; William Barry; and David Vinson. 2006. Strength of British English accents in altered listening conditions. Perception and Psychophysics 68.139–53. DOI: 10.3758/BF03193664.


Jurafsky, Daniel; Alan Bell; and Cynthia Girard. 2002. The role of the lemma in form variation. Laboratory phonology 7, ed. by Carlos Gussenhoven and Natasha Warner, 3–34. Berlin: Mouton de Gruyter.


The evolution of medial /u/ over real and remembered time


Hay
New Zealand Institute of Language, Brain and Behaviour, and Department of Linguistics
University of Canterbury
Private Bag 4800
Christchurch, 8140, New Zealand
[jen.hay@canterbury.ac.nz]

Foulkes
Department of Language and Linguistic Science
University of York
York, YO10 5DD, United Kingdom
[paul.foulkes@york.ac.uk]