

## Brief article

## Tracking word frequency effects through 130 years of sound change

Jennifer B. Hay<sup>a,b,\*</sup>, Janet B. Pierrehumbert<sup>c,d,a</sup>, Abby J. Walker<sup>e</sup>, Patrick LaShell<sup>a</sup><sup>a</sup> New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, New Zealand<sup>b</sup> Department of Linguistics, University of Canterbury, New Zealand<sup>c</sup> Department of Linguistics, Northwestern University, United States<sup>d</sup> Northwestern Institute on Complex Systems, United States<sup>e</sup> Department of English, Virginia Polytechnic Institute and State University, United States

## ARTICLE INFO

## Article history:

Received 16 December 2013

Revised 24 February 2015

Accepted 28 February 2015

Available online 23 March 2015

## Keywords:

Sound change

Lexical frequency

Exemplar theory

New Zealand English

Chain shift

## ABSTRACT

Contemporary New Zealand English has distinctive pronunciations of three characteristic vowels. Did the evolution of these distinctive pronunciations occur in all words at the same time or were different words affected differently? We analyze the changing pronunciation of New Zealand English in a large set of recordings of speakers born over a 130 year period. We show that low frequency words were at the forefront of these changes and higher frequency words lagged behind. A long-standing debate exists between authors claiming that high frequency words lead regular sound change and others claiming that there are no frequency effects. The leading role of low frequency words is surprising in this context. It can be elucidated in models of lexical processing that include detailed word-specific memories.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Living languages are always in flux. Not only do our inventories of words change, but also how we say the same words. Eventually, sound changes affect all relevant words the same way; at the completion of the change, all words containing the sound are affected. This generalization allowed the 19th century Neogrammarian linguists to reconstruct Proto-Indo-European by comparing lists of cognate words in different languages, laying the foundations of historical linguistics as a science.

The sound changes under investigation in this paper are cases of what Labov (2004) terms ‘regular sound change’. That is, they can be described as the gradual transformation of the phonetic realizations of phonemes in a continuous

phonetic space. It is now more than a century following the insights of the Neogrammarians, and yet the detailed dynamics of such regular sound changes are still not well understood. In particular, it is not agreed whether sound changes affect all words concurrently and uniformly, or whether different groups of words are differentially affected. Many scholars claim that frequent words lead gradual phonetic sound changes and that less frequent words follow (Bybee, 1985, 2000; Pierrehumbert, 2001). This claim is contested by prominent scholars in mainstream historical linguistics and sociolinguistics. They maintain that regular changes proceed uniformly, and that documented cases of word-specific effects actually involve mechanisms other than gradual phonetic sound change (see Labov, 2004, 2010, and reviews by Kiparsky (2014) and Garrett (2014)).

Deep understanding of this controversial issue has been severely limited by the lack of speech recordings at historical time scales. Here, we investigate this question using the longest historical archive of transcribed speech now

\* Corresponding author at: New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch, New Zealand.

E-mail address: [jen.hay@canterbury.ac.nz](mailto:jen.hay@canterbury.ac.nz) (J.B. Hay).

extant: the Origins of New Zealand English (ONZE) corpus. Using this archive, we investigate frequency effects in a set of interconnected regular sound changes – the short front vowel chain-shift that has unfolded over 130 years in New Zealand English. Our analysis is a considerable advance on previous work on frequency effects because it (a) considers how word frequency effects evolve over time and (b) also considers how word frequency effects manifest in interlinked sound changes involving two or more categories. The results reveal that words of different frequencies behave differently in a surprising way that can be elucidated by psycholinguistic theory.

## 2. Lexical frequency and regular sound change

According to theories in which people remember the detailed phonetic properties of individual words (such as Bybee, 2000; Phillips, 1984, 2006; Pierrehumbert, 2001, 2002), it is expected that frequent words should lead regular sound changes. The memory of each word's pronunciation is updated at a rate that is proportional to the frequency with which the word is encountered.<sup>1</sup> The distribution of remembered pronunciations affects subsequent productions of the word. Assuming that words are subject to a persistent bias in pronunciation when a change is in progress – and that the bias is constant – then frequent words should change faster simply because they undergo the bias more often (Pierrehumbert, 2001).

Indeed, for sound changes that involve the deletion or reduction of speech sounds, it has often been claimed that frequent words display the innovative variant more at any given time than infrequent words do. For example, a highly influential study of the loss of final /t/ claims that frequent words are leading the change, on the basis that the /t/ is more often absent in frequent words such as *just* or *plant* than in less common words like *rust* or *lint* (Bybee, 2000, 2002). However, this result (and others like it) is also consistent with a stable effect where frequent words simply have more deletion, independent of the ongoing change. It is well documented that frequent words tend to be articulated with less effort, even in the absence of sound change (e.g. Gahl, 2007; Jurafsky, Bell, & Girand, 2002; van Son, Bolotova, Lennes, & Pols, 2004; Wright, 1979). The presence of an effect during periods of both change and stability is not sufficient to demonstrate a relationship between the change and that effect. Indeed, the pervasiveness of these effects has been used to argue that frequency does not actually play a role in regular sound change, because “structured variation is not in itself sound change; it can persist for centuries and even millennia” (Kiparsky, 2014:71). Dinkin (2008) argues that the ‘real effect’ of frequency on sound change is to exert a simple, stable lenition effect. In the terminology of mathematics and statistics (see Priestley, 1988), these authors suggest

that a stationary process (one which is invariant with regard to shifts in time) is responsible for the fact that the distributions of phonetic realizations for frequent words typically include more lenited variants than those for rare words.

The claim that words of different frequencies are affected differently by regular sound changes amounts to the claim that there is a non-stationary effect, i.e. the effect of frequency is different during periods of language change than during stable periods, a fact that can be overlooked by researchers on both sides of the debate. As outlined above, there has been a strong tendency to interpret the presence of stationary frequency effects as indicative of a role of frequency in sound change. Conversely, the absence of stationary frequency effects has been interpreted as indicating that lexical frequency is not involved in sound change. For example, Labov (2010) investigates a range of sound changes in progress, finding negligible evidence for frequency effects at work. His conclusion is that frequency is not involved in sound change. However he also does not look for non-stationary effects. A non-stationary effect would reveal itself as a statistical interaction between word frequency and the progress of the change, and not as a constant effect.

If frequent words actually lead in a sound change, then the effect of word frequency should be largest when the change is progressing most rapidly, and smallest during stable periods in the language. A new study of the lenition of medial intervocalic /t/ in the ONZE corpus addresses this prediction (Hay & Foulkes, in press). Originally a voiceless plosive, /t/ in this position has become increasingly voiced as New Zealand English has developed. In the early stages of the change, there was a small difference in voicing rates between lower and higher frequency words, and this difference increased once the change was well underway. This result provides empirical support for the predictions of the models described above, in which the memory of a word's pronunciation is updated at a rate that derives from the frequency with which the word is encountered (Pierrehumbert, 2001).

Low frequency words are known to lead in a different type of linguistic change: analogical change (Bybee, 1985; Lieberman, Michel, Jackson, Tang, & Nowak, 2007; Phillips, 1984). In analogical changes, irregular words in a grammatical paradigm are replaced by words conforming to more regular patterns. For example, in the English past tense, irregular forms such as *wept* tend to become regularized (cf. *weeped*) by analogy with regular past tense forms. Low frequency irregular forms can be relatively quickly replaced with regular forms, whereas more frequent irregular forms (such as *kept*) have remained in use long after many rare irregular forms were lost. This happens because irregular high frequency forms are extremely well learned and resist competition from regular patterns. A related phenomenon, demonstrating the entrenchment of high frequency forms in memory, is the finding that high frequency words are slower to be replaced with other lexical forms (Pagel, Atkinson, Calude, & Meade, 2013; Pagel, Atkinson, & Meade, 2007). But this explanation cannot extend to phonetically regular sound changes, as they do not involve any analogical force.

<sup>1</sup> Note that updating a distribution need not involve storing separate memories for the encountered token. As outlined in Pierrehumbert (2001), the phonetic parameter space is assumed to be granularized, and groups of tokens which are similar with respect to the granularization will be encoded as identical (see Kruschke, 1992).

Several other types of low frequency-leading changes have been reported. According to Phillips (2006), changes that involve abstract generalizations over the phonological structures of the word types in the lexicon may act like analogical changes in affecting low frequency words first. She contrasts such changes with those that only involve the phonetic realizations of phonemes (such as vowel shifts), which she describes as affecting more frequent words first. Though intriguing, these observations are limited by the fact that they rest on written historical records and synchronic experimental studies, rather than on longitudinal data. As a result, the gradualness of the changes can only be inferred. Ogura (2012) also suggests that changes may behave differently depending on whether they are driven by perceptual, rather than productive forces. She claims that, while high frequency words generally lead change, low frequency forms can dominate perceptually driven change because: “perceptually or cognitively unfavourable forms can be learned and maintained in their unfavourable forms if they are high frequency in the input” (Ogura, 2012, p.438).

For phonetically gradual sound changes that are not structurally complex, we are not aware of any reported case of low frequency forms leading. For this type of sound change, it is either argued that high frequency forms should lead (Bybee, 2000), or that there should be no frequency effect (Labov, 2004). Almost no work has investigated how frequency effects pattern over time during the course of such change. Only recently, with the creation of large transcribed and aligned longitudinal speech archives, has it even become possible to rigorously test for such patterns. The present study and Hay and Foulkes (in press), both using the ONZE corpus, are the only two studies we are aware of that carry out such a test. Hay and Foulkes explore a consonant lenition. Here we explore a different type of change: a set of interlinked vowel changes that shifts the vowels without leniting them.

### 3. The New Zealand English short front vowel shift

Modern New Zealand English (NZE) is notable for the unusual quality of three short front vowels. These vowels are so different from their counterparts in other dialects as to cause misunderstandings: NZE *bat* often sounds to speakers from elsewhere like *bet*, NZE *bet* sounds like *bit*, and NZE *bit* often sounds like *but*. We henceforth refer to the three sets of short front vowels as BAT vowels, BET vowels, and BIT vowels.

This pattern resulted from gradual changes over more than a century. Over the history of NZE, the vowels in BAT and BET words came to be closer and more front, and BIT words became more open and back. Gordon et al. (2004) provide evidence that the BAT words may have initially shifted in response to the encroaching vowel in words like *bart*, which are pronounced with long open vowels in the *r*-dropping New Zealand dialect. This is corroborated by the analysis presented here, which shows that BAT vowels before voiced sounds are most advanced in the change (see Table A1). These are exactly the vowels that would be most confusable with advancing BART

words, due to their durational overlap (c.f. Chen, 1970; Mack, 1982). The later shifts in BET and BIT words occurred as a reaction to the advance of the BAT words.<sup>2</sup> The end result is that the vowels have rotated in acoustic space but have remained distinct from each other.

The rotation of these three vowels is a regular sound change, which eventually affected all words containing the vowels. It is characterized as a *chain-shift* because the changes are interlinked, and the original set of lexical contrasts is maintained. Specifically, it is a *push chain* because the lower vowels moved first, and the higher vowels moved away when the lower vowels encroached on them. The typology of chain shifts dates from Martinet (1952), who contrasted push chains with ‘pull chains’, in which categories move to occupy phonetic space previously vacated by the retreat of other categories. Although historical records with sufficient detail to distinguish these types of chain shift are rare, Lubowicz (2011) cites 5 other examples of push chains in addition to the New Zealand English Vowel Shift.

Chain shifts rotate vowels without any loss of lexical contrast. In this sense, they differ from /t/-deletion and many other leniting sound changes which can neutralize lexical contrasts. A major limitation of the model discussed above (Pierrehumbert, 2001) is that it in fact has no treatment of lexical contrasts. As a result, it does not provide a general mechanism for chain shifts, in which the system of lexical contrasts is actually preserved. In push chains, a category with a systematic bias moves first, and encroaches on another category which then retreats. According to Pierrehumbert’s model, the result of such an invasion would be neutralization of the categories. Indeed, no existing model integrates a treatment of word-specific phonetic patterns with an up to date treatment of lexical competition.

When one vowel encroaches on another, a region of phonetic ambiguity between the two vowels is created. Our suggestions about the mechanism for push chains builds on findings that word frequency effects are greater for poor or ambiguous speech signals than clear ones (Norris & McQueen, 2008). In the ambiguous region near the category boundary, a low frequency word is more vulnerable to being misunderstood than a high frequency word. The best understood tokens of low frequency words will be those far from the boundary. We presume that neutralization is avoided because chain shifts proceed by such small increments that the region of phonetic ambiguity between adjacent vowels at any given time is also small, in relation to the total within category variation. The consequences of this situation will be discussed in more detail in Section 6.

<sup>2</sup> All vowels have F2 voicing effects in which the pre-voiced (longer) vowels are fronter. This can be interpreted as stemming from the well-documented effect that longer vowels provide more opportunity for an articulatory target to be reached, and so can have more peripheral realizations (Lindblom, 1963). However the effect of voicing on F1 of BAT vowels cannot be interpreted in this way. The pre-voiced vowels are the higher (and thus more centralized) tokens, and so would actually take less effort to produce.

#### 4. Methods

Our data-set comprises a total of 2741 word types (80646 word tokens) from the speech of 549 different speakers, with birth-dates spanning 136 years (1851–1987). The large size of our data-set gives us more statistical power than other studies of sound change. The historical time-depth provides a unique opportunity to track sound change over many generations of speakers.

This time-depth is made possible by the unique corpus from which the data is drawn. The Origins of New Zealand English (ONZE) corpus contains recordings of speakers born between 1851 and 1987 (Gordon, MacLagan, & Hay, 2007). These have been orthographically transcribed, and time-stamped at the utterance-level. The HTK-Toolkit (Young et al., 2006) was used to force-align the speech to the phoneme level, using phonological representations extracted from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). Using a LaBB-CAT interface (Fromont & Hay, 2008, 2012) first and second formants (F1 and F2) were automatically extracted at the midpoint of the vowel, using Praat's standard settings (Boersma & Weenink, 2005).<sup>3</sup> F1 and F2 are the first and second resonances of the vowel, which can be estimated from the acoustic signal and which effectively characterize dimensions of the vowel quality. F1 captures the open/close dimension. Vowels with a higher F1 are more open. F2 captures the front/back dimension. Vowels with a higher F2 are more front. Formants were converted to the Bark scale, a transformation that reflects the perceptual scaling of the two dimensions.

The range of formant values varies across speakers, partially as a function of vocal tract length. Some sociolinguistic work attempts to normalize these values, for example by adjusting with reference to the midpoint of the speaker's values for each formant, combined across vowels (see overview in Adank, Smits, & van Hout, 2004). We do not do this because normalization can make changes in one vowel manifest as adjustments to another vowel's formants. The adjustment could also be unduly influenced by frequent words in a way that would obscure potential frequency effects. To avoid such artifacts, we deal with the varying formant ranges across speakers statistically, by using mixed effects models, as described below.

The data-set was restricted to measurements of lexically stressed vowels occurring in monosyllabic and disyllabic content words. Vowels occurring before /l/ were eliminated because New Zealand English has a merger of BET and BAT vowels before /l/ (Hay, Drager, & Thomas, 2013). Because many of the recordings were made in field conditions using early recording technology, extensive precautions were taken to identify and remove values indicative of errors in the force-alignment or formant tracking. Specifically, tokens with F1 or F2 measurements beyond 3 vowel  $\times$  gender standard deviations were removed. Extreme value cut-offs were set individually

based upon each speaker's mean F1 and F2 vowel formants. Distributions of F1 and F2 values for each individual speaker were manually inspected. Multimodal and non-bivariate normal formant distributions were then further checked for measurement error in the source recordings. After remediation, speaker vowel means were recalculated and the extreme value cut-offs were adjusted appropriately. The analyzed data-set comprises 18,464 tokens of BAT words (868 types), 36,736 of BET words (1132 types), and 25,446 of BIT words (741 types) from 549 speakers (269 females, 280 males). The frequency of each word type was drawn from the full corpus. Word counts ranged from 1 to 4905 (mean 29.48, sd 201.10). The log of these values was taken as an estimate of lexical frequency.<sup>4</sup>

Following established practice in sociolinguistics (see Bailey, 2008) we use speaker year of birth as a proxy for tracking the progress of the change.<sup>5</sup>

Linear mixed effects models (see Baayen, 2008) were fit separately to Bark F1 and Bark F2 of each vowel, using the *lme4* library in R (Bates, Maechler, & Bolker, 2011; R Core Team, 2012). Each model contained random intercepts for word and speaker. Fixed effects of following voicing, gender, log lexical frequency (centered), and year of birth (centered) were tested, as well as interactions between these factors. Non-linearities in the effect of year of birth were modeled using *restricted cubic splines* from the *rms* library (Harrell, 2013). Model selection was guided by  $\chi^2$  likelihood tests, Akaike Information Criterion (Akaike, 1974) and Bayesian Information Criterion (Schwarz, 1978). The best fit models are reported. Models which included linguistic factors (frequency and/or voicing) as fixed effects also included these factors as random slopes for speaker. If these factors were not justified in the model, they were dropped from both the fixed and random effects. The final models were then rerun with uncentered factors

<sup>4</sup> A reviewer asks about the relationship of these corpus-derived frequencies to CELEX frequencies (Baayen et al., 1995). Our analysis includes 84 types that do not contain any entries in CELEX (mainly proper nouns). Omitting these forms, a Spearman's correlation between the log corpus frequency and log CELEX wordform frequency returns a highly significant correlation ( $r_s = .71$ ,  $p < .0001$ ). Examples of words over-represented in the corpus, as opposed to CELEX, include *dredges*, *paddock*, and *lambling*, all farming terms associated with the New Zealand rural economy. Underrepresented words include *cent*, *weapons*, and *actors*. We take the corpus frequencies to be a closer, more context-relevant approximation of word frequency distributions as experienced by the speakers.

<sup>5</sup> Using speech-date as a predictor rather than year of birth would not be practical, as it does not provide enough range, and would simply cluster the speakers into three main groups relating to when the data was recorded. The earlier born speakers were older when they were recorded than the later born speakers. If we assume that speakers may continue to move in the direction of change throughout their lifetime (Harrington, Palethorpe, & Watson, 2000), then using year of birth as a proxy for change will slightly under-estimate the speed of change (Gordon et al., 2004). In the case of vowels – formant values are affected as speakers age, so the fact that older speakers are over-represented amongst the earliest speakers will also affect the apparent speed of change in this respect. Effects of ageing on F1 (Harrington, Palethorpe, & Watson, 2007), for example, may mean that our data underestimate the speed of BAT and BET height changes, and over-estimate the BIT changes. Taken together, these factors tell us to interpret the exact apparent speed and timing of movement of vowels in our data with some caution. However none of these factors interfere with the distribution of the change over different word frequencies. Any observed word frequency effects should be independent of the effects of this common simplifying assumption.

<sup>3</sup> The choice of the midpoint was a pragmatic one, in that it can objectively and automatically identified for large numbers of vowels. Because all of the vowels involved are largely monophthongal in NZE, there is no reason to believe that a location based on target identification would have different results.



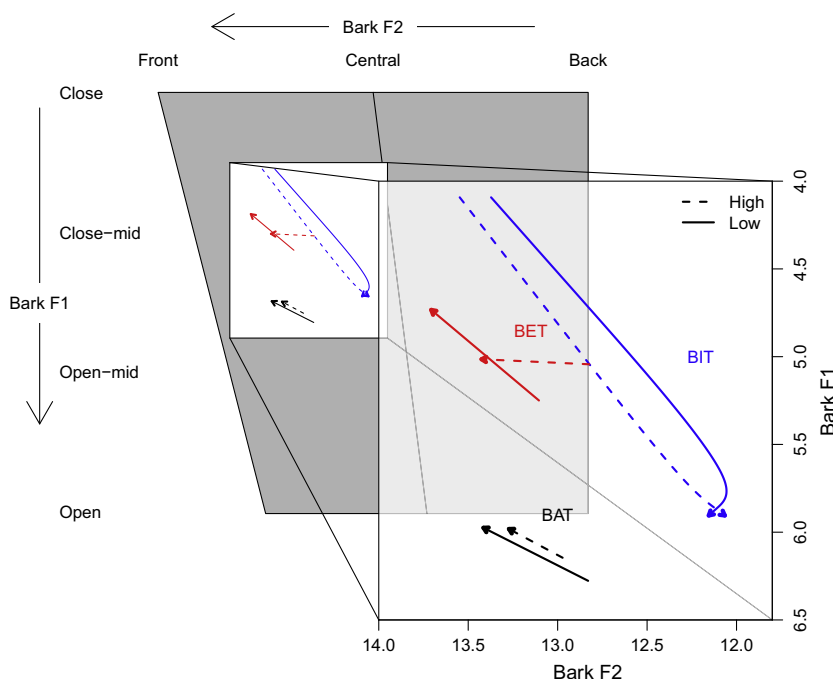
to aid in interpretability. Significance levels were calculated using [Satterthwaite's \(1946\)](#) approximations for the degrees of freedom using the *lmerTest* library ([Kuznetsova, Brockhoff, & Christensen, 2013](#)).

## 5. Results

For all three vowels, there was a significant interaction between year of birth and lexical frequency (in BAT F1 and F2, BET F1, and BIT F2). In each case, the interaction term goes in the opposite direction from the year of birth coefficient, indicating that high frequency forms are not advancing as quickly as low frequency forms. That is, we observe a non-stationary effect, with low frequency words are leading the change. A non-linear effect of year of birth was statistically justified for BIT but not for the other vowels. The coefficients from the final statistical models are shown in [Table A1](#).

This outcome is the opposite of that predicted by the models of sound change outlined above. [Figs. 1 and 2](#) display the changes in the short front vowels, as described by our analysis. For clarity, we show the values for female participants only. Note, however, that no models include interactions containing gender. The model predictions for males are therefore uniformly shifted in the direction of the gender effect, and otherwise identical.

In [Fig. 1](#), all three vowels are plotted on a traditional vowel quadrant, and the modeled trajectory of movement from the beginning of our analyzed time-period to the end is indicated by separate arrows – one for the highest frequency words, and one for the lowest frequency words.



**Fig. 1.** The New Zealand front short vowel shift, as represented in the speech of female participants. The shaded area shows how the vowels are positioned with respect to the overall vowel space. The area immediately surrounding the vowels is expanded for clarity. Model predictions for overall changes in F1 and F2 for female speakers are shown, with the direction of shift indicated by the arrows (see model details in [Table A1](#)). Solid lines summarize the model predictions for the lowest frequency words in the dataset. Dashed lines summarize the model predictions for the highest frequency words. All arrows cover the same time period. Low frequency words move more quickly.

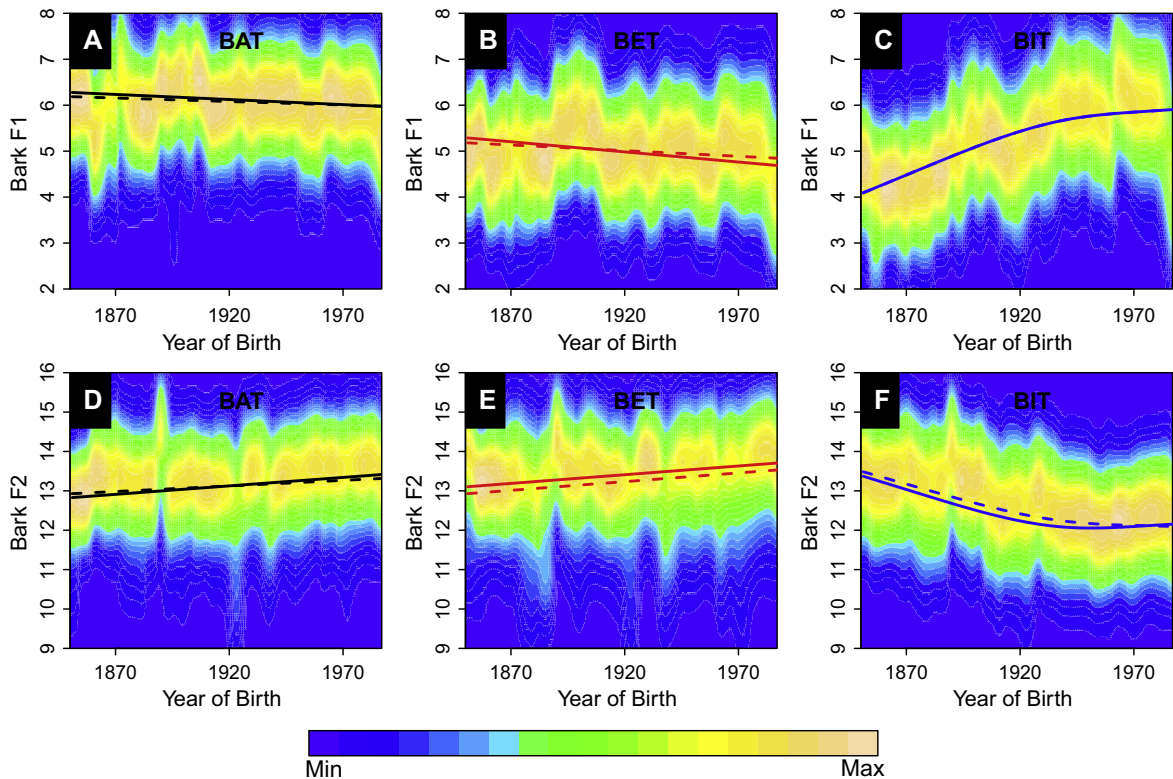
The lower frequency words move further, over the same time period. In [Fig. 2](#), the separate models for F1 and F2 of each vowel are plotted over heatmaps, illustrating the underlying distributions over which they are modeled. The frequency effect is again illustrated by showing separate trajectories for the highest and lowest frequency words.

## 6. Discussion

We have found significant interactions between year of birth and frequency for all three vowels. These interactions provide evidence for the involvement of lexical frequency in regular sound change. For all three vowels, low frequency words lead the way. These effects appear surprising in the context of the sound-change literature, but a conjecture about the mechanism can be formulated using theories in which listeners have detailed phonetic memories for specific words (e.g. [Goldinger, 1998](#); [Pierrehumbert, 2006](#); [Wedel, Kaplan, & Jackson, 2013](#)).

The factors that set push-chains in motion are not well understood, though it seems likely that social factors as well as cognitive ones must be involved. However, once a push chain has been initiated, the encroachment of one vowel on the acoustic space of the next can be viewed as an advancing front of sound change. The phonetic overlap between the two distributions creates the potential for ambiguity between the categories.

How does the existence of a region of ambiguity between two categories cause the category encroached on to be repelled? A mechanism is sketched in [Fig. 3](#).



**Fig. 2.** Contour plots showing the distributions of Bark F1 and Bark F2 values over time, for each of the three vowels in the speech of female participants. The hotter colors in the figure (yellow tones) are more frequent formant realizations than the cooler colors in the figure (blues). The relative frequency data was generated using the `bkde2D` function from the `KernSmooth` package in R. The frequency data were plotted as a heatmap using a modified form of the `filled.contour` function from the `graphics` package in R. Note that formant values on the y-axis are laid out from low to high, in contrast to the traditional vowel polygon in Fig. 1. Model predictions are laid over the data. Solid lines summarize the model predictions for the lowest frequency words. Dashed lines summarize the model predictions for the highest frequency words. 2C shows an effect of year of birth but no effect of word frequency. All other panels show a significant frequency effect. For F2 of bet words, low frequency words are consistently ahead of high frequency words. In all other panels (2A, 2B, 2D, 2F) there is a significant interaction between year of birth and word frequency. For all three vowels, low frequency words move more quickly (i.e. have a steeper slope) in one or both formants.

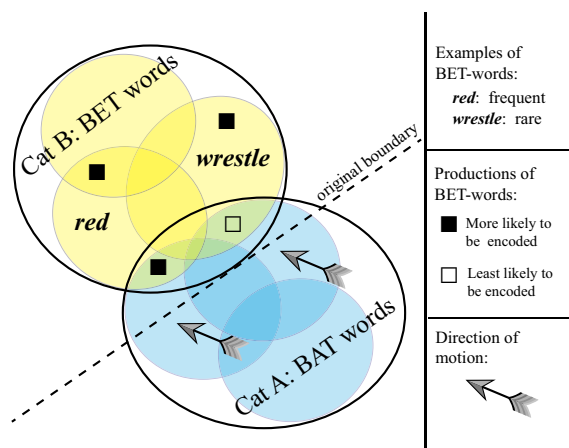
The mechanism relies on the idea that if tokens are difficult to understand, they are less likely to be encoded in memory. Indeed, there is evidence that poor or non-canonical examples of words are less likely to be robustly stored even if they can be understood (e.g. Sumner & Samuel, 2009).

In Fig. 3, the advancing front of Category A (here, the BAT words) has created a phonetic overlap with a neighbouring Category B (here, the BET words). Examples of BET words that are well away from the ambiguous region are more reliably encoded and remembered than those falling in the ambiguous region. The loss of tokens from ambiguous region (in combination with random variation in production that extends the distribution of B on the side away from Category A) causes an overall shift of the B category away from the encroaching A (see also Blevins & Wedel, 2009; Martinet, 1952; Wedel, 2006).

This category-level interpretation makes a further prediction regarding the distribution of phonetic variants across the words in the category, because ambiguity places low frequency words at a particular disadvantage. High frequency words are much more easily accessed than low frequency words (Forster & Chambers, 1973). The speech

processing system resolves ambiguous speech signals by bias toward the more frequent lexical candidate (Norris & McQueen, 2008) and listeners have a greater tendency to mishear low frequency words as high frequency words than vice versa (Savin, 1963). The effects of phonological ambiguity are not confined to close lexical competitors. In lexical decision tasks, low frequency words are also more likely to be misinterpreted as nonwords (Luce & Pisoni, 1998). Thus, pronunciations of high-frequency words that lie near the advancing front of a nearby vowel should be easily and accurately recognized, on the average. We assume that these tokens will then be robustly stored and affect subsequent productions. Pronunciations of low frequency words that lie near the advancing front, on the other hand, are less easily recognized. They may be misheard as a competing high frequency word, or as a nonword. Or they may be recognized, but not reliably stored.<sup>6</sup>

<sup>6</sup> Non-canonical pronunciations may, paradoxically, have an advantage in being stored if they are socially salient (Sumner, Kim, King, & McGowan, 2014). This factor would not affect Category B in Fig. 3, which represents a stage in the sound change in which Category B has no active socially triggered bias.



**Fig. 3.** Schematic diagram of the mechanism for word frequency effects in a push chain. A shifting distribution of the low front vowel in BAT words (Category A) encroaches on the distribution for the mid front vowel in BET word (Category B). The colored ellipses represent three examples of lexical items within each category. Their placement and shape is arbitrary. Once a region of ambiguity is created (shown here by the region of the BAT distribution that has crossed the original category boundary), productions of BET words are encoded with variable reliability. Productions of BET words that are far from the region of ambiguity are reliably encoded. Productions of BET words that fall in the region of ambiguity are more reliably encoded for high frequency words, such as “red”, than for low frequency words, such as “wrestle”. The noise in the production system that enables the BET distribution to retreat from the encroaching BAT distribution is not illustrated.

On the average, this leads to more attrition for low-frequency words than for high-frequency words in the ambiguous region created by the sound change.

These factors conspire to mean that the distribution of remembered low frequency words should be repelled more quickly from the advancing front of a sound change. Importantly, we are not suggesting that the speaker actively enhances contrast in cases of potential ambiguity. The frequency effects simply follow from general mechanisms of word perception and storage.

The effects of word frequency in our study are small in comparison to the within-category variation for each vowel. This fact helps to explain why such a large sample has been needed to find the effects. It is consistent with Pierrehumbert’s (2006) suggestion that word-specific phonetic patterns are subtle, within-category, effects. It is of course possible for words to acquire multiple, categorically distinct phonological spell-outs, as shown for example in studies of how frequent words are produced in spontaneous speech (Kemps, Ernestus, Schreuder, & Baayen, 2004). However, for the New Zealand short front vowel shift, the movement within a single generation was much smaller than the separation between categories. Differential effects of word frequency are therefore even smaller.

Importantly, it does not directly follow from our findings that such a frequency effect would definitely be observable in any push-chain relationship. The effect of frequency on confusability will be influenced by a range of factors, including the shape and size of the distributions, the degree of overlap and phonetic confusability, and the speed of change.

## 7. Conclusion

We have tracked three interrelated vowel changes as they unfolded over the history of New Zealand English. Our mixed effects models of automatically extracted formant values reveal robust frequency effects in all three changes, with the lowest frequency words moving most quickly. For the vowels being pushed in the chain shift that we investigated, the low frequency words were in advance. This finding is very surprising in the context of contemporary sociolinguistic debate about the role of frequency in regular sound change. However it follows naturally from speech processing models in which people remember detailed phonetic properties of words, and in which ambiguous tokens of low-frequency words are vulnerable to misperception.

Claims that word frequency can play no role in regular sound change are therefore wrong. However, competing claims that high frequency words always lead are also wrong. The particular nature of lexical frequency effects in a given sound change can only be predicted and interpreted in the light of a detailed understanding of the mechanisms underlying the change itself.

## Author contributions

J.B.H. and J.B.P. conceived the study and were involved in the theoretical interpretation. A.J.W. and J.B.P. were involved in the acoustic analysis. P.L. and J.B.H. performed the statistical analysis. P.L. created Figs. 1 and 2. J.B.P. created Fig. 3. The manuscript was drafted by J.B.H. and J.B.P., with input from A.J.W. and P.L.

## Acknowledgments

This work has been supported by a Royal Society of New Zealand Marsden Grant and a Rutherford Discovery Fellowship to the first author, and a University of Canterbury Erskine Fellowship to the second author. In addition, this project was made possible through a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. The ONZE data was collected by the Mobile Disc recording Unit of the NZ Broadcasting Service, Rosemary Goodyear, Lesley Evans, members of the NZ English class of the Linguistics Department, University of Canterbury, and members of the ONZE team. The work done by members of the Origins of New Zealand English Project (ONZE) in preparing the data, making transcripts, and obtaining background information is also gratefully acknowledged. The Corpus was created and supported with funding from the following sources: University of Canterbury, Foundation for Research, Science and Technology (the New Zealand Public Good Science Fund), the Royal Society of New Zealand, The New Zealand Lotteries Board Fund, and the Canterbury History Foundation. We are particularly grateful to Robert Fromont for his work programming LaBB-CAT – the ONZE Corpus search engine and interactive interface. This

**Table A1**

Summary of statistical models.

BAT	Bark F1				Bark F2			
	Estimate	Std Error	t value	p	Estimate	Std error	t value	p
Intercept	10.5660	1.4116	7.4850	<0.001***	4.5110	1.5360	2.9370	0.003**
Year of birth	−0.0023	0.0007	−3.1820	0.002**	0.5268	0.1441	3.6550	<0.001***
Voiced = Y	−0.0775	0.0165	−4.6940	<0.001***	0.0045	0.0008	5.6800	<0.001***
Freq	−0.2637	0.1048	−2.5150	0.012*	0.1074	0.0253	4.2440	<0.001***
Gender = M	−0.5592	0.0429	−13.024	<0.001***	−1.0110	0.0478	−21.1310	<0.001***
Freq × year of birth	0.0001	0.0001	2.4610	0.014*	−0.0003	0.0001	−3.6960	<0.001***
<b>BET</b>								
Intercept	12.9900	2.3910	5.4310	<0.001***	5.1002	1.4618	3.4890	0.001***
Year of birth	−0.7972	0.1092	−7.3020	<0.001***	−0.0330	0.0090	−3.6520	0.019*
Voiced = Y	−0.0041	0.0012	−3.3630	0.003**	0.0043	0.0008	5.7870	<0.001***
Freq	x	x	X	x	0.0858	0.0308	2.7850	0.008**
Gender = M	−0.5103	0.0694	−7.3500	<0.001***	−0.9759	0.0503	−19.4150	<0.001***
Freq × year of birth	0.0004	0.0001	7.4170	<0.001***	x	x	x	x
<b>BIT</b>								
Intercept	−33.7809	3.5628	−9.4820	<0.001***	46.6108	3.6139	12.8980	<0.001***
rcs(year of birth, 3)'	0.0205	0.0019	10.9310	<0.001***	−0.0180	0.0019	−9.4530	<0.001***
rcs(year of birth, 3)''	−0.0089	0.0019	−4.8170	<0.001***	0.0115	0.0019	5.9350	<0.001***
Gender = M	−0.7782	0.0499	−15.5940	<0.001***	−0.8420	0.0393	−21.4520	<0.001***
Freq	x	x	X	x	−0.4691	0.3998	−1.1730	0.241
Voiced = Y	x	x	X	x	0.0861	0.0422	2.0390	0.042*
Freq × rcs(year of birth, 3)'	x	x	X	x	0.0003	0.0002	1.2590	0.209
Freq × rcs(year of birth, 3)''	x	x	X	x	−0.0006	0.0002	−2.8730	0.004**

Significance levels were calculated using [Satterthwaite's \(1946\)](#) approximations for the degrees of freedom using the lmerTest library ([Kuznetsova et al., 2013](#)). Broad significance classes are indicated with \* (<.05), \*\* (<.01), and \*\*\* (<.001). Four models include significant frequency × year of birth interactions. In each case, the interaction coefficient goes in the opposite direction from the year of birth coefficient, indicating that high frequency forms are not advancing as quickly as low frequency forms. Note that for BIT words, the effect of year of birth on formant values was modeled as a non-linear effect using a restricted cubic spline (RCS) from the rcs() function in the R package rms ([Harrell, 2013](#)). Non-linear effects were also tested for year of birth for other vowels and found non-significant. RCS applies a function to the predictor similar to the nonlinearity provided by an exponential function. RCS functions are described by the number of endpoints (called knots) in the function, which are the number of subsets used plus one. For both Bark F1 and F2, an RCS with 3 knots was used, which corresponds to a quadratic polynomial ( $x^2$ ). The first year of birth parameter is labeled rcs(year of birth, 3)' and is applied to the first subset of the data. The second parameter is labeled rcs(year of birth, 3)'' and is applied to the second subset of the data. The interactions of year of birth with word frequency are similarly labeled. A Bark scaling is chosen because it more accurately reveals the perceptual magnitude of the modeled shifts. Because this is a simple transformation, this choice does not have any meaningful influence on the shape of the final models.

manuscript has benefited from feedback from Cynthia Clopper, Donald Derrick, Matt Goldrick, Peter Racz and three anonymous reviewers.

## Appendix A

See [Table A1](#).

## References

- Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116, 3099–3107.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2, CD-ROM), LDC catalogue No.: LDC96L14, Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bailey, G. (2008). Real and apparent time. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change*. Oxford, UK: Blackwell Publishing Ltd.
- Bates, D., Maechler, M., & Bolker, B. (2011) lme4: Linear mixed-effects models using S4 classes. *R package* (Version 0.999375-42). <<http://CRAN.R-project.org/package=lme4>>.
- Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, 26, 143–183.
- Boersma, P., & Weenink, D. (2005). *Praat: Doing Phonetics by Computer* (Version 4.3.14) <<http://www.praat.org/>>.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning & form*. Philadelphia/Amsterdam: John Benjamins.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation & Change*, 14, 261–290.
- Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford: CSLI.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, 129–159.
- Dinkin, A. (2008). The real effect of frequency on sound change. *Penn Working Papers in Linguistics*, 14(1), 97–106.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behaviour*, 12, 62–635.
- Fromont, R., & Hay, J. (2008). ONZE miner: The development of a browser-based research tool. *Corpora*, 3(2), 173–193.
- Fromont, R., & Hay, J. (2012). LaBB-CAT: An annotation store. *Proceedings of the Australasian Language Technology Workshop*, 113–117.
- Gahl, S. (2007). "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. *Language*, 84(3), 474–496.
- Garrett, A. (2014). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics*. New York: Routledge.
- Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Gordon, E., Campbell, L., Hay, J., MacLagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origins & evolution*. Cambridge: Cambridge University Press.
- Gordon, E., MacLagan, M., & Hay, J. (2007). The ONZE corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating & digitizing language corpora. Diachronic databases* (2, pp. 82–104). Basingstoke: Palgrave Macmillan.



- Harrell, F.E. (2013). Rms: Regression modeling strategies. *R package* (Version 3.6-3). <[http://CRAN.R-project.org/package=rms\(2013\)](http://CRAN.R-project.org/package=rms(2013))>.
- Harrington, J., Palethorpe, S. & Watson, C.I. (2007). Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers. In *the Proceedings of Interspeech 2007*. Antwerp.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the queen speak the Queen's English? *Nature*, 408, 927–928.
- Hay, J. & Foulkes, P. (in press). The evolution of medial /t/ over real and remembered time. To appear in *Language*.
- Hay, J., Drager, K., & Thomas, B. (2013). Using nonsense words to investigate vowel merger. *English Language and Linguistics*, 17(2), 241–269.
- Jurafsky, D., Bell, A., & Girand, C. (2002). The role of lemma in form variation. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 3–34). Mouton de Gruyter.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 117–127.
- Kiparsky, P. (2014). New perspectives in historical linguistics. In C. Bowerman & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 64–102). New York: Routledge.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. (2013). lmerTest. *R package version 2.03*. <<http://cran.r-project.org/web/packages/lmerTest/index.html>>.
- Labov, W. (2004). *Principles of linguistic change, internal factors*. Malden: Wiley-Blackwell.
- Labov, W. (2010). *Principles of linguistic change. Cognitive & cultural factors* (Vol. III). Oxford: Wiley Blackwell.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713–716.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773–1781.
- Lubowicz, A. (2011). Chain shifts. In M. van Oostendorp, C. Ewen, B. Hume, & K. Rice (Eds.), *Companion to phonology* (pp. 1717–1735). Oxford: Wiley-Blackwell.
- Luce, P., & Pisoni, D. B. (1998). Recognizing spoken words. The neighborhood activation model. *Ear & Hearing*, 19, 1–36.
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, 71, 173–178.
- Martinet, A. (1952). Function, structure, and sound change. *Word*, 8(1), 1–32.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395.
- Ogura, M. (2012). The timing of language change. In J. M. Hernandez-Campoy & J. C. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics* (pp. 427–450). Oxford: Wiley-Blackwell.
- Pagel, M., Atkinson, Q. D., Calude, A. S., & Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences*, 110(21), 8471–8476.
- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, 449(7163), 717–720.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60, 320–342.
- Phillips, B. S. (2006). *Word frequency and lexical diffusion*. New York: Palgrave, Macmillan.
- Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, 34(6), 516–530.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter.
- Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic Press.
- R Core Team (2012). *R: A language & environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, <http://www.R-project.org/>.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114.
- Savin, H. B. (1963). Word frequency effects and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35, 200–206.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4(1015).
- Sumner, M., & Samuel, A. G. (2009). The role of experience in the processing of cross-dialectal variation. *Journal of Memory and Language*, 60, 487–501.
- van Son, R.J.J.H., Bolotova, O., Lennes, M., & Pols, L.C.W. (2004). Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian). In: *Proceedings of Interspeech 2004* (pp. 1277–1280).
- Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23, 247–274.
- Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*, 128(2), 179–186.
- Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition*, 7(6), 411–419.
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., et al. (2006). *The HTK Book (for HTK Version 3.4)*. Cambridge: Cambridge University Engineering Department.