Phonological representation: Beyond abstract versus episodic.

Janet B. Pierrehumbert
University of Oxford

March 1, 2016

Preprint of: Pierrehumbert (2016), Annual Review of Linguistics 2, 33-52.

1 Introduction

Phonology deals with systematic patterns in the sound structure of languages. Language scientists propose phonological representations with the goal of developing general and predictive understanding of these patterns. We want our understanding of phonology to be general, capturing the implicit knowledge that different speakers of the same language share, as well as similarities and differences across languages. We want it to be predictive, accurately characterizing what previously unseen words and phrases might occur in the future. The goal of having a predictive theory already means that the theory must be synoptic. Only by advancing generalizations is it possible to make predictions about examples that have not yet been seen.

Many linguists impute phonological representations to individuals. Of course, the brain does not carry out derivations and calculations in the exact form that scientists write them. The assumption that people have phonological representations amounts to the assumption that when people acquire highly virtuosic capabilities – such as our rapid and adaptable ability to process speech and learn the forms of new words and expressions – they implicitly acquire generalizations that are effective in predicting new data. Scientific theories of phonology in turn succeed insofar as they categorize and generalize along the same lines that people do. The representations used in phonological theory should therefore be well-aligned with the way phonological information is encoded in the mind.

Phonological representations are abstract. The representational apparatus of any successful scientific theory – from the theory of evolution to electromagnetism – is abstract. Even the acoustic theory of speech production – an area of phonetics where many linguists feel the rubber really hits the pavement – is actually very abstract, involving as it does a highly idealized schema of the vocal tract and voice source, as well as the mathematical apparatus for analyzing resonances. And in fact the mammalian auditory processing system implicitly incorporates these abstractions in representing vowels as a map defined by the lowest formants (Pierrehumbert, 2000).

However, when linguists talk about phonology being abstract, they normally mean something more specific. The controversy about abstractness in phonology has its roots in assumptions about levels of representation and modularity in linguistic theory. The seminal debate about abstractness in the phonology (Kiparsky, 1973; Hyman, 1970; Stampe, 1979) targets a highly modular linguistic theory in which the phonology manipulates lexical representations that are discrete (contrasting with the continuous domain of articulatory and acoustic phonetics), and underlying representations of words have been minimized in the sense of being stored using the fewest featural specifications possible, equivalent to minimizing the number of bits needed to store the representation. Minimizing the underlying representations means avoiding listing multiple, similar, forms for the same word as well as avoiding extraneous detail within each form. In this target theory, abstractness carries additional conceptual baggage beyond the ordinary meaning of the word. Notably, the abstract representations tend to be opaque (in the sense that the specifications lack any simple or direct acoustic or articulatory correlates). One cause of this opacity is the simple fact that speech as a physical process naturally involves many constraints and correlations amongst its parameters; minimizing the mental representations requires finding a more abstract set of dimensions that remove these correlations. A second cause is variability across contexts and speakers in how any given lexical contrast is realized.

Figure 1 illustrates this second problem using phonetic measurements from a study on allophones of /r/ and /t/ in American and Glaswegian English (German et al., 2013). It shows probability density functions for the third formant (F3) in words like *notice* and *worries*, where the target consonant appears in intervocalic position under falling stress. In this context, the American English /t/ is predominately produced as [ɾ] and /r/ is produced as [ɹ]. The primary acoustic difference between the two sonorant approximants is the value of F3; [ɹ] has a low F3 whereas the [ɾ] has a higher F3. The exact value of F3 varies across speakers and across phonological contexts. The mapping of the categories to the phonetic values would still, however, be transparent if all of the F3 values for [ɹ] fell in one range and all of those for [ɾ] fell in a distinct range. But this is not what we see. There is a strong effect of gender, with the modal value for female [ɹ]s falling right on the boundary between male [ɹ] and [ɾ]. There is also substantial variability within each gender, as shown by the range of optimal decision thresholds for individuals graphed in the figures. For values of F3 ranging from 1883 Hz to 2640 Hz (encompassing 55% of the total sample), an optimal decision of which phoneme has occurred is dependent on who is speaking. This means that the phonemic information is abstract, depending on the separability of social and lexical influences on the acoustic signal. Another point to note is the bump towards the left of the male distribution for /t/, also mirrored by a shoulder in the distribution for females. This group of examples with very low F3 values for /t/ arises in the context of dorsal /l/, which is a source of long-distance co-articulation effects (Kelly & Local, 1986). This provides a little example of the effect of phonological context on the phonetic realization of a category.

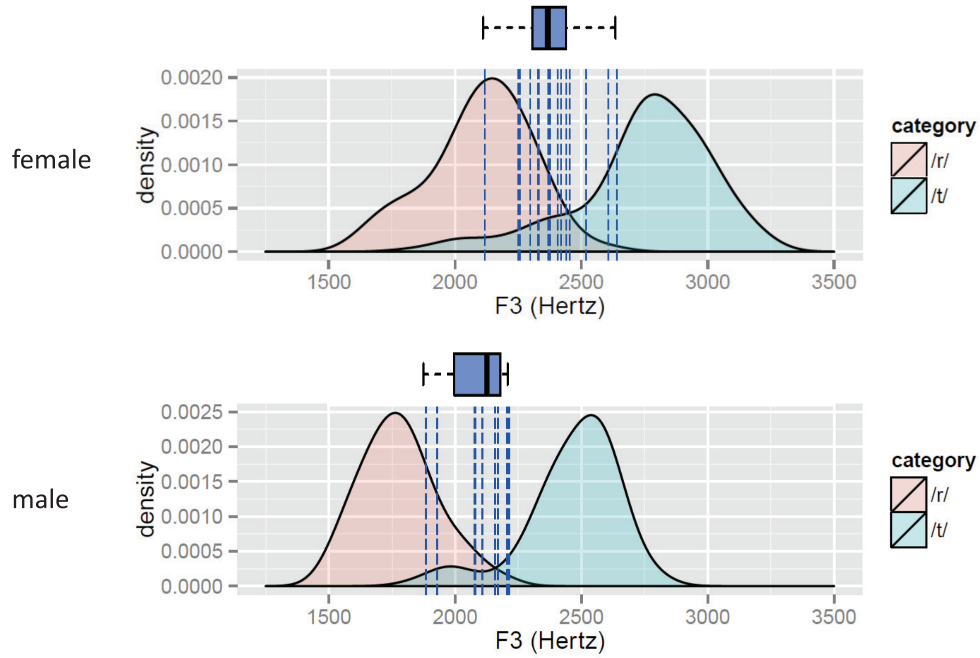

Figure 1: **Distributions of F3 for American English [ɹ] and [ɾ]** Plotted from original data collected for German et al. (2013). Vertical lines indicate optimal decision boundaries between [ɹ] and [ɾ] for individual speakers. Horizontal boxplots summarize the distribution of decision boundaries for each gender.

The distinctive F3 of American English [ɹ] actually provides one of the most studied and most transparent cases of the relationship between phonology and acoustic phonetics. In general, this relationship is transparent insofar as the competing categories are convex (so that a token that falls in between any two examples of a category is also an example) and discriminable (they exhibit little overlap, so that setting the correct decision threshold yields a low error rate in identification). Putative categories that are phonetically neutralized with other categories on the surface, that correspond to null phonetic events, or that group together dissimilar types of events, are more contentious (Kiparsky, 1973; Hyman, 1970; Stampe, 1979). In other domains, notably the processing of visual scenes, categories with such formal properties are difficult to learn. Episodic theories of phonology, however, provide ways that such categories could come to exist because they include social and contextual parameters that can in principle distinguish between speech sounds that are identical in their acoustics and articulation, or group together speech sounds that are phonetically dissimilar. In further discussion of the same study below, we will see how the dialectal dimension further reduces the transparency of the distinction between /r/ and /t/.

The within-category variability in Figure 1 could be captured in a cognitive model using very detailed representations, as proposed in Pierrehumbert (2002); Hawkins (2003); Johnson (2006); Wade & Möbius (2010). Or it could be captured using general principles that map more discrete and invariant lexical forms onto phonetic outcomes, as a function of the context. Generative linguists initially pursued this second approach, because it is so general, providing an obvious account of how a newly learned word can immediately be subject to contextual variation. For example, as soon as an American speaker learns the neologism *macket*, they would be likely to produce a [ɾ] in *My macket is all all wet now...*, producing that [ɾ] with an F3 value typical for their own voice and the phonological environment.

The best successes of this approach were cases in which the variation in the realization of a phonological category is triggered by the phonological context and seems to apply to all words equally (Pierrehumbert, 2006a). For example, Klatt (1976) succeeded in formulating a set of four rules for vowel duration that explained 97% of the variance in a large set of nonsense words of varying length that had been recorded in carrier sentences. In a study of 250 words containing /t/ or /d/, Zue & Laferriere (1976) found that /t/ was glottalized in 95% of the tokens of words such as *potent, sentence, kitten* that contain /t/ before a syllabic /n/. However, these successes concealed some methodological limitations and challenges. Relatively few studies in the 1970's and 1980's used large numbers of different words to exemplify any single phonetic realization pattern. Before the rise of mixed effects models (see (Baayen, 2008)), the equivalence of different words containing the same target sequence was often assumed rather than examined. Lastly, the approach concealed the challenges in explaining how the patterns of variation are represented and learned. Admittedly, there are strong typological tendencies in principles of phonetic realization, often grounded in facts about motor control and audition (Pierrehumbert, 2000). However, the exact relationship between phonological categories and phonetic realizations is language-specific (Mielke, 2008). As an example, there is a universal tendency for vowels to be nasalized before nasal stops, but the time course and extent of nasal co-articulation varies across languages, and can even change over time in the same language (Zellou & Tamminga, 2014).

Clearly, acquiring phonetic realization patterns requires a mechanism for phonetic details to impact representations in memory. A way to conceptualize this mechanism is to assume it accumulates detailed phonetic memories, and that generalizations are formed by further processing. Alternatively, one might propose that the representations are parametric representations of phonetic distributions, and that learning incrementally updates the values of the parameters. These concepts share the assumption that a great deal of phonetic detail is represented in memory, and that a more discrete level of representation (the units of phonological encoding) is linked to a less discrete level (a parametric map of the phonetic space).

In its original meaning, episodic memory referred to the human capability to form complex mental representations of sequences of events and to consciously recall them (Tulving, 2002). Telling an autobiographical story, or recalling an incident as if it were a movie in one's mind would be a typical manifestation of episodic memory. This ability shows that the informational capacity of human memory is remarkably large – undercutting the assumption that lexical representations need to be minimized. It also reveals a capacity for representing gradient information: not only the flow of time, but also other gradient information such as location, emotion, or attitude. The concept of episodic memory in the speech processing literature diverges from this initial conception because little about speech processing is conscious. Iconic memories of individual utterances are rare. The workhorse of phonological learning is the automatic and unconscious acquisition of implicit knowledge. But the theory adopts the claims that memory capacity is large, that representations in memory are extremely detailed, and that they include time and many other non-speech properties. The information included in episodic memories in the original sense of the term is essentially unrestricted. Proposing episodic memory in the lexicon opens the door to the inclusion of many, many other dimensions of word contexts, ranging from the topic of conversation and who the participants were to features such as status relationships and emotional valency.

A word is an association between a wordform and a meaning. The meanings of words indeed incorporate traces of the physical, social, and emotional contexts in which the words were experienced. All of these factors shape the connotations of a word, and the productive command of such connotations is revealed in many linguistic behaviors, such as insults, flattery, or white lies. Meanings of lexicalized compounds provide a striking illustration of how much pragmatic context can become encapsulated in the semantic representations of lexical items. For example, *slapstick* and *muckrake* have retained their emotional connotations of humour and disgust, respectively, even after the literal meanings of their components have been lost. What is less obvious is the extent to which the episodic effects shape the cognitive representations of wordforms, and of phonology in general.

In this article, I first review the motivations for positing abstract representations. Then, I summarize the different kinds of contextual factors that have been shown to exert systematic effects on how words are pronounced. Next, I turn to the question of how these effects, and the contextual factors that induce them, are reflected in long-term mental representations of words, as shown by experimental, sociolinguistic, and historic studies. Episodic effects in morphophonology provide an additional, and relatively unexplored, area for discussion. Finally, these results come together to support a viewpoint in which highly detailed memories of wordforms, including word-specific phonetic properties and indexical information associated with word variants, are complemented by more abstract representations that support productivity and generalization.

2 What abstract representations do for us

According to the phonological principle, the wordforms in the lexicon of any language can be characterized using a small number of coding elements. The number is small in the sense that it is much smaller than the number of words in the lexicon; the elements are meaningless in themselves, but in legal combinations they are associated with meanings. By creating a new combination of elements, speakers of a language can create a word with a new meaning, which need not be related in any particular way to the meanings of pre-existing wordforms. For example, the neologism *ferbir* need not have any meaning relationship to other words that begin in *fer*, such as *fertile*. This principle is shared in all seminal work in phonology, such as work by Trubetskoy, Jakobson, Saussure, Sapir, Bloomfield, Hockett, and Halle; these scholars differ in their views about the extent to which phonetics shapes phonology, but they do not differ in the idea that phonology involves abstract coding units that can be recombined in many different ways.

Phonology defines equivalence classes amongst speech events that occurred at different times and places, and whose physical properties are objectively different. These equivalence classes are needed to explain the productivity and adaptability of phonological behavior. People can understand words spoken by strangers who have their own phonetic idiosyncracies. They can invent new words that conform to the general patterns of their native language. Other people can encode, remember, and later reproduce these novel words.

To understand and evaluate models of this cognitive capability, it is useful to think in terms of the tradeoff between precision and recall. These are test statistics used in natural language processing to evaluate how well a model trained on a finite data sample (as all models are) performs in the subsequent processing of data from the same domain. Precision is the ratio of correctly accepted examples to all accepted examples. Recall is the ratio of correctly accepted examples to all examples that should have been accepted. These two measures are of interest because high-performance models need both to accept all bona fide examples, and to reject false examples. Since human phonological processing is a high-performance system, we must assume that it is also quite successful in achieving both of these goals.

Here is a tutorial example for readers who are unfamiliar with these concepts. The example is based on an idealized Malayo-Polynesian language, in which the syllable structure requires strict alternation between consonants and vowels. The model is trained on 4 words of the language, and it is tested on a data set that includes these 4 words, 4 additional words, and 8 examples of words from other languages that should be rejected. Table 1 lists these sets.

Training set	Test set			
	Good		Bad	
VC	VC	CV	VV	CC
CVC	CVC	VCV	CCV	VVC
CVCV	CVCV	VCVC	VVCV	CCVC
VCVCV	VCVCV	CVCVC	CCVCV	VVCVC

Table 1: Training set and test set for tutorial example

Table 2 displays the test performance for a model that memorizes the training set, without forming any generalizations. The model has perfect precision, but it has poor recall because it does not accept any of the novel good examples in the test set.

	Memory list model			
	Accept	Reject		Recall
Possible	VC CVC CVCV VCVCV	CV VCV VCVC CVCVC		$4/(4+4)$ $= 0.5$
Impossible		VV CCV VVCV CCVCV	CC VVC CCVC VVCVC	
Precision	$4/(4+0) = 1.0$			

Table 2: Test performance of the pure memorization model

Table 3 displays the test performance for a model that generalizes aggressively, taking the phonological grammar to be all combinations of the single symbols observed in the training set. This model has perfect recall, but it has poor precision because it accepts all of the bad examples in the test set.

	Unigram Model			
	Accept		Reject	Recall
Possible	VC CVC CVCV VCVCV	CV VCV VCVC CVCVC		$8/(8+0)$ $= 1.0$
Impossible	VV CCV VVCV CCVCV	CC VVC CCVC VVCVC		
Precision	$8/(8+8) = 0.5$			

Table 3: Test performance of the over-generalizing unigram model

In Table 4, we see the performance of the Goldilocks model, which is neither too strict nor too permissive. By forming generalizations at the correct scale (in this example, the biphone scale), it accepts all the good forms, including both the previously seen ones and the novel ones, while rejecting all the bad forms.

	Digram Model				
	Accept		Reject		Recall
Possible	VC CVC CVCV VCVCV	CV VCV VCVC CVCVC			8/(8 + 0) = 1.0
Impossible			VV CCV VVCV CCVCV	CC VVC CCVC VVCVC	
Precision	8/(8 + 0) = 1.0				

Table 4: Test performance of the Goldilocks digram model

In natural language processing, systems that fit the data in an excessively detailed way, as in Table 1, are called "over-trained" systems. They are avoided using methods that prune details while retaining significant patterns. In short, they build abstractions. Humans also build abstractions in order to avoid the pitfalls of over-trained mental representations. By their nature, abstractions omit many details.

Abstractionist theories of phonology obviously provide mechanisms – in the form of a phonological grammar – for generating or processing novel forms. Often overlooked by the abstractionists is the fact that episodic theories such as Johnson (2006); Goldinger (1998); Wade & Möbius (2010) do as well. The simplest mechanism discussed in these articles supports generalization through the cumulative force of examples that are similar to each other in a high-dimensional experiential space. Under this approach, a highly rated nonword such as /zæmpi/ (Hay et al., 2004) would be a high-scoring neologism due its many points of similarity to known words, such as *Zack*, *amp*, and *skimpy*. Dimensions that are not functionally relevant can be omitted from the calculation of similarity, providing a way to ignore extraneous details. The granularity in the implicit generalizations that are formed can be captured by adjusting the parameters that describe which examples count as nearby and which count as far away in the multi-dimensional space. Markedness effects that take the form of asymmetric similarity judgments can be captured in such models through mechanisms exploiting frequency differences (Nosofsky, 1992). An example would be a finding that /k/ acts as less distinct from /t/ than /t/ from /k/), in relation to speech errors, wordlikeness judgments, or any other evidence about phonological processing. Generally speaking, the approach overcomes the problem displayed in Table 1. It even lends itself to generalizations that involve novel phonemes using features that are attested in other, known phonemes.

However, some major findings in phonology are more of a challenge for the idea that apparent abstractions are ephiphenomal consequences of surface similarity in many dimensions. One is the finding that the productivity of lexical patterns strongly depends on their type frequency, e.g. the number of different words that display the pattern (Hay et al., 2004; Frisch et al., 2001; Richtsmeier, 2011). Token frequency (how often the pattern occurs in running speech) is also important, especially in relation to pre-lexical learning by infants (Daland & Pierrehumbert, 2011), and to peripheral speech processing. But for higher-level phonological processing by adults, type frequency is more predictive than token frequency. Type frequency can only be defined by forming generalizations over an abstract phonological code, rather than directly over the surface realizations.

A second line of motivation for abstraction in phonology arises from incongruities between the information needed to capture different constraints or processes. For example, in Arabic verbal roots, identical consonants in the 2nd and 3rd positions act like a single consonant in relation to the strong constraints against homorganic sequences. However, they still act like two consonants for the process of phonetic realization (McCarthy, 1986; Frisch et al., 2004). In an analysis of mismatch negativity data obtained with an oddball paradigm, Eulitz & Lahiri (2004) find that the same acoustic-phonetic difference has different consequences for lexical access depending on the direction of the comparison, with the marked phonemes differing less from default phonemes than the other way around. The effect must be relate to the logical structure of the system rather than to frequency, because it is still found if frequency is controlled (Cornell et al., 2011). In general, accurate and general descriptions of such findings requires a minimum of three levels of representation (a phonetic level, a categorical encoding level, and an underlying or morphophonemic level). A classic work showing how much can be achieved with just these three levels is Lakoff (1993).

A further motivation for abstraction in phonology is findings that new phonetic categories are learned very slowly in comparison to the rate at which established categories can be adapted to new communicative requirements (Sankoff & Blondeau, 2007; Cutler et al., 2010; German et al., 2013). Finally, abstract mechanisms are also needed to explain how regular sound changes come to sweep through the whole vocabulary. Without them, it is not obvious why words with few lexical neighbors are not stranded behind in regular sound changes. Recognition of such results, taken together with results favoring highly detailed cognitive representations, has given rise to hybrid models that include both types of information (Pierrehumbert, 2002; Hawkins, 2003; Wedel, 2012; German et al., 2013; Sumner et al., 2014; Ernestus, 2014).

3 Context affects phonetic details

An episode is a sequence of events in time. By this definition, even a short phrase counts as an episode. But episodes can also be defined at longer time scales (such as the discourse segment or the conversation). They can also defined in a richer manner, to include nonlinguistic dimensions of experience. Thus, the issue of episodic traces in phonological representation comes down to the effects of context. These effects arise in many ways.

3.1 Phrasal context

When sentences are constructed, words are assembled in a buffer and a prosodic structure is built for the whole sequence. The locations of phrasal stress depend in a complicated way on the information structure of discourse (German et al., 2006; Schwarzschild, 1999). This means that a given word may be very prominent in one sentence, and highly subordinated in another sentence. This factor has a major impact on the f0 contour and timing of the word (Shattuck-Hufnagel & Turk, 1996). It also affects how much effort is devoted to articulating the word, with pervasive effects on the segmental characteristics (Pierrehumbert & Talkin, 1991; Keating et al., 2004).

3.2 Frequency/predictability

On the average, frequent words are produced in a less effortful manner than more unusual words (Wright, 1979; Jurafsky et al., 2002). In addition, if word frequency is controlled, words prove to be more reduced if they are highly predictable from other words in the immediate context (Aylett & Turk, 2004; Ernestus, 2014). These two effects are probably manifestations of the same underlying mechanism, one which allocates articulatory effort on the basis of how accessible words are in context. Accessibility is in turn a cumulative function of probabilistic effects at different time scales, with the overall word frequency representing the longest time scale (that of the speaker's linguistic experience to date) and word transition probabilities representing the shortest time scale. Effects on word likelihood at intermediate time scales also exist, because words are much more frequent in relation to some topics than others (Church & Gale, 1995; Altmann et al., 2009), and this factor also affects the degree of word reduction (Heller & Pierrehumbert, 2011; Baker & Bradlow, 2009).

3.3 Different voices and dialects

People's voices differ because of anatomical differences in the vocal folds and vocal tract. Even controlling for biological sex, significant differences amongst individuals still exist. For example, in a study of /r/ articulation, individuals used different articulatory strategies to achieve the characteristic low F3 value, but these strategies cause individual differences in F4 (Espy-Wilson et al., 2000). An additional layer of differences arises due to linguistic exposure during phonological learning, since phonological learning means internalizing ambient phonological patterns well enough. But a degree of conformity that is sufficient to understand and produce speech stills leaves statistical room for individual variation.

The origin of some additional individual characteristics is not fully understood. For example, people differ in their overall speech rate (Johnson et al., 1993), as well as details of timing such as consonant/vowel duration ratios and the timing of off-glides (Remez, 2010). A study of individual variation in the mixed-dialect context in the Shetland Islands during the North Sea oil boom found that two sons of the same English parents differed in their VOT patterns (Scobbie, 2006). In evaluating such differences, it can be difficult to distinguish effects of language exposure from effects of attitude and emotion. For example, brothers might have the same home environment, but differ in who they like or admire. These attitudes could cause them to imitate different people. But the same attitudes could cause them to socialize with different people and thereby experience different linguistic input outside the home.

3.4 Indexical information

In linguistic theory, there is a long-standing distinction between referential information and indexical information. Referential information is what the speaker is talking about, whereas indexical information is information about the speaker, the social context, or the physical context[1]. If individual

[1] Indexical information is also taken by many researchers to include information about the discourse context. The discourse context can affect on the referents of words and thereby affect the truth conditions for sentences; these effects obviously comprise part of the referential information. However, other parts of the discourse context, such as the

differences in speech production are systematic and reproducible, then these differences can be used by the listener to make inferences about which specific person, or what type of person, was speaking. Similarly, if differences relating to the social context or nature of the interaction are systematic, the listener can also use these to follow the social aspects of a linguistic interaction. Indexical information in speech is the main focus of socio-phonetics, and much is now known about it (Hay & Drager, 2007).

Numerous studies have examined the correlation of voice characteristics with gender. Although often grounded in anatomical difference (men have a longer pharynx, a longer vocal tract overall, and heavier vocal folds than women), differences in some languages are greater than anatomical differences would explain (Johnson, 2006), and they can emerge at younger ages than the anatomical differences are observable (Perry et al., 2001). Further, some differences have no anatomical grounding, but differ in arbitrary ways from dialect to dialect (Foulkes & Docherty, 2006), These differences are widely assumed to occur as young girls preferentially imitate adult women, and boys imitate adult men. Remarkably, Foulkes & Docherty (2006) identified an additional mechanism in their field study in Newcastle: Caregivers speak differently to female infants than to male ones, even before the infants begin to talk. Gender identity also affects pronunciation patterns. The vowel spaces of men and women who self-identify as GLB (gay, lesbian, or bisexual) exhibit statistically significant differences from those of people who self-identify as straight (Pierrehumbert et al., 2004). Boys with the (rather controversial) diagnosis of Gender Identity Disorder (Zucker & Bradley, 1995), who are extremely likely to be gay or transsexual as adults (Wallien & Cohen-Kettenis, 2008), already display systematic differences in their vowels from typically developing boys (Munson et al., 2015).

Age also affects pronunciation patterns. In addition to the physical effects of aging (Harrington et al., 2007), there are differential effects due to linguistic changes in progress. Speech patterns are most plastic when people are young. This fact is exploited in the use of *apparent time* in sociolinguistics research, in which a synchronic sample of speakers of different ages in a speech community used as a proxy for the progress of linguistic changes over time (Bailey, 2008). Detailed studies, however, have established the existence of some phonetic plasticity in mature speakers. Queen Elizabeth's speech has shifted towards the rising Southern Estuary standard, as revealed by acoustic measurements of decades of Christmas broadcasts (Harrington et al.,

emotional tone or the cooperativeness of the interaction, would naturally be included in the indexical information.

2000). In a longitudinal panel study on the shift in /r/, in Montreal French, Sankoff & Blondeau (2007) found that most adult speakers were stable but a sizeable minority made changes. Such findings mean that apparent time provides a conservative estimate of rates of change in language, but that on the average, detailed phonetic patterns are systematically related to the age of the speaker. Lastly, it has been suggested that people use different stylistic registers at different points in their lives (Holmes, 1992; Sankoff, 2004).

Dialect and social class are also major indexical features. Languages are constantly in flux, human social networks tend to have strong clustering, and people are disposed to imitate the speech patterns of the people they interact with. These facts alone predict that different systems will emerge over time in social groups that are isolated from each other. Beginning in the 1960's, field studies established correlations between social networks and pronunciation patterns (Labov, 1972; Milroy & Milroy, 1985). Very importantly, differences can persist even between groups that are in contact with each other, such as students in the same high school (Eckert, 1989; Mendoza-Denton, 2008). Some (though not all) such differences become conventionalized as indicators of social identity. This is revealed by the behavior of speakers, for example by what dialectal features are involved in style-shifting between speech with close personal associates and speech with outsiders. It is also revealed in the behavior of hearers. Purnell et al. (1999) showed that three of the principal California dialects (African American Vernacular, Chicano English, and Standard American English) could be identified well above chance just from the word *hello*, and the AAVE and ChE speakers faced discrimination in the housing market solely on the basis of the dialect used in a brief, standard, phone inquiry. Perrachione et al. (2010) show that phonological features of AAVE, and not anatomical correlates of race, are the primary determinant of how people perceive the race of speakers from voice recordings.

Indexical features interact with other sources of variation in production. Mendoza-Denton (2008) found that jocks and gang members shifted critical vowels in different directions under emphasis. Clopper & Pierrehumbert (2008) found that the vowels of words that are semantically predictable from the previous context are not only shorter, but also more dialectally shifted, than words in unpredictable contexts.

These examples indicate that dialectal and social features are not merely epiphenomenal consequences of general processes of phonological learning and accommodation. Instead, they constitute a channel of information that is conveyed concurrently with the referential information. The ability to perceive indexical features, to produce them more or less in different contexts, and to generalize them to new words and new interlocutors means that they must be cognitively represented.

4 Storage in memory

The contextual effects I have just discussed occur at relatively short time scales, as speakers use words in different contexts and listeners encounter different speakers. "Phonology" stands for cognitive representations that integrate information over many contexts, and therefore it has a longer characteristic time scale. Research about episodic effects in phonology considers two types of effects at long time scales. The first is effects of context on the way that phonological representations are formed. A second, and more speculative, possibility is the inclusion of contextual factors in these representations.

Explorations of the first idea already yield striking and significant results. These focus on evidence for storage of word-specific information that could not have been predicted solely from the abstract phonological representation of a word, but only from the range of contexts in which it appears. The results undercut previous claims that phonological representations are minimized, preserving information that is invariant across contexts while discarding information that varies.

Goldinger's landmark study of automatic imitation behavior (Goldinger, 1998) found that the pronunciations of low-frequency words were more affected by exposure during the study phase of the experiment than those of high-frequency words, an effect that he attributed to rare words having less entrenched phonetic representations than frequent words. An alternative interpretation is that rare words are more salient or noticeable than frequent words, and so more likely to have their representations updated during the experiment; but this alternative still entails that two contextual factors (how often the word is experienced in general, and the phonetic details during the experiment) impact the lexical representation.

Seyfarth (2014) explored the reduction patterns of content words in the Switchboard corpus. Controlling for overall word frequency, he found that words that generally appear in more predictable contexts (as indexed from word transition probabilities) are more reduced than other words, even when they occur in unpredictable contexts. Walker & Hay (2011) compare lexical access times for words statistically associated with different ages of speakers in the ONZE corpus. The stimuli were produced an older speaker (age 50), and a younger speaker (age 22). Lexical access for older words such as *knitting* was facilitated when produced by the older speaker, and access for younger words such as *sexist* was faciliated when produced by the younger speaker.

Schweitzer et al. (2015) studied the variability of intonation patterns on multi-word sequences in relation to the contextual probability of the accented word. In German and English, intonation is assigned at the phrasal level, communicating pragmatic dimensions of meaning. One would expect that any word sequence, including collocations, could in principle be produced with any pattern. Schweitzer et al. (2015) show, however, that collocations have less variable f0 contours than unpredictable sequences, indicating that differential use of intonational patterns is stored in association with collocations. The data are not sufficient to pinpoint exactly what is stored (whether the phonetic form of the intonation, its phonological representation, or even the components of pragmatic meaning). All of these alternatives, however, represent contextually triggered information that goes beyond a standard phonological representation of a word in an intonational language.

Further evidence about indexical effects on memory formation comes from experiments on processing of dialectal forms. In a repetition-memory task, Sumner & Samuel (2009) found that non-rhotic primes did not produce as much lexical facilitation as rhotic primes for American English listeners who spoke a rhotic dialect, but were familiar with the non-rhotic dialect. In a study of listeners tested in Ohio, Clopper et al. (under review) found that words spoken in a Northern Cities dialect failed to produce any repetition benefit even though they were just as intelligible as words spoken in the more standard and contextually relevant Midlands dialect. This was true even for the Northern Cities speakers in the participant pool. The processing difficulty for the less standard forms (also evidenced by longer reaction times) evidently weakened the ability to form memories. Results like these mean that phonological representations are not just generalizations over experienced speech. Instead, the experience is filtered by prior experience, which shapes what new experiences count as more familiar or contextually relevant.

Experienced speech also proves to be filtered by social evaluations. Sumner & Kataoka (2013) found better facilitation for the British non-rhotic dialect, which is prestigious for speakers of American English, than for the mid-Atlantic dialect, which is less prestigious. In an ingenious study involving loanwords for beer and gelato, Lev-Ari & Peperkamp (2014) showed that the extent of phonological adaptation of the wordforms depended on the relative prestige of the donor language for the semantic domain of the particular product.

A new study by Hay et al. (2015) addresses the long-standing dispute about whether regular (or "Neogrammarian") sound changes affect all words equally and concurrently, or whether some words lead and others lag while the change is in progress. The study explores the short front vowel rotation in New Zealand English (NZE), a notable feature of NZE that can cause misunderstandings by speakers of other dialects: NZE *bat* often sounds to speakers from elsewhere like *bet*, NZE *bet* sounds like *bit*, and NZE *bit* often sounds like *but*. Using acoustic measurements from the Origins of New Zealand English database (Gordon et al., 2007), which encompasses speakers with birthdates spanning 130 years, they identify a statistical interaction between word frequency and speaker's year of birth (which is a conservative proxy for time in tracing the sound change). While the change is in progress, low frequency words are in the lead. Conforming to claims in Pierrehumbert (2006a), the effect is small in comparison to the within-category variation for each vowel, which is why such a large sample is needed to find it. The authors propose a mechanism based on the fact that push chains create regions of ambiguity when one vowel advances into the space of the other. Low frequency words are known to be at a disadvantage in recognition if the signal is ambiguous. Because of attrition in storing the low frequency words near the frontier of the change, the center of mass for memories of these words is further from the frontier than for high frequency words, as illustrated in Figure 3.

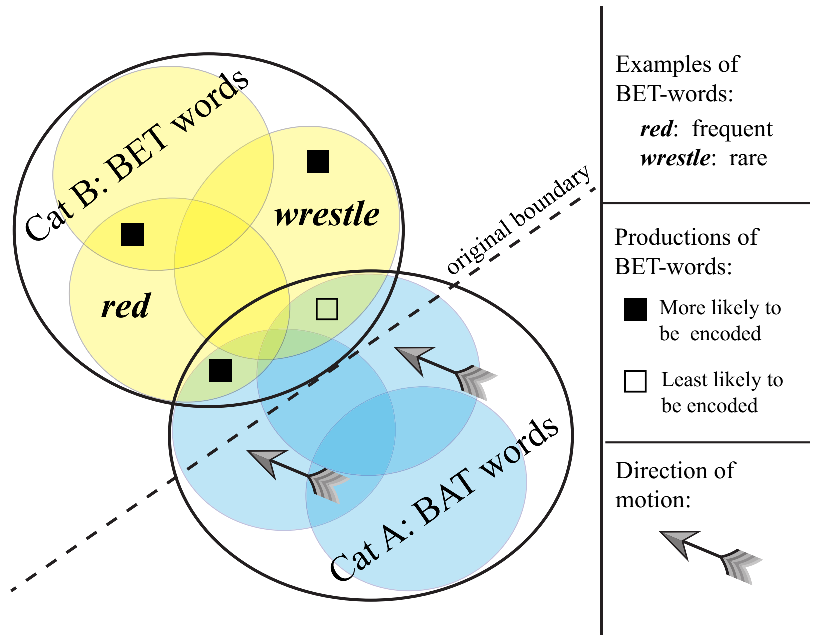

Figure 2: **Schematic diagram of the mechanism for word frequency effects in a push chain.** A shifting distribution of the low front vowel in BAT words (Category A) encroaches on the distribution for the mid front vowel in BET word (Category B). The colored ellipses represent three examples of lexical 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. (Reproduced from Hay et al. (2015))

The pattern for this contrast-preserving push chain differs from that for a leniting change in the same database, in which high-frequency words lead the change (Hay & Foulkes, 2015). Subtle anti-homophony effects in historical change are also reported in Wedel (2006, 2012).

It is impossible for classic, modular, theories of language sound structure to capture any of the findings I have just reviewed. In such theories, the phonetic realization of a word is entirely determined by its phonological representation. There is no mechanism for word-specific phonetic memories to arise and play a role in subsequent processing. The findings entail that people must have very detailed phonetic memories for specific words. Furthermore, continual updating of their mental representations is influenced by episodic factors: how often a word is encountered, the characteristics of the social and discourse context in which it was encountered, and also experiences with other words having the same phoneme or a contrasting phoneme.

A stronger claim about episodic traces in lexical representations is the claim that episodic information is encoded and stored in association with phonological information. Classic findings about the interaction between talker identification and word identification provide a point of departure for evaluating this claim. Nygaard et al. (1994) show that training people to identify a set of voices results in better word recognition for novel words presented in noise produced in the same voices. Nygaard & Pisoni (1998) extended this finding to show that listener training using full sentence materials is needed to obtain an advantage for processsing novel sentences. Bradlow et al. (1999) showed that recognition memory for words is better for words produced in the same voice than in a different voice. At first blush, such findings could merely be taken to mean that phonetic memories are extremely detailed, that tokens of a word spoken by the same speaker are more similar to each other than tokens produced by someone else, and that this greater similarity facilitates lexical retrieval. And indeed, facilitory effects have been established for voices that are similar to each other (e.g in dialect or gender), but not identical. However, it is already doubtful that the words produced by each speaker in Nygaard's experiment are all similar enough to each other (in comparison to acoustic differences amongst words) to yield the observed effects. Recently, Pufahl & Samuel (2014) has used the same methodology to show that even ambiant sounds, such as car horns or bird calls, are associated with experienced words in memory.

Other recent studies ask more directly whether activating indexical information impacts speech perception or production in ways that cannot be captured by acoustic similarity alone. A series of experiments by Hay and colleagues have explored how non-speech indexical cues affect speech processing. Hay & Drager (2010) found that New Zealand listeners shifted their vowel classification patterns depending on whether stuffed kangaroos and koalas (associated with Australia) or stuffed kiwis (associated with New Zealand) were present in the room. In a corpus study and controlled word-reading task, Sanchez et al. (2015) found that the production of New Zealand lax front vowels is affected by the presence of words associated with Australia (for example, place names) in the context. Drager et al. (2010) showed that New Zealanders produced vowels in words like *kit* differently depending on whether they had been exposed to positive or negative information about Australia, or neutral information. People's attitudes about Australia interacted with the contextual information provided, showing that the processing of indexical information is not solely a matter of topical priming, but depends on social factors as well.

German et al. (2013) explore the interaction of word-specific and general learning in a study in which Midwestern college students attempted to imitate the dialect of English spoken in Glasgow, which was completely unfamiliar to them. In the training phase, they tried to imitate 48 different sentences spoken by a Glaswegian speaker. A test phase explored the level of generalization to new materials. The participants returned a week later for a retest. Unbeknownst to the subjects, the features of interest were the allophones of /t/ and /r/, with the training materials including 12 sentences exemplifying each of four cases: /r/-initial words, /t/-initial words, /r/-medial words, and /t/-medial words. In Glaswegian, medial /t/ is aspirated instead of flapped, and /r/ is a flap rather than an approximant. In the retest phase, memories of the Glaswegian speaker were reactivated by the use of filler blocks comprised of sentences without any /t/s or /r/s whatsoever.

All participants could modify the probability of a known allophone, aspirated /t/, reliably and immediately, achieving over 95% accuracy for words like *notice* in the first training block. Using a [ɾ] in words like *worries* is more difficult, as it involves remapping the relationship between an allophone and a phoneme. Accuracy improved from the first to the second training block, and some participants were not able to adapt in this way. However, both of these adaptations were extremely rapid and successful compared to learning a new phonological category. None of the participants who undertook to create a novel category did so accurately. Possibly the most important results of this study come from the retest results a week after the initial training. When participants returned to the lab, they were quite successful at generalizing the patterns to a fresh set of new materials despite encountering no new training items. There was a statistically significant, but relatively small, advantage for training materials over new items. The fact that this advantage exists indicates that participants formed indexical associations to allophonic patterns for specific words. The fact that it was small indicates greedy generalization from words to the phonological grammar.

Like the perception study of Cutler et al. (2010), this production study demonstrates the relevance of episodic information to phonological processing. The system is adapted in response to the patterns produced by a particular speaker; the adaptation is not a permanent change to the system, but can be triggered afresh by new exposure to the same speaker. At the same type, both studies indicate that fast adaptation is dominated by processing at a general, abstract level. The word-specific effects in Hay et al. (2015) were small, and so were the word-specific effects in German et al. (2013).

5 Episodic effects in morphophonology

The debate about episodic versus abstractionist approaches to phonology has mainly focused on the relationship between the phonological encoding of words and their fine phonetic details. But phonetic realization is far from the whole of phonology. Phonology also encompasses morphophonology: the similarities and differences between morphologically related words, and the surface representations of words created by combining morphemes in novel ways. Although many morphological alternations are predictable from phonotactic constraints, others are specific to particular morphemes and must be learned as generalizations over pairs or paradigms of words (Pierrehumbert, 2006b). This means that episodic influences on what words people know, and what indexical meanings they associate with these words, provides another potential avenue for episodic effects in phonology.

Individual differences in vocabulary are a mainstay of research on authorship attribution in the digital humanities and in natural language processing. These include differences in the frequencies of particular words (notably, function words) and differences in more general lexical properties such as word length and lexical richness (Tweedie & Baayen, 1998; Zheng et al., 2006). Starting from the point of view of words in on-line discussion groups, Altmann et al. (2011) show that most words are not used with the same frequencies by most of the speech community. Instead, different people talk about different topics, and people also prefer to use different words to talk about any given topic.

These very evident differences in vocabulary lead one to expect that people and groups of people would differ in their morphological systems. This prediction is confirmed by studies of dialect and language change. Differences between African American Vernacular morphology and general American morphology have been extensively studied; many educated AAVE speakers can code-switch between the two systems depending on the social situation (Labov, 1972; Green, 2002). Socio-geographical variation in morphological productivity is also found in Dutch (Keune et al., 2006), and no doubt in other languages as well. Effects of gender, social class, and speech register have also been documented (Plag et al., 1999; Nevalainen & Raumolin-Brunberg, 2003).

The consequences of this situation for the cognitive representation of phonology are rather unexplored. By their nature, such effects would be both episodic (because they would incorporate indexical information or other contextual information) and abstract (because they would affect the existence and productivity of alternations amongst phonological categories). Some groundwork for approaching this issue is provided by two recent experiments on social factors in morphological productivity: Rácz et al. (2014) and Rácz et al. (under review). Rácz et al. (2014) obtained forced-choiced judgments about the past tense forms for a large number of nonce verbs that were designed to phonologically resemble irregular English verbs, such as *sing/sang; burn/burnt; keep/kept; drive/drove; cut/cut*. The mathematical modeling compares the role played by lexical gang effects (phonological similarity to existing clusters of forms) and general rules. A systematic difference of gender was also found, with males categorizing verbs more strongly on the basis of lexical support than females to. These results imply that the productivity of minor phonological alternations (such as those involved in the irregular verb forms) might be used as a source of indexical information.

Rácz et al. (under review) reports an artificial language learning experiment in which participants were exposed to a diminutive formation pattern that varied according to the phonological context; the sex, age or ethnicity of the interlocutor; or the interlocutor's physical orientation of the interlocutor. Participants were quite successful in learning associations with phonological context and gender, age or ethnicity. They were unsuccessful in learning associations the view of the interlocutor. In the phonological condition, generalization to new words was observed. In the other conditions, generalization of specific words to new people with the same social characteristics was found, but for the previously seen interlocutors, there was no significant generalization to new words. These results can be compared to results reviewed above regarding the effects of speaker and dialect familiarity on word recognition. Both results indicate that indexical information is associated with wordform information, for individual words. But the association of indexical information with general word formation patterns is an additional step, that needs to be evaluated separately.

6 Conclusions

Any utterance incorporates both propositional information and contextual information, including information about the speaker and information about the circumstances in which the speech occurred. The debate about abstractionist versus episodic theories of phonology is reminiscent of the debate between Gestalt and compositional theories in visual perception, insightfully reviewed in Wagemans et al. (2012). By using modifications of the Garner paradigm (Garner, 1974), classic studies in speech processing showed that lexical and indexical dimensions are not fully separable (Bradlow et al., 1999; Goldinger, 1996; Nygaard et al., 1994; Nygaard & Pisoni, 1998); see also the extensive review in Johnson (2005). Incomplete separation of these dimensions is found not only in perception, but also in memory and production. Further relevance of Gestalt principles to the theory of language sound structure is suggested by the many cases in which phonological properties prove to be configurational, when considered in their full range of variability. Such cases include the distinction between /r/ and /t/, and the vowels involved in the New Zealand English push chain.

Experimental studies of speech perception and speech production, as well as sociolinguistic field studies and analysis of archival recordings, have provided unequivocal evidence that mental representations of phonological forms are extremely detailed. They include word-specific phonetic characteristics that have arisen from contextual factors, as well as traces of individual voices or types of voices. These effects cannot be captured in strongly abstractionist models, in which phonological information should be completely separable from indexical information and other sorts of contextual information. At the same time, word-specific effects appear to be small in relation to the overall variability within phonological categories. In tasks that require generalizing from experienced words to new words, the fastest and most reliable effects involve adapting pre-existing abstract phonological categories. Such tasks include adapting to new dialects, a challenge for which the first line of attack is recycling known abstract categories rather than learning new ones.

Taken together, these results indicate that early abstractionist theories of phonology were wrong in thinking that mental representations of phonology are minimized. The mental representations include a vast amount of detail, including word-specific phonetic properties and general phonetic patterns relating to indexical features. As a corollary, the representations must in fact be highly redundant. At the same time, the early abstractionists were to some extent correct with their ideas about modularity. Positing abstract representations is crucial to understanding people's ability to generate and process novel forms. However, the modules are leaky. Leakage crosses levels of representation (associating the lexical representations of words to their phonetics), and also associates indexical features with words and phonological patterns.

7 Acknowledgments

I'm very grateful to James German for creating Figure 1 from the original data for German et al. (2013). I'm also grateful for support from John Templeton Foundation Award No. 36617. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

References

Altmann, EG, JB Pierrehumbert & AE Motter. 2009. Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE* 4. doi:DOI: 10.1371/journal.pone.0007678.

Altmann, EG, JB Pierrehumbert & AE Motter. 2011. Niche as a determinant of word fate in on-line groups. *PLoS ONE* 6(5). doi: 10.1371/journal.pone.0019009.

Aylett, M & AE Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47. 31 – 56.

Baayen, RH. 2008. *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press.

Bailey, G. 2008. Real and apparent time. In J K Chambers, P Trudgill & N Schilling-Estes (eds.), *The handbook of language variation and change*, Blackwell Publishing Ltd.

Baker, RE & AR Bradlow. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52(4). 391–413.

Bradlow, AR, LC Nygaard & DB Pisoni. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics* 61(2). 206–219.

Church, K & WA Gale. 1995. Poisson mixtures. *Nat Lang Eng* 1. 163–190.

Clopper, CG & JB Pierrehumbert. 2008. Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America* 124(3). 1682–1688.

Clopper, Cynthia G, TN Tamati & JB Pierrehumbert. under review. Variation in the strength of lexical encoding across dialects .

Cornell, SA, A Lahiri & C Eulitz. 2011. What you encode is not necessarily what you store: Evidence for sparse feature representations from mismatch negativity. *Brain Research* 79–98.

Cutler, A, F Eisner, JM McQueen & D Norris. 2010. How abstract phonemic categories are necessary for coping with speaker-related variation. In *Laboratory phonology 10*, 91 – 113. De Gruyter Mouton.

Daland, R & JB Pierrehumbert. 2011. Learning diphone-based segmentation. *Cognitive Science* 1. 119 – 155.

Drager, K, JB Hay & AJ Walker. 2010. Pronounced rivalries: attitudes and speech perception. *Te Reo* 53. 27 – 53.

Eckert, P. 1989. *Jocks and burnouts: Social categories and identity in the high school*. New York: Teachers College, Columbia University.

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

Espy-Wilson, CY, SE Boyce, M Jackson, S Narayanan & A Alwan. 2000. Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America* 108. 343 – 356.

Eulitz, C & A Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16. 577 – 583.

Foulkes, P & G Docherty. 2006. The social life of phonetics and phonology. *J. Phonetics* 34. 409–438.

Frisch, SA, NR Large, B Zawaydeh & DB Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. In J Bybee & P Hopper (eds.), *Proceedings of the symposium on frequency effects and emergent grammar*, pp. 159180. Amsterdam.

Frisch, SA, JB Pierrehumbert & MB Broe. 2004. Similarity avoidance and the ocp. *Natural Language and Linguistic Theory* 22(1). 179–228.

Garner, WR. 1974. *The processing of information and structure*. Erlbaum.

German, JS, K Carlson & JB Pierrehumbert. 2013. Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics* 41(3). 228 – 248.

German, JS, JB Pierrehumbert & S Kaufmann. 2006. Evidence for phonological constraints on nuclear accent placement. *Language* 82(1). 151–168.

Goldinger, SD. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology-Learning Memory and Cognition* 22(5). 1166–1182.

Goldinger, SD. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251 – 279.

Gordon, E, M Maclagan & JB Hay. 2007. The ONZE corpus. In *Creating & digitizing language corpora, Diachronic databases, 2*, 82104. Palgrave Macmillan.

Green, LJ. 2002. *African American English: A linguistic introduction*. Cambridge University Press.

Harrington, J, S Palethorpe & CI Watson. 2000. Does the queen speak the Queen's English? *Nature* 408. 927–928.

Harrington, J, S Palethorpe & CI Watson. 2007. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Proceedings of interspeech 2007*, .

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

Hay, JB & K Drager. 2007. Sociophonetics. *Annual Review of Anthropology* 36. 89–103.

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

Hay, JB & P Foulkes. 2015. The evolution of medial /t/ in real and imagined time. *Language* .

Hay, JB, JB Pierrehumbert & ME Beckman. 2004. Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in Laboratory Phonology VI*, 58–74. Cambridge Univ. Press.

Hay, JB, JB Pierrehumbert, AJ Walker & PJ LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139. 83–91.

Heller, J & JB Pierrehumbert. 2011. Word burstiness improves models of word reduction in spontaneous speech. In *Amlap 2011*, Architectures and Mechanisms for Language Processing (AMLaP) 2011.

Holmes, J. 1992. *An introduction to sociolinguistics.* Longman.

Hyman, L. 1970. How concrete is phonology. *Language* 46(1).

Johnson, K. 2005. Speaker normalization in speech perception. In DB Pisoni & R Remez (eds.), *The handbook of speech perception*, 363–389. Blackwell Publishers.

Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34(4). 485– 499.

Johnson, K, P Ladefoged & M Lindau. 1993. Individual differences in vowel production. *Journal of the Acoustical Society of America* 94. 701 – 714.

Jurafsky, D, A Bell & C Girard. 2002. The role of lemma in form variation. In *Laboratory phonology VII*, Mouton de Gruyter.

Keating, P, T Cho, C Fougeron & C-S Hsu. 2004. Domain-initial articulatory strengthening in four languages. In *Papers in Laboratory Phonology VI*, 145–162. Cambridge Univ. Press.

Kelly, J & JK Local. 1986. Long domain resonance patterns in english. In *International conference on speech input/ output: Techniques and applications. Conference publication no. 258*, 304–9. Institute of Electrical Engineers.

Keune, K, R Van Hout & RH Baayen. 2006. Socio-geographical variation in morphological productivity in spoken Dutch: a comparison of statistical techniques. In *Proceedings JADT 2006.*, .

Kiparsky, RPV. 1973. How abstract is phonology? In Osamu Fujimura (ed.), *Three Dimensions of Linguistic Theory*, TEC.

Klatt, DH. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59(5). 1208–1221.

Labov, W. 1972. *Language in the inner city: Studies in the Black English Vernacular*. University of Pennsylvania Press.

Lakoff, G. 1993. Cognitive phonology. In JA Goldsmith (ed.), *The last phonological rule*, 117–145. University of Chicago Press.

Lev-Ari, S & S Peperkamp. 2014. An experimental study of the role of social factors in language change: The case of loanword adaptations. *Laboratory Phonology* 5(3). 379 – 401.

McCarthy, JJ. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17(2). 207–263.

Mendoza-Denton, N. 2008. *Home girls*. Oxford, UK and Cambridge, MA: Blackwell.

Mielke, J. 2008. *The emergence of distinctive features*. Oxford University Press.

Milroy, J & L Milroy. 1985. Linguistic change, social network, and speaker innovation. *J Linguistics* 21. 339 – 384.

Munson, B, L Crocker, JB Pierrehumbert, A Owen-Anderson & KJ Zucker. 2015. Gender typicality in children's speech: A comparison of the speech of boys with and without gender identity disorder. *J Acoust Soc Am* 137.

Nevalainen, T & H Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Longman.

Nosofsky, RM. 1992. Similarity scaling and cognitive process models. *Annual Review of Psychology* 43. 25–53.

Nygaard, LC, MS Sommers & DB Pisoni. 1994. Speech perception as a talker-contingent process. *Psychological Science* 5. 42–46.

Nygaard, Lynne C & DB Pisoni. 1998. Talker-specific learning in speech perception. *Perception and Psychophysics* 60(3). 355–376.

Perrachione, TK, JY Chiao & PCM Wong. 2010. Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition* 114(1). 42 – 55.

Perry, TL, RN Ohde & DH Ashmead. 2001. The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America* 109. 2988–2998.

Pierrehumbert, JB. 2000. The phonetic grounding of phonology. *Bulletin de la communication parlée* 5. 7 – 23.

Pierrehumbert, JB. 2002. Word-specific phonetics. In *Laboratory Phonology VII*, 101–139. Mouton de Gruyter.

Pierrehumbert, JB. 2006a. The next toolkit. *J. Phonetics* 34(6). 516–530.

Pierrehumbert, JB. 2006b. The statistical basis of an unnatural alternation. In *Laboratory Phonology VIII: Varieties of phonological competence*, 81–107. Mouton de Gruyter.

Pierrehumbert, JB, T Bent, B Munson, AR Bradlow & JM Bailey. 2004. The influence of sexual orientation on vowel production. *J Acoust Soc Am* 116(4). 1905–1908.

Pierrehumbert, JB & D Talkin. 1991. Lenition of /h/ and glottal stop. In *Papers in Laboratory Phonology II*, 90–117. Cambridge Univ. Press.

Plag, I, C Dalton-Puffer & RH Baayen. 1999. Productivity and register. *Journal of English Language and Linguistics* 3. 209–288.

Pufahl, A & AG Samuel. 2014. How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology* 70(1). 1–30.

Purnell, T, W Idsardi & J Baugh. 1999. Perceptual and phonetic experiments on american english dialect identification. *Journal of Language and Social Psychology* 18. 10–30.

Rácz, P, C Beckner, JB Hay & JB Pierrehumbert. 2014. Rules, analogy, and social factors codetermine past-tense formation patterns in English. In *Proceedings of the 52nd annual meeting of the Association for Computational Linguistics*, .

Rácz, Peter, JB Hay & JB Pierrehumbert. under review. Sociolinguistic variation and social language learning: an experimental study .

Remez, RE. 2010. Spoken expression of individual identity and the listener. In E Morsella (ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, 167 – 181. Psychology Press.

Richtsmeier, PT. 2011. Word-types notword-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology* 2. 157–183.

Sanchez, K, JB Hay & E Nilson. 2015. Contextual activation of australia can affect new zealanders' vowel productions. *Journal of Phonetics* 48. 76 – 95.

Sankoff, G. 2004. Adolescents, young adults and the critical period: Two case studies from seven up. In C Fought (ed.), *Sociolinguistic variation: Critical reflections*, 121–139. Oxford University Press.

Sankoff, G & H Blondeau. 2007. Language change across the lifespan; /r/ in Montreal French. *Language* 83(3). 560 – 588.

Schwarzschild, R. 1999. AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7. 141–177.

Schweitzer, K, M Walsh, S Calhoun, H Schütze, B Möbius & et al. 2015. Exploring the relationship between intonation and the lexicon: Evidence for lexicalized storage of intonation. *Speech Communication* 66. 65–81.

Scobbie, J. 2006. Flexibility in the face of incompatible vot systems. In *Laboratory Phonology VIII*, 367 – 392. Mouton de Gruyter.

Seyfarth, S. 2014. Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. *Cognition* 133. 140 – 155.

Shattuck-Hufnagel, S & AE Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25. 193–247.

Stampe, D. 1979. *A dissertation on natural phonology.* Garland.

Sumner, M & R Kataoka. 2013. Effects of phonetically-cued talker variation on semantic-encoding. *J Acoust Soc Am* 134. EL485 – EL491.

Sumner, M, S-K Kim, E King & KB McGowan. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology* 4. doi:doi: 10.3389/fpsyg.2013.01015.

Sumner, M & AG Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60. 487–501.

Tulving, E. 2002. Episodic memory: From mind to brain. *Annual Review of Psychology* 53. 1–25.

Tweedie, FJ & RH Baayen. 1998. Measures of lexical richness in perspective. *Computers and the humanities* 32. 323 – 352.

Wade, T & B Möbius. 2010. Detailed phonetic memory for multi-word and part-word sequences. *J Lab Phonology* 1. 283–294.

Wagemans, J, J Feldman, S Gepshtein, JR Pomerantz, Peter A van der Helm & Cees van Leeuwen. 2012. A century of Gestalt psychology in visual perception ii: conceptual and theoretical foundations. *Psycho Bull* 138(6). 1218 – 1252.

Walker, A & JB Hay. 2011. Congruence between word age and voice age facilitates lexical access. *Laboratory Phonology* 2. 219–237.

Wallien, MSC & PT Cohen-Kettenis. 2008. Psychosexual outcome of gender dysphoric children. *Journal of the American Academy of Child and Adolescent Psychiatry* 47. 1413–1423.

Wedel, AB. 2006. Exemplar models, evolution and language change. *The linguistic review* 23(3). 247–274.

Wedel, AB. 2012. Lexical contrast maintenance and the organization of sublexical systems. *Language and Cognition* 4(4). 319 – 355.

Wright, CE. 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.

Zellou, G & M Tamminga. 2014. Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics* 47. 18–35.

Zheng, R, J Li, H Chen & Z Huang. 2006. A framework for authorship identification of on-line messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology* 57(3). 378 – 393.

Zucker, KJ & SJ Bradley. 1995. *Gender identity disorder and psychosexual problems in children and adolescents*. Guilford Press.

Zue, VW & M Laferriere. 1976. Acoustic study of medial /t,d/ in English. *Journal of the Acoustical Society of America* 66(4). 1039–1050.