Papers in Historical Phonology
http://journals.ed.ac.uk/pihph
ISSN 2399-6714
Volume 7, 81-101
DOI: 10.2218/pihph.7.2022.8914

Frequency-predicted shifts independent of word-specific phonetic details

Chelsea Sanker
Stanford University

Abstract

Some sound changes seem to proceed at different rates depending on lexical frequency; these are often interpreted as reflecting phonetically detailed exemplar memories, with changes spreading via lexical diffusion (Pierrehumbert 2002; Bybee 2012). However, such patterns do not necessarily require word-specific phonetic details. Variation associated with lexical frequency also exists when there is no evidence for a change in progress, which might be explained by the process of lexical access: Higher lexical frequency facilitates activation, causing faster and more reduced productions (Gahl et al. 2012; Kahn & Arnold 2012; Jurafsky et al. 2002). This work examines how repeated exposure to word-specific VOT manipulations influences listeners' category boundary between aspirated and unaspirated stops in those words. Listeners' VOT category boundary is lowered after exposure to shortened VOT stimuli and also after exposure to lengthened VOT stimuli. These results suggest that frequency-related sound change can largely be explained by frequency directly influencing reduction in phonetic implementation and perceptual access. The size of the effect differed based on the acoustic characteristics of the exposure stimuli; this may suggest a role of word-specific phonetic details, but could also reflect different levels of activation due to the prototypicality of the stimuli.

1 Introduction

Some sound changes seem to depend on lexical frequency, potentially reflecting word-specific exemplar memories and lexical diffusion (Pierrehumbert 2002; Bybee 2012). However, there is also synchronic variation across words, with greater reduction in higher frequency words (Pluymaekers et al. 2005; Gahl 2008). This synchronic variation does not always align with sound changes in progress; frequency effects can be stable. Rather than depending on word-specific acoustic details, frequency conditioning might be explained by higher frequency facilitating activation, causing more reduction (Kahn & Arnold 2012; Gahl et al. 2012; Jurafsky et al. 2002). Some effects of a word's frequency or informativity persist even when contextual factors are accounted for (Tang & Shaw 2021; Seyfarth 2014), and listeners more accurately identify reduced high-frequency words than reduced low-frequency words (Connine et al. 2008; Pitt et al. 2011). These patterns might suggest that effects of frequency can become part of the word's representation if they are sufficiently consistent in a speaker's experience of that word. However, it is also possible that higher frequency creates higher resting activation for a word, independent of context, leading to faster activation and thus more reduction in production and more tolerance for poor acoustic matches in perception. Experimental data can help test whether word-specific acoustic details are necessary to account for apparent frequency-conditioned sound changes, or if an alternative account can also capture these patterns.

Previous work has shown that voice onset time (VOT) is shorter in higher frequency words than in lower frequency words (e.g. Yao 2009; Chodroff & Wilson 2017). This study investigates whether frequency itself is impacting expected VOT or if listeners are remembering the particular VOT that they have heard for each lexical item, by manipulating both (a) local frequency, based on repeated exposure and (b) the VOT in exposure items. Subsequent testing of listeners' perceptual category boundaries reveals an effect of both of the manipulated factors.

1.1 Word-specific acoustic details

Some phonological models allow individual words to be associated with particular phonetic details. However, most of the observed patterns are also consistent with alternative explanations based on speed of access. Much of the evidence used in support of word-specific phonetic details comes from patterns that are correlated with lexical frequency, which will be the focus here, though there are also patterns associated with factors like morphological complexity (Plag et al. 2017; Seyfarth et al. 2018) and emotional valence (Nygaard et al. 2009; Nygaard & Lunders 2002), which might reflect different underlying processes.

It is well established that aspects of speech production like duration are correlated with lexical frequency (e.g. Pluymaekers et al. 2005; Gahl 2008) and informativity more broadly (Seyfarth 2014; Tang & Shaw 2021). However, it is not clear how these patterns ought to be analyzed. Some analyses attribute these differences to listener-oriented speech planning;

listeners do not need as many cues to identify the presence of a high probability word (Aylett & Turk 2004; Pate & Goldwater 2015). Reduction of high frequency words might alternatively be an automatic effect of retrieval; higher frequency words quickly reach high activation, and higher activation results in faster productions (Gahl et al. 2012; Kahn & Arnold 2012; Jurafsky et al. 2002). These accounts are important in providing an explanation for how effects of frequency arise. Some accounts combine automatic effects of informativity with word-specific phonetics; words have typical acoustic characteristics that are produced by context, but these acoustic details eventually become associated with the particular word (Seyfarth 2014; Tang & Shaw 2021).

Several perception experiments demonstrate that listeners make lexical decisions more accurately when an acoustic token has the characteristics that are typical of the word, e.g. presence or absence of schwa (Connine et al. 2008) and the realization of an underlying /t/ (Pitt et al. 2011). However, these effects do not necessarily indicate word-specific phonetics. The main result in these studies is that listeners are more tolerant of reduction in higher frequency words than in lower frequency words; there is no evidence that reduced forms are perceived more accurately or more quickly than unreduced forms, even in high-frequency words where the reduced form is common. This tolerance for reduction in higher frequency words can be predicted by higher resting activation levels. Higher resting activation means that high frequency words are retrieved more rapidly (Luce & Pisoni 1998), which produces a bias towards identifying ambiguous acoustic input as being higher frequency (Howes 1957; Savin 1963; Connine et al. 1993). Consistent with this bias being an effect of resting activation levels rather than listeners having a broader range of pronunciations reflected in the exemplar memories of higher frequency words, the response bias towards higher frequency words is increased in masking noise (Sommers et al. 1997). That is, the frequency bias is stronger when the acoustic input is more ambiguous and thus has less of an influence on the activation of different words.

Another line of evidence that is interpreted as reflecting word-specific phonetic details comes from convergence. Some studies have found more convergence in lower frequency words than in higher frequency words (Goldinger 1998; Babel 2010; Nielsen 2011; Dias & Rosenblum 2016). This is usually explained with exemplar models (e.g. Goldinger 1998): The exemplars from the task make up a large proportion of the overall exemplars for a lower frequency word, but a small proportion of the overall exemplars for a higher frequency word, so they have a larger impact on lower frequency words. However, other studies have failed to replicate the effect of lexical frequency on convergence (e.g. Pardo et al. 2013; Pardo et al. 2017). When the acoustic characteristics of exposure stimuli are consistent (e.g. increased VOT in the initial stop of all words), shifts are extended to the same sound in words that were not part of the exposure (e.g. Nielsen 2011); this does not disprove the possibility of word-specific phonetics, but does provide strong evidence for shared acoustic targets at a phoneme level. The role of this shared phonological level must be considered when interpreting variation in measured convergence for different words in experiments which present all exposure stimuli with the same acoustic manipulation. Sanker (2021) directly tests word-specific convergence by manipulating different words in different directions (e.g. raised F2 in one word and lowered F2 in another); this experiment finds no evidence for word-specific convergence. Mere repetition increases similarity between talkers more for lower frequency words than for higher frequency words, which could create the apparent correlation between lexical frequency and convergence, providing an alternative to the explanation based on word-specific phonetics (Sanker 2021). On the other hand, Rochet-Capellan & Ostry (2011) were able to elicit distinct shifts in three different words based on altering listeners' auditory feedback for F1 differently for each word. This result might depend on the small number of words and the large number of repetitions for each; it is unclear if a study using more words with fewer repetitions could find a similar pattern.

1.2 Regularity of sound change

The status of word-specific phonetic details has implications for sound change; if individual words can have distinct acoustic targets, irregular sound changes should be common. While there is debate about how well the regularity assumption holds up and whether certain diachronic developments might best be explained as irregular sound changes, the majority of reconstructed sound changes are regular (Ringe & Eska 2013:79-82), and other cases can usually be explained with analogy, contact between dialects of the same language, or other processes (Harrison 2003; Garrett 2015). However, some diachronic developments do pose issues for the Neogrammarian hypothesis (e.g. Blevins & Wedel 2009).

Under some analyses, changes in progress are irregular as they spread through the lexicon but ultimately produce regular outcomes most of the time (Harrison 2003; Bybee 2002; Wang & Cheng 1977). Such analyses often point to lexical frequency as evidence, because some changes seem to progress at different rates based on lexical frequency, appearing first in high frequency words (Bybee 2002; Hay et al. 2015; Pierrehumbert 2002) or occasionally appearing first in low frequency words (Todd et al. 2019). Hay et al. (2015) analyze frequency effects as reflecting detailed phonetic memories of individual words: Ambiguous tokens are more likely to be misperceived and less likely to be stored, and higher frequency words are more likely to be accurately perceived despite potential ambiguity, so lower frequency words are more resistant to shifts that encroach on another phonological category. On the other hand, when a category is moving away from another, then greater recognizability of ambiguous high frequency words would cause low frequency words to shift more rapidly (Todd et al. 2019).

However, frequency effects exist even when there is no evidence for a change in progress (e.g. Gahl 2008; Bell et al. 2009) and independently of the direction of change among changes in progress (Dinkin 2008). As discussed above, speech production is influenced by factors like lexical frequency (Pluymaekers et al. 2005; Gahl 2008) and predictability in context (Seyfarth 2014; Tang & Shaw 2021; Jurafsky et al. 2002), which produce variation across the lexicon. The existence of variation does not necessarily indicate that the category is shifting or splitting. Reduction-driven shifts may first be apparent in high-frequency words, but if this reduction leads to a shift, the prototype for the entire phonological category is shifting based on these items changing the center of the distribution, rather than the target phonetic details changing just for specific words. The phonological category can still have a consistent target across words, with differences caused by processes in lexical access (Gahl et al. 2012; Kahn & Arnold 2012; Jurafsky et al. 2002). Thus, apparent frequency-sensitive sound changes can be explained without requiring each word to have independent phonetic details.

What different predictions are made by different models for frequency effects in sound change? If phonetic details are stored individually for each word (Goldinger 1998; Pierrehumbert 2002; Bybee 2012), then it should be relatively common for each word to undergo a shift not shared with other words, particularly if phonology is analyzed as emergent rather than an underlying structure that links phonemes shared across words (e.g. Johnson 1997; Goldinger 1998; Arnold et al. 2017). If variation correlated with frequency is an automatic consequence of lexical storage and retrieval and no word-specific phonetic memories exist (Gahl et al. 2012; Kahn & Arnold 2012; Jurafsky et al. 2002), then sound changes should all be regular. Ringe & Eska (2013) show that most sound changes are indeed regular, and that most apparent exceptions can be explained with analogy or other processes. However, some exceptions are a challenge for alternative explanations. Blevins & Wedel (2009) examine several diachronic developments which seem to require an analysis as irregular sound changes, in which a merger is resisted in words where the merger would produce homophones that are not usually disambiguated by context. They analyze these sound changes using an exemplar model with exemplars both at the phoneme level and the lexical level; speakers are less likely to use productions that would create lexical ambiguity, and thus the typical realization of words at risk of homophony may resist a shift that is reflected in other words. Such a model of homophony avoidance in exemplar selection may be supported by experimental work showing that the acoustic correlates of phonological contrasts are larger in words with a minimal pair for that contrast, e.g. the VOT for the initial stop in *cod* (cf. *god*) vs the initial stop in *cop* (Baese-Berk & Goldrick 2009). However, rather than attributing this pattern to exemplars with word-specific acoustic details, Baese-Berk & Goldrick (2009) interpret their results as reflecting hyperarticulation due to words with more neighbors requiring higher activation to inhibit their competitors.

1.3 This paper

This work examines evidence for the phonological representation and sound change processes that result in higher frequency words exhibiting more reduction than lower frequency words, using an experiment in which listeners are exposed to words with different VOT manipulations. If reduction in higher frequency words is due to ease of retrieval, then recent exposure should result in shorter VOT category boundaries, regardless of the VOT in the exposure items. If listeners store word-specific acoustic details, then their VOT category boundary should increase for words heard with long VOT and decrease for words heard with short VOT.

2 Methods and Materials

Participants were 96 native speakers of American English (mean age 27.8; 36 male, 58 female, 2 nonbinary). The study was run online, with participants recruited and paid through the Prolific system and the experiment presented through Qualtrics.

Participants were instructed that they would complete two tasks. In the first task, they would hear English words and categorize the vowel in each one as being long or short in duration. In the second task, they would hear English words and identify each as matching one of two associated written response options. The stimulus items were presented as a list;

listeners clicked on an audio player icon to hear each stimulus. Responses were given by clicking on one of the written options given under the icon for the stimulus. Within a block, the order of items was randomized.

There were two parts in this experiment: an exposure block and a testing block. Stimuli were made from recordings of one female American English speaker reading monosyllabic English words containing an onset stop, produced in randomized order and recorded in a quiet room with a stand-mounted Blue Yeti microphone in the Audacity software program and digitized at a 44.1 kHz sampling rate with 16-bit quantization. A list of all words is given in Table 1.

high frequency	**low frequency**
came/game	cob/gob
cap/gap	coo/goo
cold/gold	cull/gull
could/good	kale/gale
pan/ban	peep/beep
pat/bat	perch/birch
pet/bet	pike/bike
punch/bunch	pudge/budge
teen/dean	tame/dame
tip/dip	tomb/doom
too/do	torque/dork
town/down	tusk/dusk

Table 1: List of words used in the study, by lexical frequency.

First, listeners completed the exposure block. In this block, they categorized the vowel of each word as being long or short in duration. This task was aimed at ensuring that participants listened closely to the exposure items. Listeners heard 8 words during this block, 4 with lengthened VOT (mean 137 ms) and 4 with shortened VOT (mean 51 ms), made from naturally produced words beginning with voiceless aspirated stops. The mean VOT was the same for both high frequency and low frequency words. For each participant, the manipulation was consistent for all instances of a word, e.g. *town, pan, pet, could* with long VOT, and *tomb, peep, cob, cull* with short VOT. Each word presented during exposure was heard with 3 different vowel durations (the naturally produced vowel, a 20% shortened vowel, and a 20% lengthened vowel), and each of these was repeated 3 times, for a total of 9 appearances of each word. This exposure served to establish the local frequency of each item and potentially establish an expected VOT. There were three exposure conditions for these target words: (a) lengthened VOT, (b) shortened VOT, (c) no exposure (control). The words that appeared in this block were balanced across participants; there were 6 versions of the exposure block, covering each of the 24 words in each exposure condition.

The exposure block did not include any orthographic presentation of each word, in order to avoid possible visual priming. While this makes it possible that listeners sometimes perceived the stimulus as a word different from what was intended, the naturally produced F0 and other correlates of aspirated stops make it unlikely that misperception was common. Note that if the shortened VOT items were perceived as beginning with unaspirated stops, the predicted result would be that the VOT category boundary might be lengthened or unchanged for words in this condition; as will be seen subsequently, this is not the observed result.

Second, listeners completed a testing block, which tests the effect of exposure type. In this block, listeners heard words in isolation and identified each one as either starting with an aspirated stop or an unaspirated stop (e.g. *town* vs *down*). Both of the words in each minimal pair roughly matched in lexical frequency; half of the pairs were high-frequency words and half were low-frequency words. Listeners heard all 24 items manipulated along a 3-step VOT continuum (72 total stimuli), created from recorded items with ambiguous VOT produced by a trained phonetician. The small number of steps was used to reduce the possibility that additional exposure might obscure effects of the exposure manipulation.

These testing stimuli were created first by copying 10 ms of noise from the middle of the aspiration of the naturally produced item to produce the longest step and removing 10 ms of noise from the middle of the aspiration of the naturally produced item to produce the shortest step, while the naturally produced item served as the middle step. These stimuli were tested with a pilot group of 10 participants; if the middle step was identified consistently by all participants, the continuum was shifted by 10 ms in the direction of the category that was not selected, in order to move it towards ambiguous range that might be susceptible to effects of the exposure condition. For example, the middle step for *town/down* was consistently identified as *town*, so this became the longest VOT step, decreasing the VOT of each *town/down* stimulus by 10 ms. If all steps were consistently identified by all participants, the continuum was shifted by 20 ms. After running half of the participants (balanced across the 6 versions of the exposure block), 4 words were still identified with high consistency, and were shifted another 10 ms for the rest of the participants.

The stimuli had a mean VOT of 14 ms for bilabials, 32 ms for alveolars, and 38 ms for velars.

This method of creating the testing stimuli was necessary in order to make it possible to use only three steps in the testing continuum while including a substantial number of stimuli that would be ambiguous. However, it means that the VOT of the testing stimuli differed by word, so the absolute proportion of aspirated vs unaspirated responses based on inherent characteristics of the word (e.g. lexical frequency, place of articulation of the initial stop) will not be interpretable. The key aspect of the results is how the proportion of responses differed by exposure condition.

Statistical results are from a logistic mixed effects model, calculated with the lme4 package in R (Bates et al. 2015); p-values were calculated by the lmerTest package (Kuznetsova et al. 2015).

2.1 Hypotheses and predictions

There are two main hypotheses for how listeners will be impacted by exposure to manipulated VOT:

Hypothesis A: Listeners' lexical representations might include word-specific acoustic details, and will thus be impacted by the specific acoustic characteristics they have heard for each word presented during exposure. Under this analysis, listeners' expectations will align with the VOT heard for each word during exposure; relative to words not heard during exposure, they will have a category boundary with longer VOT for words heard with lengthened VOT and a shorter VOT for words heard with shortened VOT.

Hypothesis B: The frequency of a word might impact its accessibility and thus its expected VOT. Under this analysis, recent repeated exposure to a word will make listeners retrieve it more quickly, resulting in shorter expected duration (including VOT), regardless of whether the exposure tokens had shortened or lengthened VOT.

3 Results

The results are for the testing phase, in which listeners made decisions between minimal pairs differing in the aspiration of the initial stop.

Table 2 presents the summary of a mixed effects logistic regression model for the proportion of aspirated responses relative to unaspirated responses (e.g. *town* rather than *down*). The fixed effects were exposure type (lengthened VOT, shortened VOT, no exposure) and VOT continuum step. There were random intercepts for participant and by word. Recall that each listener encountered an equal number of words with lengthened VOT and shortened VOT; the exposure conditions differed across words (balanced across participants), not across participants.

	β	SE	z value	p value
(Intercept)	-0.582	0.194	-2.99	0.00276
Exposure Lengthened	0.22	0.0833	2.65	0.00818
Exposure Shortened	0.506	0.085	5.95	< 0.0001
ContinuumStep	1.19	0.0411	28.9	< 0.0001

Table 2: Regression model for responses of the unaspirated category. *Reference Levels: Exposure = None*

Listeners accepted aspirated stops as having a shorter VOT in words that they had recently been exposed to, both when the exposure stimuli had shortened VOT and when exposure stimuli had lengthened VOT. This tolerance for shorter VOT in aspirated stops was greater for words that had been heard with shortened VOT during exposure than those which had been heard with lengthened VOT during exposure. Figure 1 illustrates the proportion of aspirated responses for each exposure condition.

The VOT continuum step was also a significant predictor of responses; listeners were more likely to identify stimuli as having voiceless initial stops when the VOT was longer.

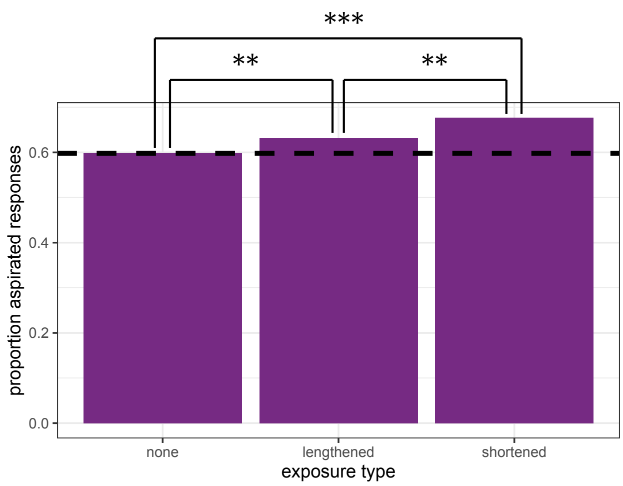

Figure 1: Proportion of aspirated responses. The dashed line marks the proportion of aspirated responses for words with no exposure, to facilitate comparison across conditions.

Table 3 presents the summary of a mixed effects logistic regression model for the proportion of aspirated responses relative to unaspirated responses when including lexical frequency as a factor. The fixed effects were exposure type (lengthened VOT, shortened VOT, no exposure), lexical frequency category (high frequency, low frequency), VOT continuum step, and the interaction between exposure type and frequency. There were random intercepts for participant and by word.

	β	SE	z value	p value
(Intercept)	-0.55	0.237	-2.32	0.0201
Exposure Lengthened	0.355	0.126	2.81	0.00497
Exposure Shortened	0.651	0.125	5.23	< 0.0001
FrequencyCategory High	-0.0632	0.272	-0.233	0.816
ContinuumStep	1.19	0.0411	28.9	< 0.0001
Exp Lengthened : FreqCat High	-0.26	0.182	-1.43	0.153
Exp Shortened : FreqCat High	-0.296	0.184	-1.61	0.109

Table 3: Regression model for responses of the unaspirated category. *Reference Levels: Exposure = None, FrequencyCategory = Low*

The main effect of exposure type is the same as in the simpler model; listeners accepted aspirated stops as having a shorter VOT in words that they had recently been exposed to, both when the exposure stimuli had shortened VOT and when exposure stimuli had lengthened VOT.

The VOT continuum step was also still a significant predictor of responses; listeners were more likely to identify stimuli as having voiceless initial stops when the VOT was longer.

There was no significant interaction between lexical frequency and the effect of exposure. However, it may be noteworthy that the effect of exposure had a trend towards being smaller for high frequency words; this might become significant with a larger dataset. There was also no significant main effect of lexical frequency, but recall that the manipulation differed by word, so this lack of effect is not interpretable. Figure 2 illustrates the proportion of aspirated responses for each exposure condition and lexical frequency category.

4 Discussion

4.1 Synchronic phonology

The results provide evidence for an effect of frequency itself, producing shortened VOT category boundaries for words in both the shortened VOT

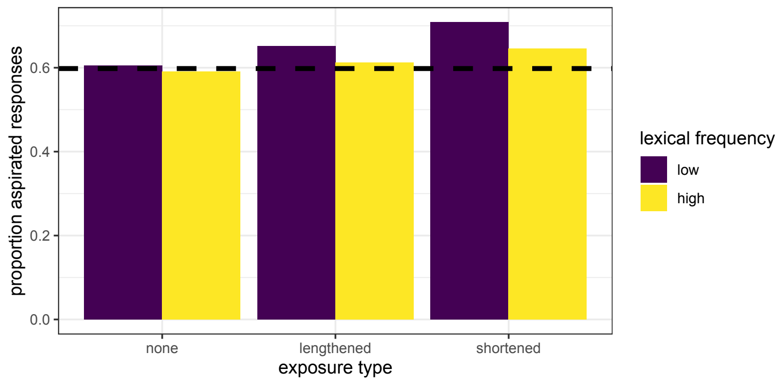

Figure 2: Proportion of aspirated responses. The dashed line marks the proportion of aspirated responses for words with no exposure, to facilitate comparison across conditions.

exposure condition and the lengthened VOT exposure condition. There was a significantly shorter VOT category boundary in the shortened VOT exposure condition than in the lengthened VOT exposure condition, which could provide evidence for word-specific acoustic details but could also reflect differences in activation of each word based on the prototypicality of the stimulus items.

The category boundary is lowered in both VOT manipulation conditions, suggesting that this shift is is due to recent exposure increasing the frequency of these words in the local context. Differences in VOT are based on this predictability. Salient words are accessed more easily (Dahan et al. 2001; Dufour et al. 2013), which results in faster access across a range of tasks, e.g. naming (McRae et al. 1990; Forster & Chambers 1973) and lexical decision (Whaley 1978; Goh et al. 2009), as well as shorter durations in production (Gahl 2008; Seyfarth 2014; Tang & Shaw 2021). These differences in access are also likely to impact category boundaries in perception, with listeners expecting shorter durations for words that are more rapidly retrieved. Previous work does suggest that listeners will more readily accept reduced forms for higher frequency words than for lower frequency words (Connine et al. 2008; Pitt et al. 2011). Under this analysis, lexical frequency itself impacts the realization of VOT and expectations in perceptual access. The shift is not due to the particular realization of VOT in the exposure stimuli, just the presence of the lexical item during exposure. Such an effect does not depend on individual words having distinct VOT targets; the underlying phonetic targets can be uniform across words, with differences in production explained by later processes.

The shorter category boundary for recently heard words can also be explained by priming. Listeners are more likely to identify ambiguous stimuli as matching a previously heard word rather than a word that hasn't already been heard before during the task (Ratcliff et al. 1997; Masson 2002). Increased activation due to recent exposure would make participants more likely to perceive incoming stimuli as containing acceptable tokens of voiceless aspirated stops because the higher activation of these words means that the acoustic input doesn't need to add as much to the activation in order to result in that word being selected. This account differs from the preceding account in what it predicts for the impact of exposure on the increased acceptability of different acoustic variants. If higher local frequency sets expectations for shorter duration based on faster access, then this sort of exposure should only increase the acceptability of short VOT or other reduced forms. In contrast, if higher local frequency results in generally relaxed thresholds for acoustic matches, then other variants such as hyperarticulated forms (e.g. longer VOT in short-VOT stops) will also be more accepted. Under this account, the perceptual category boundary results might not be paralleled in production.

Although both shortened VOT exposure and lengthened VOT exposure resulted in accepting shorter VOTs as falling into the aspirated category, the effect was larger for the shortened VOT exposure condition, which might suggest that listeners do have word-specific VOT targets, in addition to frequency itself affecting VOT. Recall that all participants encountered an equal number of words with lengthened VOT and shortened VOT, so any differences between the lengthened VOT and shortened VOT exposure conditions must be due to lexical effects rather than reflecting shifts in phoneme-level targets. Under hybrid exemplar models, there are exemplar clouds linked across instances of the same phoneme as well as across instances of the same word (Goldinger 1998; Pierrehumbert 2002). Because there are far more tokens for a phonological category (e.g. /t/) than for the realization of a specific word (e.g. the /t/ in *town*), shifts just in a particular word are less likely to occur than a shift in a phonological category. However, it might be possible for an individual word to shift when it has recurring patterns of its realization in natural usage (Tang & Shaw 2021; Seyfarth 2014), or consistent repeated form in the experimental context used in the current study.

Alternatively, the difference between conditions in this study might be due to the prototypicality of the VOTs used in each condition; Andruski et al. (1994) demonstrated that words with more prototypical pronunciation of their component phonemes are more strongly activated. The shortened VOT condition in the present study had a mean VOT of 51 ms, which is well within the range of typical naturally produced long-VOT stops in English. In contrast, the lengthened VOT condition had a mean VOT of 137 ms, which is substantially longer than typical naturally produced long-VOT stops in English (Allen & Miller 1999; Yao 2009). Because the exposure items in the lengthened VOT condition are less typical of the English long-VOT category, they would produce weaker activation of the words heard with these long VOTs, particularly given that completing the exposure task (categorizing vowels as long or short) did not require lexical access. The less prototypical exposure items in this condition would result in a smaller increase in activation for these words than the more prototypical exposure items in the shortened VOT condition, resulting in a smaller effect on subsequent access.

There were no significant effects of a word's lexical frequency on responses, though there was a trend towards more of an effect of exposure with lower frequency words, which might become significant in a larger dataset. An effect might be predicted based on how overall frequency and local frequency interact; the salience and accessibility of a word can be increased, but cannot be decreased in the same way. The salience of a high-frequency word is not decreased by a few minutes in a task without exposure to that word, whereas a low frequency word can be made salient by recent exposure during an experimental task. It is possible that the effects of repeated recent exposure within the task partially obscure effects of lexical frequency. Connine et al. (1993) found a bias towards identifying stimuli with ambiguous VOT as higher frequency words (e.g. *best* rather than *pest*), in a study with a 4-step VOT continuum. In contrast, Politzer-Ahles et al. (2020) did not find an effect of lexical frequency on responses, in a study with a 9-step VOT continuum. The difference between their results might be due to the larger number of times that each word was heard during testing in Politzer-Ahles et al. 2020.

4.2 Implications for sound change

These results show that a change in frequency itself impacts category boundaries, separately from the acoustic form of the stimuli that listeners were exposed to. Based on these results, apparent frequency-sensitive sound changes can be explained without requiring each word to have independent phonetic details.

Reduction-driven shifts may first be apparent in high-frequency words (e.g. Bybee 2002; Hay et al. 2015), but that does not mean that the target phonetic details in these words have changed while other words have not. It is possible that productions vary around a shared prototype, and the prototype for the entire category might shift if that phoneme is often produced in a particular way, e.g. based on how it is reduced in high-frequency words (Pierrehumbert 2002; Bybee 2012). This is consistent with frequency effects that are observed even when there is no evidence that a change is in progress (e.g. Gahl 2008; Bell et al. 2009), which can be explained by processes in lexical access (Gahl et al. 2012; Kahn & Arnold 2012; Jurafsky et al. 2002). Under this analysis, word-specific acoustic details are not necessary for explaining synchronic variation correlated with frequency nor for explaining shifts along those axes of variation, which also explains why most sound changes are regular (Ringe & Eska 2013). The results from the current experiment support this analysis by providing evidence for phonetic differences arising as a direct result of lexical frequency, based on recent repeated exposure decreasing VOT category boundaries even when the exposure stimuli had lengthened VOT.

On the other hand, the existence of an alternative explanation that can account for many apparent word-specific sound changes correlated with frequency does not necessarily mean that word-specific acoustic details do not exist. Within these results, there is a significant difference between the long-VOT and short-VOT exposure conditions; this might suggest that listeners' representations do include memories of word-specific acoustic detail, and that those details can be shifted with sufficient exposure to sufficiently consistent tokens of those words (cf. Rochet-Capellan & Ostry 2011). However, it is possible that the results of the current study can be explained instead by the influence of phonological prototypicality on lexical activation levels. Word-specific acoustic details may be useful for explaining the existence of irregular phonological developments that cannot be explained by analogy or inter-dialectal borrowing, such as the cases identified by Blevins & Wedel (2009), though there are relatively few sound changes in this category. It is also possible that those cases might develop due to patterns of activation interacting with lexical competition (cf. Baese-Berk & Goldrick 2009), rather than depending on word-specific acoustic details.

5 Conclusion

Manipulating local frequency with repeated exposure demonstrates that lexical frequency itself impacts listeners' category boundaries; listeners are more willing to accept aspirated stops with shorter VOT in a word that they have heard repeatedly than in a word that they did not hear before within the task, even if the tokens heard during exposure had long VOT. These results suggest that effects of frequency can largely be explained by speed of lexical access (Gahl et al. 2012; Kahn & Arnold 2012), both for synchronic variation and diachronic developments; just because higher frequency words have different realizations does not mean that they have different acoustic targets in their phonological representations.

The size of the effect differed based on the acoustic characteristics of the exposure stimuli, with shorter VOT category boundaries for words heard with short VOT than for words heard with long VOT. This might be similarly explained as an automatic effect of lexical access, based on the lengthened VOTs being less prototypical than the shortened VOTs. The less prototypical stimuli would produce weaker activation of the stimulus words, and thus have less of an effect on subsequent access. However, the difference between exposure conditions could also be explained with word-specific phonetic details.

Comments invited

PiHPh relies on post-publication review of the papers that it publishes. If you have any comments on this piece, please add them to its comments site. You are encouraged to consult this site after reading the paper, as there may be comments from other readers there, and replies from the author. This paper's site is here:
https://doi.org/10.2218/pihph.7.2022.8914

Acknowledgements

I would like to thank all the participants in the Fifth Edinburgh Symposium on Historical Phonology for thought-provoking discussions.

Author's contact details

Chelsea Sanker
Stanford University
Margaret Jacks Hall, Building 460
Stanford, CA 94305
USA

sanker@stanford.edu

References

Allen, J. Sean & Joanne Miller. 1999. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America* 106(4). 2031–2039.

Andruski, Jean, Sheila Blumstein & Martha Burton. 1994. The effect of subphonetic differences on lexical access. *Cognition* 52(3). 163–187.

Arnold, Denis, Fabian Tomaschek, Konstantin Sering, Florence Lopez & R. Harald Baayen. 2017. Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PloS One* 12(4). Article e0174623.

Aylett, Matthew & Alice 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(1). 31–56.

Babel, Molly. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39. 437–456.

Baese-Berk, Melissa & Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24(4). 527–554.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26(2). 143–183.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14(3). 261–290.

Bybee, Joan. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: Perception, production and social factors*, 211–234. John Benjamins.

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47.

Connine, Cynthia, Larissa Ranbom & David Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70(3). 403–411.

Connine, Cynthia, Debra Titone & Jian Wang. 1993. Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(1). 81–94.

Dahan, Delphine, James Magnuson & Michael Tanenhaus. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42(4). 317–367.

Dias, James & Lawrence Rosenblum. 2016. Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception, & Psychophysics* 78. 317–333.

Dinkin, Aaron. 2008. The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics* 14(1). 8.

Dufour, Sophie, Angèle Brunellière & Ulrich Frauenfelder. 2013. Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive Science* 37(3). 489–507.

Forster, Kenneth & Susan Chambers. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12(6). 627–635.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496.

Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66. 789–806.

Garrett, Andrew. 2015. Sound change. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 227–248. Routledge.

Goh, Winston, Lidia Suárez, Melvin Yap & Seok Hui Tan. 2009. Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review* 16(5). 882–887.

Goldinger, Stephen. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279.

Harrison, S. P. 2003. On the limits of the comparative method. In Brian Joseph & Richard Janda (eds.), *The handbook of historical linguistics*, 213–243. Blackwell.

Hay, Jennifer, Janet Pierrehumbert, Abby Walker & Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change.

Cognition 139. 83–91.

Howes, Davis. 1957. On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America* 29(2). 296–305.

Johnson, Keith. 1997. The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics* 50. 101–113.

Jurafsky, Daniel, Alan Bell & Cynthia Girand. 2002. The role of the lemma in form variation. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology VII*, 3–34. Berlin: Mouton de Gruyter.

Kahn, Jason & Jennifer Arnold. 2012. A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language* 67. 311–325.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2015. *lmertest: Tests in linear mixed effects models*. https://CRAN.R-project.org/package=lmerTest. R package version 2.0-29.

Luce, Paul & David Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19(1). 1–36.

Masson, Michael. 2002. Bias in masked word identification: Unconscious influences of repetition priming. *Psychonomic Bulletin & Review* 9(4). 773–779.

McRae, Ken, Debra Jared & Mark Seidenberg. 1990. On the roles of frequency and lexical access in word naming. *Journal of Memory and Language* 29(1). 43–65.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39. 132–142.

Nygaard, Lynne, Debora Herold & Laura Namy. 2009. The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science* 33. 127–146.

Nygaard, Lynne & Erin Lunders. 2002. Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition* 30(4). 583–593.

Pardo, Jennifer, Kelly Jordan, Rolliene Mallari, Caitlin Scanlon & Eva Lewandowski. 2013. Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language* 69. 183–195.

Pardo, Jennifer, Adelya Urmanche, Sherilyn Wilman & Jaclyn Wiener. 2017. Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* 79. 637–659.

Pate, John & Sharon Goldwater. 2015. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language* 78. 1–17.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology VII*, 101–140. Berlin: Mouton de Gruyter.

Pitt, Mark, Laura Dilley & Michael Tat. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics* 39(3). 304–311.

Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53. 181–216.

Pluymaekers, Mark, Mirjam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118(4). 2561–2569.

Politzer-Ahles, Stephen, Ka Keung Lee & Lue Shen. 2020. Ganong effects for frequency may not be robust. *Journal of the Acoustical Society of America* 147(1). EL37–EL42.

Ratcliff, Roger, David Allbritton & Gail McKoon. 1997. Bias in auditory priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(1). 143–152.

Ringe, Don & Joseph Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge University Press.

Rochet-Capellan, Amélie & David Ostry. 2011. Simultaneous acquisition of multiple auditory-motor transformations in speech. *Journal of Neuroscience* 31(7). 2657–2662.

Sanker, Chelsea. 2021. Convergence doesn't show lexically-specific phonetic detail. In Ryan Bennett, Richard Bibbs, Mykel Brinkerhoff, Max Kaplan, Stephanie Rich, Amanda Rysling, Nicholas Van Handel & Maya Wax Cavallaro (eds.), *Supplemental proceedings of the 2020 Annual Meeting on Phonology*, .

Savin, Harris. 1963. Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America* 35(2). 200–206.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155.

Seyfarth, Scott, Marc Garellek, Gwendolyn Gillingham, Farrell Ackerman & Robert Malouf. 2018. Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience* 33(1). 32–49.

Sommers, Mitchell, Karen Iler Kirk & David Pisoni. 1997. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing* 18(2). 89.

Tang, Kevin & Jason Shaw. 2021. Prosody leaks into the memories of words. *Cognition* 210. Article 104601.

Todd, Simon, Janet Pierrehumbert & Jennifer Hay. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185. 1–20.

Wang, William & Chin-Chuan Cheng. 1977. Implementation of phonological change: The Shuāng-fēng Chinese case. In William Wang (ed.), *The lexicon in phonological change*, Mouton.

Whaley, Charles. 1978. Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17(2). 143–154.

Yao, Yao. 2009. Understanding VOT variation in spontaneous speech. *UC Berkeley PhonLab Annual Report* 5.