

JOURNAL ARTICLE

The perceptual filtering of predictable coarticulation in exemplar memory

Jonathan Manker

Department of Linguistics, Rice University, Houston, TX, US
jonathan.manker@rice.edu

Exemplar models of word representations have remained ambivalent or impressionistic as to precisely what veridical auditory information is stored in individual word exemplars. Earlier models (Johnson, 1997b) suggest all perceived information was stored in memory, whereas more recent proposals (Pierrehumbert, 2002; Goldinger, 2007) suggest some degree of abstraction occurs in storing particular exemplars. Findings from the phonetic accommodation paradigm (Goldinger, 1998; Nielsen, 2011, etc.) suggest that the accumulation of new exemplars may drive the spread of sound change. At the same time, some theories of sound change suggest that perceptual biases serve as a starting point for change (Ohala, 1981, 1983). The current study investigates how perceptual biases, such as the predictability of coarticulation, can shape the contents of exemplars. The experimental results suggest that an *expected* phonetic alteration, such as f₀ raising on vowels following voiceless consonants—a predictable coarticulatory effect—is more likely to undergo some degree of abstraction when stored in exemplar memory, whereas *unexpected* phonetic detail (e.g., f₀ raising following voiced consonants) is more faithfully stored or maintained for longer in memory. These findings suggest perceptual biases that could shape pools of exemplars, leading to different expectations for conditioned versus unconditioned sound changes.

Keywords: Exemplar theory; discrimination; speech perception; coarticulation; sound change

1. Background: Exemplars

1.1. Exemplars store fine phonetic details

Phonologists have long attempted to describe and model how words are represented in the mental lexicon. One of the most enduring proposals is that words are stored as abstract, formal representations, composed of strings of contrastive segments, a principle borrowed from generative phonology (Chomsky & Halle, 1968) and applied to models of speech perception involving normalization (Gerstman, 1968; Tranmüller, 1981, etc.). Contemporaneous to these developments, phoneticians were becoming aware of the vast acoustic variability of the speech signal. Liberman et al. (1967) demonstrated the variability of particular phonemes produced by speakers due to coarticulatory effects, while Stevens (1972) and Klatt (1979) noted the variability due to idiosyncrasies of a speaker's voice, gender, and vocal tract size (even among speakers from the same speech communities). This acoustic variability proved problematic without an additional system by which phonological representations of words could be extracted from the messy acoustic signal. The solution was in the normalization of the speech signal: Coarticulatory effects were undone, and the idiosyncratic features of talkers' voices were stripped away by the listener, allowing the activation of an invariant phonological form.

In the following decades, exemplar models challenged the concept of abstract phonological forms as the mental representations of words. The core principle of exemplar-based

models is that percepts—in the case of speech perception, words—are stored in memory as collections of individual labeled instances, rather than as a single abstraction. Goldinger (1996) provided evidence that exemplars were used in speech perception, finding that subjects were more accurate in identifying whether a word had been repeated or not if the repetitions were produced in the same voice. This suggests not only that voice information was stored in memory, but that this information was an integrated part of the percept along with the phonological form of the word itself. If only an abstract representation had been activated, with voice information stripped away, no voice effect should have been found.

These findings led to a line of research in *phonetic accommodation* which extended and further corroborated the predictions of the exemplar model. The theory of phonetic accommodation asserts that an individual's pronunciation will drift towards that exhibited by his or her interlocutors after being exposed to it. Phonetic accommodation is predicted by exemplar theory because if prior stored instances of words are what comprise an individual's mental representation of a word, a new instance will then (slightly) shift the mean pronunciation of future utterances of that word. Goldinger (1998) observed just such an effect in an AXB task, such that speakers' repetitions of words (B) heard after the stimulus (X) sounded more like the stimulus than the original pronunciation (A). Furthermore, Goldinger found other effects that suggest the structure of exemplar clouds. Immediate shadowing of the stimuli produced a stronger imitative effect than delayed shadowing, suggesting new exemplars fade from memory with time. Secondly, more repetitions of words resulted in stronger accommodation, suggesting a greater number of new exemplars have a greater ability to shift the previous production averages. Lastly, low frequency words also displayed greater accommodation, suggesting a smaller pre-existing cloud of exemplars would be more prone to change than those containing more prior exemplars. While Goldinger's (1998) accommodation study judged similarity to the model based on the qualitative impressions of a panel of judges, later studies were able to quantify phonetic drift due to accommodation, measuring small changes in particular phonetic features such as VOT (Shockley, Sabadini, & Fowler, 2004; Nielsen, 2011) and vowel quality (Tilsen, 2009).

While many studies have established strong evidence for the existence of word exemplars and their relevance in speech perception, it is less clear precisely what information is stored in one. Generally speaking, a standard exemplar model assumes that exemplars are fairly veridical representations, being stored “as they occur, without any abstraction at all” (Johnson, 2007, p. 27). However, the abstraction that Johnson mentions refers to the transformation of newly perceived percepts into prototypical forms, as needed for speech recognition, a separate phenomenon from exemplars themselves undergoing some sort of perceptual transformation or abstraction while stored in memory as part of clouds of instances of particular words. Nevertheless, studies by Johnson (1997a) and (1997b) propose exemplars with purely veridical information, though with varying degrees. Johnson (1997a) considers reduced representations of vowel exemplars composed only of formant values. However, Johnson (1997b) argues for the storage of ‘auditory spectra’ containing all auditory information that the ear gleans from the acoustic signal. He states that this more detailed representation “is realistic because it is based on psychoacoustic data, and it also avoids making assumptions about which of the many potential acoustic features should be measured and kept in an exemplar of heard speech” (2007, p. 34), though he concedes that more data-driven evidence could lend support for a more compact representation.

Others propose models which simultaneously contain veridical exemplars and abstract phonological representations. For example, Pierrehumbert (2002, 2016) argues in favor of a hybrid model containing both phonological encoding and exemplar representations. This helps to account both for generalizations such as Neogrammarian type sound changes

which affect all instances of a particular phoneme, as well as word-specific phonetics that might be influenced by word frequency, a problem better handled by exemplar representations. McLennan and Luce (2005) and Luce and McLennan (2005) find that both abstract representations and exemplars may co-exist, being activated at different stages in speech perception. While such hybrid models include both phonological and exemplar levels of representation, it is unclear if exemplars in these models would have purely veridical information or if they may include some degree of abstraction as well.

Goldinger (1996) does not directly define the contents of exemplars, although he describes the perceptual process of encoding voice information in episodes as being automatic, which could suggest there is no specificity in what details are stored. Nevertheless, he also notes that episodic encoding of phonetic details may occur “only to the extent that they matter in original processing,” and that these memory traces will usually “emphasize elements of meaning, not perception” (p. 1180). Goldinger (2007) further considers the question of what exemplars encode, asserting that any ‘raw data’ will necessarily undergo some abstract transformation, such that “each stored ‘exemplar’ is actually a product of perceptual input combined with prior knowledge, the precise balance likely affected by many factors” (p. 50). Hawkins (2003, 2010), proposes that exemplars with fine phonetic detail are first stored in memory, but are processed for signal-to-structure mapping, only to the extent needed to achieve an understanding of the linguistic meaning. This can be achieved a number of ways depending on the context, whereas processing of the individual words, phonemes, or subphonemic details may not be necessary depending on the context. In some cases, top-down processing will be faster in identifying words and phonemes as opposed to actual processing of the speech signal, in which case the veridical phonetic details may not have been used to understand the linguistic meaning. This suggests a possible bias in the storage and maintenance of different phonetic information present in the speech signal which might in some way be modulated by contextual information.

Some previous work has similarly indicated that the availability of top-down information could be a factor in the storage and maintenance of phonetic detail. Nye and Fowler (2003) is a rare case of an accommodation study which used sentence rather than word stimuli, and as such, provided insight into how top-down information influences phonetic accommodation. In their experiment, subjects were presented with nonce word stimuli that varied in how closely they approximated English, particularly with regard to phonotactic similarity—e.g., low orders of approximation grossly violated English phonotactic structure, such as [ə tʌ ɔɪmɛkænd prən vʊʃəl], whereas higher order stimuli closely resembled English, such as [hɪz ə pɪnto ænd hi fɒtəgræs wændəfɒlli]. The results showed that subjects more closely imitated the *lower* orders of approximation, which suggested that more information about the higher order stimuli may have undergone some degree of abstraction due to high level linguistic knowledge of English, allowing for less detailed processing of higher order stimuli.

Manker (2019) directly assessed the effect of contextual knowledge on the storage of phonetic details in exemplar memory. In a discrimination task, subjects heard sentences with one of the words repeated, and were asked to determine if the word sounded exactly the same or somewhat different when repeated. The results showed that subjects displayed better discrimination when the words had been heard in an *unpredictable* context, e.g., “Joe turned and saw the *cabins*,” as opposed to a *predictable* context, e.g., “Pioneers built log *cabins*.” In a second accommodation experiment with the same stimuli, it was revealed that subjects also displayed better VOT and pitch contour accommodation for words that had been heard in unpredictable context, corroborating the results of the discrimination task. Ultimately, this suggested that more detailed, veridical exemplars were stored or maintained in memory for unpredictable words, because more reliance on processing of

fine phonetic detail was needed in the first place in order to identify these words, whereas the recognition of predictable words could be achieved via top-down processing using contextual information.

Thus, these studies showed that semantic contextual information is one possible bias affecting what details might be stored in exemplar memory. Other linguistic information may also make certain details of the speech signal prone to being ignored or abstracted. The current study considers the effect of coarticulation in a similar way: If a particular coarticulatory effect is expected, does it need to be processed, stored, and maintained in exemplar memory when such detail could easily be abstracted or reconstructed based on higher level linguistic knowledge?

To answer this question, I conducted a discrimination task that compared listeners' abilities to store and maintain the acoustic details of expected coarticulation compared to unexpected acoustic detail. This should test whether coarticulatory predictability modulates the details that are stored in exemplar memory, by either stripping them away or abstracting them. I specifically examined the coarticulatory phenomenon of f_0 raising versus lowering following voiceless and voiced consonants respectively, a phenomenon which can lead to tonogenesis following neutralization of consonant voicing (Hombert, Ohala, & Ewan, 1979). I hypothesized that subjects would show better discrimination of stimuli that differ in acoustic detail that would not arise from coarticulation, which would suggest more veridical details of the auditory signal are stored in exemplar memory for unpredictable speech.

2. Methodology

2.1. Voicing and F_0

Phoneticians have established that a consonant's voicing can slightly perturb the f_0 of the following vowel (House & Fairbanks, 1953; Lehiste & Peterson, 1961; Hombert & Ladefoged, 1976; Hombert et al., 1979; Kingston, 1989; Kingston & Diehl, 1994; Kingston, Diehl, Kirk, & Castleman, 2008; Kingston, 2011; Ratliff, 2015). Following a voiceless consonant, the f_0 of the vowel tends to be raised 5–10 Hz briefly before sloping down to its intended target, whereas voiced consonants cause a similar pattern of pitch lowering (Figure 1). According to Hombert et al. (1979), this happens either due to the increased

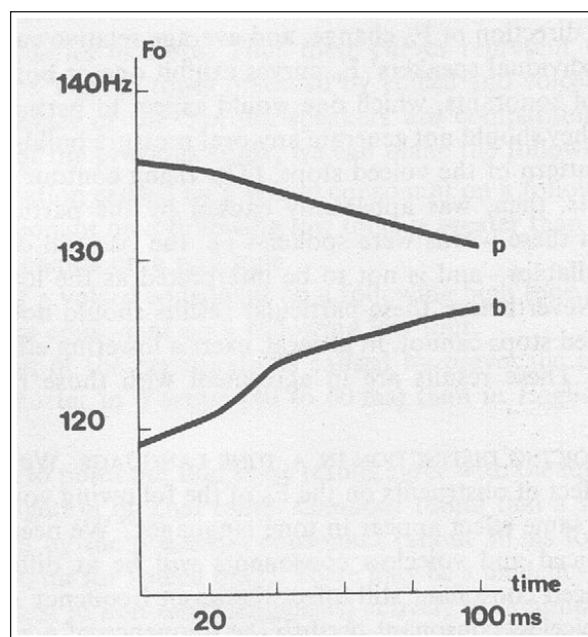


Figure 1: Effect of voicing and voicelessness on f_0 , reproduced from Hombert et al. (1979).

oral pressure which results from voicing, leading to lower pressure in the vocal folds and thus lower pitch, or due to the vocal cord tension needed for consonant voicing which affects the tension of the vocal folds during the following vowel as well. Ultimately, this phenomenon often sows the seeds of tonogenesis, where an older voicing contrast, e.g., /pa/ ~ /ba/, is enhanced with pitch differences, thus /pá/ versus /bà/, and is then lost with the survival of the new tone contrast, /pá/ ~ /pà/ (Hombert et al., 1979). Such a development, whether completed or not, has been observed in a diverse body of languages, including several Mon Khmer languages (Svantesson, 1991), Chadic languages (Wolff, 1987), Punjabi (Gill & Gleason, 1972), Cham (Thurgood, 1999), and many others.

Given these observations, in the current experiment, I investigated whether listeners more successfully discriminated a word and its repetition with an f0 contour that would not arise from coarticulation as opposed to a word and repetition with a coarticularly expected f0 contour. That is to say, listeners were expected to store the f0 details of a syllable /bà/ (indicating a slightly raised f0 at the beginning) better than /pà/, where such a pitch contour is expected (Figure 2). The raised f0 contour will be the predictable result of coarticulation from the preceding voiceless stop, and such details may then not survive in exemplar storage.

2.2. Participants

One hundred subjects in two counterbalanced groups of 50 were recruited via Amazon Mechanical Turk, an online platform also used for conducting the experiment (Yu & Lee, 2014 find similar results in speech perception experiments run in person versus using Amazon Mechanical Turk). Subjects were required to be native speakers of English located in the United States and without speech or hearing disorders. They were compensated with \$3.50 for the approximately 20-minute experiment.

2.3. Procedure

The experiment consisted of 200 trials for each subject with two basic types of stimuli. Ninety-six of the stimuli (8 examples × 6 different phonemes × 2 repetition conditions, ‘same’ or ‘different’) were AX discrimination trials in which the subject heard a word spoken in isolation, followed by a one-second pause, two seconds of multi-speaker babble (as a distractor), and then a repetition of the initial word (Figure 3). When repeated, the

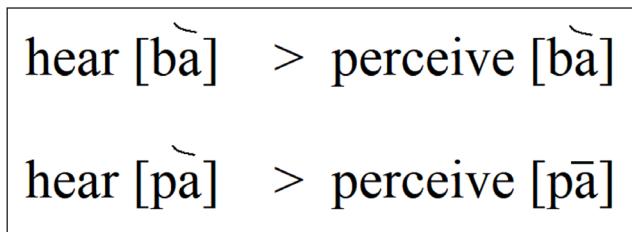


Figure 2: Demonstration of hypothesized predictability-based perceptual bias.

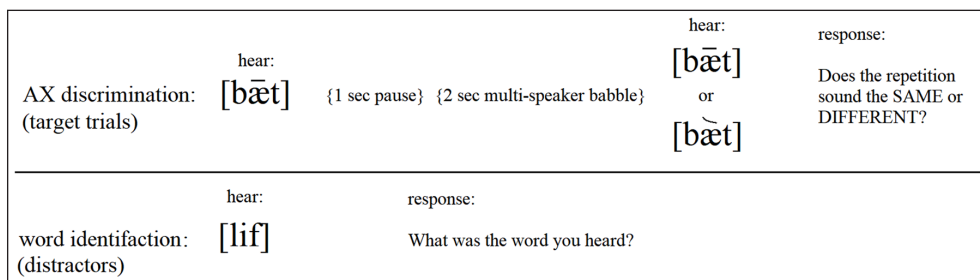


Figure 3: Demonstration of both trial types.

word had either the exact same f0 contour as the initial hearing, or the beginning was raised slightly mimicking the effect of pitch raising following a voiceless consonant (see Section 2.3. for more details). Then, the subject was asked whether the repetition sounded exactly the ‘same’ or ‘different’ than the initial hearing.

Of these 96 discrimination stimuli, there were 48 which were in fact different when repeated, and 48 which were the same. Each of these 48 then was evenly divided into 24 stimuli with target words beginning with voiceless stops and 24 stimuli with voiced stops. These included eight tokens for each of six different phonemes at three different places of articulation, /p t k b d g/.

The other 104 stimuli, which thus occurred slightly more than 50% of the time, were word identification trials. After hearing the target word, there was a pause and subjects were asked to type the word they heard. It is important to note that subjects would not know which type of stimulus they were presented with until after hearing the initial word, at which time they either heard a repetition or not. These trials were intended to be distractors, to encourage the subjects to listen to speech more ‘naturally’—that is, to identify words rather than focusing on subphonemic details. For example, there could be some concern that the subjects might become aware that ‘different’ always means a raised pitch contour, and they would start listening at a more acoustic level, possibly overriding or weakening any higher-level perceptual bias that might arise in a natural listening setting. However, this is mostly a concern of achieving a false negative, as such a phenomenon would result in no difference between how the voiced and voiceless stimuli are perceived. Any significant difference then between the perception of the voiced and voiceless stimuli would indicate an effect of the initial consonant’s voicing. A similar approach was used in Manker (2019), though it is unclear whether contextual level effects would not occur without the distractors. The procedure involved in the two types of stimuli are represented visually in **Figure 3**, while a breakdown of the different stimuli conditions is shown in the tree in **Figure 4**.

The experiment was run on the SurveyGizmo online questionnaire platform. Before beginning the experiment, subjects provided informed consent and were presented with a short training session showing the types of questions they would encounter. Subjects were instructed that ‘different’ responses indicated subtle differences in pronunciation,

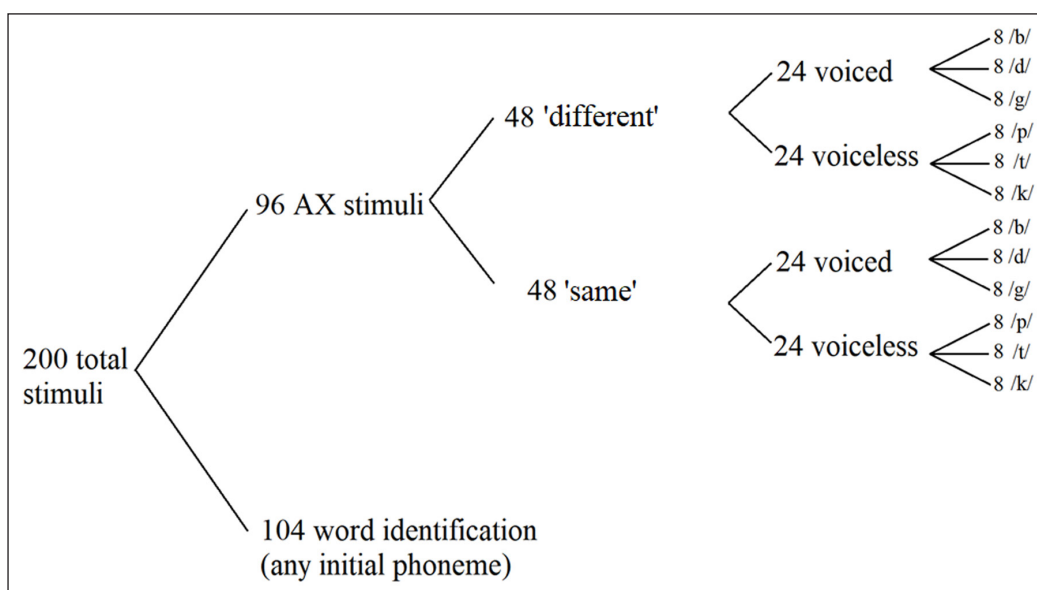


Figure 4: Tree showing the breakup of different stimuli.

and not whole word, sound, or speaker differences. Subjects were also instructed to use headphones in a setting free of distractions.

2.4. Stimuli selection and manipulation

The target word list included 96 unique words beginning with one of six voiced and voiceless stop consonants. All words were a single syllable with no initial consonant clusters, of the form /CV(C)(C)/. Each of these words occurred with a reversed-voicing minimal pair counterpart—e.g., base/pace, two/do, goat/coat—which ensured that the phonological environments after the initial consonant was controlled for voiced versus voiceless stimuli as a whole (e.g., if there would happen to be an effect of the following vowel). Each of 96 words was presented to each subject once, in which case it was either the same when repeated, or different, with a raised f_0 contour. Two groups of 50 subjects each were presented with exactly half of the stimuli in order to have ‘different’ stimuli for each of the 96 target words. Subjects in Group A, for example, heard “buy” *different* and “back” the *same*, whereas subjects in Group B heard “buy” the *same* and “back” *different*.

The stimuli were produced by a phonetically-trained male in his 30s, recorded in a quiet location with a Zoom H4n Handy recorder. The voiceless stimuli were produced with aspiration, while the voiced stimuli were produced with voicing during the stop closure. The Praat Manipulate tool was used to alter the pitch contours for the stimuli. A neutral f_0 contour was extracted from a vowel-initial word and was applied to all the base utterances. For the ‘different’ repetitions, the f_0 was manipulated to begin 18 Hz higher than the base utterance and gradually slope downward, reaching the same pitch as the base utterance after about 125 ms. The base f_0 contour and the raised different f_0 contour are shown in **Figure 5**.

3. Results

Data was collected from a total of 100 subjects who completed the experiment. All data was kept, even in cases when the subjects did not do better than chance at the discrimination task. Thus, with 96 AX discrimination stimuli, there were a total of 9600 responses. Of these responses, 58.4% were correct (in noting either sameness or difference), with 39.9% incorrect responses, and 1.6% stimuli left unanswered. This suggests some difficulty with the discrimination task, yet clearly indicates that given the number of subjects and

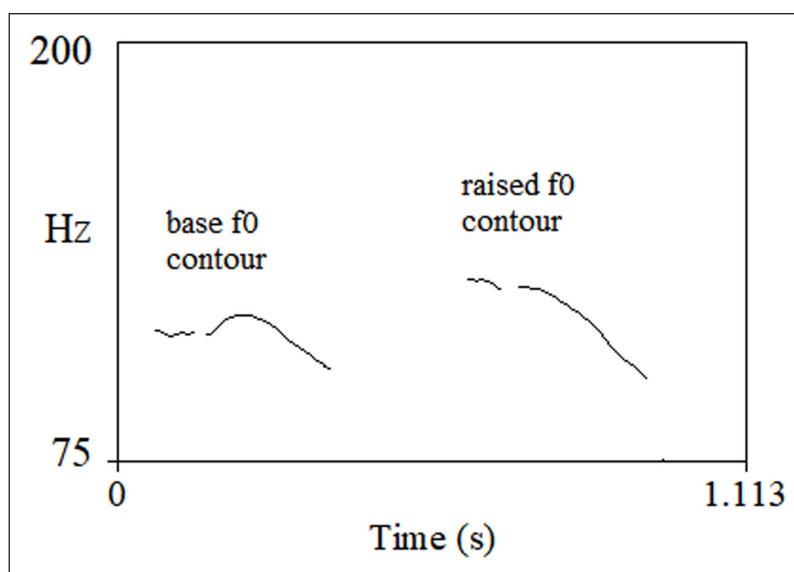


Figure 5: Manipulated F0 contours of stimuli.

trials, subjects did significantly better than chance (50%, with a binomial test yielding $p < 0.0001$) and thus fully understood the nature of the experiment as a whole. There was also a clear response bias towards believing the stimuli sounded the ‘same’ when repeated, evidenced by a 70.25% success rate for the ‘same’ tokens compared to only 46.6% success for the ‘different’ tokens. This again suggests the acoustic difference in the repeated word was subtle.

By examining only the ‘different’ stimuli, we can determine whether subjects were more likely to notice the difference for voiced consonant stimuli—as is hypothesized—than for voiceless consonant stimuli. Here we see a clear bias, as predicted.

Looking at just the 4800 ‘different’ stimuli we can begin to assess whether subjects were more likely to notice an increased pitch contour on voiced stimuli as opposed to voiceless ones. Subjects showed a 52.9% success rate in noticing the raised f_0 contour on the voiced stimuli (1269/2400) compared to only a 40.4% success rate in noticing the raised f_0 contour on the voiceless stimuli (970/2400). This demonstrates a much higher success rate in noticing the raised f_0 contour for the voiced stimuli, where such a contour would not arise from coarticulation.

3.1. *D'* statistical analysis

The sensitivity index, d' , was calculated to assess the statistic significance of the data. This statistical measure is useful for determining a subject’s ability to perceive similarity or difference while taking into account the possibility of response biases—for example, subjects who primarily think everything sounds the ‘same.’ Thus, this statistic takes into account not only a subject’s ‘hits’—correctly identifying when the word repetition is different—but also ‘false alarms,’ where the subject believes the stimuli sounded different when they were in fact the same. Thus, a subject with a 100% hit rate but also a 100% false alarm rate would have a very low d' score, whereas a 100% hit rate and a 0% false alarm rate would result in a high score, indicating a high level of sensitivity in detecting the acoustic differences in the stimuli (MacMillan & Creelman, 2005). Two d' values were calculated for each subject, one quantifying sensitivity towards differences in the voiced stimuli and the other for the voiceless stimuli, using the same-different d' equation (rather than ‘yes-no’) via the *sensR* package in R. Thus, we can compare over all subjects to see if their d' scores are significantly higher for the voiced tokens, as is hypothesized.

Additionally, since d' is less accurate for very high hit and false alarm rates approaching 100% or 0% due to resulting in infinite z -scores (Stanislaw & Todorov, 1999), I followed a similar method to the log-linear approach detailed in Hautus (1995). In order to avoid 0% and 100% rates, a value of 1 was added to each of the false alarm, correct rejection, miss, and hit totals for each subject. Thus, a perfect hit rate of 24/24 and 0/24 misses would be corrected to 25/26 and 1/26 respectively.

D' was calculated for all 100 subjects in both the voiced and voiceless stimuli conditions. The mean d' for the voiced stimuli over all subjects was 1.59, whereas the mean for the voiceless stimuli was only 0.96. Subjects also showed a bias towards responding ‘same,’ rather than ‘different,’ reflected in a criteria location value, c , of 0.5938, the positive value here indicating a higher miss rate than false alarm rate. In order to assess statistical significance, a paired, two-tailed t -test was conducted comparing the individual voiced versus voiceless stimuli scores for each of the 100 subjects. The results show subjects’ d' for the voiced stimuli was significantly higher than the d' for voiceless stimuli ($p < 0.0001$, $t = 6.37$, $df = 100$), corroborating the assessment that subjects more readily could perceive the raised pitch contour on voiced rather than voiceless stimuli. The scores for all subjects are shown below in **Figure 6**, with d' values for voiceless stimuli along the y-axis and values for voiced stimuli along the x-axis. Each dot represents one of the 100

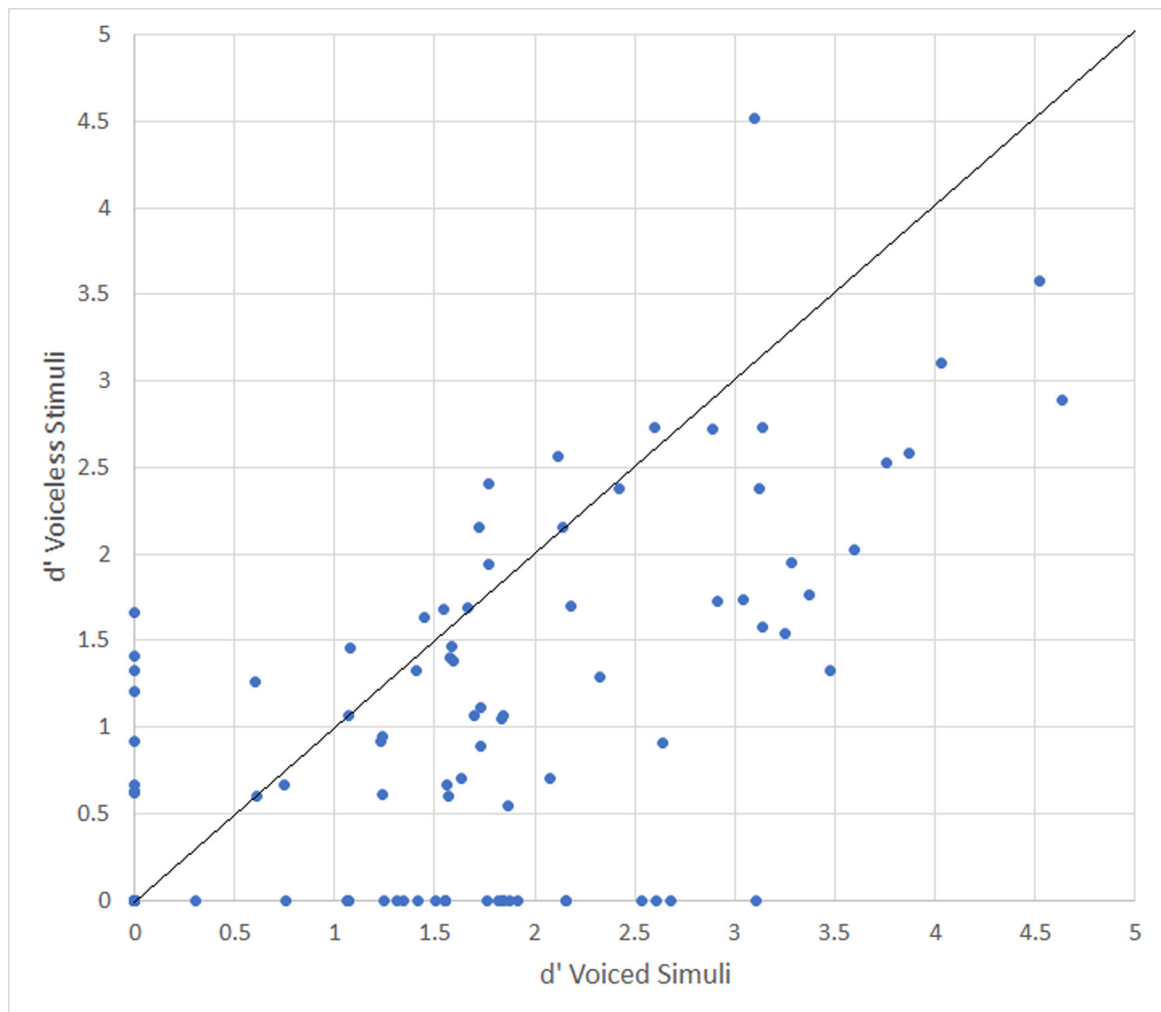


Figure 6: d' scores for voiced and voiceless stimuli over all subjects. Each dot represents a single subject.

participants, and the diagonal line represents an equal d' score for both the voiced and voiceless stimuli—thus those falling above the line demonstrated better discrimination of the voiceless tokens, while those falling below the line demonstrated better discrimination of the voiced tokens. The greater number of dots below the line reflects the bias towards better discrimination of the voiced tokens among individual subjects.

3.2. Principal Components Analysis

A post-hoc Principal Components Analysis (PCA) was also conducted in order to determine if subjects used different listening and/or response strategies. The PCA was run using the `prcomp` function in R. The model included twelve variables, which were the total number of correct responses (out of eight) for each of the six phoneme stimuli (/p t k b d g/) repeated either the same or different (6 phonemes \times 2 conditions). The results revealed just two principle components that accounted for more than 5% of the data.

As shown in the biplot in **Figure 7**, PC1 explains 57.1% of the variation in the data, whereas PC2 accounts for 18.4%. The red arrows show the contribution of each variable to the two PCs. For example, higher accuracy of the ‘different’ stimuli led to a higher PC1 score (thus the leftward points of the ‘same’ variables and the rightward points of the ‘different’ variables). Thus, we can conclude that PC1 is the component indicating either a ‘same’ or ‘different’ response bias. For example, subject 6 responded ‘different’ to all 96 stimuli, whereas subject 10 responded ‘same’ to all but one of the 96 stimuli.

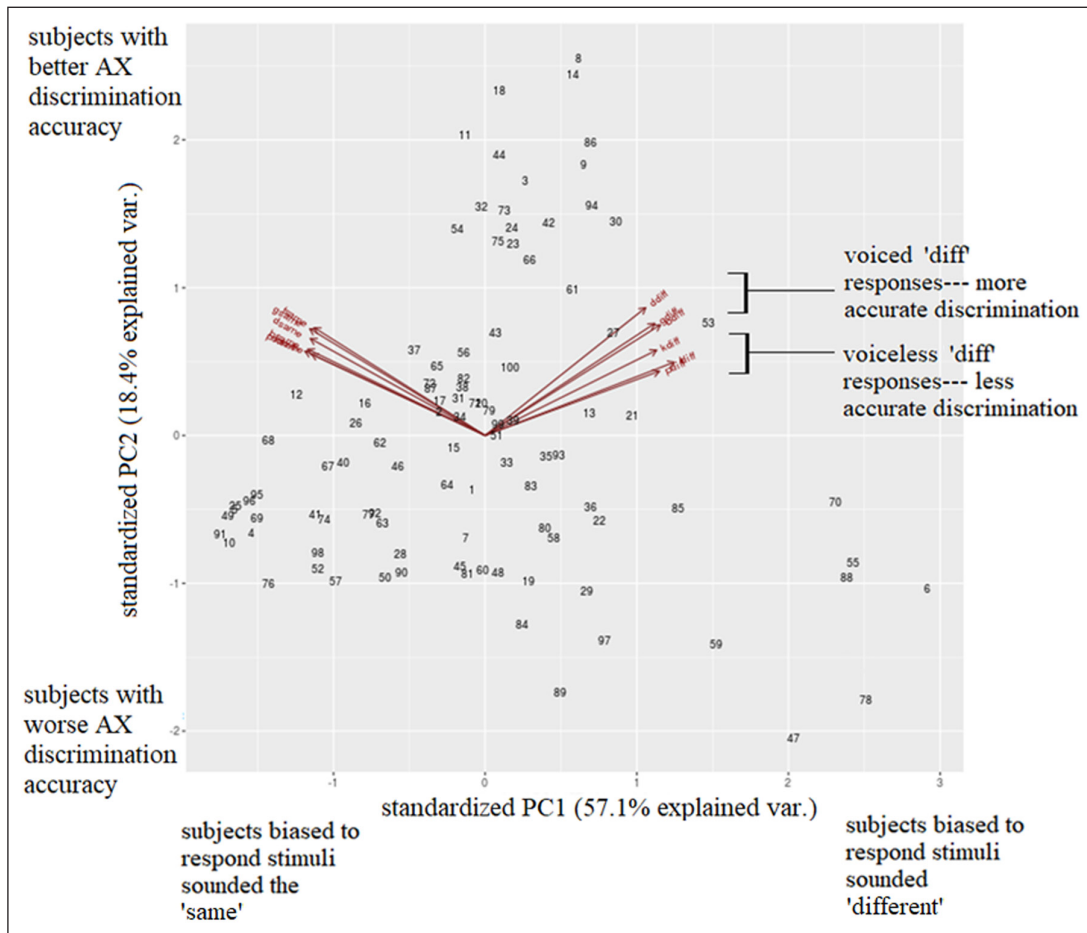


Figure 7: Principal Components Analysis of subjects.

Higher accuracy of *all* stimuli contributed to a higher PC2 score, suggesting that PC2 correlates with how accurately the subjects responded overall. For example, subject 8, with the highest PC2 score, correctly discriminated 89 of the 96 stimuli. The overall triangular shape of the subjects’ spread is indicative of the fact that the ones who did not have a strong ‘same’ or ‘different’ response bias usually did better overall. Further inspecting PC2, we see that the three voiced ‘different’ variables contributed more to a higher PC2 since subjects were more successful in discriminating those stimuli (thus the higher upward tilt of those arrows).

4. Discussion

The results show that subjects were better at discriminating target word stimuli when the only acoustic difference was variation that would *not* be an expected result of coarticulation, as compared to variation that would be expected to result from coarticulation. That is to say, subjects more readily noticed, stored, and/or maintained a raised f0 contour in memory following *voiced* consonants, but did not do this to the same extent following *voiceless* consonants, where such coarticulation would be expected. The interpretation of these results suggests consequences for both our current knowledge of exemplar theory as well as sound change.

4.1. Consequences for exemplar theory

The main findings from this study suggest that either some acoustic detail is filtered from exemplar memory or that certain details fade more rapidly from those exemplars. This departs from some other exemplar models in suggesting some degree of abstraction occurs

in the process of storing exemplars. For example, Johnson (1997b) suggests that the entire auditory spectrum is stored in memory. The current findings suggest that, contrary to this, that not all information is stored in its raw, veridical form. However, Johnson's assertion cannot be ruled out entirely. It could be the case that the raw auditory spectrum is briefly stored in a sort of working memory before being committed to exemplar storage. All that can be deduced for sure is that different features of the auditory signal, particularly those which are more predictable in some way, fade from exemplar memory more quickly. Johnson's (1997a) suggestion of reduced exemplars that store only formant information also requires some amendment. It is unclear that there is any automatic means of compressing the auditory data for storage in memory; rather, the listener's attention guides what is stored, and details that are less predictable in a particular context are more likely to be committed to longer term storage of the veridical details. Generally speaking, these findings align well with hybrid exemplar models (e.g., Pierrehumbert, 2002), which suggest listeners store both exemplars and abstract representations in memory. However, it remains unclear whether both types of exemplars exist, or if instead exemplars are typically a patchwork quilt of veridical and abstracted information. While further study is needed to understand the relationship between veridical and abstracted detail in memory, it seems probable that details of individual exemplars will continue to fade from memory and become abstracted over time. At some point, the memory trace may primarily contain phonemic information, or further abstract to represent mere words or ideas. However, this does not preclude the possibility that individual exemplar clouds—containing thousands of traces of varying degrees of abstraction—may be linked to some sort of fully abstract representation of the word.

The current findings also fit well with Goldinger's (2007) statement that "each stored 'exemplar' is actually a product of perceptual input combined with prior knowledge" (p. 50), as well as Hawkins' (2003, 2010) observations that the speech signal is processed only to the extent needed for extracting linguistic meaning. This also concurs with Goldinger & Azuma's (2003) application of Adaptive Resonance Theory (ART), which suggests that there is no fixed unit of speech perception—speakers adaptively process speech units in whatever way achieves the quickest comprehension of the linguistic meaning of the speech signal. We could further apply the current findings to this model to suggest that the details processed in speech perception are what end up being encoded in the memory traces themselves. Ultimately, acoustic information that is predictable based on coarticulation or context (such as found in Manker, 2019) may not survive in the exemplar.

It should also be noted that the current findings do not suggest that any automatic process of abstraction must act in transforming the speech signal before word recognition, as is typical in models of speech perception including normalization (Gerstman, 1968; Tranmüller, 1981, etc.). Rather, those details that are particularly predictable and/or redundant in speech perception may be most likely to fade from exemplar memory. In fact, some subphonemic information is often facilitative in speech recognition (Johnson, 1997a), and in such cases, I would expect these details would more likely survive in exemplar memory. Further study will examine this question more closely.

While the results strongly demonstrate an effect of perceptual salience of coarticulatory details in a given phonetic context, an alternative analysis could challenge whether the observed perceptual bias is relevant to exemplar storage at all. For example, perhaps listeners at some point in the experiment became aware that 'different' always meant a raised pitch contour, at which point they began to ignore the initial utterance and only focus on the repetition. In this case, the difference in perceptual salience was all that motivated the observed bias, with no bias in what was originally stored or maintained in exemplar memory. An experimental design including initial utterances with raised pitch

which is then repeated would be able to rule out or confirm this possible explanation. However, I believe this alternative account is unlikely, primarily due to the inclusion of the filler ‘word-identification’ stimuli. In these cases, no word was repeated at all, so it encouraged listeners to pay close attention to the initial utterance of the word and not only its (possible) repetition. Additionally, the dual tasks would likely distract from subjects’ attempts to determine any patterns in the repetitions. In any case, further research can explore and disambiguate this competing interpretation.

4.2. Relation of findings to compensation for coarticulation

Compensation for coarticulation is a phenomenon whereby listeners perceptually ‘undo’ coarticulatory effects of neighboring sounds in order to determine the intended underlying segments. This was famously observed in Mann and Repp (1980), in which subjects were more likely to perceive an acoustically ambiguous fricative as the sound [s] rather than [ʃ] following [u], arguably due to the coarticulatory effect caused by the rounded vowel [u], which tends to cause the neighboring sounds to lower in frequency. The mechanics of this process suggest something similar to the effect found in the current paper—that listeners perceptually remove an initial pitch raise following voiceless sounds as an expected effect of coarticulation, thus stripping away this detail in exemplar memory. However, Holt, Lotto, and Kluender (2001) find sensitivity to the coarticulatory relationship between F0 and voicing in Japanese quail, suggesting awareness of this relationship is not rooted in human speech perception. As a result, it is not clear if the phenomenon observed in the present paper is the result of compensation for coarticulation or a distinct phenomenon rooted in more general acoustic predictability, though the perceptual consequences may be quite similar. Future research and analysis are needed to investigate the relationship of predictability-modulated acoustic awareness and compensation for coarticulation.

4.3. Consequences for sound change

Ohala’s (1981, 1983) account of perceptual correction, and other studies of compensation for coarticulation (Yu, 2010; Yu, Abrego-Collier, & Sonderegger, 2013), have considered the role of perceptual biases in sound change. For example, Ohala (1981) claims that hypocorrection is one source of sound change—when speakers notice certain coarticulatory details but do not attribute them to their phonological environment. Yu (2010) found that female subjects with low Autism Quotient scores demonstrated less compensation for coarticulation, being more likely to notice certain articulatory effects (e.g., hearing [ʃ] before [u] instead of [s] and not attributing the lowered fricative frequencies to coarticulation). The results of the current study could be applied to either of these accounts, following the proposal that such ‘misperception’ results from encoding predictable coarticulatory effects in exemplar memory.

The current results make some additional predictions, however. New variation that is phonologically predictable, such as coarticulation, even when exaggerated a bit beyond what listeners are used to hearing, is more likely to evade notice and fail to be retained in memory, at least shortly after perceiving these details. However, new variation that is phonologically unconditioned would be more likely to be stored and maintained in exemplar memory. If, following the phonetic accommodation paradigm, we assume that new exemplar traces inform future productions, then we might expect conditioned versus unconditioned sound changes to spread differently within a language. For example, a conditioned change like the nasalization of vowels before nasal consonants (followed by their eventual loss) should more likely evade the notice of listeners since it is a predictable result of coarticulation, though a change of this sort is under the constant articulatory pressure that causes coarticulation in the first place. On the other hand, an unconditioned

change, like a chain shift causing /p t k/ to become /p^h t^h k^h/ in all phonological environments should be more likely to be encoded into listeners' exemplar memories since the change is not phonologically predictable. However, whereas we might predict its spread from speaker to speaker may happen more rapidly, it is not clear how strong the motivation is to begin in the first place, since unconditioned changes may be influenced more by phonological considerations (e.g., pressures within the sound system) rather than articulatory pressure.

Additionally, it is not clear how the effect observed in this study would be maintained over longer periods of time, necessary for eventually permanent changes in a language. For example, predictable detail, while stored in memory less faithfully and possibly abstracted in some way, may actually be retained for longer, whereas the more veridical memories of unpredictable memory may fade more quickly. In any case, the results may suggest some differences in the way that conditioned and unconditioned sound changes spread, though much more work is needed to understand the relevance, if any, of predictability-based perceptual biases in exemplar storage on sound change.

4.4. Future research

Several important questions remain in order to understand the nature and contents of exemplars. First of all, it is necessary to continue to survey how different acoustic cues are stored in exemplar memory, and the various perceptual biases that facilitate or impede the storage of various auditory information. One question of particular interest will be whether or not certain coarticulatory cues are in fact stored in memory. While I have suggested that predictable coarticulatory cues may be ignored and perceptually filtered, there is a lot of research showing that coarticulation can be used to facilitate speech recognition (Johnson, 1997a; Beddor, Krakow, & Lindemann, 2001). Beddor et al., for example, state that “listeners use coarticulatory variation as information about sounds that are further up or down the speech stream” (p. 56). If coarticulation aids in speech recognition in this way, following my previous proposal that details that are used in speech recognition will more likely survive in exemplar memory, this should result in better storage of these details. However, it is unclear how the events of auditory storage and maintenance unfold. For example, perhaps once phoneme or word recognition occurs, such coarticulatory details are rapidly lost from memory or abstracted. Alternatively, there could be some difference in anticipatory versus confirmatory coarticulation. In the present study, the f₀ modulation occurred after the voicing and VOT cues had provided ample evidence as to the initial sounds in the target words (e.g., ‘bath,’ ‘path,’ etc.). The f₀ cue only served as an expected confirmation. On the other hand, nasality on a vowel that cues an upcoming nasal consonant occurs at a point in time when the listener does not know the upcoming sound (such as is shown for English by Lahiri & Marslen-Wilson, 1991). Thus, anticipatory coarticulation of this sort might be more faithfully stored and maintained in memory.

Further research should also consider additional aspects of predictability and expectation. For example, predictability of a person's voice may lead to lower awareness of certain acoustic information—more acoustic information might be stored when listening to different speakers produce stimuli, similar to the stronger imitative effect found in Goldinger (1998) when subjects heard multiple model speakers. In addition to the contextual predictability effect found in Manker (2019), which was based on word priming, we could consider whether general situational context also results in lower attention and less faithful storage of the auditory signal. For example, if objects are visually presented before they are referred to, will listeners store less auditory detail of these words? Finally, we might consider how the predictability of semantic context interacts with other forms of predictability, such as phonologically predictable coarticulation.

Lastly, further work in the phonetic accommodation paradigm will complement the findings of the current study. For example, as in Manker (2019), should we expect to find greater accommodation of predictable coarticulation compared to unpredictable phonetic variation? Zellou, Scarborough, and Nielsen (2016), for example, did in fact find imitation of ‘hyper-nasalized’ vowel coarticulation, such that speakers did in fact increase their own vowel nasality after hearing greater vowel nasality produced. From this study alone, it is not clear whether such imitation of coarticulation is weaker in magnitude than imitation of unconditioned acoustic variation. Secondly, this is again a case of an anticipatory coarticulation effect, so would the effect be weaker for perseverative coarticulation? Further research will help to interpret the growing body of literature in phonetic accommodation as a whole and its relevance in the storage and maintenance of detail in exemplar memory.

5. Conclusion

The present study addresses the question of whether predictable coarticulatory detail is stored and maintained in exemplar memory to the same degree as unpredictable acoustic variation. The results of an AX discrimination task show that subjects did a significantly better job at discriminating tokens that differed in phonologically unpredictable ways as opposed to those that merely displayed expected coarticulatory detail. This suggests some degree of filtering or abstraction occurs in exemplar storage and is modulated by the predictability of the variation. Future research will continue to address the phenomenon of predictability and expectation, how it shapes the contents of exemplars, and its relevance in sound change.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** List of stimuli used for this experiment. This includes the target words, which included voiced-voiceless minimal pairs, as well as fillers, which were single syllable words with no other phonological restrictions. DOI: <https://doi.org/10.5334/labphon.240.s1>

Competing Interests

The author has no competing financial, professional, or personal interests that might have affected the objectivity or integrity of this publication.

References

- Beddor, P., Krakow, R., & Lindemann, S. (2001). Patterns of perceptual compensation and their phonological consequences. In E. Hume & K. Johnson (Eds.), *The Role of Speech Perception in Phonology* (pp. 55–78). San Diego: Academic Press.
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.
- Gerstman, L. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics, AU-16* (pp. 78–80). DOI: <https://doi.org/10.1109/TAU.1968.1161953>
- Gill, H., & Gleason, H. (1972). The salient features of the Punjabi language. *Pakha Sanjam, 4*, 1–3.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166–1183. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.4638&rep=rep1&type=pdf>. DOI: <https://doi.org/10.1037/0278-7393.22.5.1166>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>

- Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 49–54). Retrieved from: <http://icphs2007.de/conference/Papers/1781/1781.pdf>
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*, 305–320. DOI: [https://doi.org/10.1016/S0095-4470\(03\)00030-5](https://doi.org/10.1016/S0095-4470(03)00030-5)
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, *27*, 46–51. DOI: <https://doi.org/10.3758/BF03203619>
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3–4), 373–405. DOI: <https://doi.org/10.1016/j.wocn.2003.09.006>
- Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. *Laboratory Phonology*, *10*, 479–510. DOI: <https://doi.org/10.1515/9783110224917.5.479>
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, *109*, 764–774. DOI: <https://doi.org/10.1121/1.1339825>
- Hombert, J., & Ladefoged, P. (1976). The effect of aspiration on the fundamental frequency of the following vowel. *Journal of the Acoustical Society of America*, *59*, S72 (abstract). DOI: <https://doi.org/10.1121/1.2002863>
- Hombert, J., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. *Language*, *55*, 37–58. Retrieved from: http://linguistics.berkeley.edu/~ohala/papers/phonet_expl_tones.pdf. DOI: <https://doi.org/10.2307/412518>
- House, A., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *10*, 105–113. DOI: <https://doi.org/10.1121/1.1906982>
- Johnson, K. (1997a). Speech perception without speaker normalization: An exemplar model. In Johnson & Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press. Retrieved from: <http://linguistics.berkeley.edu/~kjohnson/papers/SpeechPerceptionWithoutSpeakerNormalization.pdf>
- Johnson, K. (1997b). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics*, *50*, 101–113. Columbus, Ohio. Retrieved from: <http://linguistics.berkeley.edu/~kjohnson/papers/Johnson1997.pdf>
- Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M. J. Solé, P. Beddor & M. Ohala (Eds.), *Experimental Approaches to Phonology. In Honor of John Ohala* (pp. 25–40). Oxford University Press.
- Kingston, J. (1989). The effect of macroscopic context on consonantal perturbations of fundamental frequency. *Journal of the Acoustical Society of America*, *85*, S149 (abstract). DOI: <https://doi.org/10.1121/1.2026802>
- Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, J. Ewen Colin, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. Chapter 97). Malden, MA & Oxford: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444335262.wbctp0097>
- Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language*, *70*, 419–454. DOI: <https://doi.org/10.1353/lan.1994.0023>
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, *36*, 28–54. DOI: <https://doi.org/10.1016/j.wocn.2007.02.001>

- Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243–288). Hillsdale, NJ: Erlbaum. DOI: [https://doi.org/10.1016/S0095-4470\(19\)31059-9](https://doi.org/10.1016/S0095-4470(19)31059-9)
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3), 245–294. DOI: [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- Lehiste, I., & Peterson, G. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, 33, 419–425. DOI: <https://doi.org/10.1121/1.1908681>
- Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461. DOI: <https://doi.org/10.1037/h0020279>
- Luce, P., & McLennan, C. (2005). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 591–609). Malden, MA: Blackwell. Retrieved from: <https://pdfs.semanticscholar.org/1eb3/da8316bf9a3804f32c6d1414a68331c7404e.pdf>. DOI: <https://doi.org/10.1002/9780470757024.ch24>
- MacMillan, N., & Creelman, C. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Manker, J. (2019). Contextual Predictability and Phonetic Attention. *Journal of Phonetics*, 75, 94–112. DOI: <https://doi.org/10.1016/j.wocn.2019.05.005>
- Mann, V., & Repp, B. (1980). Influence of vocalic context on the perception of the [ʃ]-[s] distinction. *Perception and Psychophysics*, 28, 213–228. DOI: <https://doi.org/10.3758/BF03204377>
- McLennan, C., & Luce, P. (2005). Examining the Time Course of Indexical Specificity Effects in Spoken Word Recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31(2), 306–321. DOI: <https://doi.org/10.1037/0278-7393.31.2.306>
- Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. DOI: <https://doi.org/10.1016/j.wocn.2010.12.007>
- Nye, P., & Fowler, C. (2003). Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31(1), 63–79. Retrieved from: <http://www.haskins.yale.edu/Reprints/HL1279.pdf>. DOI: [https://doi.org/10.1016/S0095-4470\(02\)00072-4](https://doi.org/10.1016/S0095-4470(02)00072-4)
- Ohala, J. (1981). The listener as the source of sound change. In C. Masek, R. Hendrick & M. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago: Chicago Linguistics Society. Retrieved from: http://linguistics.berkeley.edu/~ohala/papers/listener_as_source.pdf
- Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag. Retrieved from: <http://linguistics.berkeley.edu/~ohala/papers/macn83.pdf>. DOI: https://doi.org/10.1007/978-1-4613-8202-7_9
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110197105.101>
- Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2, 33–52. DOI: <https://doi.org/10.1146/annurev-linguistics-030514-125050>
- Ratliff, M. (2015). Tonoexodus, tonogenesis, and tone change. In P. Honeybone & J. Salmons (Eds.), *Handbook of Historical Phonology* (pp. 245–261). Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199232819.013.021>

- Shockley, K., Sabadini, L., & Fowler, C. (2004). Imitation in shadowing words. *Perception and Psychophysics*, 66, 422–429. DOI: <https://doi.org/10.3758/BF03194890>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149. DOI: <https://doi.org/10.3758/BF03207704>
- Stevens, K. (1972). The quantal nature of speech: Evidence from articulatory acoustic data. In D. Denes, (Ed.), *Human communication: A unified view*. New York: McGraw-Hill.
- Svantesson, J.-O. (1991). Hu: A language with unorthodox tonogenesis. In J. Davidson (Ed.), *Austroasiatic Languages: Essays in Honour of H. L. Shorto* (pp. 67–79). London: SOAS.
- Thurgood, G. (1999). *From Ancient Cham to Modern Dialects: Two Hundred Years of Language Contact and Change*. Honolulu: University of Hawai'i Press.
- Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37(3), 276–296. DOI: <https://doi.org/10.1016/j.wocn.2009.03.004>
- Tranmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustic Society of America*, 69, 1465–1475. DOI: <https://doi.org/10.1121/1.385780>
- Wolff, E. (1987). Consonant-tone interference in Chadic and its implications for a theory of tonogenesis in Afroasiatic. In D. Barreteau (Ed.), *Langues et cultures dans le bassin du Lac Tchad* (pp. 193–216). Paris: ORSTOM.
- Yu, A. (2010). 'Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change.' 2010. *PLoS ONE*, 5(8). DOI: <https://doi.org/10.1371/journal.pone.0011950>
- Yu, A., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLOS ONE*, 8(9), e74746. DOI: <https://doi.org/10.1371/journal.pone.0074746>
- Yu, A., & Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals. A cross-validation study. *Journal of the Acoustical Society of America*, 136(1), 382–388. DOI: <https://doi.org/10.1121/1.4883380>
- Zellou, G., Scarborough, R., & Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *Journal of the Acoustical Society of America*, 140, 3560–3575. DOI: <https://doi.org/10.1121/1.4966232>


How to cite this article: Manker, J. 2020 The perceptual filtering of predictable coarticulation in exemplar memory. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1):20, pp. 1–17. DOI: <https://doi.org/10.5334/labphon.240>

Submitted: 18 October 2019

Accepted: 28 September 2020

Published: 19 November 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 