

Context, Predictability and Phonetic Attention

Jonathan Manker

Department of Linguistics

University of California, Berkeley

Berkeley, California, United States of America

[jtmanker@berkeley.edu](mailto:jtmanker@berkeley.edu)

Abstract

Lindblom et al. (1995) proposed two modes of listening to speech: a “what” mode, in which listeners focus on meaning, and a “how” mode, where listeners attend to details of pronunciation. This theory fits with Hickok and Poeppel’s (2004, 2007) more recent dual stream model of speech perception. What conditions then are necessary for modulating the use of one listening mode or the other? Following observations concerning the effect of higher level linguistic information on speech perception (Cole & Jakimik 1980, etc.), I will detail the results of two experiments which consider how structural and semantic context (word predictability) interact with the listener’s attention to phonetic details. The experiments use the phonetic accommodation or imitation paradigm (Goldinger 1998, etc.) as a tool to determine what phonetic details subjects noticed after hearing target words in a variety of contexts. The first experiment compares the degree of accommodation in isolated phrase vs. sentence context. The second experiment considers how the variable of word predictability within the context of a sentence influences the degree of accommodation. The results suggest listeners attend more closely to sub-phonemic details of pronunciation when less structural and semantic context is present and that contextual predictability modulates phonetic attention.

Keywords: *speech perception, imitation, accommodation, attention, listening modes, predictability, context*

## **1. Introduction**

### *1.1. Background: Modes of Listening*

Decades of research in speech perception have supported the theory that word recognition in fluent speech is achieved through a combination of bottom-up and top-down processing; that is, listeners make use of the acoustic signal as well as linguistic knowledge, such as phonotactics, morphosyntactic rules, semantics, and pragmatics in order to understand fluent speech (Cole & Jakimik 1980). Early studies such as Pollack and Pickett (1963) showed that subjects had significant difficulty in recognizing isolated words excised from sentence productions. Warren (1970) famously documented the *phoneme restoration effect*, whereby listeners fail to notice a phoneme is missing or replaced by a noise. Warren presented listeners with a word such as “legislatures” within a sentence context, but replaced the /s/ with a cough.

Subjects were unable to determine the location of the cough, suggesting that the context helped fill in missing phonetic cues.

Several other studies suggest that listeners pay attention to phonetic details differently depending on the context. Marslen-Wilson & Welsh (1978) observed subjects who were asked to shadow words (repeated immediately after hearing). Some of the model words contained errors, but were “fluently restored” such that the subject corrected the error immediately with no disruption. The results showed that fluent restoration was more likely to occur in the third syllable of a word than the first, suggesting that subjects were paying closer attention to the phonetic details at the beginning of the word, but by the time the word was recognized, these details were more likely to pass unnoticed. Cole, Jakimik, & Cooper (1978) found a similar result, such that speakers were more likely to notice a mistake in the onset of a word (‘made’ pronounced ‘nade’) as opposed to in the coda (‘time’ pronounced ‘tine’). Samuel (1981) revisited Warren’s (1970) phonemic restoration effect, considering the phenomenon specifically with respect to context. His findings show a more robust restoration effect in contextually predictable sentences, leading to his statement that “restoration is a function of context” (p. 481). Exploring this idea further, Samuel (1987) investigates how the location of *uniqueness points* in words can influence phonemic restoration. A uniqueness point is a location in a word at which the word no longer has any competitors for word recognition that would match the string of phonemes to that point. Samuel found that words with early uniqueness points had stronger phonemic restoration effects when occurring late in the word, suggesting an interaction between the top down information involved in word recognition and how listeners pay attention to phonetic details.

Lindblom et al. (1995), building on observations made by Ohala (1981, 1983), proposed two modes of listening which may be active under different circumstances. The ‘what’ mode is the standard means of listening to speech, in which listeners focus on the content and meaning of what is being said. In this mode, listeners make full use of all higher level knowledge (phonology, syntax, semantics) for aiding word recognition. In the ‘how’ mode, speakers are more concerned with the phonetic detail of speech, where these details might include phenomena such as coarticulation and reduction. Ohala (1981) proposed that situations in which a single speaker may adopt a new variant may arise from misperception when a listener fails to undo a coarticulatory effect, and does not reconstruct the speaker’s intended form. Building on this, Lindblom (1995) suggests this could be when a listener processes speech in

the “how” rather than the “what” mode, and the acoustic signal is processed in its raw, literal form, causing a listener to fail to normalize a coarticulatory effect.

More recent literature in neurolinguistics has provided evidence of different neural pathways that might be involved in ‘what’ and ‘how’ listening. Hickok and Poeppel (2004, 2007) define two streams of processing involved in what they refer to as *speech recognition* and *speech perception*. The first stage of processing the acoustic signal is bilateral and occurs in the superior temporal gyrus, however, further processing progresses along one of two streams. The *ventral* stream involves projection ventro-laterally toward the inferior posterior temporal cortex bilaterally. It is here that lexical retrieval, sound to semantic mapping, occurs (*speech recognition*), interfacing with various regions of the brain where conceptual information (visual, motor, etc.) is stored. The *dorsal* stream diverges from the superior temporal gyrus, projecting dorso-posteriorly in the region of the posterior Sylvian fissure and progressing finally towards the frontal regions, primarily oriented on the left side. The dorsal stream is involved both in motor planning as well as the sublexical processing of acoustic input (*speech perception*) such as phoneme identification and rhyming tasks; this dual function suggests a link between production and perception. The ventral stream is roughly analogous to Lindblom’s ‘what’ mode, while the dorsal stream corresponds to the ‘how’ mode.

### *1.2. Background: Phonetic Accommodation*

Phonetic accommodation, or imitation, is a phenomenon whereby a speakers’ pronunciation is subtly, and often subconsciously, influenced by perceiving the phonetic details of other people’s speech. The phonetic accommodation paradigm is rooted in exemplar-based theories of speech (Johnson 1997, Pierrehumbert 2002) which propose that individual tokens of words are stored in memory. Recognizing words may be achieved without the need to normalize speech to abstracted, phonological forms of words (though exemplars may be used alongside abstracted forms, as proposed by Pierrehumbert 2002). Goldinger (1996) found that listeners use details of a speaker’s voice to aid in speech recognition, rather than discarding these details for the purpose of normalization.

Sancier & Fowler (1997) was an early study which examined the commonly held observation that a person’s speech may change after exposure to speech of another dialect or language. Their study observed the speech of a bilingual speaker of Brazilian Portuguese and English after spending prolonged amounts of time in either Brazil or the United States. After

hearing the relatively longer VOT (voice onset time) of American English stops for several months, the subject's Brazilian English stops were produced with significantly longer VOT than normal as well. Goldinger (1998) was among the earliest phonetic accommodation studies which attempted to induce the phenomenon in the lab over short amounts of time. Subjects were asked to produce words before and after hearing the same words produced by a model speaker. If speakers were shifting their speech in the direction of the model, it would indicate that subjects were processing and storing the phonetic details, at least for a short amount of time. To judge the similarity of the subjects' productions to that of the model, other subjects were given an AXB task in which they were asked whether a given subject's pre-stimulus (A) production or post-stimulus production (B) was closer to that of the stimulus (X). Goldinger found that speakers did in fact sound more like the model after hearing the model, and thus were storing detailed word exemplars, even when not given instructions to imitate. Additionally, Goldinger found that immediate repetition yielded a stronger imitation effect than delayed repetition, suggesting a particular exemplar will become less influential over time. Also, low frequency words are better imitated than higher frequency words, perhaps due to that fact that speakers have more stored exemplars of high frequency words which compete with the stimulus. Finally, Goldinger showed that multiple repetitions of the stimulus also yielded a stronger imitation effect, presumably adding to the listener's pool of exemplars.

The years following Goldinger's study have witnessed a blossoming of phonetic accommodation studies, in both methodology and scope. Goldinger's (1998) study utilized the AXB perception task for determining similarity to the stimulus, which has the strength of being able to account for similarity along any number of dimensions and for any phonetic cues that the judges happen to notice. However, this model has weaknesses in being impressionistic and imprecise in determining just what cues are being imitated by the subjects. Later studies involved lab-induced imitation in which researchers obtained quantitative measurements of specific linguistic features and were able to pinpoint the features subjects were imitating. Shockley, Sabadini, & Fowler (2004) presented listeners with stimulus words with artificially lengthened VOT, and found that subjects stored this subphonemic information in their exemplars, evidenced by a significant and measurable increase in VOT after hearing the stimulus. Other studies have shown imitation to other phonetic features such as vowel quality (Tilsen 2009, Honorof et al. 2011, Babel 2010, Babel 2012), F<sub>0</sub> (Mixdorff et al. 2012, Heath 2014), and vowel nasalization (Zellou et al. 2013).

Other studies of phonetic accommodation investigated various aspects of the mechanisms behind the phenomenon. Pickering & Garrod (2004) proposed the *interactive alignment* account in which speakers automatically “align” at all linguistic levels--- from phonetics to semantics--- for the purpose of facilitating communication in dialogue. Shockley, Sabadini, & Fowler (2004) propose that humans have “a fundamental disposition to imitate” (p. 422) suggesting the process of phonetic accommodation is automatic and occurs without any intention to mimic the speech that we hear. Lewandowski (2012) found that speakers accommodated even when explicitly told *not* to sound like their interlocutors. Nevertheless, others have explored mechanisms that may impede the process. Babel (2010) found that subjects’ attitudes towards a particular group of people and their dialect influenced the phenomenon of accommodation. In this case, New Zealanders who held negatives of opinions of Australians showed less convergence to the model’s Australian English, while those holding more positive views showed a greater degree of accommodation. Babel (2012) showed that physical attractiveness may also play a role in phonetic accommodation. Additionally, she found that subjects accommodated more to low rather than high vowels, suggesting specificity in which phonetic characteristics were accommodated. Pardo (2006) found that in conversational pairs, both the gender and conversation role (the one either giving or receiving the instructions) also facilitated or impeded the process of accommodation.

Phonetic accommodation has been shown to be sensitive to and influenced by phonological factors. Nielsen (2011) considered the relationship between exemplars and abstracted representations of sounds. Her study found that hearing target words with lengthened VOT affected the production of other words that had not been heard as part of the stimulus. That is to say, hearing /p/ produced with lengthened VOT in certain words affected other instances of /p/, and even new productions of /k/ (being another voiceless aspirated stop), though to a lesser effect, suggesting the presence of both exemplar and abstracted representations. Such results suggest an extra dimension of complexity in the model not considered in earlier studies of phonetic accommodation.

Furthermore, Nielsen (2011) showed that speakers accommodated lengthened VOT of voiceless stops, but did not accommodate reduced VOT of voiceless stops. The restriction may be rooted in phonological considerations. VOT is a salient phonetic cue which distinguishes voiced and voiceless stops, having relatively shorter and longer VOT respectively. Lengthening the VOT of voiceless stops may help enhance their distinctiveness compared to their voiced

stop counterparts. On the other hand, shortening the VOT of voiceless stops causes them to become more like voiced stops, encroaching on the perceptual territory of voiced stop VOT. Mitterer & Ernestus (2008) also found sensitivity of phonetic accommodation to a language's phonological system. Their study found that subjects would not accommodate to categorical phonological differences such as the type of trill used in Dutch--- alveolar or uvular--- suggesting unintentional accommodation occurs mostly along gradient, phonetic dimensions. Overall, research in the area of phonetic accommodation in the last decade has revealed a number of factors--- social and phonological--- which may facilitate or impede the phenomenon from occurring.

The vast majority of phonetic accommodation studies have only observed the phenomenon in the production of isolated words. Isolated words are devoid of any additional context and their predictability would be identical to their frequency. As a result, few studies have considered how context and predictability, or other top down factors, interact with phonetic accommodation. One exception is Nye & Fowler (2003), which investigated how “order of approximation” to English influenced the degree of accommodation. Sentences of nonce words were constructed that range in how closely they resembled English, primarily in phonotactics. Higher orders of approximation included many words that closely resemble English words, whereas lower orders included words that grossly violated English phonotactic rules (lax vowels in open syllables, words lacking vowels, etc.). The results of the study showed higher degrees of imitation to lower orders of approximation to English. This suggested that speakers drew more heavily from their own linguistic experience whenever possible, indicating an interaction with higher level linguistic knowledge (phonotactics, the lexicon, etc.). This observation harmonizes with Goldinger's findings that subjects accommodate more to words of lower frequency, suggesting the new exemplars have less competition with the older exemplars. In any case, Nye & Fowler (2003) used an AXB perception task to determine the similarity of the pre- and post-stimulus productions to the model. Thus, it is unclear what features and what words were being accommodated within the sentence framework, and whether the structure and semantics supplied by the sentence interacted with the degree of accommodation.

More recent work in exemplar theory (beyond studies of phonetic accommodation) has also led to findings that higher level linguistic information can modulate attention to details, affecting what it stored in memory. Goldinger (2007) proposes that “each stored exemplar is

actually a product of perceptual input combined with prior knowledge...” (p. 50) suggesting that exemplars are not merely raw acoustic data, but are filtered by a listener’s cognitive experience. Such a proposal may be reflected in the findings of Pierrehumbert (2006) which shows that speakers attend more to informative socio-indexical features. Likewise, Maye (2007) demonstrates that a listeners’ exemplars are shaped by past linguistic experience, where ones’ L1 has trained the listener to cue in to specific phonetic details.

Thus, results from previous phonetic accommodation studies reveal some findings hinting at an interaction between higher level knowledge of the language and the occurrence or degree of imitation observed (particularly related to word frequency). This is also paralleled by more recent investigations concerning the mechanisms of exemplar theory, suggesting that higher level information may act to filter what phonetic details are stored and processed. Additionally, strong observations from the speech perception literature suggest that higher level linguistic knowledge has a more robust effect on speech perception with more context to draw from, whereas reliance on the acoustic signal is more crucial when no context is present. Thus, it can be hypothesized that listeners may attend to phonetic details more strongly in words as opposed to sentences, or when a word is unpredictable in a given sentence context. Listeners may utilize the “what” or ventral mode in the presence of context and higher level information, and “how” or dorsal mode in its absence, when exemplars may be stored more as raw, unfiltered data. The listening mode may then be modulated by the structural context of the utterance (isolated phrases vs. within sentences) or the contextual predictability. The experiments in this study address these questions, considering how inducing different listening modes through the presence or absence of context can lead to different degrees of phonetic attention and thus phonetic accommodation.

## **2. Experiment #1: Structural Context and Phonetic Attention**

### *2.1. Goal / Hypothesis*

The first experiment largely follows the design of imitation experiments such as Shockley, Sabadini, and Fowler (2003) and Nielsen (2011), which showed subjects imitated lengthened VOT after hearing the stimulus. However, this experiment compares the effect of hearing the target word within a context-free phrase<sup>1</sup> (‘the *pelican*’) as opposed to being embedded in a sentence (‘The *pelican* is flying over the beach’). Our hypothesis is that

sentential context, full of higher level syntactic and semantic information, will activate ventral ‘what’ listening, while isolated words or phrases will be processing using dorsal ‘how’ listening. As a result, a less phonetic accommodation should be observed in isolation as opposed to in sentential context.

## *2.2. Method*

### *2.2.1. Stimuli*

All target words in this experiment begin with the target phoneme /p/ (see Appendix A for a full list of the stimuli used in this experiment). All target words had initial stress, were from one to three syllables in length with the first syllable following a CVC(C) structure, avoiding any clusters with /p/. Following the observations of Goldinger (1998), we avoided words of very high frequency in order to encourage phonetic accommodation. The model for the experiment was a female speaker of Australian English. While most of the potential subjects were likely to be speakers of American English, we felt the use of a speaker with some phonetic differences from the subjects might be more likely to induce closer phonetic listening in both isolated phrase and sentential context.

The target sentences were recorded in a sound booth in the UC Berkeley PhonLab and were read at a natural pace. After recording, the VOT of the initial /p/ sounds of the target words were artificially lengthened to twice their initial length, or to a minimum of 120 ms in the case the tokens did not pass this threshold by doubling their initial VOT. A stable portion of the VOT, beyond the burst but before any hint of voicing, was spliced from the original token, copied, and inserted between the burst and onset of voicing. Some care had to be taken to ensure the tokens were natural sounding, which in some cases required revision. Splicing at zero-crossings improved the naturalness of the tokens.

### *2.2.2. Procedure*

The experiment consisted of three blocks: (1) baseline reading; (2) immediate shadowing; (3) post-exposure reading, lasting approximately 20 minutes. The experiment was conducted in a sound booth in UC Berkeley’s PhonLab, where subjects were recorded with an AKG C3000 microphone (recorded as .wav files at a sampling rate of 22.1 KHz) and heard stimuli via AKG K271 headphones, adjusted to a comfortable volume. The experiment was run on a Lenovo ThinkCentre desktop computer using a template Python script developed for administering imitation and accommodation experiments.

As shown in Table 1, the first block, the baseline reading, consisted of 60 two-word phrases, both nouns and verbs, read off the screen in a random order. The noun phrases were of the form *determiner* + *noun* (e.g., ‘the pelican,’ ‘the porch,’ ‘the pantry’) while the verb (infinitive) phrases were of the form ‘to’ + *verb* (e.g., ‘to perish,’ ‘to publish,’ etc.). Of the 60 words in the baseline reading, 20 were target /p/ noun phrases that would end up being heard in immediate shadowing block, along with 20 additional “novel” /p/ noun phrases and 20 “novel” verb infinitive phrases which would not be heard in the immediate shadowing block. We used this method in order to replicate the carryover effect found in Nielsen (2011), where the effect of VOT lengthening in the target words was carried over to words that had not been heard as part of the stimulus, but began with the same or similar phoneme. In the second block, the immediate shadowing, participants in two different conditions heard either (A) 20 target sentence stimuli containing the target phrases, played twice in random order for a total of 40 repetitions, or (B) 20 stimuli consisting only of the target noun phrases, also played twice in random order for a total of 40 repetitions. In the case of condition (A), the subjects were told to listen for the /p/ initial word in the sentence, and repeat only that word and the word immediately before it (the determiner). For condition (B), the target phrases heard in the stimulus were extracted from the original recordings of the entire sentences, such that condition (A) and (B) were both hearing the exact same recording of the exact same target phrase, but either in isolation or in sentence context. The third block, the post-exposure reading, was identical to the first block but allowed for measuring the effect of the stimulus on the target words. Crucially, the only difference between conditions A and B was the structural context of the stimulus heard in the immediate shadowing block, where the phrase was heard inside a sentence for condition A and in isolation in condition B; the production tasks were the same for both conditions and in all three tasks.

Table 1: Experiment #1 groups and blocks

condition	baseline reading		immediate shadowing	post-exposure reading	
A	<b>isolated phrases</b>	20 target NPs (‘the pelican’)	heard <b>sentences</b> consisting of the 20 target NPs: ‘ <i>The pelican</i> is flying over the beach.’	<b>isolated phrases</b>	20 target NPs (‘the pelican’)
		20 “novel” NPs (‘a partner’)			20 “novel” NPs (‘a partner’)

		20 “novel” VPs (‘to perish’)	repeated <b>isolated phrases</b> : <i>‘The pelican.’</i>		20 “novel” VPs (‘to perish’)
B	<b>isolated phrases</b>	20 target NPs (‘the pelican’)	heard <b>isolated phrases</b> consisting of the 20 target NPs: <i>‘The pelican’</i>	<b>isolated phrases</b>	20 target NPs (‘the pelican’)
		20 “novel” NPs (‘a partner’)			20 “novel” NPs (‘a partner’)
		20 “novel” VPs (‘to perish’)			20 “novel” VPs (‘to perish’)

### 2.2.3. Subjects

Condition A consisted of 11 subjects (8 female, 3 male, an additional subject’s data being eliminated due to a recording malfunction) while condition B consisted of 12 subjects (9 female, 3 male). Subjects were recruited primarily from UC Berkeley’s undergraduate student population. Subjects provided informed consent and were compensated \$5 for their participation.

## 2.3. Baseline to Immediate Shadowing

### 2.3.1. Measurements

VOT was measured using a hybrid method which combined automated VOT measurements with thorough quality control. A python script using the pyalign subprocess (Sprouse & Johnson 2016), utilizing the Penn Phonetics Lab Forced Aligner (Yuan & Liberman 2008), created textgrids for each .wav file for their analysis in Praat, v. 6.0.14 (Boersma & Weenink 2014). Another Python script was used to identify the location of burst for the /p/ in each target word. A new textgrid was produced for each .wav file including point tiers for the location of the burst and the onset of the vowel. Finally, each textgrid-aligned .wav file was inspected by the researchers using a Praat script for rapid editing of textgrids. Aberrant burst or vowel onset measurements could be corrected in the textgrids by dragging the point tier to the correct location. VOT was calculated as the difference between the burst and vowel onset.

2.3.2. Statistical Analysis

In order to quantify the effect of structural context on phonetic accommodation, a mixed-effects regression model was fitted to the data in R using *lmer()* function in the lme4 package. The response variable, RELVOTDIFF, was the difference in VOT divided by word length from baseline to immediate shadowing, in order to normalize for rate ( $\text{vot}_{\text{post-exposure}} / \text{wordlength}_{\text{post-exposure}} - \text{vot}_{\text{baseline}} / \text{wordlength}_{\text{baseline}}$ ), as it was observed that subjects tended to speak more quickly as the experiment progressed. SUBJECT and TARGETPHRASE were included as random effects in the model in order to account for variation among subjects and in behavior towards specific phrases. The model included the fixed effect CONDITION, which was either A (shadowing sentences) or B (shadowing isolated phrases).

2.3.3. Results

Analysis of the results indicates that CONDITION, whether or not the target phrase was heard in sentence context or in isolation, was only marginally significant ( $\chi^2 = 2.7563$ ,  $df = 1$ ,  $p = 0.09687$ ). Over all subjects in condition A (sentence listeners), VOT decreased by 6.29 ms on average (median value of 6.8 ms) as shown in Table 2, with some decrease observed even in the relative duration of VOT to vowel length (an average -0.25% decrease). On the contrary, over all subjects in condition B (isolation listeners), VOT showed only the slightest amount of increase (mean = 0.06 ms, median 0.51 ms), with a relative lengthening of VOT to vowel length (0.87% increase). Thus, for all subjects this suggests a slight decrease in VOT for sentence listeners, and a slight increase in VOT for isolation listeners, although with an only marginally significant difference.

Table 2: VOT increase from baseline to shadowing by group

	VOT increase mean	VOT increase median	RelVOT mean (vot/wordlength * 100)	RelVOTdiff median (vot/wordlength * 100)
A (sentence listeners)	-6.29 ms	-6.80 ms	-0.25%	-0.41%
B (isolation listeners)	+0.06 ms	+0.51 ms	+0.87%	+0.85%

A clearer picture emerges when looking at changes by subjects. Figures 1 and 2 show the individual and group variability in change from baseline to immediate shadowing VOT and VOT:WordLength (relativized for speech rate) respectively. In Figure 1, we see that most of subjects in condition A (sentence listeners) had lower VOT in the immediate shadowing block, while most of the subjects in condition B (isolation listeners) had higher VOT in the immediate shadowing block. Figure 2, showing the ratios of VOT to word length, which normalized for any change in speech rate, shows a similar situation. In general, more subjects showed an increase in VOT:WL ratio compared to VOT length alone, suggesting subjects were in fact speaking quicker by the second block. While nearly half in condition A (sentence listeners) do show an increase in relative length of VOT to word length, Figure 2 shows that all but two of the subjects in condition B (isolation listeners) showed an increase in VOT:WL. These two outlier subjects in condition B, curiously enough, showed a greater decrease in VOT and VOT:WL than any subjects in either group. The presence of these outliers greatly decreases the significance of the effect observed in the data, which otherwise shows a regular difference between subjects' degree of accommodation between the two conditions, with greater accommodation when hearing an isolated stimulus as opposed to a stimulus within a sentence.

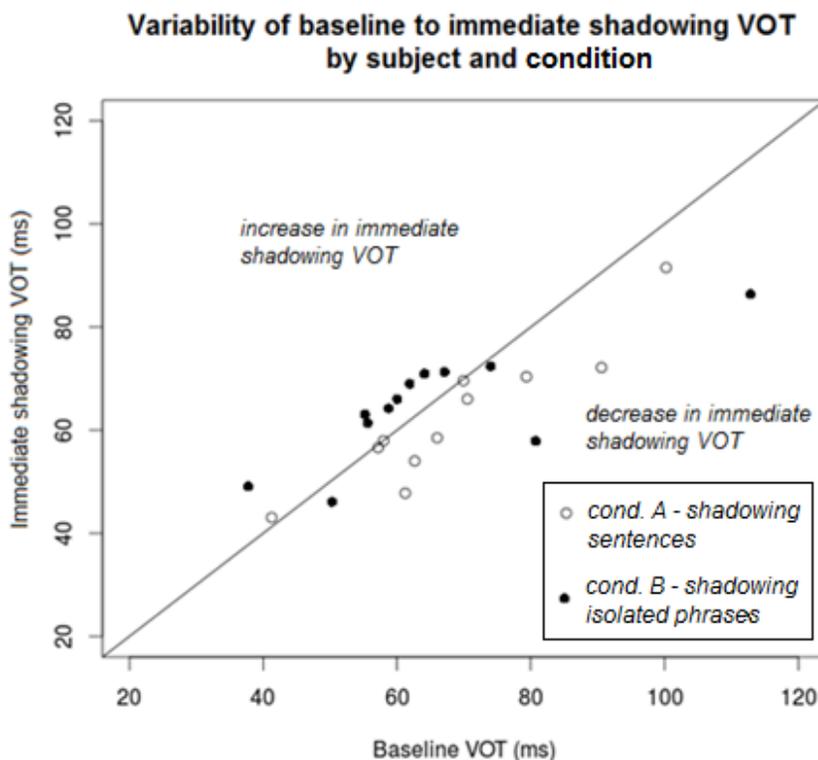


Figure 1: Variability of baseline to immediate shadowing VOT by subject and condition: Each dot represents a subjects' mean baseline VOT compared to their mean immediate shadowing VOT. Subjects falling to the left of the line indicate an increase in VOT from the baseline to the immediate shadowing task, while subjects falling to the right of the line indicate a decrease in VOT.

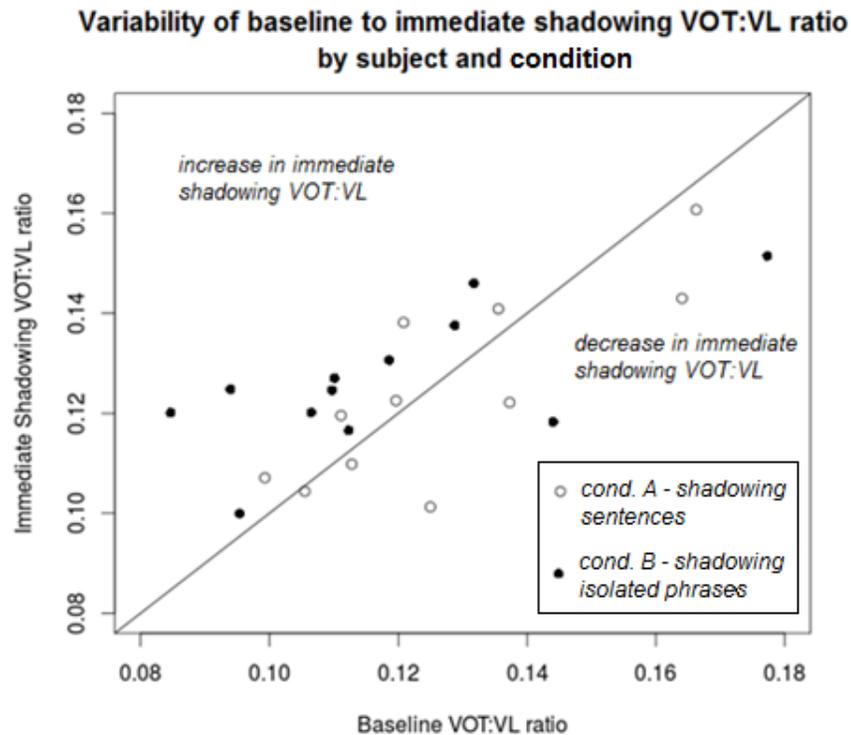


Figure 2: Variability of baseline to immediate shadowing VOT to word length ratio by subject and condition: Each dot represents a subjects' mean baseline VOT:WL ratio compared to their mean immediate shadowing VOT:WL ratio. Subjects falling to the left of the line indicate an increase in VOT relative to word length from the baseline to the immediate shadowing task, while subjects falling to the right of the line indicate a decrease in VOT:WL.

#### 2.4. Baseline to Post-Exposure

Along with the change in baseline compared to the immediate repetition shadowing, the change in VOT from the baseline to post-exposure reading was also considered. Following

the findings of Goldinger (1998) Nielsen (2011), new exemplars may persist even after a shadowing task, and those exemplars may affect other words that share similar sounds. For analyzing this carry-over effect, and whether the difference still held between the two group conditions, a mixed-effects regression model was once again fitted to the data with R's *lmer()* function in the *lme4* package. As with before, the response variable RELVOTDIFF was the difference in VOT divided by word length from baseline to post-exposure, to control for changes in speech rate ( $\text{vot}_{\text{postexposure}} / \text{wl}_{\text{postexposure}} - \text{vot}_{\text{baseline}} / \text{wl}_{\text{baseline}}$ ). SUBJECT and TARGETPHRASE were included as random effects in the model, while CONDITION, A (sentence listeners) or B (isolation listeners) was the fixed effect.

#### *2.4.1. Results*

Surprisingly, most subjects showed not only a decrease in VOT going from baseline to post-exposure reading, but a decrease in VOT relative to word length. While a faster speech rate might be expected, it is not clear why the VOT of /p/ was actually shorter with respect to vowel length after hearing the stimulus. For all words, condition A (sentence listeners) subjects showed a larger mean decrease of -0.77% in the VOT to word length ratio, while those in condition B (isolation listeners) showed a decrease of -0.44% in VOT:WL ratio. For just the subset of 20 words heard as the stimuli in the immediate shadowing block, condition A subjects showed a mean VOT:WL decrease of -0.93% compared to a decrease of -0.37% for condition B subjects. While this suggests longer VOT for the isolation listeners, as expected, the mixed-effects model reveals no significant difference between the two conditions (all words:  $\chi^2 = 0.33$ ,  $df = 1$ ,  $p = 0.5674$ ; only stimuli words subset:  $\chi^2 = 1.09$ ,  $df = 1$ ,  $p = 0.2958$ ).

#### *2.5. Discussion of Experiment #1*

The results from the first experiment found marginally significant differences in the degree of phonetic accommodation that occurred among subjects listening to speech within different structural contexts. Subjects hearing a target word within sentence showed little to no accommodation of lengthened VOT, while subjects hearing the target word within an isolated phrase showed a small degree of accommodation. Two outlier subjects in the isolated phrase listening condition (B) exhibited the strongest decrease in VOT of all subjects, which strongly affected the significance of the effect. Given the regularity of the effect for the rest of the subjects in this condition, the distribution of these subjects may actually be bimodal, while the two outliers represent phonetic divergence from, rather than convergence to the model

speaker. Following the findings of Babel (2010), this divergence may have been caused by the attitudes of these two subjects towards Australian speakers such as our model. While we had hoped a non-local speaker would cause increased attention to phonetic details, a speaker of a local dialect may have avoided social impediments in the subjects' accommodation.

A particularly unexpected occurrence in this experiment was the general decrease in VOT following the baseline reading, both in the immediate shadowing and post-exposure blocks. Some of this decrease was likely due to rate change, as subjects may have become tired throughout the course of the experiment, or were repeating increasingly primed and familiar phrases. However, the results suggest a decrease in VOT:WL ratio, particularly in the final block of the experiment in both conditions. It is possible this was the result of the overuse of the /p/ gesture, which was not dispersed among a variety of other filler phrases beginning with other sounds. While we had anticipated a stronger accommodation effect with a more overt phonetic target (keeping every target word beginning with /p/ with no fillers), it may have resulted in gestural fatigue, obscuring the degree of accommodation.

The results of this experiment leave open the possibility of multiple analyses. One possibility is that the presence of syntactic structure, requiring a certain amount of processing in order to group constituents and assign semantic roles for arguments, etc., resulted in decreased attention to phonetic details. Additionally, the presence of meaningful semantic context may aid the listener in word recognition by supplying top-down information, such that less reliance on the acoustic signal is necessary. If either of these is the case, it is still unclear whether the presence of syntactic structure and semantic context has the effect of "flipping a switch" from the "how" to the "what" mode of listening, or rather, if the phonetic attention given to particular words is merely a direct function of the amount of context available to that particular word. The present experiment did not control for contextual predictability, with words occurring in a variety of locations within the sentence, with presumably a range in how much context aided in word recognition. The second experiment controls for these variables in order to address this particular question.

Another possibility is that, rather than being caused by the mere presence of structure or context, that simply having longer segments of speech to listen to decreased the phonetic attention given to a particular feature of a particular target word. Since listeners in condition A heard the target within an entire sentence, there were far more details for them to attend to, while condition B had much shorter utterances to process. Unfortunately, this confounding

variable may be impossible to control for, as more syntactic structure will usually result in more phonetic information.

### **3. Experiment #2**

#### *3.1. Goal/Hypothesis*

The first experiment yielded only marginally significant results in showing a difference in phonetic attention when listening to sentences versus isolated phrases. Additionally, the effect of syntactic as opposed to semantic context was not controlled for. The second experiment considers the effect of semantic context, specifically how predictable a given word is in its context, and how this affects the listener's phonetic attention with regard to that word. For example, in the sentence 'The pioneers made log *cabins*,' the final word is more predictable than in the sentence 'Joe turned and saw the *cabins*,' where 'cabins' in the first sentence is primed by 'pioneers,' and 'log' immediately before it. Following the observations of phoneme restoration (Warren 1970, Samuel 1981) and fluent restoration in shadowing (Cole & Jakimik 1978) we expect more attention to be given to the phonetic details of a word when there is no context to fill in what the intended sound should be. Thus, Lindblom's (1995) "how" mode should be activated (or more active) when hearing a word that is not supplied by the context, while the "what" mode should dominate when hearing a predictable word. The second experiment utilizes the phenomenon of phonetic accommodation once again to determine what phonetic details the listeners has attended to. Thus, a higher degree of accommodation is expected for unpredictable words, where less accommodation is expected for predictable words.

#### *3.2. Method*

##### *3.2.1. Stimuli*

All target words in this study began with the phoneme /k/ and were two syllables with initial stress. Target words occurred as the final word in all sentences, such that predictability could be determined entirely from the preceding context. The sentences containing the /k/ initial target words averaged approximately 10 syllables in length. The model speaker for this experiment was a male speaker of American English from northern California. The stimuli were recorded in a sound booth in a casual speaking style.

Each /k/ target word had two features digitally manipulated. The VOT of the /k/ was approximately doubled to a minimum of 100 ms. In addition, the pitch of the first syllable was raised by approximately 20 Hz, overall giving a sense of stronger prominence to the first syllable. For this experiment, acoustic manipulations were achieved using the pitch and duration manipulation tools in Praat v. 6.0.14 (Boersma and Weenink 2014). By our own impression this was a more successful means of creating natural sounding lengthened VOT than the copy-and-paste technique used in experiment #1. Lastly, the manipulated words were copied from their occurrence in the predictable sentences and pasted into the unpredictable contexts. By doing this, listeners would hear the exact same recording of the word (e.g., ‘cabins’) in either a predictable or unpredictable context. The model speaker may have pronounced the words subtly different based on the context (e.g., hyperarticulated when in an unpredictable context to serve the needs of the listener, as theorized in Lindblom 1995, Alyett & Turk 2004). Thus, this copying technique assured that listeners would not hear a hypo- or hyperarticulated form corresponding with the contextual environment.

### *3.2.2. Measuring Predictability*

In order to ensure that listeners would in fact be able to predict the /k/ initial target words, an online survey was conducted using Amazon Mechanical Turk. Subjects were compensated \$2 for a 15 minute survey which included all 60 sentences with target words believed to be predictable from the surrounding context, along with 40 filler sentences to prevent subjects realizing most of the words began with /k/ sounds. Participants read each sentence off the screen, with a blank indicating the target word, as shown below:

Kings and queens live in \_\_\_\_\_.

Subjects were asked to fill in the blank with the first word they thought of, which must be grammatical and would be a word others would likely think of as well. Overall, all 60 target predictable words were correctly guessed by between nine and 34 out of 34 total subjects, and were always the most commonly guessed words for each sentence (see Appendix B for a full list of the stimuli and their predictability values).

### *3.2.3. Procedure and Subject Groups*

The entire experiment was run twice with 20 subjects each. For both experiments, the subjects’ task was to repeat sentences presented aurally. In the first version of the experiment,

subjects were given no instruction to imitate, as is the practice in the phonetic accommodation paradigm (Goldinger 1998, etc.). However, considering the findings of the experiment #1, which suggested the possibility that little to no imitation may occur in sentence shadowing, the second group was asked to “sound more like the model speaker in some way.” The intention was to increase the imitation effect and remove any social barriers that might cause divergence (Babel 2010) that might obscure the effect if it was weaker in sentence context.

Simultaneously, it allowed the researchers to manipulate the type and level of attention given to the stimuli; with no instruction to imitate, subjects might listen in the default “what” mode, whereas when told to imitate, may listen more in the “how” mode. While it was hypothesized that those told to imitate would show greater imitation than those given no instruction, there was no strong hypothesis going into the experiment about which group (told to imitate or not) would exhibit a greater difference in imitating predictable versus unpredictable words. While those given no instruction to imitate may show little to no imitation (as seen in the first experiment) for either set of words, the absence of instructions to imitate might induce a more natural “what” mode of listening, where phonetic details might be ignored in the presence of stronger contextual cues.

Among the 20 subjects in each experiment, 10 subjects formed counterbalanced groups A and B as shown in Table 3. For all groups in the study, the experiment consisted of a single block in which the subjects were asked to shadow 100 sentences, consisting of 60 sentences containing target words. Of the 60 target words, group A heard 30 of these in a predictable context and 30 in an unpredictable context. No target word was heard twice by any participant, in order to avoid priming the words and affecting the predictability. Group B heard the same 60 target words, but the 30 that were predictable in group A were heard in an unpredictable context in group B, while the 30 unpredictable words in group A were heard in a predictable context. Counterbalancing the groups was necessary in case there were biases in the words selected in a particular group, such that, for example, the 30 predictable words just happened to be of a significantly different frequency than the unpredictable words, or if there were differences in neighborhood densities, etc. Thus if both groups behaved similarly, it is a good indication of sensitivity to contextual predictability as opposed to other possible variables. Additionally, all groups also heard 40 filler stimuli sentences which contained no final word with an initial /k/ sound and were not in any way manipulated, to obscure the prominent final /k/ word pattern at least partially.

Table 3: Conditions, groups, and blocks for experiment #2

	No instruction to imitate	Told to imitate
Group A	30 predictable: “The pioneers made log <i>cabins</i> .”	30 predictable: “The pioneers made log <i>cabins</i> .”
	30 unpredictable: “The first thing Mary saw was the <i>coffins</i> .”	30 unpredictable: “The first thing Mary saw was the <i>coffins</i> .”
	40 fillers	40 fillers
Group B – counterbalanced (reverse predictability)	30 predictable: “The vampires are sleeping in <i>coffins</i> .”	30 predictable: “The vampires are sleeping in <i>coffins</i> .”
	30 unpredictable: “Joe turned and saw the <i>cabins</i> .”	30 unpredictable: “Joe turned and saw the <i>cabins</i> .”
	40 fillers	40 fillers

### 3.2.4. Subjects

Both experiments had a total of 20 subjects consisting of two counterbalanced groups of 10 subjects. The 20 subjects given no instruction to imitate included 8 males and 12 females, while the 20 subjects told to imitate included 9 males and 11 females. Most of the subjects were UC Berkeley undergraduates. Subjects provided informed consent and were compensated \$5 for their participation.

### 3.3. Results

#### 3.3.1. Measurements

A similar method to what was used for measuring the data in the first experiment was also used for the second experiment. The boundaries for the burst, vowel onset, and end of the vowel in the first syllable were generated and then adjusted in a Praat textgrid, and durations were extracted using a Python script. Pitch measurement was achieved using a Python script

using the Entropic Signal Processing System (a package of UNIX environment speech processing and analysis tools) routine `get_f0` (Talkin & Lin 1996) which was fed the vowel onset and end values from the `textgrid`. These boundaries were used as endpoints for measuring the mean pitch of the vowel in the target syllable.

### *3.3.2. Statistical Analysis*

A mixed-effects regression model was used to determine the effect of contextual predictability on phonetic accommodation. The model was run in R using the `lmer()` function in the `lme4` package. Response variables included `VOTDIFF` which was the difference in VOT between the model and the subject for each token ( $\text{vot}_{\text{subj}} - \text{vot}_{\text{model}}$ ), `RELVOTDIFF2` which was the difference in VOT in relation to the vowel length between the model and subject, thus normalized for rate ( $\text{vot}_{\text{subj}}/\text{vl}_{\text{subj}} - \text{vot}_{\text{model}}/\text{vl}_{\text{model}}$ ), as well as `RELITCHDIFF` which was the difference in pitch between the model and subject, in this case being the pitch of the target syllable (the first syllable of the target words which had lengthened VOT) divided by the average pitch of the entire target utterance ( $\text{target.pitch}_{\text{subj}}/\text{utterance.pitch}_{\text{subj}} - \text{target.pitch}_{\text{model}}/\text{target.pitch}_{\text{model}}$ ). This was necessary given the model speaker was male, and had a lower speaking voice than most of the female subjects. All of these variables involve subtraction of the model's values from the subjects' values in order to show whether the subject exceeded (i.e., a positive difference) or fell short of (i.e., a negative difference) the model's performance.

Random effects in the model included `SUBJECT` and `TARGETWORD`. Fixed effects included `PREDICTABILITY`, whether or not the target word was predictable or unpredictable, `ORDER`, the number of the trial within the experiment from 1 to 100, `GENDER` (male or female), `GROUP` (A or B, with counterbalanced target word lists) and `CONDITION`, referring to whether the subjects were told to imitate or not. Additionally, interactions among all combinations of these variables were included. Because `CONDITION`, and interactions of other response variables with `CONDITION`, are frequently significant (e.g., for, `CONDITION` is significantly different for `VOTDIFF`  $\chi = 10.3571$ ,  $df = 1$ ,  $p = 0.00129$  \*\*, and for `RELVOTDIFF`,  $\chi = 21.3049$ ,  $df = 1$ ,  $p = 3.917e-06$  \*\*\*), we will consider the results of the two experiment conditions separately.

### *3.3.3. VOT: no instruction to imitate*

When given no instruction to imitate, the independent variable `PREDICTABILITY` showed a highly significant effect for both response variables, `VOTDIFF` and `RELVOTDIFF`, (`VOTDIFF`:  $\chi =$

29.4319,  $df = 1$ ,  $p = 5.792e-08$  \*\*\*; RELVOTDIFF:  $\chi = 25.1808$ ,  $df = 1$ ,  $p = 5.22e-07$  \*\*\*), shown in Tables 5 and 6. The mean and median of VOTDIFF and RELVOTDIFF according to the predictability conditions can be seen in Table 4. VOTDIFF is shown to be greater (further from zero) for predictable words and lower (closer to zero) for unpredictable words. This means there is greater difference from the model in predictable words, and that subjects were closer to the models' VOT when shadowing contextually unpredictable words. The same pattern occurs with RELVOTDIFF where the lower percentages for unpredictable words indicate the subjects were closer in imitating the VOT to vowel length ratio of the model.

Few other predictors were significant, although there is some significance of the GENDER:PREDICTABILITY interaction (as seen in Table 5). Otherwise, both men and women performed similarly with respect to VOT. The lack of any ORDER effect suggests subjects are not getting closer to or further from the model's VOT throughout the course of the experiment. The lack of any GROUP effect suggests the two counterbalanced groups (10 subjects each, where the predictability of the target words was reversed) were not significantly different in any way that might have affected the significance of the PREDICTABILITY effect.

#### *3.3.4. VOT: Told to imitate*

When told to imitate, subjects' VOT averaged 14.77 ms closer to the model over all subjects as opposed to when given no instruction to imitate. Additionally, the effect of PREDICTABILITY was realized quite differently. Unlike in the no instruction to imitate condition, there was not a significant effect of PREDICTABILITY over the entire duration of the experiment, as shown in Table 5. However, there was a strongly significant ORDER effect. A Pearson's product-moment correlation test reveals a positive correlation between both ORDER and VOTDIFF ( $r = 0.1169$ ,  $t = 4.0626$ ,  $df = 1191$ ,  $p < 0.0001$ ) and ORDER and RELVOTDIFF ( $r = 0.0862$ ,  $t = 2.9876$ ,  $df = 1191$ ,  $p = 0.002869$ ). The positive correlation indicates an approach towards the model's VOT and VOT to vowel length ratio over the course of the experiment.

A mixed-effects model also reveals a significant interaction between PREDICTABILITY and ORDER (for both response variables, VOTDIFF and RELVOTDIFF). This suggests that imitation of predictable versus unpredictable words may change in different ways over time. A Pearson's product-moment correlation test shows no significant correlation between either VOTDIFF and ORDER ( $r = 0.0373$ ,  $t = 0.9081$ ,  $df = 591$ ,  $p = 0.3642$ ) or RELVOTDIFF and ORDER ( $r =$

0.0402,  $t = 0.9782$ ,  $df = 591$ ,  $p = 0.3284$ ) for predictable words only. However, for unpredictable words alone there is a significant positive correlation between both VOTDIFF and ORDER ( $r = 0.1934$ ,  $t = 4.8193$ ,  $df = 598$ ,  $p < 0.0001$ ) and RELVOTDIFF and ORDER ( $r = 0.1266$ ,  $t = 3.1214$ ,  $df = 598$ ,  $p = 0.001887$ ). This suggests that when told to imitate, subjects' VOT becomes closer to the model only for unpredictable words, while VOT for predictable words shows no change over time. Thus, while PREDICTABILITY is not a significant predictor of VOT imitation in the told-to-imitate condition, the hypothesized difference emerges only by the last quarter (tokens 75-100) of the experiment (as shown in fig. 4). Density distributions for both conditions for both predictable and unpredictable words is shown in Figure 3.

Table 4: VOTDIFF and RELVOTDIFF means and medians by condition and predictability

		NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
		Predictable words	Unpredictable words	Predictable words	Unpredictable words
VOTDIFF	<i>mean</i>	-59.0 ms	-54.1 ms	-42.8 ms	-40.7 ms
	<i>median</i>	-58.6 ms	-52.0 ms	-45.4 ms	-41.9 ms
RELVOTDIFF	<i>mean</i>	-56.98%	-50.15%	-45.65%	-43.23%
	<i>median</i>	-55.49 %	-49.09%	-45.76%	-44.08%

Table 5: Linear mixed-effects model for response variable RELVOTDIFF<sup>3</sup>

Analysis of Deviance Table (Type II Wald chisquare tests)						
Response: RELVOTDIFF ~ PREDICTABILITY * GROUP * ORDER * GENDER + (1 SUBJECT) + (1 WORD)						
	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	Chisq	Df	Pr(> Chisq)	Chisq	Df	Pr(> Chisq)
predictability	24.6517	1	<0.0001 ***	1.3089	1	0.25260
Group	0.0004	1	0.98408	0.4939	1	0.48218
Order	0.0087	1	0.92588	26.0099	1	<0.0001 ***
gender	0.4647	1	0.49545	0.4912	1	0.48338
predictability:group	0.1158	1	0.73362	0.0743	1	0.78523

predictability:order	0.0780	1	0.77999	4.0696	1	0.04366 *
group:order	1.0147	1	0.31377	1.6115	1	0.20428
predictability:gender	5.7389	1	0.01659 *	1.1275	1	0.28832
group:gender	0.0150	1	0.90246	0.0051	1	0.94314
order:gender	0.0360	1	0.84960	0.9583	1	0.32762
predictability:group:order	0.6107	1	0.43454	0.2397	1	0.62439
predictability:group:gender	0.0911	1	0.76279	0.2023	1	0.65285
predictability:order:gender	0.0666	1	0.79631	0.2683	1	0.60444
group:order:gender	0.0680	1	0.79420	0.0094	1	0.92266
predictability:group: order:gender	2.7375	1	0.09802 .	0.5681	1	0.45101
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

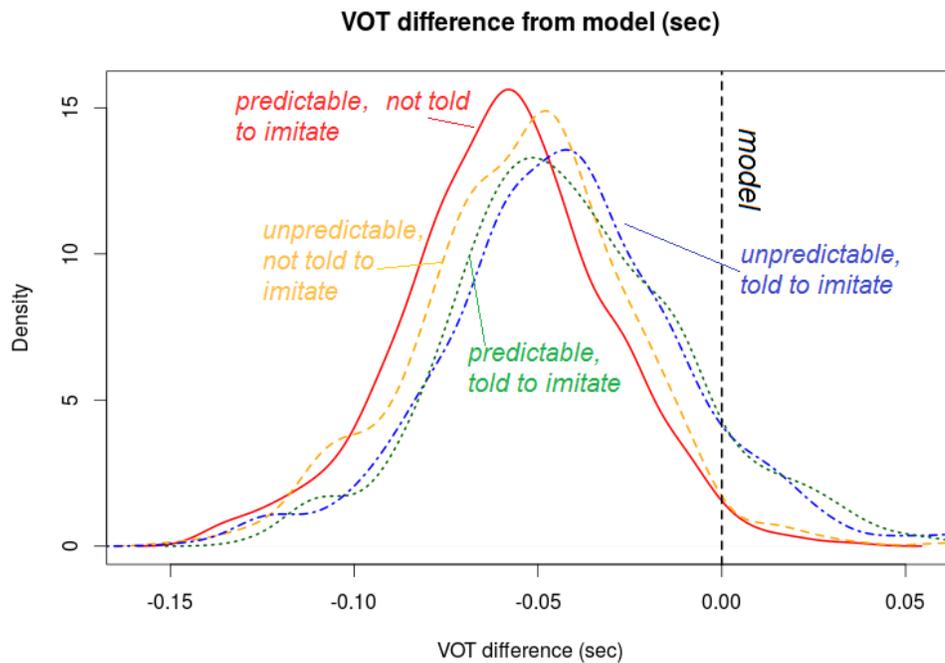


Figure 3: Density distributions by condition and predictability

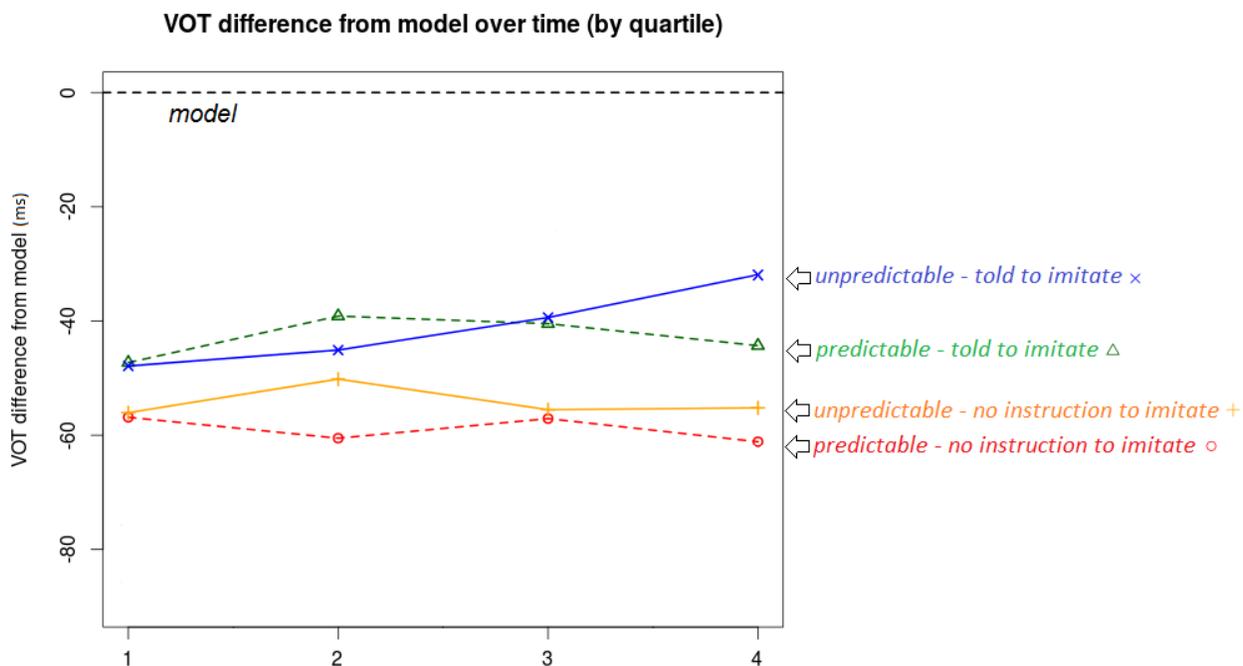


Figure 4: VOT difference from the model over time, from the 1<sup>st</sup> to 4<sup>th</sup> quartiles of the experiment, in each of the four instructional and predictability environments.

### 3.3.5. Pitch

The results for pitch suggest a similar pattern as with VOT imitation in predictable and unpredictable words either when given no instruction to imitate or when told to imitate, as shown in Table 6. PREDICTABILITY is a highly significant predictor of RELPITCHDIFF (difference in the ratio of the target syllable’s pitch divided by the utterance pitch, compared to the model). Another significant predictor was GENDER. While pitch of the target syllable was normalized compared to the mean pitch of the utterance, men and women may imitate the pitch of a male model differently. Men, having a similar range to the male model, may have been imitating absolute pitch while women were imitating relative pitch. Based on this observation, RELPITCHDIFF will be considered separately for men and women.

Table 6: Linear mixed-effects model for response variable RELPITCHDIFF

Analysis of Deviance Table (Type II Wald chisquare tests)
Response: RELPITCHDIFF ~ PREDICTABILITY * GROUP * ORDER * GENDER + (1 SUBJECT) + (1 WORD)

	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	Chisq	Df	Pr(> Chisq)	Chisq	Df	Pr(> Chisq)
predictability	48.3855	1	<0.0001 ***	13.1558	1	<0.0001 ***
group	3.8333	1	0.05024 .	0.2029	1	0.6524086
order	0.0034	1	0.95354	8.4875	1	0.0035760 **
gender	6.0356	1	0.01402 *	3.3086	1	0.0689184 .
predictability:group	0.1866	1	0.66576	2.2074	1	0.1373489
predictability:order	0.1555	1	0.69335	1.5153	1	0.2183251
group:order	1.4441	1	0.22948	0.4264	1	0.5137769
predictability:gender	2.5693	1	0.10896	2.5155	1	0.1127340
group:gender	2.4190	1	0.11987	0.8756	1	0.3494137
order:gender	2.5687	1	0.10900	1.8854	1	0.1697238
predictability:group:order	0.0226	1	0.88062	1.5965	1	0.2064069
predictability:group:gender	2.0753	1	0.14970	0.0041	1	0.9489500
predictability:order:gender	0.0725	1	0.78771	4.0510	1	0.0441456 *
group:order:gender	0.9858	1	0.32076	0.3084	1	0.5786795
predictability:group:order:gender	0.0035	1	0.95264	1.7176	1	0.1900028
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

3.3.5.1. Pitch – by gender, no instruction to imitate

When analyzing pitch separately for both men and women when given no instruction to imitate, both genders still show a strongly significant effect of PREDICTABILITY on the degree of pitch imitation (Tables 7 and 8). Table 8 shows the mean and median values for RELPITCHDIFF for both men and women. Here, 0% would indicate having a target syllable pitch to utterance pitch ratio identical to the model, while negative ratios indicate the target pitch was not as high relative to the rest of the utterance as it was for the model speaker. When given no instruction to imitate, both genders show a higher RELPITCHDIFF in unpredictable words (and closer to model), which indicates the target syllable was pronounced with higher pitch. Women actually displayed positive values for RELPITCHDIFF for unpredictable words; this suggests they may have overshoot the model’s pitch contour for target unpredictable words.

Men show a significant effect of ORDER on RELPITCHDIFF, while women do not. A Pearson’s product-moment correlation test reveals a small *negative* correlation between RELPITCHDIFF and ORDER, -0.1041 ( $t = -2.1934$ ,  $df = 439$ ,  $p = 0.0288$ ), suggesting men actually drifted further from the model over time, which is unexpected. Both genders show some significance of GROUP--- (GROUP for women and a GROUP:ORDER interaction for men) suggesting some small differences in the word and sentence selection between Group A and B may have manifested in the realization of pitch imitation. However, even with these predictors present PREDICTABILITY remains a strongly significant variable.

*3.3.5.2. Pitch—by gender, told to imitate*

When told to imitate, PREDICTABILITY is significant only for women, but remains not quite significant for men (tables 9 and 10). In any case, the difference in imitation between predictable and unpredictable words is greater when not told to imitate, while target syllables of unpredictable words are only slightly closer to the contour displayed by the model (higher, closer to 0%) when told to imitate (for both men and women). Men show no significant effect of ORDER when told to imitate, while women do; this ORDER effect was strongly significant for VOT imitation when told to imitate. On the contrary, men do show a strongly significant PREDICTABILITY by ORDER interaction, just as all subjects displayed for VOT when told to imitate. A Pearson product-moment correlation test shows that male subjects show a small and non-significant negative correlation between RELPITCHDIFF and ORDER for predictable words ( $r = -0.065$ ,  $t = -1.0548$ ,  $df = 262$ ,  $p = 0.2925$ ), as opposed to a positive correlation between RELPITCHDIFF and ORDER ( $r = 0.201$ ,  $t = 1.3513$ ,  $df = 264$ ,  $p = 0.1778$ ) for unpredictable words, although also not significant. As with VOT when told to imitate, this suggests men may be becoming more like the model in imitating pitch, however only with unpredictable words. Women do not show a significant PREDICTABILITY by ORDER interaction. Density distributions for men and women’s pitch differences from the model for both conditions in predictable and unpredictable words are shown in Figures 5 and 6 respectively.

Table 7: Mean and median values for RELPITCHDIFF by condition and predictability

	NO INSTRUCTION TO IMITATE		TOLD TO IMITATE	
	Predictable words	Unpredictable words	Predictable words	Unpredictable words

UC Berkeley Phonetics and Phonology Lab Annual Report (2016)

RELITCHDIFF (MEN)	<i>mean</i>	-23.82%	-17.00%	-9.06%	-7.18%
	<i>median</i>	-19.83%	-15.83%	-6.37%	-4.64%
RELITCHDIFF (WOMEN)	<i>mean</i>	-3.55%	6.88%	-2.30%	3.50%
	<i>median</i>	-5.01%	9.66%	0.06%	4.15%

Table 8: Linear mixed-effects model for response variable RELITCHDIFF (MEN)

Analysis of Deviance Table (Type II Wald chisquare tests)						
Response: RELITCHDIFF (MEN)~ PREDICTABILITY * GROUP * ORDER * GENDER + (1 SUBJECT) + (1 WORD)						
	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	Chisq	Df	Pr(> Chisq)	Chisq	Df	Pr(> Chisq)
predictability	12.931 0	1	0.0003232 ***	2.4758	1	0.115612
group	0.0024	1	0.9607905	0.5514	1	0.457750
order	7.1177	1	0.0076326 **	1.4566	1	0.227467
predictability:group	1.3622	1	0.2431623	2.3606	1	0.124433
predictability:order	0.1243	1	0.7244663	9.0632	1	0.002608 **
group:order	5.6837	1	0.0171231 *	0.0388	1	0.843813
group:order:predictability	0.1653	1	0.6843370	5.3744	1	0.020435 *
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Table 9: Linear mixed-effects model for response variable RELITCHDIFF (WOMEN)

Analysis of Deviance Table (Type II Wald chisquare tests)						
Response: RELITCHDIFF (WOMEN)~ PREDICTABILITY * GROUP * ORDER + (1 SUBJECT) + (1 WORD)						

	NO INSTRUCTION TO IMITATE			TOLD TO IMITATE		
	Chisq	Df	Pr(> Chisq)	Chisq	Df	Pr(> Chisq)
predictability	25.4992	1	< 0.0001 ***	10.9250	1	0.0009487 ***
group	5.1045	1	0.02386 *	0.5848	1	0.4444412
order	1.6495	1	0.19903	5.9768	1	0.0144956 *
predictability:group	0.0015	1	0.96921	1.7714	1	0.1832120
predictability:order	0.2155	1	0.64247	0.2184	1	0.6402586
group:order	0.0206	1	0.88600	0.1866	1	0.6657810
group:order:predictability	0.0025	1	0.95991	0.0332	1	0.8555027

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

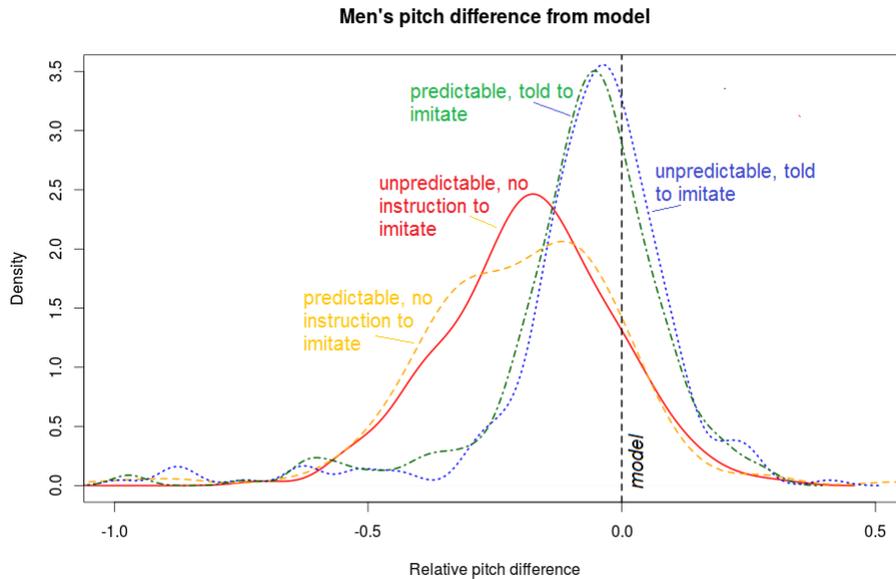


Figure 5: Density distributions for men's pitch difference from model by condition and predictability

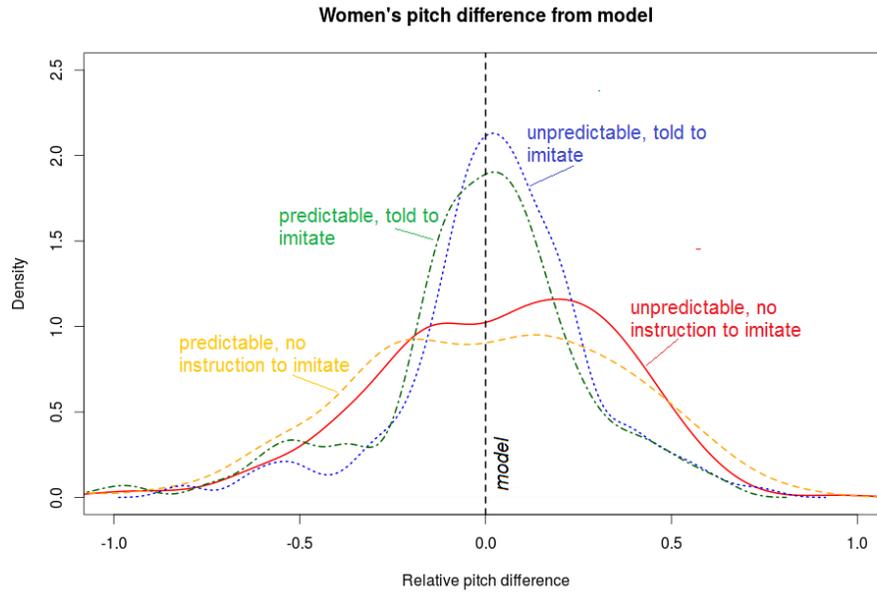


Figure 6: Density distributions for women’s pitch difference from model by condition and predictability

### 3.4. Predictability and imitation condition

The results suggest differences in the two conditions, whether subjects are given no instruction to imitate or are told explicitly to imitate, particularly in the independent variables PREDICTABILITY, ORDER, and the PREDICTABILITY by ORDER interaction. Tables 10, 11, 12, and 13 show the results of a mixed-effects model, showing only these variables and their interactions with CONDITION (told to imitate vs. no instruction) for the three different response variables.

Table 10: Linear mixed-effects model for response variable VOTDIFF by CONDITION

Analysis of Deviance Table (Type II Wald chisquare tests)			
Response: VOTDIFF ~ CONDITION*PREDICTABILITY * GROUP * ORDER*GENDER + (1 SUBJECT) + (1 WORD)			
	Chisq	Df	Pr(> Chisq)
condition	8.5911	1	0.003378 **

condition:predictability	3.7412	1	0.053087 .
Condition:order	13.4274	1	0.000248 ***
Condition:order:predictability	3.5992	1	0.057809 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 11: Linear mixed-effects model for response variable RELVOTDIFF by CONDITION

Analysis of Deviance Table (Type II Wald chisquare tests)			
Response: RELVOTDIFF ~ CONDITION*PREDICTABILITY * GROUP * ORDER*GENDER + (1 SUBJECT) + (1 WORD)			
	Chisq	Df	Pr(> Chisq)
condition	1.3565	1	0.2441473
condition:predictability	4.9669	1	0.0258376 *
condition:order	13.0133	1	0.0003093 ***
condition:order:predictability	2.1798	1	0.1398295
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 12: Linear mixed-effects model for response variable RELPITCHDIFF (MEN) by CONDITION

Analysis of Deviance Table (Type II Wald chisquare tests)			
Response: RELPITCHDIFF (MEN) ~ CONDITION*PREDICTABILITY * GROUP * ORDER + (1 SUBJECT) + (1 WORD)			
	Chisq	Df	Pr(> Chisq)
condition	8.9864	1	0.0027199 **
condition:predictability	1.8070	1	0.1788631
condition:order	4.6745	1	0.0306128 *
condition:order:predictability	2.4658	1	0.1163470
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Table 13: Linear mixed-effects model for response variable RELPITCHDIFF (WOMEN) by CONDITION

Analysis of Deviance Table (Type II Wald chisquare tests)			
Response: RELPITCHDIFF (WOMEN) ~ CONDITION*PREDICTABILITY * GROUP * ORDER + (1 SUBJECT) + (1 WORD)			
	Chisq	Df	Pr(> Chisq)
condition	0.0938	1	0.75943
condition:predictability	4.5828	1	0.03229 *
condition:order	0.5148	1	0.47306
condition:order:predictability	0.0049	1	0.94417
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

The results of the mixed-effects models confirm that subjects are influenced by PREDICTABILITY, ORDER, and to some degree, the PREDICTABILITY by ORDER interaction in different ways in the two instructional conditions. PREDICTABILITY has a significantly stronger effect on the degree of imitation when given no instruction to imitate, while ORDER is more influential when told to imitate. The PREDICTABILITY by ORDER interaction only emerges when told to imitate, although is only marginally significantly different in the two conditions for the VOTDIFF response variable.

#### 4. Discussion

Experiment #1 considered whether stronger imitation would occur when listening to isolated speech as opposed to sentences. The results were marginally significant in suggesting this was the case, however there were several confounding variables (syntax, semantics, total length of sentences vs. isolated phrases) making it unclear what was causing the difference in the two groups. The second experiment considered only the effect of the contextual predictability of words within sentences. The results of the two conditions (no instruction to imitate, vs. told to imitate), found the predictability of the target word to be a significant factor in the degree of imitation, where unpredictable words were more closely imitated than their predictable counterparts, particularly when given no instruction to imitate. Predictability

was still significant when told to imitate, however this effect was realized as a predictability by order interaction.

These findings suggest a speaker's listening mode may be modulated by semantic and syntactic factors such as contextual predictability. However, considering the results of the two conditions in experiment #2, it seems the situation may be more complicated than simply having two categorical modes--- the 'what' and 'how' mode (as proposed by Lindblom). Listening in the "how" mode--- perhaps activated in single word imitation studies, devoid of sentential context, as well as when particular words are unpredictable within sentences--- may cause listeners to be more likely to process the actual exemplars being perceived, such that these will more greatly affect future productions of those particular words. This mode of listening fits nicely with the exemplar-based theories of Johnson (1997) and Pierrehumbert (2002). On the other hand, listening in the 'what' mode may result in the activation of abstracted, phonological forms of words, where phonetic details of the particular exemplars are largely ignored and discarded. In any case, 'what' and 'how' may represent two extreme ends of the listening spectrum, such that both particular exemplars and abstracted forms may be activated under most conditions, but with different relative weightings for their influence on perception and production.

In the second study of the current paper, subjects were either given no instructions to imitate or were told explicitly to imitate. The results showed a stronger effect of predictability when no instruction to imitate was given, while the difference in predictability was less profound and non-significant, when told to imitate. These findings suggest that when given no instruction to imitate, this likely induced a more natural style of listening to sentences, and given the abundance of syntactic and semantic context, this would suggest speakers could recognize and process the sentences using the 'what' mode. However, when the context fails to give clues that might help identify a subsequent word, listeners process the phonetic details of these unpredictable words. As a result, a larger perceptual weight is given to exemplars of unpredictable words, which is reflected in the speech of these listeners turned speakers.

When told to imitate, subjects were closer to the model in VOT and pitch as opposed to when given no instruction to imitate (whether for target predictable or unpredictable words). Additionally, there was a smaller effect of predictability. This suggests that speakers were consciously tuning in to the phonetic details more than they normally would for merely being able to identify and repeat the words in a sentence. Thus, this suggests something closer to the

'how' mode was induced when told to imitate. Subjects were largely able to override the more typical 'what' mode in this task, and listened for phonetic details whether or not the speech they heard was predictable or not. The stronger order effects seen in purposeful imitation suggest that speakers were in fact trying to sound like the model and became better at doing so after hearing and producing more speech. The predictability by order interaction--- the main effect of predictability observed in the told-to-imitate condition--- is a more surprising result. While being told to imitate seems to induce listening more in the 'how' mode, it is possible that over time in this particular experiment that subjects became accustomed to hearing words with initial stress. For predictable words, not only was the word was predictable, but the stress pattern may have become predictable as well. In imitating the predictable words, subjects merely had to make sure the word was in fact the word they anticipated, and that the initial syllable was stressed, and then in their own production applied their own abstracted notion of stress rather than imitating the phonetic details of the particular exemplar. On the other hand, when hearing unpredictable words, the word itself could not be anticipated (even if the stress pattern was predictable). Thus, more thorough phonetic processing would occur in order to identify the word, resulting in more precise imitation of the actual perceived exemplar. Subjects would become no better at anticipating unpredictable words throughout the course of the experiment, unlike with the learnable predictability of the stress pattern, yielding the difference in behavior observed for predictable and unpredictable words over time. Figure 7 shows a more detailed account of how abstracted forms or exemplars may be activated depending on the instructional conditions and predictability of the words.

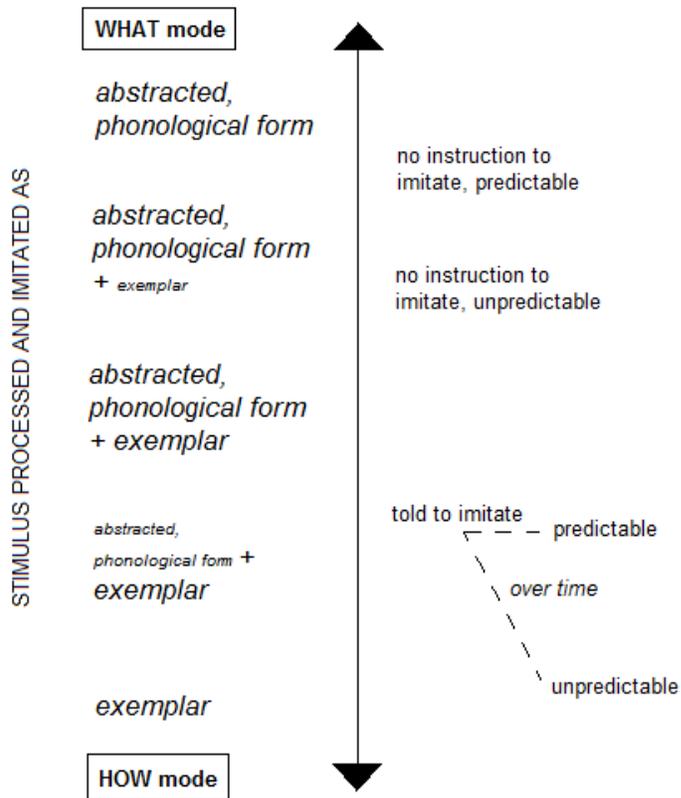


Figure 7: Proposed relative weight of exemplar vs. phonological form by condition and predictability

Lastly, the results of both experiments add new insight to the phenomenon of phonetic accommodation. As with the findings of Pardo (2006) and Babel (2010, 2012), who found social conditions modulating accommodation, as well as Nye & Fowler (2003) and Nielsen (2011) who found elements of linguistic structure also facilitating or impeding accommodation, the present study finds an interaction between higher level linguistic information and accommodation, in particular contextual predictability. While our findings show that accommodation does in fact occur in words placed in a sentence context, it is not yet clear whether accommodation fails to occur at all in certain conditions, such as when shadowing predictable words. In any case, accommodation does not seem to be a purely automatic process, but interacts in complex ways with social factors and linguistic information at many structural levels.

## 5. Conclusion

The results from these experiments demonstrate that higher level linguistic information such as contextual predictability modulate the listener's attention of phonetic details. Even within fluid speech, listeners adjust the degree of phonetic processing needed for word recognition depending on the predictability of anticipated words. Other studies (Lindblom et al. 1995, Garrett & Johnson 2012) have considered how processing speech in different listen modes (such as the 'what' or 'how' mode) may be relevant in sound change. The findings of this paper suggest that details of more predictable words may go unnoticed, while these phonetic details are more likely noticed when hearing contextually unpredictable words. Ultimately, this may begin to shape the pool of exemplars differently for words that more commonly occur in predictable or unpredictable contexts, leading to a bias for certain sound changes to affect only one group of words or another. While contextual predictability may merely be an additional factor to consider alongside frequency within the theory of lexical diffusion (Bybee 2001, Phillips 2006), it may also be relevant when considering the differences between broad classes of words, such as function versus content words. Function words (and morphemes), in being both fewer in number and often being grammatically required in certain contexts, should be more contextually predictable than content words, which are far more numerous. Given the dichotomy in sound changes affecting content and function words (such as the widespread reduction and lenition favored in function words), this distinction may arise from differences in attention and listening modes when hearing function and content words. Thus, it remains an open question as to how phonetic attention affects sound change, particularly in word classes differing in contextual predictability.

### **Acknowledgements**

Special thanks to Dr. Keith Johnson, Dr. Susan Lin, the Berkeley PhonLab, and the Phonology Phorum for their helpful comments and suggestions. Also thanks to Libby Perfitt for assisting in collecting the data for this experiment.

### **Notes**

<sup>1</sup> Phrases were used instead of isolated words because the experiment was originally designed for investigating whether part of speech is a variable in impeding the carry-over imitation effect (from target to novel words). Thus, the stimuli needed a preceding word to

unambiguously mark the part of speech--- such as the verb ‘to pin’ as opposed to the potential noun ‘the pin’.

<sup>2</sup> RelVOTDiff in the second experiment normalized VOT to vowel length, rather than word length as in the first experiment, due to technical reasons involving the scripts that extracted this information from the textgrids.

<sup>3</sup> Table is not included for the response variable VotDiff, which had the same pattern of significance as RelVotDiff.

## **Bibliography**

- Alyett, M. and Turk, A. (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, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(04), 437-456. doi: [dx.doi.org/10.1017/S0047404510000400](https://doi.org/10.1017/S0047404510000400)
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189. doi: [dx.doi.org/10.1016/j.wocn.2011.09.001](https://doi.org/10.1016/j.wocn.2011.09.001)
- Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 6.0.14.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Cole, R. and Jakimik, J. (1980). A model of speech perception. In Cole, R. (ed.) *Perception and Production of Fluent Speech*, (pp. 133-163). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cole, R., Jakimik, J., and Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 64, 44-56. doi:[dx.doi.org/10.1121/1.381955](https://doi.org/10.1121/1.381955)

- Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change. In Alan Yu (ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford: Oxford University Press.
- 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>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. doi:dx.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 17<sup>th</sup> International Congress of Phonetic Sciences*, 49-54. Retrieved from: <http://icphs2007.de/conference/Papers/1781/1781.pdf>
- Heath, J. S. (2014). Accommodation can lead to innovated variation. *UC Berkeley Phonology Lab Annual Report*, 119-145. Retrieved from: [http://linguistics.berkeley.edu/phonlab/documents/2014/Heath\\_innovated\\_variation.pdf](http://linguistics.berkeley.edu/phonlab/documents/2014/Heath_innovated_variation.pdf)
- Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99. doi: dx.doi.org/10.1016/j.cognition.2003.10.011
- Hickok, G. and Poeppel, D. (2007). Opinion – The cortical organization of speech processing. *Nature Reviews and Neuroscience*, 8(5), 393-402. doi:dx.doi.org/10.1038/nrn2113
- Honorof, D. N., Weihing, J., & Fowler, C.A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39(1), 18-38. doi:dx.doi.org/10.1016/j.wocn.2010.10.007
- Johnson, K. (1997). 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/SpeechPerceptionWithoutSpeakerNormalizati.on.pdf>

Lewandowski, N. (2012). Automaticity and consciousness in phonetic convergence (abstract). *Proceedings of the Listening Talker Workshop*, 71. Retrieved from: <http://listening-talker.org/workshop/abstracts/poster/A15.pdf>

Lindblom, B., Guion, S., Hura, S., Moon, S-J., and Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, 7(1), 5-37.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

Maye, J. (2007). Learning to overcome L1 phonological biases. *Proceedings of the 17th International Congress of Phonetic Sciences*, 63-66. Retrieved from: <http://www.icphs2007.de/conference/Papers/1783/1783.pdf>

Mitterer, H. & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168-173. doi:[dx.doi.org/10.1016/j.cognition.2008.08.002](https://doi.org/10.1016/j.cognition.2008.08.002)

Mixdorff, H., Cole, J., & Shattuck-Hufnagel, S. (2012). Prosodic similarity: Evidence from an imitation study. *Proceedings of the 6th International Conference on Speech Prosody, Sp 2012*, 2, 571-574. Retrieved from: [https://www.researchgate.net/publication/266459323\\_Prosodic\\_Similarity\\_-\\_Evidence\\_from\\_an\\_Imitation\\_Study](https://www.researchgate.net/publication/266459323_Prosodic_Similarity_-_Evidence_from_an_Imitation_Study)

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142. doi:[dx.doi.org/10.1371/journal.pone.0074746](https://doi.org/10.1371/journal.pone.0074746)

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>

Ohala, J. (1981). The listener as the source of sound change. In Masek, C., Hendrick, R. & Miller, M. (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](http://linguistics.berkeley.edu/~ohala/papers/listener_as_source.pdf)

- Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. (ed.), *The Production of Speech* (pp. 189-216). New York: Springer-Verlag. Retrieved from: <http://linguistics.berkeley.edu/~ohala/papers/macn83.pdf>
- Pardo, J. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382-2393. doi: [dx.doi.org/10.1121/1.2178720](http://dx.doi.org/10.1121/1.2178720)
- Phillips, Betty S. (2006). *Word Frequency and Lexical Diffusion*. New York: Pargrave MacMillan.
- Pickering, M. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2),169-189. doi:[dx.doi.org/10.1017/S0140525X04000056](http://dx.doi.org/10.1017/S0140525X04000056)
- Pierrehumbert, J. (2002). Word-specific phonetics. In Gussenhoven, C., and Warner, N. (eds.), *Laboratory Phonology VII* (pp. 101-139). Berlin: Mouton de Gruyter. Mouton de Gruyter, Berlin. doi:[dx.doi.org/10.1515/9783110197105.1.101](http://dx.doi.org/10.1515/9783110197105.1.101)
- Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics* 34, 516-531. Retrieved from: <http://faculty.wcas.northwestern.edu/~jbp/publications/toolkit.pdf>
- Pollack, I. and Pickett, J. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, 6, 165-171. doi:[dx.doi.org/10.1177/002383096300600305](http://dx.doi.org/10.1177/002383096300600305)
- Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology*, 110(4), 474-494. doi:[dx.doi.org/10.1037/0096-3445.110.4.474](http://dx.doi.org/10.1037/0096-3445.110.4.474)
- Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, 26(1), 36-56.
- Sancier, M. and Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421-436. doi:[dx.doi.org/10.1006/jpho.1997.0051](http://dx.doi.org/10.1006/jpho.1997.0051)
- Shockley, K., Sabadini, L., and Fowler C. (2004). Imitation in shadowing words. *Perception and Psychophysics*, 66, 422–429. doi:[dx.doi.org/10.3758/BF03194890](http://dx.doi.org/10.3758/BF03194890)
- Sprouse, R. and Johnson, K. Forthcoming. Berkeley Phonetics Machine. To be presented at Interspeech 2016.

<http://linguistics.berkeley.edu/plab/guestwiki/uploads/Bpm.pdf>

- Talkin, D. and Lin, D. (1996). Get\_f0 Online Documentation. ESPS/Waves release 5.31. Entropic Research Laboratory.
- 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:dx.doi.org/10.1016/j.wocn.2009.03.004
- Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392-393. doi:dx.doi.org/10/3758/BF03205070
- Jiahong Yuan and Mark Liberman. (2008). Speaker identification on the SCOTUS corpus. Proceedings of Acoustics '08. Retrieved from: <http://www.ling.upenn.edu/~jiahong/publications/c09.pdf>
- Zellou, G., Scarborough, R., & Nielsen, K. (2013). Imitability of contextual vowel nasalization and interactions with lexical neighborhood density. *Proceedings of Meetings on Acoustics* 19(1). doi:dx.doi.org/10.1121/1.4805633

**Appendix A** – Experiment #1 stimulus sentences (target words underlined)

20 target NP sentences:

- 1) The pail is full.
- 2) The old man smokes a pipe.
- 3) The dogs licked their paws.
- 4) The porch is white.
- 5) The nations made a pact.
- 6) The fisherman sees the pond.
- 7) The boy watched the pandas.
- 8) The pelican flew over the beach.
- 9) The pope led the service.

- 10) A pentagon has five sides.
- 11) The man found the portal.
- 12) Eric bought a pillow.
- 13) The pagans worshipped many gods.
- 14) John is a patron of the arts.
- 15) The chef chopped the parsley.
- 16) The palace is magnificent.
- 17) Your brother is a pest.
- 18) Tom saw a panther in the mountains.
- 19) Susan is a poet.
- 20) The pantry is empty.

20 novel NP sentences:

- 1) The pulp is thick.
- 2) The girl broke the pane.
- 3) The pears are delicious.
- 4) The peas are green.
- 5) The farmer harvests the peaches.
- 6) They climbed the peak.
- 7) William is a peasant.
- 8) Singing is her passion.
- 9) Emily is wearing a parka.
- 10) The pasture is very large.

- 11) The woman sees the path.
- 12) The pauper stole bread to eat.
- 13) Mary wants to see the pageant this year.
- 14) The soldier needs the powder for his gun.
- 15) The policy is strict.
- 16) The porcupine is eating twigs.
- 17) The pigeon is walking through the city.
- 18) Each student has a partner for the game.
- 19) The children used the paste.
- 20) Anna moved the pawn.

20 novel VP sentences:

- 1) Bill published three articles this month.
- 2) Kathy poached the egg.
- 3) Steve polished his shoes.
- 4) Sally purchased a new car.
- 5) The boy panicked about the test.
- 6) The governor pardoned the criminal.
- 7) Jane pondered the meaning of life.
- 8) The Vikings pillaged a monastery.
- 9) The woman pampers her dog.
- 10) Many citizens perished from the famine.
- 11) The referee penalized the team.

- 12) The boy poked the dog.
- 13) Martha paid the cashier.
- 14) The children popped the bubbles.
- 15) Steven paused the movie.
- 16) Sally pinned the banner up.
- 17) The woman pumped the tire up.
- 18) The workers paved the street.
- 19) Susan patted the dog on the head.
- 20) Rita punished her children.

**Appendix B** - Experiment #2 stimulus sentences (target words underlined)

Sentences with contextually *predictable* target words. In parentheses the predictability values are given, as the number of respondents in the Mechanical Turk survey (out of 34) who correctly guessed the predictable word when omitted:

- 1) Pennies are made out of copper. (25/34, 73.5%)
- 2) The printer needs a new ink cartridge. (28/34, 82.4%)
- 3) When there is a blackout we light candles. (32/34, 94.1%)
- 4) A spreadsheet has rows and columns. (30/34, 88.2%)
- 5) The pioneers made log cabins. (28/34, 82.4%)
- 6) Harvard University is a prestigious college. (17/34, 50%)
- 7) The largest library in the U.S. is the Library of Congress. (27/34, 79.4%)
- 8) Root vegetables include parsnips and carrots. (9/34, 26.5%)
- 9) John stacked the plates and put them away in the cupboard. (16/34, 47.1%)

*UC Berkeley Phonetics and Phonology Lab Annual Report (2016)*

- 10) Sauerkraut is made from fermented cabbage. (29/34, 85.3%)
- 11) Kings and queens live in castles. (25/34, 73.5%)
- 12) Nobody puts baby in the corner. (17/34, 50%)
- 13) The dog is wearing a collar. (22/34, 64.7%)
- 14) The vampires are sleeping in coffins. (26/34, 76.5%)
- 15) Mary prefers milk and sugar in her coffee. (23/34, 67.6%)
- 16) The witch is brewing a potion in her cauldron. (15/34, 44.1%)
- 17) Butterfingers are my favorite type of candy. (26/34, 76.4%)
- 18) Bob spilled wine on the white carpet. (15/34, 44.1%)
- 19) Debbie drinks milk straight from the carton. (19/34, 55.9%)
- 20) The windows are covered with curtains. (10/34, 29.4%)
- 21) In the desert, we rode on camels. (25/34, 73.5%)
- 22) The great barrier reef has beautiful coral. (14/34, 41.1%)
- 23) A young cat is called a kitten. (33/34, 97.1%)
- 24) T-shirts are made out of cotton. (27/34, 79.4%)
- 25) Five nickels equal one quarter. (26/34, 76.5%)
- 26) In India, the most deadly snake is the cobra. (25/34, 73.5%)
- 27) The refrigerator and stove are in the kitchen. (32/34, 94.1%)
- 28) The needle always points north on a compass. (33/34, 97.1%)
- 29) We looked up at the night sky and saw Halley's comet. (31/34, 91.2%)
- 30) In Arizona, tourists come to the Grand Canyon. (30/34, 88.2%)
- 31) The girl dressed as a ghost for her Halloween costume. (18/34, 52.9%)
- 32) A one-person canoe is sometimes called a kayak. (19/34, 55.9%)

- 33) France, Spain, and Germany are European countries. (31/34, 91.2%)
- 34) In the desert, Anna saw a spiky Saguaro cactus. (18/34, 52.9%)
- 35) Dave likes hotdogs with mustard and ketchup. (17/34, 50%)
- 36) The photographer put film in her camera. (32/34, 94.1%)
- 37) Cats are felines while dogs are canines. (28/34, 82.4%)
- 38) The historian dug up a time capsule. (19/34, 55.9%)
- 39) At the circus they shot a man out of a cannon. (29/34, 85.3%)
- 40) A prism breaks light into separate colors. (34/34, 100%)
- 41) Bill woke up after twenty years in a coma. (26/34, 74.5%)
- 42) Six people were elected to city council. (29/34, 85.3%)
- 43) The president's words were taken out of context. (28/34, 82.4%)
- 44) Amanda won the pie eating contest. (28/34, 82.4%)
- 45) The horror movie is about a serial killer. (32/34, 94.1%)
- 46) Kathy wears glasses instead of contacts. (30/34, 88.2%)
- 47) The pitcher threw the ball to the catcher. (20/34, 58.8%)
- 48) The women are singing Christmas carols. (30/34, 88.2%)
- 49) Matt doesn't like riding big roller coasters. (27/34, 79.4%)
- 50) The Wizard of Oz is set in Kansas. (18/34, 52.9%)
- 51) Boxes are made out of cardboard. (27/34, 79.4%)
- 52) Separate words in a list with commas. (20/34, 58.8%)
- 53) The lion asked the Wizard of Oz for courage. (24/34, 70.6%)
- 54) A popular spice in Indian cuisine is curry. (14/34, 41.2%)
- 55) Sarah's favorite video game is Mortal Kombat. (34/34, 100%)

- 56) The team was down but made a huge comeback. (20/34, 58.8%)
- 57) Painters paint on a material called canvas. (29/34, 85.3%)
- 58) Jim baked chocolate chip cookies. (29/34, 85.3%)
- 59) Panthers or mountain lions are also called cougars. (14/34, 41.2%)
- 60) Art and music are part of a nations' culture. (22/34, 64.7%)

Sentences with contextually *unpredictable* target words:

- 1) The next word is 'copper.'
- 2) What I need is a cartridge.
- 3) The man is looking at the candles.
- 4) Chris said he saw the columns.
- 5) Joe turned and saw the cabins.
- 6) Linda is thinking about college.
- 7) Albert is always thinking about congress.
- 8) Molly's favorite thing is the world is carrots.
- 9) Ron decided to look at the cupboard.
- 10) Mary wishes she had more cabbage.
- 11) The woman suddenly saw the castles.
- 12) The dog stopped and looked at the corner.
- 13) The first word on the page is 'collar.'
- 14) The first thing Mary saw was the coffins.
- 15) My grandmother needs more coffee.
- 16) Matilda said that she wants a cauldron.

- 17) I have been thinking a lot about candy.
- 18) Bob is always thinking a lot about carpet.
- 19) Debbie sat and stared at the carton.
- 20) Wanda says she wants curtains.
- 21) Nick turned around and saw the camels.
- 22) Barbara turned the page and saw the coral.
- 23) The first thing I saw was a kitten.
- 24) Ashley really needs more cotton.
- 25) Paul opened his eyes and saw a quarter.
- 26) The woman kept thinking about the cobra.
- 27) Jack sat thinking about his kitchen.
- 28) The first thing the man thought of was a compass.
- 29) The people looked and saw a comet.
- 30) Don knows there is a canyon.
- 31) The first thing Julie saw was her costume.
- 32) Everybody began to stare at the kayaks.
- 33) John thinks a lot about different countries.
- 34) Anna looked to the right and saw a cactus.
- 35) Dave's favorite thing in the world is ketchup.
- 36) Lisa decided that she wants a camera.
- 37) The word at the bottom of the page is 'canines.'
- 38) The last word of the book is 'capsule.'
- 39) I know that the man has a cannon.

- 40) The boy said he saw lots of colors.
- 41) The last word bill said was 'coma.'
- 42) The man looked directly at the council.
- 43) Lately I've been thinking a lot about context.
- 44) Amanda wants to have a contest.
- 45) Frank has been thinking a lot about the killer.
- 46) Cindy wants to get more contacts.
- 47) The man turned and looked at the catcher.
- 48) Zack says he really likes carols.
- 49) The first word Matt spelled was 'coasters.'
- 50) Courtney likes to think about Kansas.
- 51) I really need to get more cardboard.
- 52) The women were talking about commas.
- 53) Ben wishes that he had more courage.
- 54) Tom's favorite thing in the world is curry.
- 55) They have been thinking a lot about combat.
- 56) What I really need now is a comeback.
- 57) Eric decided to get more canvas.
- 58) All day long Jim thought about cookies.
- 59) The boy decided to look at the cougars.
- 60) Lately everyone has been discussing culture.