

## Lexical and spectral contextual effects and auditory autonomy in the perception of neighboring vowels and consonants

Kingston, John, Daniel Mash, Della Chambless, Shigeto Kawahara (tentative order)

Bottom p.30

ALTERNATIVE TITLE: The autonomy of spectral contextual effects from lexical knowledge in the perception of neighboring vowels and consonants

Second alternative: The autonomy of auditory and lexical effects in the perception of neighboring vowels and consonants

### 1. **Introduction**

The perception of an ambiguous speech sound is often influenced by properties of adjacent sounds, which we will refer to in this paper as its “context”. For example, when given the choice between a word and a non-word, listeners are inclined to identify the ambiguous sound with the category that makes a word with its context (Ganong, 1980; Pitt & Samuel, 1993). They also choose the category that is transitionally more probable given the context (Pitt & McQueen, 1998) or phonotactically legal in that context (Massaro & Cohen, 1983; Moreton, 2002). These effects of context arise because listeners know the words of their language, the probability of one sound occurring next to another, or the legality of a sound in its context, and bring their knowledge to bear on judgments about the world. Other effects of an ambiguous sound’s context on its perception reflect instead the context’s articulatory or auditory characteristics. For example, in identifying an ambiguous speech sound, listeners compensate for coarticulation with its context (Mann & Repp, 1981) or hear that sound as contrasting auditorily with that context (Lotto & Kluender,

1998; Lotto & Holt, 2006).<sup>1</sup> Whether the ambiguous sound's context evokes linguistic knowledge, prompts compensation for coarticulation, or contrasts auditorily, the context changes the listeners' likelihood of identifying the ambiguous sound with a particular category; that is, contexts change listeners' response bias. We argue in this paper that a reliance on response bias manipulations in the research on context effects has held up resolution of long-standing disagreements about two issues: (i) when these perceptual adjustments are made and (ii) whether response biases due to top-down application of linguistic knowledge arise during a distinct module from those that arise bottom-up from compensation for coarticulation or auditory contrast. We argue in this paper that a reliance on response bias manipulations in the research on context effects makes it impossible to ascertain whether biases due to the top-down application of linguistic knowledge arise during a distinct module from those context effects arising from bottom-up compensation for coarticulation or auditory contrast.

Knowledge of what is a word—lexical knowledge—is one type of linguistic knowledge that shifts response biases (Ganong, 1980; Pitt & Samuel, 1993). In

---

<sup>1</sup> Compensation for coarticulation and auditory contrast are not independent perceptual adjustments for context but instead competing explanations for those adjustments. However, the experiments reported here were not designed to distinguish between these two kinds of explanation, so we leave open the question of which one accounts best for the effects of the context's physical characteristics on the perception of the ambiguous sound.

categorization tasks, listeners more often identify ambiguous targets as the sound that together with the context forms a word in their language, and disfavor alternative responses that do not form words. For example, listeners are more likely to identify a stop ambiguous between [g] and [k] as “g” in an *\_ift* context—which together form an English word—than in an *\_iss* context, where a word is formed instead with [k] (Ganong, 1980). One could conclude from these and similar results that lexical information feeds back onto the phonetic evaluation of the target’s acoustics, and therefore that linguistic knowledge influences pre-lexical processes, as formalized in the TRACE model of perception (McClelland & Elman, 1986; Mirman, McClelland, & Holt, 2006). Models such as TRACE are characterized as *interactive* because information flows both ways between different levels of processing, allowing for feedback of a priori knowledge about words, phonotactics, frequency, etc., on the perception of speech sounds.

An alternative bottom-up feed-forward view holds that lexical knowledge can affect the outcome of a particular task such as phoneme labeling, but does not actually influence the auditory evaluation of the signal’s acoustics (Norris, McQueen, & Cutler, 2000; Kingston, 2005). In the MERGE model of perception (Norris et al., 2000), information only feeds forward from the initial auditory evaluation to word recognition, and each stage of processing is an encapsulated module. However, MERGE permits the addition of decision modules for the purpose of carrying out particular experimental tasks

(e.g., a phoneme decision module in a phoneme decision task). These decision modules receive—but do not send—information from all perceptual levels and thereby produce the knowledge-driven response biases observed in experimental tasks. Listeners thus have a perceptual representation of a target that is not influenced by linguistic knowledge and a separate *decisional* representation that absorbs all relevant information and only determines the listener's response to a stimulus.

The strongest confirmation—as we envision it—of an autonomous model would be to demonstrate the lack of linguistic feedback to some lower/earlier stage of processing, and furthermore actively demonstrate some perceptual process that most plausibly occurs during this lower stage. To that end, we have pitted two context effects against one another, one reflecting linguistic knowledge and the other what we assume here to be auditory contrast, and we have tested listeners' discrimination as well as identification of stimuli in which both of these two context effects are at work. The identification task should establish the independence of these two effects in the following way. Stems which favor a particular sound on the basis of both lexical and contrast biases (i.e. cooperating bias stems) should elicit higher rates of response of that sound than stems which contain either a lexical bias or a contrast bias toward that same sound, but not both (i.e. conflicting bias stems). Establishing the independence of these two effects (that is, that they both exist) is a necessary first step towards arguing for their occurrence during separate stages

of processing. For assuming that categorization taps a later stage of processing, we expect to find evidence of both effects in that task. However, in a task which taps an earlier stage of processing (discrimination), we predict that lexical biases will not affect responses in any global way. Specifically, if the initial auditory evaluation of the signal is free of any feedback from linguistic knowledge, then that knowledge should not increase sensitivity to differences between stimuli, and the cumulative discriminability of stimuli should be no better when linguistic knowledge induces a response bias than when it does not (Norris et al., 2000; Kingston, 2005).

Early evidence supporting the idea that an ambiguous sound contrasts auditorily with its context can be found in results reported by Mann (1980). Her results showed that listeners identified an ambiguous stop from a [d-g] continuum as spectrally low "g" more often after spectrally high [l] than after spectrally low [r]. Lotto & Kluender (1998) replicated this finding with an experiment in which the [l] and [r] contexts were replaced by pure tones that matched the high vs low energy concentrations in the liquids' spectra. In both experiments, high or low context (both speech and non-) altered listeners' response biases: they judged a stop that was ambiguous between [d] and [g] more often as the spectrally high alternative [d] in the context of the spectrally low liquid or tone than in the context of the spectrally high liquid or tone: listeners perceived the targets as having the opposite auditory quality from its context.

Because Lotto & Kluender (1998) used a non-speech sound as context, they argued that the context's auditory quality was the source of its effect on the stop judgment.<sup>2</sup> Similar contrastive effects of non-speech contexts have since been demonstrated for vowel backness judgments by Holt, Lotto, & Kluender (2000) and for stop place judgments by Coady, Kluender, & Rhode (2003). Lotto et al (1997) found comparable contrastive effects using speech stimuli and Japanese quail as listeners. Taken together these results are evidence for a general auditory phenomenon, not specific to human speech perception: the exaggeration of the perceived difference between successive intervals that differ in some acoustic dimension, in this case the spectrum.<sup>3</sup> This effect is referred to as sequential (spectral) contrast.<sup>4</sup>

---

<sup>2</sup> Mann (1980) explains her results differently, as compensation for coarticulation with the liquid. She hypothesized that listeners responded "g" after [l] than [r] because they undo the expected coarticulatory fronting of the stop by the [l]. Listeners use the acoustic properties of signals as information about the articulatory events that caused them and adjust their labeling judgments to compensate for the common effect of coarticulation between neighboring sounds (Mann, 1981 #10126; Fowler, 1990, 2006; Fowler, Brown, & Mann, 2000). As already noted in footnote 1, we do not present evidence here that would distinguish between the auditory contrast and compensation for coarticulation accounts.

<sup>3</sup> As will be discussed in the general discussion, others have proposed that

In line with both MERGE and the auditory contrast effects just reviewed, Kingston (2005) proposed a model in which target intervals contrast perceptually with their contextual intervals during an initial, auditory processing stage of processing and that linguistic knowledge only influences category labeling at a later stage. We adopt that model, which combines auditory autonomy with sequential contrast, as our primary hypothesis here.

The hypothesis predicts that in an experimental task which taps the later stage—identification—we should observe linguistic knowledge and sequential contrast independently biasing listeners' judgments: when the two effects cooperate, we should observe larger response biases than when they conflict. The two biases should be independent because strict autonomy does not permit lexical information to feed back on earlier processes.

---

auditory contrast be interpreted as shifts in the criterion for judging whether a sound is high or low along some auditory dimension rather than exaggerating the perceived difference between the target sound and its context (see Holt, Lotto, & Kluender, 2000, Holt, 2005 in particular).

<sup>4</sup> In this paper we use “*contrast*” in two different but related ways. First, to describe the inverse relationship between two sounds: if a sound becomes perceptually less like an adjacent sound, we describe the effect as “contrastive.” Second, to refer to the proposed mechanism that explains such an effect: sequential contrast, as defined above.

If discrimination taps the workings of the hypothesized prior auditory evaluation, as extensively argued by Kingston (2005), sensitivity to stimulus differences should be unaffected by response biases induced by linguistic knowledge. Cumulative discriminability across the entire stimulus continuum (= the sum of listeners' sensitivities to all the stimulus pairs composing the continuum) should not differ between continua in which linguistic knowledge induces a response bias and those where it does not. For example, cumulative discriminability should be no better for a word-nonword continuum than a word-word continuum. As Norris, et al. (2000) observe, an interactive model in which lexical knowledge feeds back on auditory evaluation predicts that the perceived difference between adjacent stimuli will be greater when one endpoint of the continuum is a word and the other is not than when both endpoints are words (or non-words). The member of the pair that is closer to the word endpoint will be activated more strongly by such feedback than the one farther away, and that difference in activation would enhance listeners' sensitivity to the difference between them. The overall result would be higher cumulative discriminability of pairs from the word-nonword continuum than similar pairs from continua whose endpoints have the same lexical status.<sup>5</sup> The

---

<sup>5</sup> Our autonomous model also predicts that purely auditory spectral contrast *does* improve cumulative discriminability across a continuum. Unlike lexical biases, sequential contrast should exaggerate the perceived difference between neighboring stimulus intervals and in doing so make a high-low

failure to find such differences would disconfirm a positive prediction of such interactive models.

The autonomous alternative does not predict that lexical knowledge will not alter sensitivity locally within a continuum. Our model instead predicts that lexical biases will shift local peaks in discriminability in the same direction as they do the category boundaries in the identification task. Finding peak-shifts in discrimination gives meaning to null effects of lexical feedback by assuring us that discrimination is in fact sensitive to lexical biases and that listeners actually attended to the task. We also expect that local discrimination peaks will shift as a result of the response biases induced by spectral contrast and that these shifts will be independent of those induced by lexical response biases. Finding such shifts would further confirm that spectral contrast and lexical biases are independent effects of context.

An initial set of experiments was devised to test the effects of sequential spectral contrast and lexical bias on the identification of vowels and following consonants. The stimuli consisted of vowels drawn from an [i]–[u] continuum followed by stops drawn from a [t]-[p] continuum, both of which range between a spectrally high endpoint ([i] and [t]) and spectrally low one ([u] and [p]). Next to a spectrally low context an ambiguous target is predicted to sound higher if auditory contrast exaggerates the perceived difference

---

sequences more discriminable from a low-high one than a high-high sequence is from a low-low one (Kingston, 2005).

between successive intervals. Therefore, sequential contrast predicts that listeners will select “p” more often than “t” following [i] than [u] (**Experiment 1- StopPlace- C Final**). Furthermore, when judging the vowel (**Experiment 2- VowelPlace- C Final**), listeners are predicted to select “u” more often before [t] than [p].

Both predictions follow if sequential contrast exaggerates the perceived difference between neighboring sounds. If it does, then its effects should operate both backward and forward: a sound with an intermediate spectrum will be perceived as having a higher spectrum whether it precedes or follows a sound with a low spectrum.<sup>6</sup> To create lexical biases, we concatenated the vowel and stop continua with initial consonants that formed word-nonword continua in which the lexical biases conflicted or cooperated with the contrast biases. Lexical biases should incline listeners to choose the category that makes a word with the context, and could therefore add to or subtract from the contrast biases. For example in a *keep- \*keet* continuum both wordhood and

---

<sup>6</sup> In this respect, our conception of contrast crucially differs from that developed by Holt (2000, 2006) where a sound’s context shifts the criterion for judging whether it is high or low. As Fowler (2006) observes such a model of contrast effects predicts that only preceding context would contrastively influence a target, because only past experience is expected to bring about such shifts. We will elaborate further on this debate in the General Discussion.

spectral contrast favor a “p” response, since *keep* is a word and an ambiguous stop should sound more like spectrally low [p] next to the spectrally high vowel [i]. In a *group*-\**groot* continuum only the lexical bias favors “p” while the contrast bias favors “t”; although *group* is a word, the contrast bias should moderate the effect of lexical bias by inducing a spectrally high [t] percept after the spectrally low vowel [u]. We also included control continua in which both endpoints are words.

Two discrimination experiments were run using the same stimulus sets, in which listeners discriminated stimulus-pairs taken from each continuum. The autonomy hypothesis first predicts that the word-nonword continua will not improve cumulative discriminability relative to the word-word control continua. It secondly predicts that the locations of discrimination peaks should be shifted by both contrast and lexical bias independently.

To preview our results, lexical biases unsurprisingly shifted category boundaries in the expected directions: listeners identified an ambiguous segment more often as the one which made a word with its context. However, the shifts in category boundaries expected from sequential spectral contrast were only obtained when listeners identified the stops after the vowel context, where they responded “p” more often after [i] than [u]. When they instead judged the vowel target preceding the stop, their judgments *assimilated* to the

following stop, in that they responded “u” more often before [p] than [t]. That is, they more often chose the vowel that was spectrally *similar* to the stop.

This difference could reflect the order of the target and context (i.e. whether the context precedes or follows the segment being judged) or their different segmental class (i.e. whether the target is a consonant or vowel). To determine which, we reversed the order of the vowel and stop in a set of follow-up experiments. In one (**Experiment 3: StopPlace - C Initial**), listeners judged word-initial consonants from a [t]-[p] continuum that preceded either a spectrally high [e] or low [o] vocalic context. In the other (**Experiment 4: VowelPlace - C Initial**), they judged a vowel from an [e]-[o] continuum following a word-initial [t] or [p]. If order of the target and context matter, then listeners should respond “t” more often before [e] than [o] but respond “o” more after [t] than [p]. If it is the class of the stimuli that matters, they should instead respond “t” more often before [o] than [e] but “o” more often after [p] than [t]. As in the first pair of experiments, the stimuli also included lexical biases that either cooperated or conflicted with the expected effects of contrast biases.

The categorization results from this second pair of experiments confirm that the order of context and target, not their class, determines whether contrast or assimilation takes effect. Controlling for lexical bias, listeners judged an ambiguous vowel as “o” more often after [t] than after [p], and

judged an ambiguous stop as “p” more often before [o] than before [e].

Generally speaking, target judgments assimilate to a following context, but contrast with a previous context.

Though Experiments 3 & 4 were designed to address the unexpected difference between the results from Experiments 1 & 2 (backwards assimilation or misparsing and forwards contrast), they also serve as tests of the same hypotheses as Experiments 1 & 2: sequential contrast is autonomous from lexical knowledge. And, in fact, results of these experiments were otherwise very similar to those of Experiments 1 & 2. They showed that both contrast and assimilation biased responses in identification and shifted local discrimination peaks accordingly independently of lexical biases. They also showed, as predicted by autonomous models, that cumulative discriminability across the whole continuum did not significantly differ as a function of the lexical status of continuum endpoints. It was no better for word-nonword continua than for word-word or nonword-nonword continua, (Norris et al., 2000; Kingston, 2005). Contrary to the prediction of interactive models, feedback from the lexicon does not activate the stimulus that is closer to the word endpoint more than the more distant stimulus, thereby increasing sensitivity to the difference between them. This finding attests to the existence of a level of processing that receives no input from lexical knowledge. .

## **2 General Methods**

In this section we summarize those aspects of our methods that are common to all four experiments, namely, participants, equipment, location and procedures. Additional experiment-specific details regarding stimuli will be provided in forthcoming sections.

### **2.1 Participants**

Participants were adult native English speakers drawn from the University of Massachusetts, Amherst community, who earned either course credit or \$10 per hour for their participation. No listener participated in more than one experimental condition, or in any other recent or related experiment involving lexical bias. No listener reported any speech or hearing disorder.

### **2.2 Location and equipment.**

The experiment took place in a sound-attenuated chamber. Each participant sat at a PC input-output terminal connected to a computer outside the room. All stimuli were output at 16 kHz from the PCs and presented binaurally to listeners through sound-isolating Sennheiser HD 280 64 Ohm headphones, at a comfortable volume. Cedrus SuperLab Pro software (version 2.0.4) was used to present all sound stimuli and visual cues, and logged all responses. All the participants used Cedrus RB-834 response boxes to enter their responses.

## **2.3 Procedure - Identification (ID)**

### **2.3.1 Identification experiment structure**

Every experimental trial for the identification experiment had the following structure: a single stimulus drawn from the continua was played back to the listener. When the sound finished playing, two color-coded visual prompts appeared on screen, to prompt listeners to answer. The visual prompts always consisted of two vowel or consonant choices, in an appropriate alphabetic transcription (see details below). Listeners pressed one of two color-coded buttons to indicate which vowel or consonant they had heard. The prompts were displayed for 1500 ms, which was the entire interval during which listeners could enter their response—no responses were collected while the stimuli were being played. After the listener responded or after the 1500 ms interval had elapsed, the software waited an additional 750 ms (the inter-trial interval, or ITI) before presenting the next stimulus.

Listeners started with a training section that contained 10 repetitions of each continuum endpoint, presented in random order. The training trials differed from the test trials in a single regard: after listeners responded to each training stimulus (or the response-interval elapsed), feedback in the form of the correct answer would appear onscreen for 500 ms. Following that, the 750 ms ITI would pass, and the next stimulus would be presented. Performance in training was not included in any subsequent analysis of the results.

After the training section, listeners proceeded through 6 test blocks, in which they classified stimuli drawn from the entire continua. Each block comprised the entire stimulus array, with extra repetitions of the middle steps according to the ratio 3:2:1 for steps {8, 10, 12, 14}, {4, 6, 16, 18}, and {1, 20}, respectively.<sup>7</sup> Therefore, during the entire experiment, participants listened to the endpoints 6 times, the near-endpoints 12 times, and the middle steps 18 times each. Like the training block, all stimuli were presented in random order during the test blocks.

After completing each test block (and at the end of the initial training section), participants received a message telling them to take a short break, and to press a button when they were ready to continue. After completing the third block, which marked the halfway point, a different message instructed the participants to exit the testing room for a short break (5-10 minutes) for refreshments, restrooms, and rapport. They were instructed not to leave the testing room until the other participants had reached the break point so as not to distract anyone else by exiting.

Every experimental session was finished within 60 minutes, including time for written and verbal instructions, testing, breaks, debriefing, and

---

<sup>7</sup> The entire continuum in each experiment consisted of 20 steps, from which we drew 10 stimuli for the identification experiments. The finer subdivision was used in choosing stimulus pairs in the discrimination experiments.

paperwork for consent forms and receipts of compensation.

### **2.3.2 Identification instructions.**

Before starting the experiment, listeners were told they would hear a variety of single syllable speech sounds and that they were to judge the identity of one segment, for example, the word-final consonant. They were told to press one of two buttons to indicate their judgment of the given target to ("p" or "t", for example). Participants were also instructed to respond as quickly as possible once the stimulus finished playing, and to rest their forefingers or thumbs on the two response buttons to allow them to respond quickly without moving their arms or hands. They were told they would start with an initial training section during which they would receive feedback in the form of the correct answer each time they pressed a button to respond. They were told all the details of the trial structure: the duration of their response interval (1.5 seconds); that after each stimulus finished playing, two color-coded visual prompts would appear ("p" and "t" for example), matching the color and arrangement of their two buttons; etc.

### **2.3.3 Identification Conditions.**

The only experimental condition involved the positions of buttons for the two responses. Half the participants used their left button for one of the responses (e.g. "t") and right button for the other (e.g. "p"), with corresponding visual prompts, and vice versa for the remaining participants. The button conditions

compensate for any disposition participants may have had toward either the left or right response. The color-coding remained constant across both conditions: the left response was always red, and the right response always blue.

## **2.4 Procedure - Discrimination**

In the discrimination task, details of listeners, location, and equipment match those in the identification procedure described above. Separate groups of listeners participated in the identification and discrimination experiments.

### **2.4.1 Discrimination Structure.**

The discrimination task used stimuli pairs drawn from the six 20-step continua. Specifically, the pairs used compared the following steps in the continuum: 2-5, 5-8, 7-10, 9-12, 11-14, 13-16, and 16-19. The members of each pair of stimuli differed by three steps, a distance which was close enough to present a challenge but far enough apart to make the stimuli moderately discriminable for most speakers.

The experiment lasted approximately four hours total, and was conducted during two 2-hour sessions split over two different days. The large amount of time permitted us to present 24 total same-different repetitions per ordered pair per continuum, per participant. The experiment was blocked by pair, so that each portion of the experiment included only a single pair—for

example step 2 compared to step 5—with all its orderings, for all 6 continua. Each stimulus pair in the continuum was expanded into 4 ordered pairs: two same and two different, e.g. the {2, 5} block would contain stimulus presentations of the *ordered* pairs (2-2), (2-5), (5-5), and (5-2) for all stems. Each of these blocks contained three repetitions of each ordered pair, and each such block was presented twice during a session to allow for 6 total repetitions per day (3 repetitions x 4 ordered pairs x 2 days = 24 repetitions). Each such 72-trial block was repeated twice to allow for a break in between, rather than a single uninterrupted block of 144 trials.

A short unscored “training” block preceded every new stimulus-pair section; these training blocks presented a single repetition of each ordered pair for every comparison pair within the continuum (for example {2-2, 2-5, 5-5, 5-2}, plus the same ordered pairs for the other five continua). The unscored training blocks were identical to the test blocks in every respect except number of repetitions.

The discrimination experiment had breaks similar to the identification experiment, but they were more numerous because of the experiment’s length. After completing each test block, participants received a message telling them to take a short break, and to press a button to continue. Longer breaks occurred after every two discrimination-pair sections. The procedure and instructions for these longer breaks matched that of the identification

experiment.

Every experiment session was finished in approximately two hours, (ranging from 1:45 to 2:10 hours) including time for written and verbal instructions, testing, breaks, debriefing, and paperwork for consent forms and compensation receipts.

#### **2.4.2 Trial Structure.**

Each test trial began with a pair of stimuli played through the headphones, separated by a 750 ms inter-stimulus-interval. When the stimuli finished playing, two color-coded visual prompts for “same” and “different” appeared on screen, to prompt listeners to answer. The color and orientation of the visual prompt words corresponded to the arrangement and color of the two buttons on the response pad. The prompts remained on the screen for 1500 ms, which constituted the entire period during which listeners could enter their response (no responses were collected while the stimuli were playing). When listeners pressed a button to enter their response, or when the 1500 ms response interval elapsed, feedback in the form of the correct answer would appear on the center of the screen for 500 ms. After the feedback disappeared, an inter-trial interval of 750 ms would pass and the next trial would begin.

#### **2.4.3 Instructions**

Before starting the experiment, listeners were told they would be hearing a

variety of pairs of single-syllable speech sounds, and that their task was to decide whether the two members of that pair were exactly the same or slightly different and press one of two buttons to indicate which. All details of the trial and experiment structure were explained: that they would receive feedback in the form of the correct answer, that their time to respond lasted 1.5 seconds, that after each stimulus finished playing, two color-coded visual prompts would appear—"same" and "different"—matching the color and arrangement of their two buttons; etc. Participants rested their forefingers or thumbs on the two response buttons, to allow them to respond quickly without moving their arms or hands.

#### **2.4.4 Conditions.**

Each listener was run on a particular button and order condition. Button conditions differed only in which response—"same" or "different"—corresponded to the left and right button. Half the participants were assigned the left button for the "same" response and the right button for "different", with corresponding visual prompts onscreen, and vice versa for the remaining participants. Each individual experiment also presented a particular ordering of single-interval test blocks, determined by a balanced latin square, resulting in 7 or 8 (depending on experiment) separate ordering conditions; consequently, when we pooled across subjects the block for each specific discrimination interval occurred both before and after every other.

### **3 Experiments 1 & 2**

These experiments were designed to test for the existence of both contrast and lexical biases, to demonstrate their independence from one another, and finally to test the hypotheses of autonomous models (Norris, et al., 2000; Kingston (2005) that (1) lexical biases are independent from contrast biases, (2) both kinds of biases shift local discrimination peaks in the same direction as they shift category boundaries, and (3) lexical biases do not improve listeners' sensitivity to stimulus differences, as measured by cumulative discriminability across the continuum.

#### **3.1 Experiment 1 (Experiment 1- Stop Place- C Final)**

In this experiment we examine categorization of ambiguous stop consonants in word-final position as a function of two variables: the spectral content of the preceding vowel (high or low), and the lexical bias that bears on the target response. 16 native-English speakers participated in the identification experiment and 33 in the discrimination experiment.

##### **3.1.1 Stimuli and predictions**

###### **3.1.1.1 Stimulus array & predictions**

The stimuli are single-syllable words beginning with a single or complex onset, followed by a spectrally-high [i] or spectrally-low [u], followed by a member of a [t-p] continuum. The control contexts are [hi] and [hu], which both form words with a following [t] or [p] (*heat, heap, hoot, hoop*). The test contexts

differ in their onsets. When the particular onsets combine with one of the following nuclear vowels, only one final stop ([t] or [p]) forms an English word with the preceding part. The combinations of onsets and vowels used were [hi\_, hu\_, ki\_, mi\_, gru\_, ?u\_]. These combinations will be referred to as “stems”.

The control stems, [hi\_] and [hu\_], allow us to make a simple prediction regarding the effects of sequential contrast: “p” should be selected more often after [i] than after [u], because [p] is more different than [i] from the spectral standpoint. We expect no other effect of context, since both [hi\_] and [hu\_] make words with both [p] and [t].

With regard to the lexical bias stems, which unlike the controls involve lexical biases toward one or the other final consonant, we expect sequential contrast effects and lexical biases to be expressed independently in identification. We should observe effects of lexical bias as a greater proportion of “p” responses for stems that make a word with [p], for example [ki\_] and [gru\_], compared to stems that make a word with [t]. Concurrently, spectral contrast should cause the vocalic context [i] to elicit more “p” responses relative to the [u] context, because [i] is spectrally high and [p] is spectrally low.

In other words, sequential (spectral) contrast should produce a bias in target identification that either cooperates or conflicts with the lexical bias. Sequential contrast and lexical bias should cooperate and induce very many or very few spectrally-low “p” responses in the [ki\_] and [.u\_] contexts respectively. Only [p] forms a word with preceding [ki\_], and the [i] concentrates energy high in the spectrum compared to [u]. [ʔu\_] should minimize “p” responses because the word *shoot* should promote “t” rather than “p”, and the vowel [u] is spectrally low. In [mi\_] and [gru\_] contexts the two effects favor opposite responses and should induce intermediate proportions of “t” responses. For example, in [mi\_], the lexical bias favors a “t” response because *meet* is a word and *\*meep* is not, but the contrast bias favors a “p” response because [i] is spectrally high and [p] is spectrally low. Since the proposed biases conflict perceptually rather than cooperate, we predict fewer proportions of “p” responses for these stems when compared to [ki-] but more compared to [ʔu\_]. Table 1 lists the stimuli along with the predicted response biases.

Bias	Stop-Place - C Final	
None (Control)	heat - heap	hoot - heap
Cooperate	keep - *keet	shoot - *shoop
Conflict	meet - *meep	group - *groot

Predictions % "p"

hi > hu  
ki > mi, gru > .u

---

Table 1. Experiment 1 stimulus continua and predicted differences in the number of "p" responses.

For discrimination of stimuli along the p-t continuum, we predict that scores will peak at the category boundary obtained or expected in the identification task (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). For example, in the *keep*-\**keet* continuum better discrimination should occur in the t-half of the continuum, because both lexical and contrast biases favor a "p" label and should push the category boundary toward the [t]-endpoint, relative to boundaries for other contexts and because listeners should be good at discriminating pairs of stimuli that lie on either side of that boundary.

Even though we expect lexical influences to shift local discrimination peaks within continua, the autonomous model predicts no difference in the cumulative discriminability of stimulus-pairs as a function of the lexical status of their endpoints. The autonomous model does not allow lexical feedback to enhance the auditory evaluation in a way that would sharpen overall sensitivity to stimuli.

### 3.1.1.2 Stimulus creation

In this experiment the vowels [i] and [u] are attached to a preceding onset, and listeners identify and discriminate a following [t-p] continuum. The stimuli each contain a singleton consonant or cluster, a voiced transition to a vowel steady state, followed by a transition to the following consonant, followed by a silent closure interval, followed finally by a [t] or [p] burst. To strike a balance between controlling their uniformity and preserving their naturalness, we synthesized voiced intervals by interpolating between parameter values taken from naturally spoken endpoints; the voiceless [t-p] continuum was synthesized by adding waveforms of naturally spoken bursts in complementary proportions; contextual onset consonants needed no continua so were simply taken from naturally spoken tokens.

We first recorded a native English speaker (2<sup>nd</sup> author) in a sound-attenuated room and encoded the sounds as .wav files to an external desktop computer. The sampling rate was initially 44100 Hz and the resolution was 16 bits. The stimuli were downsampled to 16000 Hz before any further manipulation. We next chose a token of each type that sounded like a clear and natural example of the intended word or non-word, and that matched the other best tokens in duration, pitch, and intensity. We then isolated the voiced portions of these tokens from the surrounding consonantal noise, and extracted their formant, bandwidth, pitch, and intensity parameters using PRAAT (Boersma & Weenink, 2005). The voiced portions included vowel steady states with transitions to and from adjacent consonants, plus any initial sonorant

consonant. We manually edited many individual time-steps in the extracted parameters to smooth over values that PRAAT had mistracked.

The next steps concerned making the vowel steady-states and transitions into the following stop uniform across the six continua. Since the h\_ tokens do not have onsets that would color the vowels for place in any particular way, we used their -it, -ip, -ut, and -up intervals as the base for all other stimuli; working backward from the [t] or [p] closure, we copied the parameter values from the interval including the vowel-to-stop transition and the preceding 110 ms of the vowel's steady-state. The parameters for that portion served as replacements for the final portion of all other stimuli endpoints; in other words, we attached uniform [ip, up, it, ut] rimes to the voiced onset transitions for each stimulus endpoint. To accommodate the unique onsets (e.g. k\_, etc) we manually edited the parameters of the voiced onset transitions to form an artifact-free acoustic join with the rime.

Also for the sake of uniformity, we equalized the pitch and intensity parameters of each pair of endpoints, and tailored the duration of the initial transitions to create a total duration of approximately 170 ms when combined with the rest of the rhyme. The nasal [m\_] and the [r\_] in the [gr\_] onset consonants added 75 ms and 85 ms voiced intervals to the vowel-transition portions of the rime in their respective stimuli.

The continua between the parametric endpoints were created by interpolating 20 steps between the values. The values were then used to resynthesize the voiced intervals using the Sensyn implementation of KLSYN88 (Klatt & Klatt, 1990), yielding ip-it and up-ut continua with initial transitions appropriate for the various consonantal onsets.

We generated the t-p burst continuum using the naturally recorded noise portions of the best [t] and [p] burst in both vocalic contexts. The best tokens were chosen using the same criteria as the rhymes and onsets, and needed only modest adjustments to match their durations exactly. We then blended the [t] burst from the i\_ context with the [t] burst from the u\_ context, and likewise for the two [p] bursts, by adding their waveforms together. The blended bursts were then roughly appropriate for either vowel context, but also crucially allowed us to observe the effect of the two vocalic contexts (i\_ and u\_) on the exact same burst continuum. Otherwise, listeners would have judged different target sounds in the two contexts, confounding any comparison between them. We then added the blended [t] and blended [p] burst's wave-forms together in complementary proportions to form a 20-step [t-p] continuum. We gave the [p] endpoint a relative advantage in intensity across the continuum, because the [t] burst would otherwise overwhelm most of the intended [p] half of the continuum. Roughly balancing the continuum for perceptual rather than acoustic intensity ensures that there are a range of stimuli in the middle of the continuum that are ambiguous as to the stop's place of articulation. The

members of the burst continuum were attached to the ends of the corresponding members of the six rime continua, following a 55 ms silent interval that simulates the voiceless stop closure.

For the obstruent onset consonants [k\_, g\_, ?\_], stop bursts and fricative noise were taken from each of the best onset exemplars. To avoid any long-range coarticulatory effect of the [t] and [p] codas on our onsets, the onset candidates were originally recorded as utterances from the stimulus array without any coda, that is, as [ki, gru, ?u]. We then concatenated the best example of each type with each step of the appropriate rhyme+burst continuum. [h] onsets sounded unnatural when we added the waveforms of a naturally produced [h]s from [\_i] and [\_u] contexts, so we therefore synthesized the [h]s in the same way as we did the voiced intervals in the stimuli.

The last step in constructing the stimuli for this and all other experiments was to scale the amplitude peak of voiced portions of the continua to 0.97 of the maximum. The amplitudes of the noise portions had been adjusted previously to be suitable for the voiced portions they were attached to.

Question: WE REDID splex in July 06 with stimuli that were tweaked for some uniformity that we had failed to control for the first time around, just to make sure we got the same pattern of results (we justified not redoing DISC, only Identification). WHAT WERE THOSE TWEAKS, AND THE ORIGINAL PROBLEMS? THE ORIGINAL SP STIMULI LOOK MATCHED IN THEIR STEADY STATES (MOVING BACK FROM THE CLOSURE, at least) I CAN'T RECONSTRUCT WHAT HAPPENED.--DAN

### 3.1.2 RESULTS

#### 3.1.2.1 Identification

Figure 1 shows the total proportion of "p" responses in the two different control stems which have no-lexical bias, separated by vowel.

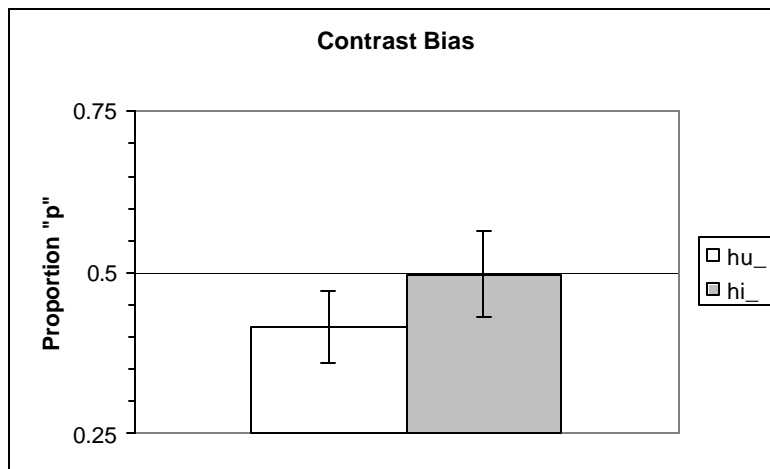


Figure 1: Mean total proportion "p" (95% confidence intervals) as a function of vowel in control stems.

We see from the figure that listeners chose the spectrally lower consonant, “p”, more often after the spectrally higher vowel [i]. This is precisely the result predicted by spectral contrast. To test the significance of this effect, we performed an ANOVA on the proportion of “p” responses using *vowel* (i vs u) as a within-subject independent variable and *button* (left or right for “p”) as a between-subject variable. The analysis reveals that the proportion of “p” responses was significantly greater following [i] than following [u] ( $F(1,14)=5.220, p = .038$ ). *Button* had a significant effect,  $F(1,14)=4.903, p=.044$ ; listeners who used their right button for the “p” response chose that response more often than the listeners who used their left. *Button* did not significantly interact with *vowel*,  $F<1$ . We next turn to the results for the stems intended to foster lexical bias in addition to effects of contrast.

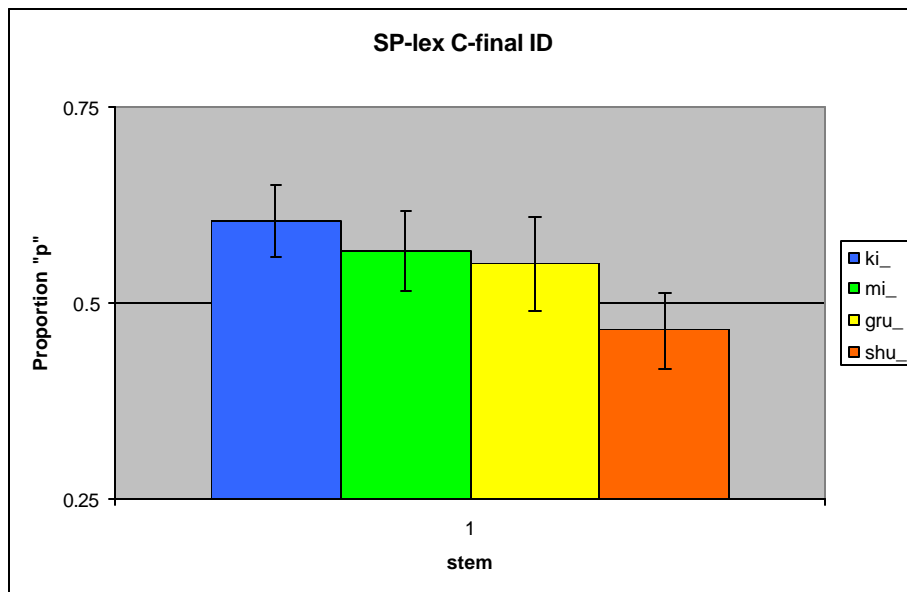


Figure 2: Mean total proportion “p” responses (95% confidence intervals) as a function of the lexical bias induced by the stem.

Figure 2 shows the greatest proportion of “p” response for the stem in which the lexical and contrast biases cooperate to favor “p” responses, [ki\_], and lowest proportion where they instead cooperate to favor “t” responses, [.u\_]. The proportion of “p” responses is in between these two extremes for the two stems in which the lexical and contrast biases conflict with one another, [mi\_] and [gru\_].

For the lexical bias stems, a mixed-design three-way ANOVA was run with *lexical bias* and *vowel* as within-subject factors, and *button* as a between-subject factor. We define *lexical bias* as the consonant (“p” or “t”) that is favored because it makes a word with the preceding stem. We define *vowel* in terms of which consonant response is favored by the vowel according to the assumption of spectral contrast, with [u] predicted to favor “t”, and [i] predicted to favor “p”. The analysis reveals main effects both of *lexical bias* ( $F(1,14)=38.927, p < .001$ ) and of *vowel* ( $F(1,14)= 11.014, p = .005$ ). There was a significant interaction between *button* and *lexical bias*, ( $F(1,14)=7.097, p=.019$ ), which reflects even greater observed effects of lexical bias differences between p-biased and t-biased stems for listeners tested on the “p-left” condition than for those tested on the “p-right” condition: listeners who used their left button for “p” gave 47.3% “p” response to the “t”-biased stems, and 54.7% to the “p”-biased stems; listeners who used their right button gave 53.4% and 61.6% “p” responses respectively. There were no other significant interactions and *button* alone had no significant effect  $F(1, 14)=2.523, p=.134$ .

### 3.1.2.2 Bias-induced shifts in discrimination performance

Here we report the evidence of category boundary shifts in the discrimination task. We compare cumulative discrimination performance in the [t] half of the continuum to that in the [p] half, for the two vocalic contexts. We expect better performance for the half of the continuum into which the listeners' category boundaries are shifted by lexical or contrast biases, because listeners should discriminate best the two stimuli that straddle their category boundary and are most likely to be labeled differently. For example, discrimination of stimuli from the *shoot*-\**shoop* continuum should be easier in the [p] than the [t] half of the continuum because both lexical and contrast biases cooperate to favor a "t" label and shift the category boundary toward the [p] endpoint. Rather than looking for peaks in the discrimination functions, we calculated the cumulative discriminability for the two halves of each continuum, across the pairs 1-5, 3-7, 5-9, and 7-11 for the [p]-half of the continuum and across pairs 9-13, 11-15, 13-17, and 15-19 for the [t]-half.

Figure 3 illustrates results for the control stems.

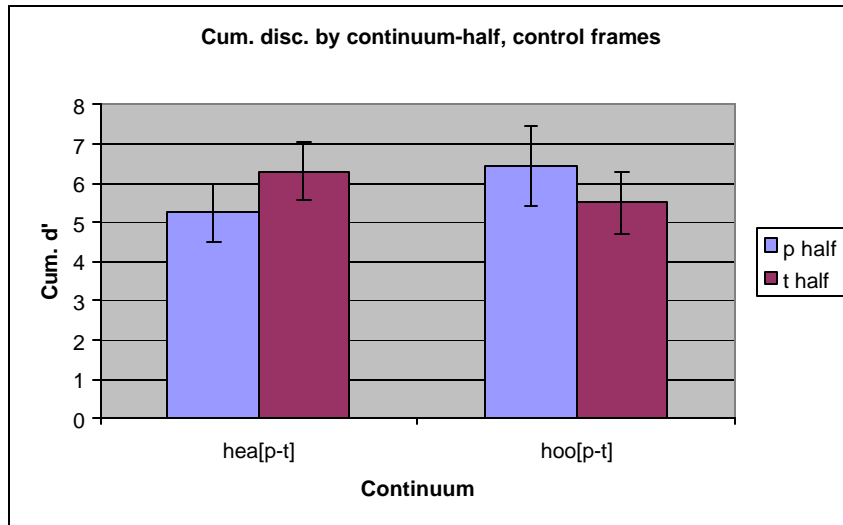


Figure 3: Mean cumulative  $d'$  values (95% confidence intervals) in the [t] vs. [p] halves of the continuum as a function of the vowels in the control stems.

The discrimination peaks correspond to the locations of category boundaries observed in identification. Cumulative discriminability is greater in the [t]-half of the continuum in the context of [i], but greater in the [p]-half in the context of [u]. This reflects categorization in that the category boundary between “t” and “p” is closer to the [t] end of the continuum after [i] compared to after [u]. An ANOVA using *continuum-half* and *vowel* as within-subjects variables showed a significant interaction between *continuum-half* and *vowel* on performance ( $F(1,31)=19.312, p<.001$ ). Neither within-subject variable had any significant effect alone (both  $F_s < 1$ ). The between-subjects variable *button* had no significant effect ( $F < 1$ ), and did not significantly interact with either of the other two variables (both  $F's < 1$ ).

Figure 4 shows results from the lexical bias stems.

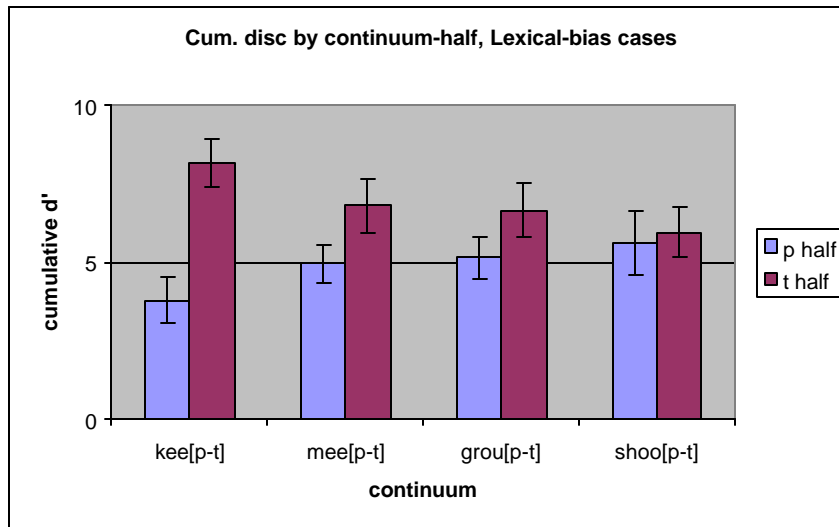


Figure 4. Mean cumulative  $d'$  values (95% confidence intervals) in the [t] vs. [p] halves of the continuum as a function of vowels and lexical biases in the lexical bias stems.

Discrimination of the lexical bias is consistent with the identification results. Within the *keep-<sup>\*</sup>keet* continuum, both lexical bias and spectral contrast favor “p”, which should cause a labeling crossover from “p” to “t” near the [t]-endpoint. As expected, discrimination performance was better in that half of the continuum. Within the *shoot-<sup>\*</sup>shoop* continuum we expect both lexical bias and spectral contrast to favor a “t” label and leave the category boundary nearest to the [p]-endpoint. In fact, *shu\_* yielded the best [p]-half performance and worst [t]-half performance relative to the other stems. Looking from left to right in the chart, toward the strongest “t”-bias context [shu\_],

performance in the [t]-half of the continuum steadily decreases, while [p]-half performance increases. The pattern is precisely what we would project from the scale of strongest to weakest “p”-bias, *kee\_ > mi\_*, *gru\_ > shu\_*.

An ANOVA reveals a significant effect of *continuum-half* ( $F(1, 31)=20.134$ ,  $p<.001$ ), and no significant individual effect of *vowel* ( $F<1$ ), or *lexical bias* ( $F<1$ ). *Continuum-half* significantly interacted with *vowel* ( $F(1, 31)=33.544$ ,  $p<.001$ ), and also with *lexical bias* ( $F(1, 31)=25.468$ ,  $p<.001$ ). *Vowel* did not significantly interact with *lexical bias* ( $F<1$ ). The three-way interaction between these variables did not reach significance  $F(1, 31)=2.852$ ,  $p>.10$ . *Button* had no significant effect by itself ( $F(1, 31)=1.311$ ,  $p=.261$ ), but contributed to two three-way interactions: *button* by *continuum-half* by *vowel* ( $F(1, 31)=4.484$ ,  $p=.042$ ), and *button* by *continuum-half* by *lexical bias* ( $F(1, 31)=4.754$ ,  $p=.034$ ). All other interactions were non-significant.

### 3.1.2.3 Cumulative discriminability across the entire continua

Figure 5 displays cumulative discriminability values across the entire continuum. These scores are calculated by simply adding all the  $d'$  values for each discrimination pair across the entire continuum. The cumulative  $d'$  score for a continuum would not change with any response-bias induced local shift in discrimination performance because an increase in sensitivity for some stimulus pairs would be offset by a decrease in sensitivity to others. However, cumulative discriminability would reach higher values for word-nonword

continua if lexical feedback increases activation of the member of each pair that is closer to the word end of the continuum.

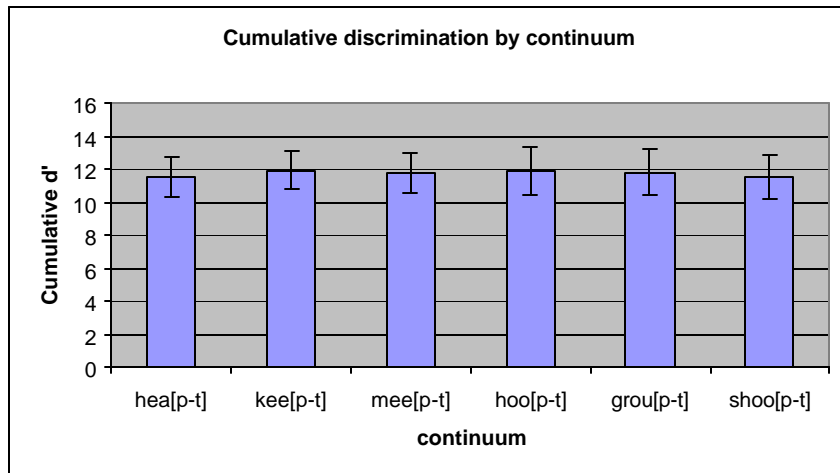


Figure 5. Mean cumulative  $d'$  values (95% confidence intervals) for control and lexical bias stems. (CHANGE Y-AXIS SCALE AS SUITABLE FOR ALL CUM. CHARTS, AFTER DISCUSSION)

The figure indicates no difference in cumulative discriminability as a function of whether there is a lexical bias ([hi\_] and [hu] vs [ki\_], [mi\_], [ʔu\_] and [gru\_] nor what the lexical or contrast biases are. An ANOVA using *stem* (controls + lexical biased stems) as the within-subjects variable and *button* as a between-subjects variable revealed no effect of *stem* ( $F(5, 155)=.343, p=.886$ ), no effect of *button* ( $F<1$ ), and no significant interaction between *stem* and *button* ( $F(5, 155)=1.478, p=.2$ ). A direct pairwise comparison between control and lexical bias stems is unnecessary here, because the absence of a significant effect of ANOVA already indicates that none of stems elicited performance that

significantly differed from any other.

The absence of any differences in cumulative discriminability between any of the continua disconfirms the interactive model's hypothesis that the auditory evaluation of these stimuli is influenced by feedback from the lexicon. Importantly, the earlier peak-shifts provide evidence that the discrimination experiment was not simply insensitive to lexical effects.

We turn next to the experiment in which the roles of target and consonant were reversed, Vowel Place - C Final.

### **3.2 Experiment 2 (VowelPlace- C Final)**

In this experiment, the stops [t] and [p] serve as context and listeners identify or discriminate members of a preceding [i-u] continuum. 21 listeners participated in the identification experiment and 31 listeners in the discrimination experiment.

#### **3.2.1 Stimuli and Predictions**

##### **3.2.1.1 Stimulus array & Predictions**

This experiment differs from Experiment 1 in that the members of a vowel continuum serve as targets, and following [p] or [t] serve as contexts. Predictions regarding the difference between cooperating and conflicting biases on categorization match those of SP-LEX. We expect spectrally-high [t]

to induce more spectrally-low “u” responses, which will add to or subtract from any lexical biases. An initial [h] serves again as the control onset, as it once again forms a word with all four possible rhymes.

Bias	Vowel Place - C Final	
None (Control)	heap - hoop	heat - hoot
Cooperate	cheap - *choop	fruit - *freet
Conflict	scoop - *sceep	sleet - *sloot
Predictions % “u”	$h\_t > h\_p$ $fr\_t > sk\_p, sl\_t > t.\_p$	

Table 2. Experiment 2 stimulus continua and predicted differences in “u” responses as a function of lexical and contrast biases.

Predictions for discrimination are identical to those for Experiment 1. For discrimination of stimuli along the [u-i] continuum, better discrimination should occur in the [i]-half of the continuum when both lexical and contrast biases favor “u” (*fruit-\*freet*, for example) and shift the category boundary closer to [i]. While local discrimination peaks should shift within continua as a function of category boundaries, the autonomous model predicts no difference in cumulative discriminability within the continua as a function of the lexical status of their endpoints.

### 3.2.1.2 Stimulus construction

The VP-lex stimuli were constructed in essentially the same way as the SP-lex stimuli. There were just a few minor differences. First, a continuum was created by interpolating between vowel place endpoints instead of stop place endpoints, which were held constant. Continua were formed between the [\_ut] and [\_it] and between [\_up] and [\_ip] rhymes. Second, the final stop bursts for each continuum were made by mixing in complementary proportions a [t] burst from an [i\_] context with one from an [u\_] context, and likewise for the [p] bursts. This produced [t] and [p] bursts whose spectra were appropriate for each step along the preceding vowel continua. We created a continuum of the onset consonants by the same method as the coda bursts. For example, waveforms of the burst and fricative noise of [tʃ]s from an [\_u] and an [\_i] context were added in complementary proportions; the same procedure was used to make the noise intervals for the other obstruent onsets. The [l]s in *sleet*-\**sloot* and [r]s in *fruit*-\**freet* were part of the voiced portion which we synthesized with the vowel, using parameters extracted from our best tokens of naturally recorded liquids in [\_u] and [\_i] contexts. All formant parameters were given appropriate transitions from the initial consonants to the steady-state [u] or [i] vowel.

### 3.2.2 Results

#### 3.2.2.1 Identification

As in Experiment 1, we present the results for the control stems separately from the lexical-bias stems. The control stems test the prediction of spectral contrast alone, while the lexical stems add effects of lexical biases to those of contrast biases. Figure 6 shows the identification results for the control stems.

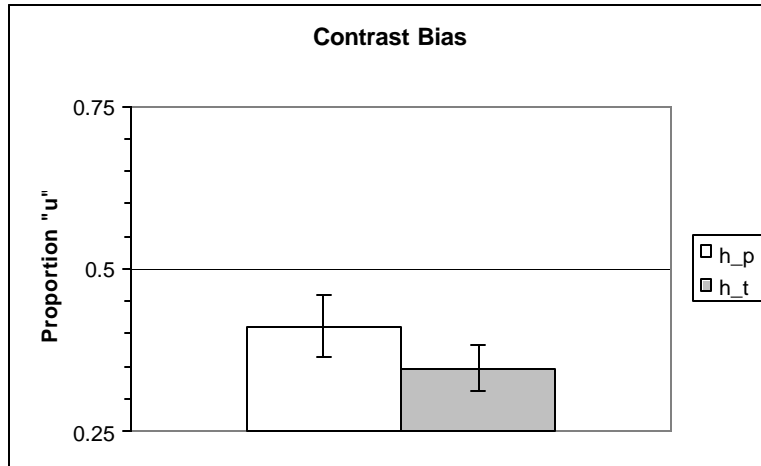


Figure 6: Mean total proportion of “u” responses (95% confidence intervals) as a function of following stop’s place in control stems.

The proportion of “u” responses was unexpectedly greater before [p] than before [t] ( $F(1, 19)=14.029, p=.001$ ). Rather than contrasting with the following stop, listeners’ judgments of the vowel *assimilated* to it. The between-subjects variable *button* had a significant effect ( $F(1,19)=5.141, p=.035$ ) and significantly interacted with *consonant* ( $F(1,19)=9.037, p=007$ ). Listeners who

used their left button for the “i” response gave that response more often than listeners who used their right button for “i”. Additionally, the two stop contexts [t] and [p] yielded a greater difference of response proportions for the listeners who used their left button “i”: the difference is .098 for [i]-left participants and .01 for [u]-right participants (in fact, the main effect of *assimilation* is largely due to the behavior of [i]-left participants).

Figure 7 displays the proportions of “u” responses obtained with the lexical bias stems.

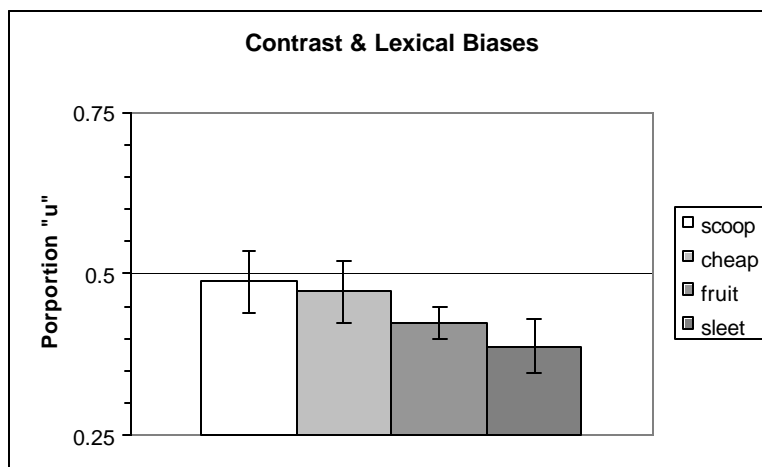


Figure 7. Mean total proportion of “u” responses (95% confidence intervals) as a function of following stop’s place and lexical bias in the lexical bias stems.

The proportions of “u” responses differed between the continua. Because listeners’ vowel place judgments again assimilated to the following stop’s place, what were predicted to be the conflicting stems, [sk\_p] and [sl\_t], turned out

instead to be the cooperating stems, while the predicted cooperating stems, instead exhibited conflict between the lexical and contrast biases. In a more novel finding, the final [p] in the [t?\_p] stem induced more “u” responses than did the lexical bias toward “fruit” in the [fr\_t] stem, which suggests that the assimilation bias induced by the final stops is stronger than the lexical biases. An ANOVA was run with *lexical bias* and *consonants* as within-subject factors, and *button* as a between-subject factor. The analysis revealed a main effect of both *lexical bias* ( $F(1,19)=4.438, p=.049$ ) and *consonant* ( $F(1,19)=18.404, p<.001$ ), and no significant interaction between them ( $F<1$ ). There was no effect of *button* and no interaction between *button* and other within-subject factors (all  $p's>.3$ ).

### **3.2.2.2 Bias-induced shifts in discrimination performance**

Figure 8 shows cumulative  $d'$  values in the [i]- and [u]-halves of the vowel place continuum in the context of following [t] vs [p].

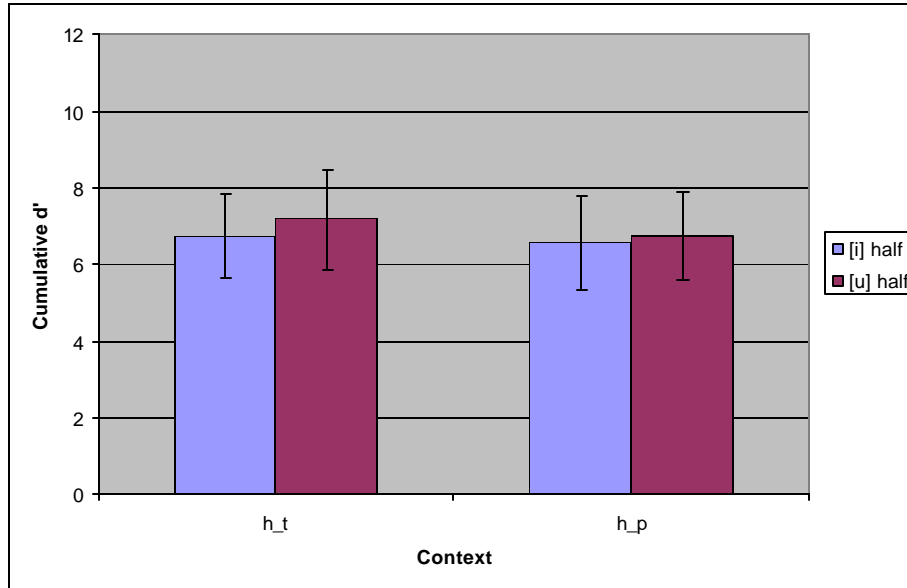


Figure 8: Mean cumulative  $d'$  values (95% confidence intervals) in the [i]- vs. [u]- halves of the continuum as a function of the stops in the control stems.

Sensitivity did not appear to differ between the [i]- and [u]-halves of the continuum, or for either context. An ANOVA reveals a non-significant effect of *continuum-half* ( $F < 1$ ) and of *consonant* ( $F(1,29) = 2.386$ ,  $p = .133$ ), and no significant interaction between them ( $F < 1$ ). The lack of a significant interaction conflicts with both the predicted contrast effect of the following stop on the category boundary for each vowel continuum, and more surprisingly with the observed assimilative effect of that stop. The between-subjects variable *button* had no significant effect and did not significantly interact with *continuum-half* or *consonant* (all  $F$ 's  $< 1$ ).

Figure 9 shows [i]- vs [u]-half discriminability for the lexical bias stems.

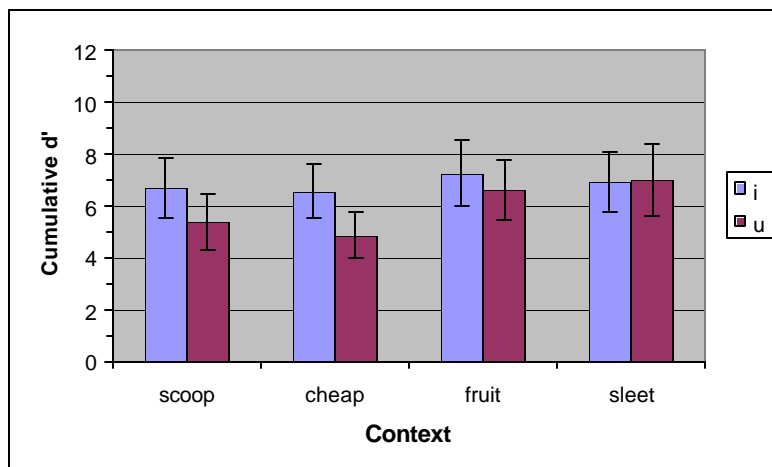


Figure 9. Mean cumulative  $d'$  values (95% confidence intervals) in the [i]- vs. [u]- halves of the continuum as a function of following stop place and lexical biases in the lexical bias stems.

From the [sk\_p] to [sl\_t] stems, discrimination improved in the [u]-half of the continuum relative to the [i]-half. The relative increase in performance is predicted from the category-boundary shifts observed in the identification data, where the combined effects of lexical bias and assimilation caused listeners to cross over from “i” to “u” responses closest to the [i] end of the continuum in the [sk\_p] stems and closest to the [u] end in the [sl\_t] stems, with the category boundaries for [t?\_p] and [fr\_t] stems lying between those extremes.

The results of an ANOVA with *continuum-half*, *consonant*, and *lexical bias* as within-subjects variables, and *button* as a between-subjects variable, revealed a marginally significant effect of *continuum half* ( $F(1, 29)=3.882, p=.058$ ).

*Lexical bias* alone was not significant ( $F(1, 29)<1$ ). *Consonant* had a highly significant effect ( $F(1, 29)=27.066, p<.001$ ). The interaction between

*continuum-half* and *lexical bias* was not significant ( $F<1$ ). *Continuum-half* and

*consonant* significantly interacted ( $F(1, 29)=12.213, p=.002$ ). The three-way interaction between these variables was significant ( $F(1, 29)=5.032, p=.033$ ). *Button* had no significant effect ( $F<1$ ), and did not significantly interact with any other variable (all  $p$ 's $>.085$ ).

### 3.1.2.4 Cumulative discriminability across the entire continua

Figure 10 displays cumulative  $d'$  values across the entire continuum for each stem.

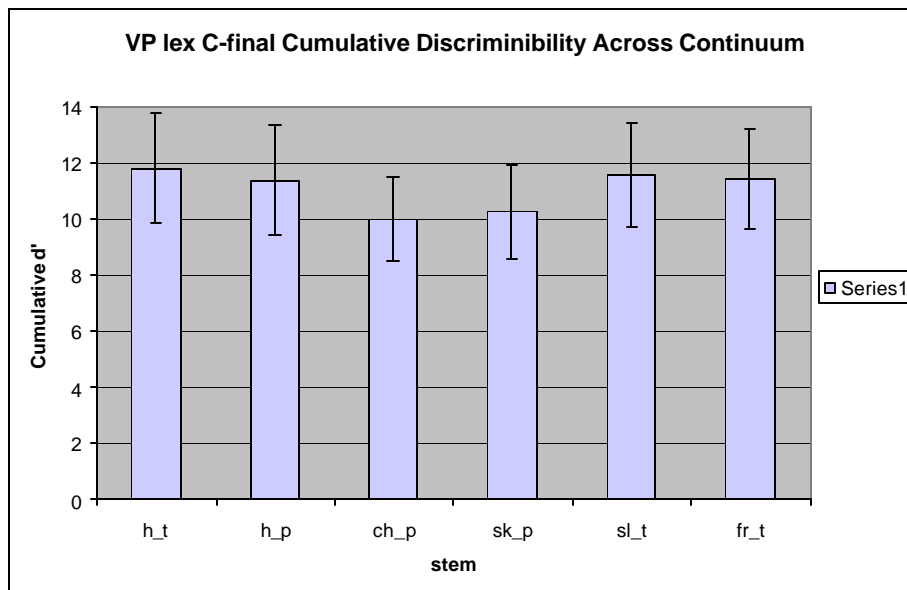


Figure 10. Mean cumulative  $d'$  values (95% confidence intervals) for control and lexical bias stems.

An ANOVA on cumulative discriminability using the six stems as within-subjects variables reveals a main effect of *stem* ( $F(5, 145)=6.439, p=.003$ ). As suggested by Figure 10, the significance seems attributable to the low discriminability of

ch\_p and sk\_p stems relative to the others. A post-hoc within-subject contrast analysis comparing these two stems and the other four reveals a significant difference ( $t(30)=2.98$ ,  $p=.005$ ). In fact, control stems were more discriminable than lexical-bias stems in general ( $t(30)=2.95$ ,  $p=.006$ ). We do not have a good explanation for why the control stems were more discriminable, but the result is at odds with the interactive model, which would predict the opposite outcome.

### **3.3 Summary of and discussion EXP 1 & 2**

#### *3.3.1 Contextual effect of neighboring segment: contrast vs. assimilation*

Listeners identified ambiguous stops from a [t-p] continuum more often as spectrally-low “p” when the preceding vowel context was spectrally-high [i] than spectrally-low [u], but they also identified ambiguous vowels from an [i-u] continuum more often as spectrally-high “i” when the following stop was spectrally-high [t] than spectrally-low [p]. The first result looks like contrast with the context, the second like assimilation to it. Local shifts in performance in the corresponding discrimination tasks were consistent with the response biases obtained in the identification tasks in Experiments 1 and 2.

Assimilation in Experiment 2 came as a surprise because we had expected to obtain contrast in both directions, and did observe it operating forwards in Experiment 1. As noted in the Introduction, there has been some

discussion in the literature as to whether contrast operates in both directions. Given our current formulation of the effect of sequential contrast as being an exaggeration of differences between adjacent intervals, it should go both ways. We suspend discussion of an alternative conception of contrast (e.g. Holt 2005) and of past empirical evidence which weighs in on the side of the former, less restrictive, version to the General Discussion.

In Experiment 1 listeners judged a stop-place continuum with the *preceding vowel* held constant as [i] or [u], while in Experiment 2 they instead judged a continuum of vowels with the *following consonant* held constant as [t] or [p]. Either differences in the order of target and context—target following context or preceding it—or in the segmental class of the target and context—consonant versus vowel—could be responsible for the different findings. Follow-up experiments 3 & 4 were designed to determine which of these two explanations are correct.

### **3.3.2 Autonomy vs. interaction**

Despite the different findings of these two experiments, the results establish that the contrastive and assimilative effects of the target sound's segmental context exist independently of the lexical effects of the entire stem. Although lexical knowledge did change discrimination performance locally, cumulative  $d'$  values across the entire continua did not differ as a function of the lexical status of continuum endpoints, conflicting with interactive models of speech processing such as TRACE in which feedback from the lexicon should sharpen

sensitivity to stimulus differences when one continuum endpoint is a word and the other not.

## 4 Experiments 3 & 4

### 4.1 Introduction & Background

If order or *directionality* of the target or context caused the results in Exp. 1 and 2, we should find that the percept of a target stop's place should assimilate to the place of the following vowel in Experiment 3 Stop Place - C initial, but that the percept of a target vowel's place should contrast with the preceding stop's place in Experiment 4 Vowel Place - C initial). We expect the opposite pattern of results if segmental class—consonant or vowel—determines whether the context has a contrastive or assimilative effect. In that case we should instead obtain assimilation when listeners judge the vowel in Experiment 4, but contrast when they judge the stop in Experiment 3.

TABLE GIVEN BELOW (it still needs explanation, if we decide it would be useful to keep).

	EXP 1 SP LEX- C Final	EXP 2 VP LEX C Final	EXP 3 SP LEX C initial	EXP 4 VP LEX C initial
	V[C]	[V]C	[C]V	C[V]
Direction: forward contrast/backward assimilation	CONTRAST	ASSIMILATION	ASSIMILATION	CONTRAST
Segment: When consonant is target, adjacent vowel yields a contrastive effect on judgment; when vowel is target, adjacent consonant yields an assimilative effect on percept of vowel.	CONTRAST	ASSIMILATION	CONTRAST	ASSIMILATION

Caption?: The target is in brackets.

## **4.2 Experiment 3 (Stop Place - C initial)**

The stimuli for the C-initial experiments include segments similar to the earlier C-final stimuli, but differ in a few ways. First, in order to match strings for various lexical statistical properties, it was necessary to substitute the mid vowels [e] and [o] for [i] and [u], respectively. The mid coronal vowel [e] concentrates energy at high frequencies much like [i] does, and the mid labial vowel [o] concentrates energy at low frequencies much like [u] does. Second, and pivotally, the [t] and [p] stops that serve as targets and contexts in this experiment precede the vowels, rather than following them.

19 listeners participated in identification, and 15 in discrimination.

### **4.2.1 Stimuli & Predictions**

#### **4.2.1.1 Predictions**

First, we expect listeners to respond “t” more often when “t” forms a word with the following context, compared to contexts that would instead form a word with “p”. Second, we expect the following [e] and [o] to exert either a contrastive or assimilative effect on the same stop judgments, depending on whether segment class or order determines the nature of the context’s effect. If the assimilation observed in Experiment 2 reflects the order of context and target, listeners should respond “t” more often before [e] than before [o]. However, if it instead reflects the segmental classes of the target and context—vowel and consonant, respectively—listeners should respond “t” more often

before [o] than before [e]. We again expect more extreme proportions of “t” and “p” responses when the contrast or assimilation biases cooperate with the lexical biases to favor the same response, and intermediate proportions when they conflict. The predictions in Table 3 are based on the assumption that the percept of the initial stop will contrast with the following vowel.

Bias	Vowel Place - C Final	
None (Control)	*teish - *peish	*toash - *poash
Cooperate	pace - *tace	toad - *poad
Conflict	tape - *pape	pope - *tope <sup>8</sup>
Predictions % “p”	$\begin{array}{l} \_e? > \_o? \\ \_es > \_ep, \_op > \_od \end{array}$	

Table 3. Experiment 3 stimulus continua and predicted differences in “p” responses as a function of lexical and contrast biases.

Note that the control stems in this experiment are all non-words rather than all words. We used non-words in this experiment, as well as Experiment 4, because we were unable to find a set of four words beginning with [p] and [t]

<sup>8</sup> The string [top] is a word, perhaps even two, *taupe* and perhaps *tope*, but both are decidedly rare (HOW RARE???) and were not known by more than ??? of the participants in this experiment or Experiment 4.

followed by [e] and [o] that had sufficiently similar lexical statistical properties.

#### 4.2.1.2 Stimulus construction

Construction of stimuli for the C-initial experiments followed the same method as for the C-final experiments, except for a few specifics.

First, since the stimulus array for the C-initial experiments doesn't already include any place-neutral [h\_] stimuli, we recorded all the necessary coda consonants in the context of h{e, o}\_, to obtain pronunciations of these consonants in the two vowel contexts free of any long-distance coarticulatory influences from the onsets. Second, our speaker recorded [pe], [po], [te], and [to] syllables without any coda consonants in order to provide for onset bursts and parameters for initial [p] and [t] transitions that would not be colored by any long-distance coarticulation with a coda consonant. **IMPORTANTLY: WHAT RECORDED CONTEXTS DID VOWEL STEADY STATE TOKENS COME FROM? Did we use a toad token's vowel because sounded best?**

smaller details: should we include consonant durations for final p/d/s/sh? The initial [p] and [t] bursts were created from the best of their recorded tokens in a manner comparable to that used to create the coda [p] and [t] bursts in Experiment 1. The [p] burst from the [\_e] context was blended with the [p] burst from the [\_o] context to create the [p] burst endpoint, and likewise for the [t] burst. Then a 20-step [p] to [t] burst continuum was created by adding the two endpoints waveforms together in complementary proportions,

with a slight advantage given to [p] per step to perceptually balance the continuum between “p” and “t” percepts. As in Experiment 1, steady-state vowels with their surrounding consonantal transitions were synthesized from parameters taken from the naturally spoken endpoints.

## 4.2.2 Results

### 4.2.2.1 Identification

Figure 11 displays the mean proportion of “p” responses as a function of the following vowel in the control stems.

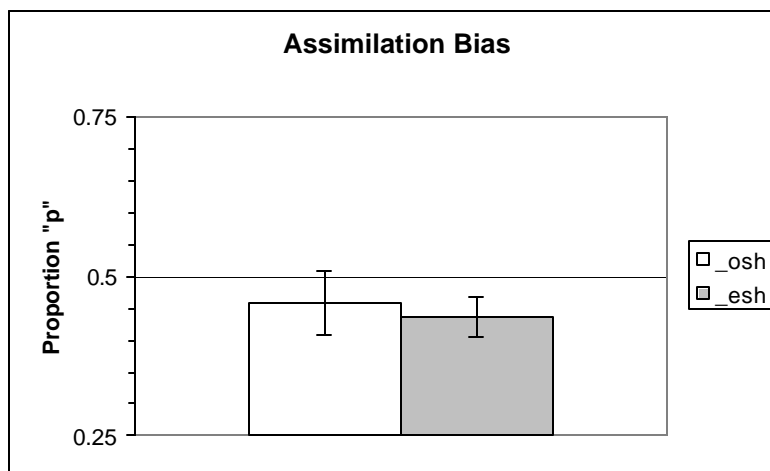


Figure 11: Mean total proportion of “u” responses (95% confidence intervals) as a function of following stop’s place in control stems.

Although there is a trend toward more “p” responses before [o] than [e], that is, toward *assimilation* to the following vowel, the proportions of “p” responses did not differ significantly between the two vowel contexts

( $F(1,17)=1.091$ ,  $p=.311$ ). *Button* had no significant effect ( $F(1,17)=1.492$ ,  $p=.239$ ), and did not significantly interact with *vowel* ( $F<1$ ).

Figure 12 shows that “p” response proportions reflect the expected lexical biases but again only weak evidence of an assimilation bias.

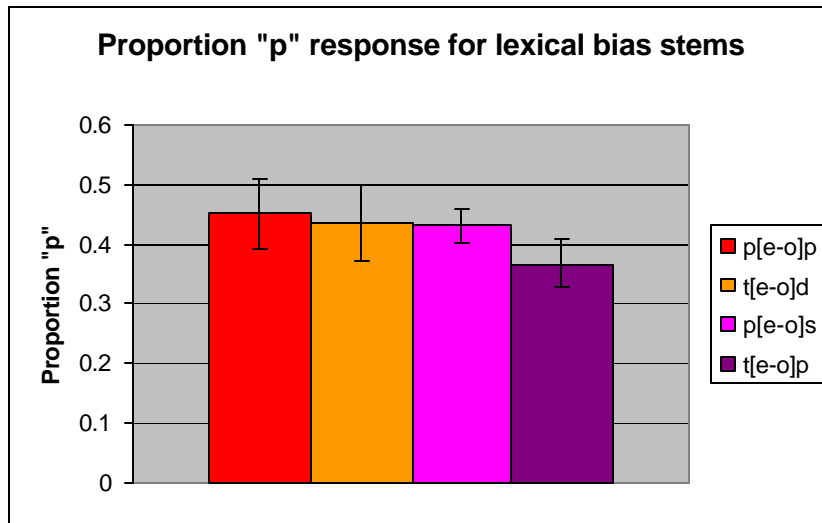


Figure 12. Mean total proportion of “p” responses (95% confidence intervals) as a function of following vowel’s place and lexical bias in the lexical bias stems.

In lexical-bias stems, we find a significant main effect of *lexical bias* ( $F(1,16)=13.191$ ,  $p<.002$ ). Listeners responded “p” more often before [\_op] and [\_es], with which it makes a word, than before [\_od] and [\_ep], with which only [t] makes a word. As with the control stems in Figure 11, the main effect of *vowel* at best weakly assimilative: listeners respond “p” more often before [o] than [e] but the effect was only marginal ( $F(1, 16)=36.017$ ,  $p =.096$ ). The effect of *vowel* interacted significantly with that of *lexical bias* ( $F(1,$

16)=17.914,  $p=.001$ ). *Button* had no significant effect alone ( $F < 1$ ), but interacted significantly with *vowel* ( $F(1, 16)=10.096, p=.006$ ); no other interaction involving *button* was significant (all  $F$ 's  $< 1$ ). Listeners who used their right button for the [e] (or "a") label showed a greater average difference in response proportions for [e] versus [o] (6% difference) contexts than did listeners who used their left button for the "ee" label (3% difference); in other words the vowel had a stronger assimilative effect for members of the former group.

#### 4.2.2.2 Discrimination

Figure 13 shows the differences in discrimination performance between the [t]- and [p]-halves of the continuum as a function of the following vowel's place for the control stems.

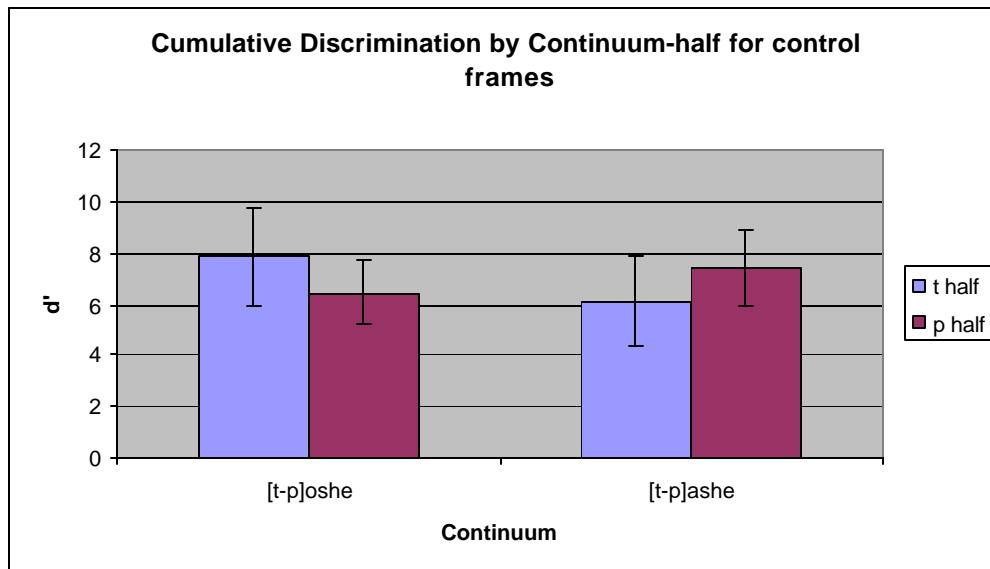


Figure 13. Mean cumulative  $d'$  values (95% confidence intervals) in the [t]- vs. [p]- halves of the continuum as a function of following vowel place.

Listeners performed better in the [t]-half of the continuum in the context [ \_o?], and better in the [p]-half of the continuum in the context [ \_e?], as expected if [e] assimilatively promotes "t" judgments of an ambiguous preceding stop. While there were no main effects of *continuum-half* ( $F < .001$ ) and *vowel* ( $F(1, 13) = 1.229$ ,  $p = .288$ ), the two factors interacted significantly,  $F(1, 13) = 6.723$ ,  $p = .02$ . *Button* had no significant effect ( $F < 1$ ) and did not significantly interact with the within-subjects variables ( $F < 1$  for all interaction terms involving *button*, except for *button* by *vowel*,  $F(1, 13) = 1.53$ ,  $p = .238$ ).

The differences in the continuum halves are much clearer evidence of an assimilation bias than the identification results displayed in Figure 11, as they indicate that listeners crossed over from "t" to "p" responses closer to the [t]-end of the continuum when the following vowel was [o] but closer to the [p] end when it was [e] (i.e. more "t" responses).

Figure 14 displays [t]- vs [p]-half discrimination performance in the stems where lexical biases are also present, and it too shows much clearer evidence of an assimilation bias than did the corresponding identification data in Figure 12.

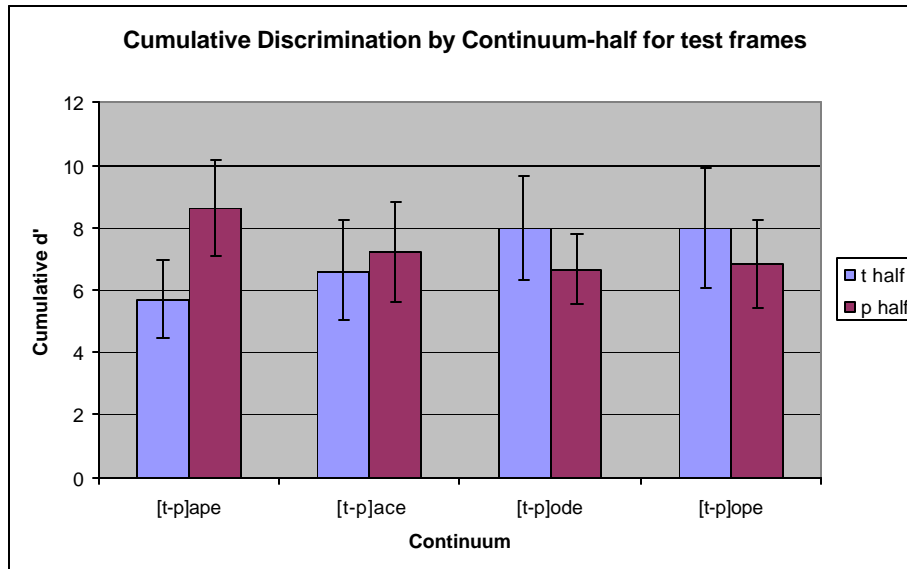


Figure 14. Mean cumulative  $d'$  values (95% confidence intervals) in the [t]- vs. [p]- halves of the continuum as a function of following vowel place and lexical biases in the lexical bias stems.

Performance in the [t]-half continuum was better relative to that in the [p]-half when the vowel context was [o] than [e], which indicates that the stop assimilated to the following vowel (i.e. more “p” responses before [o] than [e]). Performance in the [t]-half relative to the [p]-half was worst in the stem [\_ep], where the lexical and assimilation biases cooperate to induce the fewest “p” responses and best in the stem [\_op] where their cooperation instead induces the most “p” responses. For one conflicting stem, [\_as], the difference in [t]-half vs [p]-half performance shrinks as expected, but for the other, [\_od], the difference is the same as for the cooperating stem [\_op]. This last result indicates that the assimilation bias pushes percepts more toward “p” than the lexical bias pushes them toward “t”. The difference between the two

conflicting stems is the basis for the significant three-way interaction between *continuum half*, *vowel*, and *lexical bias*. A repeated-measures ANOVA was run on the cumulative discrimination scores using three within-subjects variables: *continuum half* (t or p), *lexical bias* (toward “t” or “p”), and *vowel*. None of the main effects were significant (*continuum-half*:  $F < 1$ ; *vowel*:  $F(1,13) = 1.003$ ,  $p = .203$ ; *lexical bias*  $F < 1$ ), but *continuum-half* interacted significantly with both *vowel* ( $F(1,13) = 6.564$ ,  $p = .024$ ) and *lexical bias* ( $F(1,13) = 5.959$ ,  $p = .03$ ), and the three variables interacted significantly with one another ( $F(1,13) = 12.552$ ,  $p = .001$ ). The between-subjects variable *button*—or whether listeners used their left or right button for the “same” response—had no significant effect on performance  $F(1,13) < 1$ .

Figure 15 displays cumulative  $d'$  values across the entire continua.

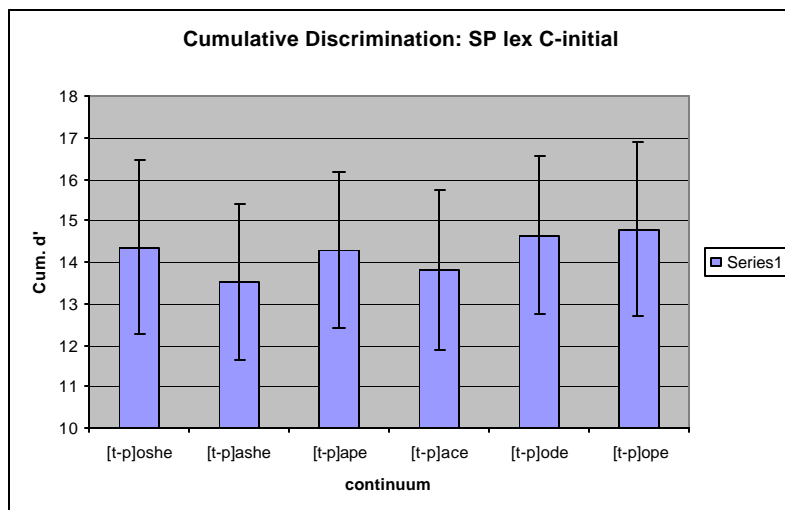


Figure 15. Mean cumulative  $d'$  values (95% confidence intervals) for control and lexical bias stems.

Listeners' sensitivity to stimulus differences was equally good for all stems, regardless of the lexical bias in the test continua relative to the controls. The six different stems do not differ from one another in their cumulative discriminability ( $F(5, 65)=1.459, p=.215$ ). *Button* had no significant effect ( $F(1, 13)=.135, p=.719$ ) and did not interact significantly with *stem*,  $F(5, 65)=1.175, p=.331$ .

### 4.3 Experiment 4 (Vowel Place - C initial)

#### 4.3.1 Stimuli and predictions

##### 4.3.1.1 Predictions

The stimulus continua were constructed to induce lexical and contrast biases on the judgments of vowel place following initial [t] and [p]. More "o" responses are expected after initial [t] than [p] if vowel percepts contrast with the stop's acoustics; lexical bias is expected to add to or hinder the visible effect of contrast. The control continua span non-word pairs, which should promote only contrast effects.

Bias	Vowel Place - C Final	
None (Control)	*teish - *toash	*peish - *poash
Cooperate	toad - *teid	pace - *poce
Conflict	tape - *tope	pope - *pape
Predictions % "p"	$\_e? > \_o?$ $\_es > \_ep, \_op > \_od$	

---

Table 4. Experiment 4 stimulus continua and predicted differences in “p” responses as a function of lexical and contrast biases.

#### 4.3.1.2 Construction

#### 4.3.2 Results

##### 4.3.2.1 Identification

Figure 16 displays the total proportion of “o” responses for the control stems.

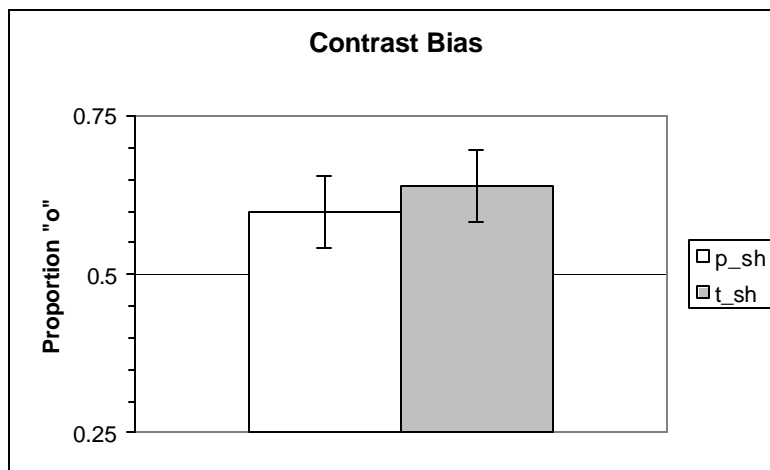


Figure 16: Mean total proportion of “o” responses (95% confidence intervals) as a function of the preceding stop’s place in control stems.

Listeners responded “o” more often after [t] than [p]. Although the difference was only marginally significant ( $F(1,17)=3.867, p=.066$ ), it indicates that the vowel percept contrasts with the preceding stop’s. *Button* had no significant effect ( $F<1$ ) and did not interact with the *consonant* effect ( $F<1$ ).

Figure 17 shows the total proportion of “o” responses for the lexical bias stems.

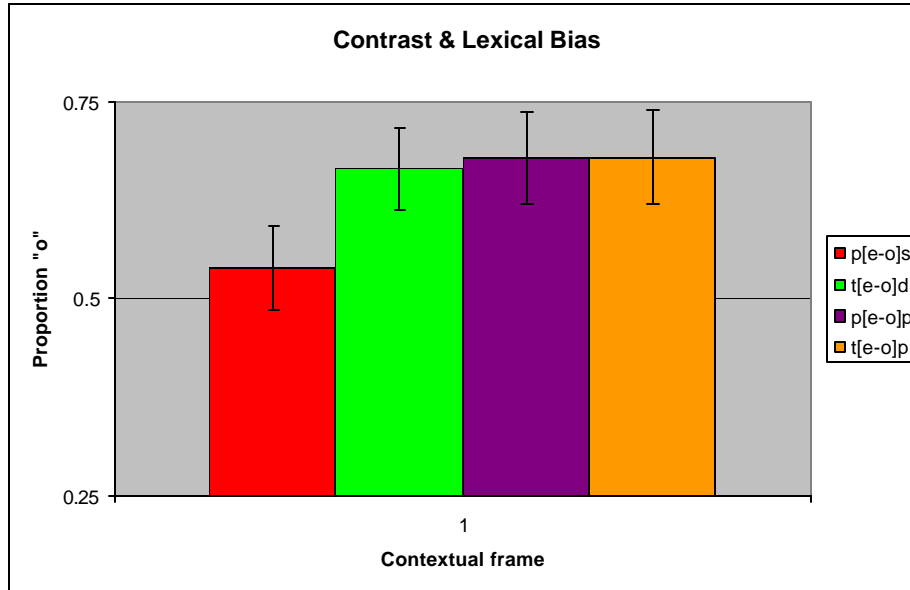


Figure 17: Mean total proportion of "o" responses (95% confidence intervals) as a function of the preceding stop's place and lexical biases in the lexical bias stems.

The figure shows that listeners responded "o" least often in the stem [p\_s], where both contrast and lexical biases cooperate to induce the fewest "o" responses, and that they responded "o" very frequently in the stem [t\_d] where the two biases should cooperate to induce the most "o" responses. However, listeners also responded "o" as frequently for the two conflicting stems, [t\_p] and [p\_p], as they did for the cooperating stem, [t\_d]. Apparently, either a contrast or lexical bias alone is sufficient to induce listeners to respond "o". An ANOVA yielded highly significant main effects of both *consonant* ( $F(1, 16)=36.017, p<.001$ ) and *lexical bias* ( $F(1, 16)=13.191, p=.002$ ). The effects of the two biases also interacted significantly with one another ( $F(1, 16)=17.914, p=.001$ ). This interaction reflects the absence of any difference in

the frequency of “o” responses between the cooperating stem [t\_d] and the two conflicting stems, [t\_p] and [p\_p], compared to the stems in which the two biases cooperate to suppress “o” responses. The between-subjects variable *button* was non-significant by itself  $F(1, 16)=.447, p=.513$ ), but interacted significantly with *consonant*,  $F(1, 16)=10.096, p=.006$ . Consonant had a greater effect for listeners who used their right button for their “o” response (9.7% difference in “o” response proportion between the t\_ and p\_ contexts, compared to 3.0% difference for listeners who used their left button for the “o” response.)

#### 4.3.2.2 Discrimination

Figure 18 displays the cumulative  $d'$  values within the [e]- and [o]-halves of the continuum for the control stems.

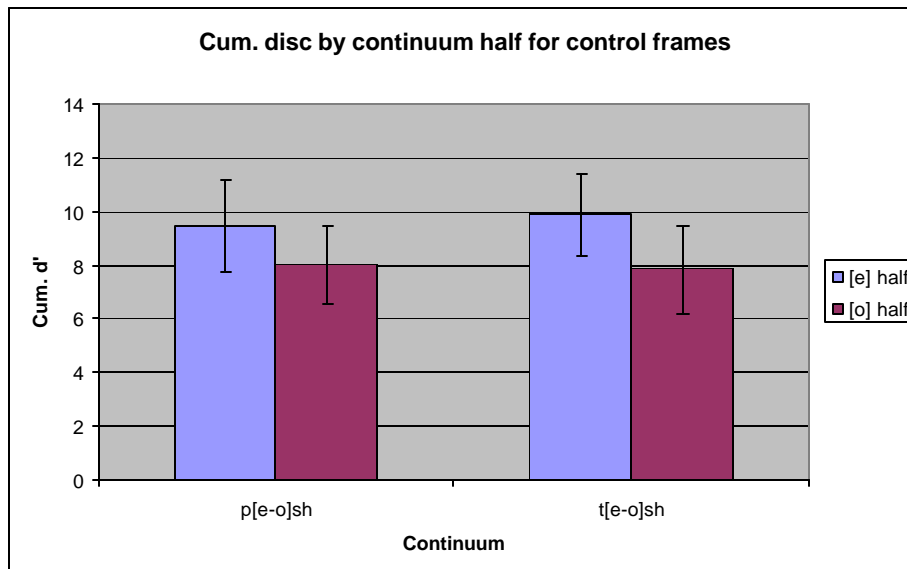


Figure 18. Mean cumulative  $d'$  values (95% confidence intervals) in the [t]- vs. [p]- halves of the continuum as a function of the preceding stop's place for the

control stems.

In both contexts, performance was better in the [e]- than the [o]-half of the continuum,  $F(1,12)=6.867$ ,  $p=.022$ . Though there was a slight trend toward better [e]-half performance after [t] relative to after [p], *consonant* did not reach significance and did not significantly interact with *continuum-half* (both  $F_s < 1$ ). The trend is expected if contrast with a preceding [t] shifts the category boundary toward the [e]-end of the continuum compared to a preceding [o]. *Button* had no significant effect alone ( $F < 1$ ) but interacted significantly with *continuum-half*,  $F(1,12)=3.749$ ,  $p=.077$ . As a whole, listeners who used their left button for the "same" response achieved an average cumulative  $d'$  of 7.09 in the [e]-half of the continuum and 10.17 in the [o]-half, while listeners who used the opposite button showed a much smaller difference between the [e]- and [o]-half: 8.77 and 9.24. No other interactions between variables reached significance, (all  $F_s < 1$ , except *continuum-half* by *button*,  $F(1,12)=3.749$ ,  $p=.077$ ).

Figure 19 shows discrimination performance in the two halves of the continuum for the lexical bias stems.

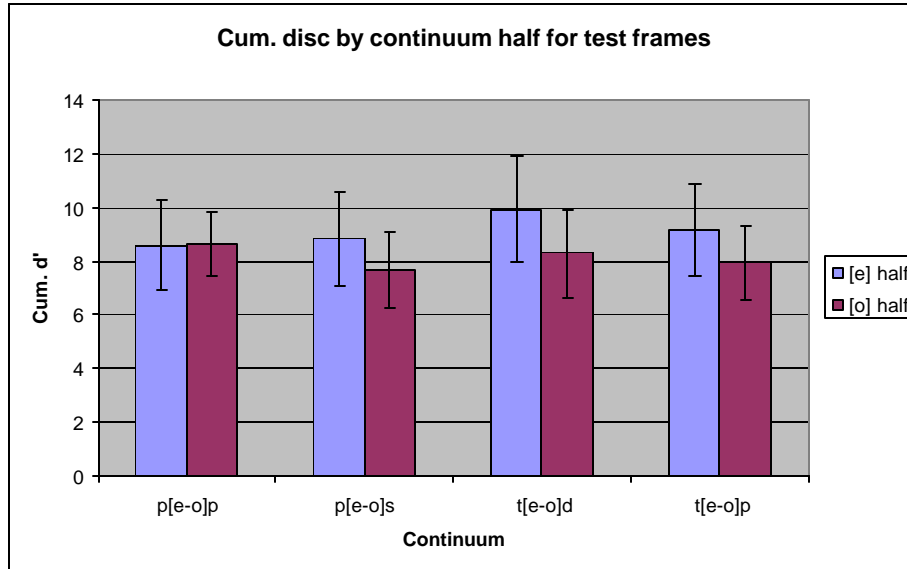


Figure 19. Mean cumulative  $d'$  values (95% confidence intervals) in the [t]- vs. [p]- halves of the continuum as a function of the preceding stop's place and lexical bias for the lexical bias stems.

The figure shows that performance is somewhat better in the [e]- than the [o]- half of the continuum for the two stems beginning with [t], as expected if contrast with the preceding stop makes the vowel sound more like [o]. However, stimuli are also more discriminable in the [e]- than the [o]- half of the continuum for the stem [p\_s], where neither the consonant nor lexical biases are expected to increase the number of "o" responses. Neither *continuum-half* ( $F(1,12)=2.395$ ,  $p=.148$ ), *consonant* ( $F(1,12)=1.894$ ,  $p=.130$ ), nor *lexical bias* ( $F(1,12)=2.736$ ,  $p=.124$ ) were significant on their own. *Continuum-half* and *consonant* interacted marginally ( $F(1,12)=3.496$ ,  $p=.086$ ), as did *continuum-half* and *lexical bias* ( $F(1,12)=4.043$ ,  $p=.067$ ). *Lexical bias* and *consonant* did not significantly interact ( $F<1$ ). The three-way interaction

between the within-subjects variables was not significant ( $F(1,12)<1$ ). The effect of *button* was not significant, and did not interact significantly with any other variable.

Mean cumulative discriminability scores across the entire continua are displayed in Figure 20.

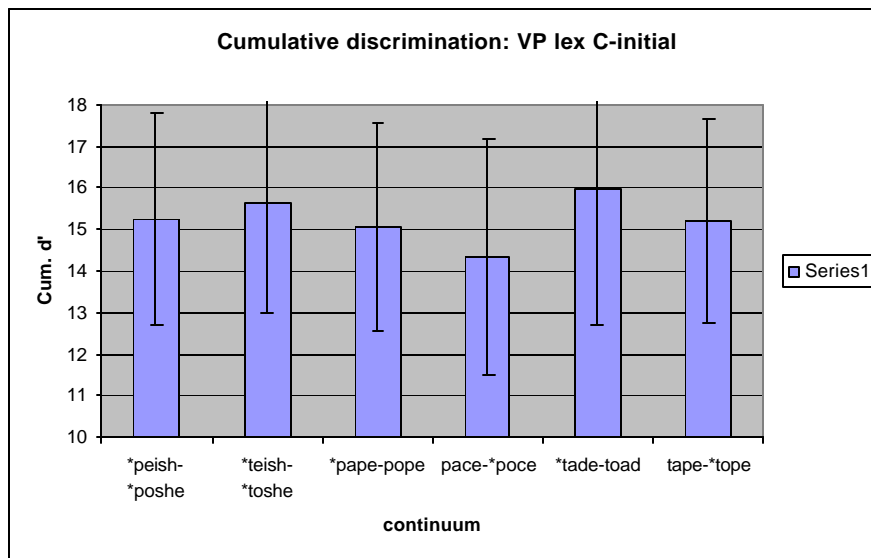


Figure 20. Mean cumulative  $d'$  values (95% confidence intervals) for control and lexical bias stems.

These values do not differ between stems in which there's a lexical bias versus those in which there is none, nor between stems with different consonant or lexical biases. A one-way repeated measures ANOVA reveals no significant effect of lexical status on cumulative discrimination in the different contexts  $F(5,60)<1$ . Listeners were no more sensitive to stimulus differences in the word-nonword continua than in the nonword-nonword control continua. *Button*

had no significant effect ( $F(1, 12)=.166, p=.691$ ), and did not interact significantly with *stem* ( $F(1, 12)=1.147, p>.3$ ).

#### **4.4 Summary of the results of Experiments 3 and 4**

In Experiment 3, listeners identified more of the stop place continuum as “p” before [o] than [e]. They discriminated stimuli better in the [t]- than the [p]-half of the continuum before [o] compared to [e], which indicates that the category boundary shifted toward [t] sooner before [o] compared to [e]. These results indicate that the percept of the stop’s place assimilated to the vowel’s. The opposite pattern of results was obtained in Experiment 4, where the roles of target and context were reversed: listeners responded “o” more often after [t] than [p] and discriminated stimuli better in the [e]- than [o]-half of the continuum after [t] compared to [p]. These results indicate that the percept of the vowel’s place contrasted with the stop’s. These experiments thus allow us to choose between the two alternative explanations of the difference between the results of Experiments 1 and 2; namely, that target percepts contrast with preceding contexts but assimilate to following ones. Segment class clearly does not matter because we obtained assimilation of consonants to vowels (Experiment 3) as well as vowels to consonants (Experiment 2), and contrast of vowels with consonants (Experiment 4) as well as of consonants with vowels (Experiment 1).

Experiments 3 and 4 also continued to demonstrate the independence of

contrast and assimilation biases from lexical biases, and they continued to show that a lexical bias does not sharpen listeners' sensitivity to stimulus differences as measured by the cumulative discriminability scores, even though lexical biases did alter which particular pairs of stimuli listeners found easiest to discriminate. Those shifts in local discriminability can be attributed to the same response biases that shift category boundaries toward the non-word end of the continuum in the identification tasks.

## **5. General Discussion**

### **5.1 Summary**

In four experiments reported here, listeners identified and discriminated members of stop and vowel place continua next to vowels or stops differing categorically in place. We expected the percepts of the members of the stop and vowel place continua to contrast with neighboring vowels or stops. Besides manipulating the target sounds' immediate segmental contexts, more distant portions of the strings were manipulated to create lexical biases. The expected lexical biases were obtained in all four experiments: listeners identified the stop or vowel more often as the category that made a word with its context than a non-word, and differences in the discriminability of particular stimulus pairs indicated that the same response biases also determined the listeners' performance in the discrimination tasks as well. However, we only obtained the expected contrast biases when the target segments followed their contexts (Experiments 1 and 4), and instead obtained assimilation biases when the

target preceded its context (Experiments 2 and 3). Both contrast and assimilation biases also altered which stimuli listeners found easiest to discriminate. All four experiments also indicated that the response biases induced by the target sounds' immediate segmental context were independent of the lexical biases, regardless of whether they were contrastive or assimilative. Finally, all four experiments showed that listeners' overall sensitivity to differences between adjacent pairs of stimuli was not better in stems for which there was a lexical bias than in those in which there was none.

## **5.2 Autonomy and auditory processing**

At the beginning of this paper, we hypothesized the existence of an initial, linguistically naïve auditory stage of processing that receives no feedback from higher, linguistically-informed levels of processing, and which exaggerates the perceived values of neighboring acoustic intervals that differ spectrally. The current experiments show that the influence of linguistic knowledge on perceptual processing is limited to biasing responses and does not enhance sensitivity across the continuum. Discrimination was often better in the region of the continuum into which lexical biases pushed the continuum category boundary, but lexical influence did not globally improve listeners' sensitivity to differences within word-nonword continua compared to word-word or non-word-non-word continua. The failure to find better cumulative discriminability in word-nonword continua compared to continua where the endpoints are both words or nonwords in all of our four experiments disconfirms a positive

prediction of interactive models of speech sound perception such as TRACE. Our results also demonstrate that the perceptual effect of a target sound's immediate segmental context can, when the target follows the context, plausibly be the product of auditory contrast. Taken together, these findings make a strong prima facie for the autonomy of an initial level of processing from a later stage when feedback from the lexicon does bias responses, and replicate a traditional contrastive contextual effect that is interpretable as auditory in nature.

### **5.3 Limits to evidence of autonomy**

Disconfirming a positive prediction of an interactive model of speech perception could be interpreted as support for an autonomous model, if interaction and autonomy are dichotomous alternatives that divide the universe of hypotheses completely between themselves. Nonetheless, finding positive evidence for a prediction of an autonomous model would strengthen our claim. Ideally, we would complement the empirical failure of linguistic knowledge to affect an earlier stage of processing with a successful demonstration that the contrast arises during an initial, linguistically naïve auditory stage of speech perception. Unlike lexical biases, sequential contrast should improve listeners' sensitivity to stimulus differences when neighboring acoustic intervals differ in their value for some acoustic dimension by exaggerating the perceived differences in value.

If such exaggeration occurs, then a sequence in which successive intervals have high and then low values for some dimension should be more discriminable from a low-high sequence than a high-high sequence is from a low-low one. The latter pair should be less discriminable because successive intervals in a sequence do not differ in value for the dimension and therefore do not contrast sequentially with one another. Most specifically, we expect that a *heap:hoot*-like pair in which the concentrations of energy in the vowels' and stops' spectra are high-low versus low-high would be more discriminable than a *heat:hoop*-like pair, in which they are instead high-high versus low-low.<sup>9</sup> Such a difference in discriminability would only arise if the discrimination task taps the initial auditory stage of processing before the intervals are categorized. After categorization of the two intervals in each sequence, the high-low versus low-high pair would be no more distinct than the high-high versus low-low pair as the categories would differ in each interval.<sup>10</sup> This prediction will be tested in a future set of experiments.

The claim that the perceptual interaction between the target sound and its immediately segmental context is a byproduct of a contrastive auditory

---

<sup>9</sup> We say "*heap-hoot*-like" and "*heat-hoop*-like" because the stimuli in a pair would be perfectly discriminable, and therefore uninformative, if they were as distinct as the words *heap*, *hoot*, *heat*, and *hoop*.

<sup>10</sup> Stephens & Holt (2003) made precisely that demonstration, justifying our gusto, but unfortunately for us they did not use our stimuli. See also Kingston (2005) for results compatible with this claim.

transformation is independent from the claim that the initial stage of processing is autonomous from later stages and linguistically naive rather than linguistically informed. The objects of perception during the initial stage of evaluation could be articulatory gestures rather than auditory qualities, as in the direct realist account of speech perception (Fowler, 1986), yet the interaction between one articulatory object and its neighbor could be as autonomous from feedback from lexical or any other kind of linguistic knowledge as an initial *auditory* stage. Because of that possibility, we are seeking evidence in favor of two, logically independent hypotheses. First, that the initial evaluation of a speech signal is not informed by the listeners' knowledge of the properties of their native language. Second, that the properties processed during that initial stage are auditory rather than articulatory. The four experiments reported here begin the accumulation of evidence in support of the autonomy hypothesis. The results of yet further experiments will extend the evidence in favor of autonomy, while others will be designed to determine whether the perceptual interactions between neighboring segments are auditory rather than articulatory. Even if those experiments show that the objects of speech perception are articulatory rather than auditory, the finding would not assail the independent demonstration of autonomy.

ACTUALLY: IF AUTONOMY IS CORRECT, AND DISCRIMINATION TAPS AN AUDITORY STAGE, SHOULDN'T MISPARSING NOT MATTER? THE DESIGN STILL SEEMS TO MAKE SENSE, EVEN IF CONTRAST ONLY WORKS FORWARD, does THE

RATIONALE FOR THE EXPERIMENT STILL HOLDS.(THIS highlight comment has been displaced but still relevant.)

#### **5.4 Contrast vs assimilation**

In identification, listeners more often gave the responses that corresponded to the spectrally low category when preceding intervals were spectrally high, and vice versa. Unexpectedly, listeners instead gave the response corresponding to the spectrally low category more often when the following interval was spectrally low than when it was high. As noted above, the first pattern is contrastive, the second pattern is assimilative. Previous literature has provided some grounds for limiting [auditory?] contrast to a forward direction, and in other cases has found both contrastive and assimilative effects of context on preceding targets.

Holt (2005) argues for a specific neural process, a criterion shift induced by stimulus-specific adaptation, as the mechanism behind auditory contrast. Although Holt does not explicitly address the issue of order of target and context, the mechanism she proposes should, by definition, restrict contrast to a forward direction, or to cases in which a context affects a following target.

Wade & Holt (2005) provides a precedent for the backward assimilation we found. When a target stop drawn from [d-g] continuum was closely by followed a pure tone differing in frequency, listeners responded with spectrally

high “d” more often when the following tone was high than when it was low. But, when the distance between consonant and tone was increased to 100 ms, a contrastive effect emerged, more “d” responses before the following low tone.

Wade & Holt suggest that when the context follows a target very closely, listeners may use information about it as information about the target. In other words, listeners may parse the stimuli in such a way that their judgments of preceding targets actually refer partly to the following context’s properties. From that perspective, it makes sense that the likelihood of the context intruding upon the locus of judgment decreases as the distance between context and target increases. In line with Wade & Holt (2005), we assume that listeners’ judgments that assimilate to a following context reflect the inclusion of the context’s information. We dub the assimilative pattern obtained in Experiments 2 and 3 “misparsing” because listeners do not judge our intended target in strict isolation from the context, but instead use a locus of judgment that spans at least part of both intervals.<sup>11</sup> Is there more to discuss?

---

<sup>11</sup> Misparsing itself is not a new novel notion. Ohala (1981) argues that misattribution of acoustic cues is a common source of sound changes. Other recent works argue that some phonological patterns owe their existence to blurrieness of sonorant boundaries. Kavitskaya (2002) argues that compensatory lengthening is due in part to misparsing of sonorant portions as belonging to the preceding vowels. Downing (to appear) partly attributes vowel lengthening before NC clusters found in Bantu languages to misparsing of nasal portions as the preceding vowels. Finally, Myers and Hanssen (2006) claim that lengthening of vowels in post-vocoid positions found in Bantu languages is due to misparsing of transitional portions as a part of the vowels.

Despite our results and those of Wade & Holt (2005), some previous literature has reported results in which a target appears to contrast perceptually with a following context (e.g. Mann & Repp, 1980; Fowler, 1984; Mann & Soli, 1991; Ohala & Feder, 1994; Mitterer, in press). Therefore, any suggestion that the mechanism responsible for contrastive response shifts would only work in a forward direction has fed into criticisms of auditory contrast as a theory. A forward-only mechanism would limit the explanatory power of auditory contrast by excluding it from consideration in the cases where context exerts a contrastive effect backward (Fowler 2006). Fowler (2006) points out that the competing theory, compensation for coarticulation, shares no such limitation in the direction of its predicted effects: listeners should compensate for coarticulation with a following sound as much as with a preceding one. But that supposed advantage for compensation for coarticulation accounts depends on one particular formulation of auditory contrast. If instead of shifting percepts because of passive neural adaptation (offered by Holt 2005) the perceptual apparatus exaggerates the perceived differences in value for some dimension between neighboring intervals, there is no reason why it should not operate backward as readily as forward.

In any case, our results support neither this bidirectional theory of contrast as exaggeration of differences nor compensation for coarticulation in this respect, since contrast was, in fact, limited by direction.

If the effect of a preceding context at least is contrastive and that contrast arises during auditory processing, then the misparsing of acoustic material from a following context as information about the current interval probably cannot be auditory, too, it seems infeasible that auditory processing could produce conflicting effects. In that case, misparsing would have to be a post-perceptual process, and would thus take place during the same stage of processing as feedback from lexical and other linguistic knowledge does. This speculation indicates the second line of research that we plan to take up in investigating these phenomena further.

In identification, listeners showed a tendency toward spectrally -low choices in the context of preceding intervals that were spectrally -high, and vice versa. But more surprisingly given our hypotheses, listeners gave assimilative judgments—or identified target choices that were spectrally more *similar* to the context—when the target preceded the context. The former contrastive pattern is consistent with the notion that the auditory system exaggerates the perceived difference between the acoustic values of two neighboring intervals that differ in a certain dimension. The latter, assimilative pattern seems to demand a post-auditory explanation since the auditory apparatus could not (?) feasibly enact both contrast and assimilation.

In line with Wade & Holt (2005), we assume listeners' judgments that assimilate to following context reflect the inclusion of the context's information. In other words they do not judge our intended target in strict isolation from the context, but instead use a locus of judgment that spans at least part of both intervals. We obtained assimilative judgments of members of an [i]-[u] continuum that preceded a final [t] or [p] context even in a follow-up experiment<sup>12</sup> using stimuli that had no formant transitions from the vowel to the consonant closure, leaving only a flat vowel steady state before the final stop context.

When we obtained our first assimilative result in Experiment 2 (VP-Lex

---

<sup>12</sup> The no-transition follow-up matched Experiment 2 VP-lex G-final in every respect except that we removed the transitions to the [t] and [p] closure in the vowel. We had previously considered that listener's may have used later information in the [i] and [u]-like targets to determine their choices, a locus which would necessarily include a portion whose F2 values approached—or assimilated to—those of the following burst.

**C-final**), we considered that listeners may have used later information in the [i]- and [u]-like target intervals to determine their choices. That locus would have necessarily included a portion of the vowel whose F2 values approached—or assimilated to—those of the following [p] or [t] burst. We therefore designed a “no-transition” follow-up that matched Experiment 2 VP-lex C-final in every detail except that we removed the vowels’ transitions to the [t] and [p] closures. The stimuli contained only a flat-contour steady-state vowel interval next to the silent closure and final stop burst. But even without transitions that colored the vowels with properties of the stop, listeners still judged the [i]-[u] continuum-members in a way that assimilated to the following [t] or [p] stops.

Ganong, W. F. III 1980 Phonetic categorization in auditory word perception, *J. Experimental Psychology: Human Perception and Performance*, 6, 110-125.

Pitt, M. A., & Samuel, A. G. 1993 An empirical and meta-analytic evaluation of the phoneme identification task, *J. Experimental Psychology: Human Perception and Performance*, 19, 699-725.

Pitt, M. A., & McQueen, J. M. 1998 Is compensation for coarticulation mediated by the lexicon? *J. Memory and Language*, 39, 347-370.

Massaro, D., & Cohen, M. 1983 Phonological context in speech perception, *Perception & Psychophysics*, 34, 338-348.

Moreton, Elio. 2002 Structural constraints in the perception of English stop-sonorant clusters, *Cognition*, 84, 55-71.

Mann, V. A. 1980 Influence of preceding liquid on stop-consonant perception, *Perception & Psychophysics*, 28, 407-412.

Mann, V. A., & Repp, B. H. 1981 Influence of preceding fricative on stop consonant perception, *J. Acoustical Society of America*, 69, 548-558.

Lotto, A. J., & Kluender, K. R. 1998 General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification, *Perception & Psychophysics*, 60, 602-619.

Lotto, A. J. & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler

(2006), *Perception and Psychophysics*, 68, 178-183.

McClelland J. L., Elman J. L. 1986 The TRACE model of speech perception, *Cognitive Psychology*, 18, 1-86.

McClelland, J.L., Mirman, D., & Holt, L.L. (2006). Are there interactive processes in speech perception? *TRENDS in Cognitive Sciences*, 10, 363-369.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-370.

Kingston, J. (2005) From ears to categories: New arguments for autonomy. In S. Frota, M. Vigario, & M. J. Freitas (Eds.), *Proceedings of the first conference on phonetics and phonology in Iberia*. Berlin: Mouton de Gruyter.

Fowler, C. A. 1990 Sound-producing sources as the objects of perception: Rate normalization and nonspeech perception, *J. Acoustical Society of America*, 88, 1236-1249.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast, *Perception and Psychophysics*, 68, 161-177.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans, *Journal of Experimental Psychology: Human Perception and Performance*, 26, 877-888.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of Acoustical Society of America*, 108, 710-722.

Coady, J. A., Kluender, K. R., & Rhode, W. S. 2003 Effects of contrast between onsets of speech and other complex spectra, *J. Acoustical Society of America*, 114, 2225-2235.

Lotto, A. J., Kluender, K. R. & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*), *Journal of Acoustical Society of America*, 113, 53-56.

**Holt 2000, Holt 2005, Holt 2006**

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.5.16) [Computer program]. Retrieved February 18, 2007, from <http://www.praat.org/>

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, 87, 820-857.

Mann & Repp, 1980; Fowler, 1984; Mann & Soli, 1991; Ohala & Feder, 1994; Mitterer, in press

Wade & holt 2005

Stephens & Holt (2003)

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer.

Coady, J. A., Kluender, K., & Rhode, W. S. (2003). Effects of contrast between onsets of speech and other complex spectra. *Journal of Acoustical Society of America*, 114, 2225-2235.

Downing, L. (to appear). On the ambiguous segmental status of nasal in homorganic NC sequences. In M. van Oostendorp & J. M. van der Weijer (Eds.), *The Internal*

- Organization of Phonological Segments*). Berlin: Mouton de Gruyter.
- Fowler, C. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. (1990). Sound-producing sources as the objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, 88, 1236-1249.
- Fowler, C. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception and Psychophysics*, 68, 161-177.
- Fowler, C., Brown, J. M., & Mann, V. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 877-888.
- Ganong, W. F. I. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.
- Holt, L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science*, 16, 305-312.
- Holt, L. (2006). The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization. *Journal of Acoustical Society of America*, 120, 2801-2817.
- Holt, L., Lotto, A., & Kluender, K. (2000). Neighboring spectral content influences vowel identification. *Journal of Acoustical Society of America*, 108, 710-722.
- Kavitskaya, D. (2002). *Compensatory Lengthening: Phonetics, Phonology, Diachrony*. New York: Routledge.
- Kingston, J. (2005). From ears to categories: New arguments for autonomy. In S. Frota, M. Vigarío & M. J. Freitas (Eds.), *Proceedings of the First Conference on Phonetics and Phonology in Iberia*. Berlin: Mouton de Gruyter.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lotto, A., & Holt, L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception and Psychophysics*, 68, 178-183.
- Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, 60(4), 602-619.
- Lotto, A., Kluender, K., & Holt, L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of Acoustical Society of America*, 113, 53-56.
- Mann, V. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, 28, 407-412.
- Mann, V., & Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *Journal of Acoustical Society of America*, 69, 548-558.
- Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Perception and Psycholinguistics*, 34, 338-348.
- McClelland, J., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

- Mirman, D., McClelland, J., & Holt, L. (2006). Are there interactive processes in speech perception? *TRENDS in Cognitive Sciences*, 10, 363-369.
- Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, 84, 55-71.
- Myers, S., & Hanssen, B. (2006). The origin of vowel-length neutralization in vocoid sequences. *Phonology*, 22, 317-344.
- Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-370.
- Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.
- Pitt, M., & Samuel, A. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 699-725.