Chapter 4

Perception of Intervocalic Voicing and Spirantization

This chapter reports the results of three experiments that were designed to explore the possible role of perception in the lenition of intervocalic stops. As discussed in the introduction, the potential effect of perceptual factors on lenition has received little attention in the literature, Kingston (2008) being a recent exception. The particular perceptual model that I examine is the P-map (Steriade 2001a,b).

Recall the basic typological fact I intend to account for – that spirantization of intervocalic voiceless and voiced stops (as in (1)) and voicing of intervocalic voiceless stops (as in (2a)) are attested, while devoicing of intervocalic voiced stops (as in (2b)) is not.

(1) a. **Attested:** Intervocalic voiceless stops targeted for spirantization
(e.g., Tiberian Hebrew)
(i) /VpV/ → [VɸV]

(ii) /VtV/ → [VθV]

(iii) /VkV/ → [VxV]

b. **Attested:** Intervocalic voiced stops targeted for spirantization *(e.g., Spanish)*

(i) /VbV/ → [VβV]

(ii) /VdV/ → [VðV]

(iii) /VgV/ → [VɣV]

(2) a. **Attested:** Intervocalic voiceless stops targeted for voicing *(e.g., Warndarang)*

(i) /VpV/ → [VbV]

(ii) /VtV/ → [VdV]

(iii) /VkV/ → [VgV]

b. **Unattested:** Intervocalic voiced stops never targeted for devoicing

(i) */VbV/ → [VpV]

(ii) */VdV/ → [VtV]

(iii) */VgV/ → [VkV]

The P-map (Steriade 2001a,b) builds on the ideas of Licensing by Cue to provide a framework for a perceptually-based understanding of typological gaps such as the absence of patterns like (2b). The core intuition of Steriade's proposal is that "[t]he aim, in any departure from the UR, is to change it *minimally* to achieve compliance with the phonotactics" (Steriade 2001a, 4). She formalizes this notion of minimality in terms of perceptibility: form A is 'closer' to form B than form C is if the A ∼ B distinction is perceptually less salient than the C ∼ B distinction, where perceptual salience is defined as mutual confusability. Knowledge of the relative perceptibility of various contrasts (however it may be manifested in actual listeners) is known as the P-map.

For example, final voiced stops are often targeted for devoicing, but never for being turned into sonorants. Steriade proposes that the latter repair is never employed because the contrast between voiced and voiceless obstruents is less perceptible in the environment V___# than the contrast between voiced obstruents and sonorants. In other words, final voiced obstruents undergo the 'smallest' change possible, where the size of a change is defined in perceptual terms. The P-map has also been used to explain phenomena such as asymmetries in consonant assimilation and the types of segments that are epenthesized (Steriade 2001b), the behavior of voicing in singleton and geminate stops (Kawahara 2006), and laryngeal co-occurrence restrictions (Gallagher 2009).

Under this approach, the explanation for the absence of intervocalic devoicing would be that devoicing is a more perceptible repair to intervocalic voiced stops than spirantization. The results of Experiments 2 – 4 have implications both for the P-map and for the traditional articulatory understanding of lenition. To the extent that the results allow the P-map to make the desired predictions, we have evidence that perception by itself is enough to account for the relevant typological patterns; it then becomes superfluous to invoke articulatory effort as an additional explanation for the same facts in the absence of more direct evidence that effort is involved. Although the sufficiency of a perceptual explanation does not completely rule out a role for articulation since a typological pattern may have multiple overlapping causes, it does mean that the purported role of articulation must be more thoroughly tested (as in chapter 3). In addition, such a result constitutes evidence in support of the P-map itself as an approach to explaining typological patterns (although not to the exclusion of other perceptual approaches; there are other models of phonological patterns that can achieve similar results).

On the other hand, to the extent that the results do *not* provide a perceptual explanation along the lines of the P-map for the relevant typological facts, we have evidence that other influences must be at work. One notable example is the fact that the P-map is meant to explain not *why* a given configuration is changed – in Optimality-Theoretic terms (Prince and Smolensky 2004[1993]), this is the role of markedness constraints – but rather *how* it changes (the role of faithfulness constraints). Thus, perception may not tell us anything at all about whatever markedness constraint drives languages to lenite in the first place; I return to this point in §5.1.

As the results of Experiment 2 show, intervocalic devoicing is a *more* perceptible change than intervocalic spirantization; by contrast, Experiment 4 shows that spirantization and voicing of voiceless stops are about equally perceptible. Therefore, the approach of the P-map seems to be on the right track in explaining the broad typology of (1) and (2). However, we will see from the results of Experiment 3 that for voiced stops, the perceptual facts differ by place of articulation in ways that do not line up neatly with the typology discussed in §2.2. Thus, while perceptual facts may be able to explain the broad outlines of intervocalic lenition, there must be other factors at work as well.

4.1 Experiment 2: Relative Perceptibility of Devoicing and Spirantization for Voiced Stops

Experiment 2 was designed to test the relative perceptibility of two logically possible repairs for intervocalic voiced stops: devoicing and spirantization. The experiment compares voiced stops at each of the three major places of articulation ([b], [d], [g]) in terms of mutual confusability with their voiceless counterparts on the one hand and spirant counterparts on the other, with the goal of determining which series is more confusable with voiced stops.

4.1.1 Design

Recording of Stimuli

Table 4.1: Perceptibility comparisons in Experiments 1 and 2

	[+voi] Spirants		**[+voi] Stops**		**[−voi] Stops**
Labials	β	∼	b	∼	p
Coronals	ð	∼	d	∼	t
Dorsals	ɣ	∼	g	∼	k
Cover Symbol	**Z**	∼	**D**	∼	**T**

The stimuli for the experiment consisted of each of the nine consonants listed in table 4.1 recorded in the environment [a___a]. Tokens were recorded by five talkers: two native speakers of Spanish (talkers 4 and 5) and three native speakers of English (talkers 1 – 3). Spanish speakers were used because Spanish has a variant of the spirantization pattern; thus, their productions of the [aZa] tokens should accurately reflect the pronunciation of lenited stops in at least one language with

Table 4.2: Elicitation of stimuli from Spanish and English talkers for Experiments 2 and 3

Stimulus	**Orthography, Block**	
	Spanish Speakers	**English Speakers**
[aba]	aba, 2	aba, 1
[ada]	ada, 2	ada, 1
[aga]	aga, 2	aga, 1
[apa]	apa, 1	apa, 1/2
[ata]	ata, 1	ata, 1/2
[aka]	aka, 1	aka, 1/2
[aspa]		aspa, 1
[asta]		asta, 1
[aska]		aska, 1
[aβa]	aba, 1	aβa, 2
[aða]	ada, 1	aða, 1
[aɣa]	aga, 1	aɣa, 2

this pattern, whether they are approximants or true fricatives. English speakers were used because (unlenited) intervocalic voiced stops are phonotactically legal in English, but not in Spanish.

The native Spanish speakers were adult L2 speakers of English who were naïve to the purposes of the experiment. They recorded the stimuli in two blocks. The first block consisted of the stimulus items [aTa] and [aZa], which are phonotactically legal in Spanish, in standard Spanish orthography (see table 4.2). Stimuli were presented to the talkers in a randomized block, with each stimulus presented 20 times. Each token was printed on a separate square of paper; talkers worked through the stack of paper at their own pace, reading each token with initial stress.[1] The second block consisted of the [aDa] stimulus items, which are

---

[1] Initial stress was used rather than final stress so that the segments of interest would not occur in the onset of a stressed syllable, a canonically 'strong' position (Beckman 2004[1998]; Smith 2004, 1441) that is expected to resist lenition. Although the Spanish spirantization process is

phonotactically illegal in Spanish (Spanish has spirantization of voiced stops intervocalically, among other environments). For this block, talkers were instructed to pronounce the consonants as they would be pronounced in English (i.e., as stops); the talkers and the experimenter discussed how the English and Spanish pronunciations differ to ensure that the talkers understood what was being asked of them.

The native English speakers were linguistically trained American students in their 20s who were naïve to the purposes of the experiment. Again, the stimuli were recorded in two blocks. The first block consisted of the stimulus items [aTa], [aDa], and [aða], which are phonotactically legal in English (but not [aβa] or [aɣa], which are not). Subjects were requested to avoid flapping in [ata] and [ada]; since flapping is optional, these pronunciations were still phonotactically legal, although perhaps in a more formal register. Stimuli were presented in IPA (see table 4.2). The first block also contained the additional stimulus items [asTa], which were recorded in hopes of obtaining reduced aspiration on the voiceless stops. Stimuli were presented and recorded as described above. The second block consisted of the stimuli [aβa] and [aɣa], which contain segments absent from the English inventory; in addition, the [aTa] stimuli were presented again, and talkers were asked to avoid aspirating the voiceless stops.

Selection of Stimuli

Naturally produced stimuli were used because the purpose of this experiment is to determine how well listeners can distinguish between the relevant sounds, not to identify what cues they use to do so. However, there are three ways in

---

not sensitive to stress, other lenition processes are (such as English flapping).

which the production of these stimuli might bias the results of the experiment:

1. English voiced and voiceless stops contrast in aspiration, and not merely voicing. This property is especially pronounced at the beginning of stressed syllables, but seems to be present even intervocalically (see figure 4.1 below). Thus, the voiced and voiceless stops produced by the English talkers might be easier to discriminate than the voicing contrast in a language that does not use aspiration to the same extent.
2. In producing the [aβa] and [aɣa] stimuli, the English speakers tended to produce relatively long initial vowels (see figures 4.2 and 4.8), presumably because these stimuli were non-native and required extra attention. This property of those talkers' stimuli could be used by subjects as a cue to the Z $\sim$ D distinction that may not be found in natural speech, thus artificially increasing the salience of that distinction. This length difference could also interfere with subjects' perception of the T $\sim$ D distinction, where length of the preceding vowel is a common cue for voicing (see Kingston and Diehl (1994) and references therein).
3. As noted above, for all of the talkers, some of the stimuli involved non-native segments, phonotactics, or both. It is possible that the vowels in these non-native stimuli were distorted, thus providing an additional (artificial) cue to the relevant distinctions.

Of the 20 tokens of each stimulus produced by each talker, approximately 10 tokens were selected for use in the experiment. The number of tokens that were selected depended on the quality of the tokens and ranged between 6 and

13; in most cases (33 out of 45), the number was between 8 and 10. Tokens were analyzed in Praat (Boersma and Weenink 2007) and selected so as to maximize the naturalness of the tokens and minimize the potential confounds 1 and 3 discussed above, as follows:

1. English speakers' tokens of [aTa] were selected from the ordinary [aTa] stimuli in the first block. The [asTa] tokens were not used because of the effects of the coronal [s] on the formants of the first vowel. The unaspirated [aTa] tokens were not used because they were difficult for the English speakers to produce naturally.

2. All tokens with any obvious abnormality were excluded (e.g., tokens with stress on the final syllable, closure during a spirant, lack of closure during a stop, and so on).

3. All [aZa] tokens (of both English and Spanish speakers) were rated for naturalness by two native speakers of Spanish (not the same Spanish speakers who recorded stimuli). These speakers were asked to evaluate the tokens as Spanish nonsense words, paying attention only to the consonant. For each talker and place of articulation, the tokens with the highest ratings from both raters were selected.

4. For the English speakers, the tokens of the [aTa] stimuli were selected that had the shortest period of aspiration after the stop.

5. One of the Spanish speakers (talker 4 in the graphs below) produced tokens of the [aTa] stimuli with unusually long stop closures. For this talker, the [aTa] tokens with the shortest closures were chosen. [aTa] tokens for

the other Spanish speaker (talker 5) were selected like the [aDa] tokens, as described below.

6. After filtering by these criteria, the F1 and F2 values of the vowels of the remaining stimuli were measured. For each talker and place of articulation, the [aDa] tokens were chosen that were closest to the [aZa] and [aTa] tokens for that talker and place of articulation (as measured by the formants of the first vowel). The goal of this procedure was to ensure that the vowels of the [aDa] tokens were not systematically more similar to those of the [aZa] tokens than the [aTa] tokens, or vice versa. Thus, we can be reasonably sure that listeners' sensitivity to the relevant pairs is truly grounded in differences among the consonants, not accidental differences among the vowels.[2]

Selected tokens were trimmed, leaving 200 ms of silence on either side of the word, and amplitude-normalized using a Praat script that set the peak of each file at .8. The acoustic characteristics of the selected tokens, as relevant to the possible confounds described above, are given below with the results of the experiment.

Participants

24 native speakers of English participated in the experiment; all were students in undergraduate linguistics courses at the University of California, Santa Cruz. Subjects received either monetary compensation or extra credit, depending on the course. Subjects ranged in age from 18 to 23 years old and none were native speakers of a language other than English.

---

[2] Of course, some of the cues to the voicing and continuancy distinctions of interest are found in the vowels as well as the consonants. Since the formant values were measured at the midpoints of the vowels, it is reasonable to assume that they were affected by the adjacent consonants as little as possible.

Procedure

All instructions and stimuli were presented, and subjects' responses recorded, using SuperLab 4.0.5. The experiment involved a 'same'-'different' task: subjects were presented with a pair of tokens and asked to indicate whether they had heard two different words or two repetitions of the same word. The stimulus presentation was purely auditory and subjects responded via colored buttons on a button box (Cedrus Response Pad, model RB-620); thus, the stimuli were not represented to the subjects orthographically in any way during the experiment. This avoidance of written representations was intentional; in English, for example, [ɡ] and [k] have standard orthographic representations that subjects would be familiar with, while [ɣ] does not. Subjects were told that the stimuli were "not words from any particular language"; they were also informed that the vowels were the same in all of the words and were instructed to pay attention only to the consonants.

All pairs of tokens that were presented were within-talker and within-place. Within each condition, there were three types of 'same' trials (Z $\sim$ Z, D $\sim$ D, and T $\sim$ T) and four types of 'different' trials (Z $\sim$ D, D $\sim$ Z, T $\sim$ D, and D $\sim$ T). Each token appeared in one 'same' trial and one 'different' trial (two 'different' trials in the case of the [aDa] tokens), with some repetitions when a given talker had more tokens for one of the stimuli paired in a 'different' trial than the other; however, each combination of tokens was presented only once. 'Same' trials were pairs of distinct tokens, never two repetitions of the same token. In total, there were approximately 70-80 trials for each pairing of talker and place of articulation, depending on how many tokens for that talker and place were available.

The first part of the experiment consisted of a practice session, during which subjects were trained on the relevant contrasts. After each trial, subjects were told whether they had responded correctly. The practice session ended after subjects had heard at least 10 trials at each place of articulation and after either they had answered 8 out of 10 trials in a row correctly, or after 5 minutes, whichever came first.

The main part of the experiment was presented in four blocks. The first block consisted of all trials. The second block consisted of only the labial trials, the third block of only the coronal trials, and the fourth block of only the dorsal trials. Subjects were given no feedback during these blocks. For each trial, both the response and the reaction time[3] were recorded. Subjects were given the opportunity to take a break before each block. The entire experiment lasted between 45 minutes and an hour.

4.1.2 Results

Acoustic Properties of the Stimuli

Figure 4.1 shows the duration of aspiration in the [aTa] stimuli. Unsurprisingly, the tokens produced by English talkers that were used in the experiment had significantly more aspiration (average 68 ms) than the tokens produced by Spanish talkers (average 23 ms; $p = 4.5 \times 10^{-49}$). In addition, the tokens produced by English talkers that were selected for use in the experiment had significantly less aspiration than all of the tokens that were actually produced by English talkers

---

[3] Reaction times were recorded from the onset of the second stimulus, then recalculated from the end of the second stimulus (including the 200 ms of silence). Responses that occurred before the end of the second stimulus are omitted; these trials account for less than .2% of all trials in the experiment. Measuring reaction times from the end of the consonant produces almost identical results, as discussed with the results below.

(average 76 ms; $p = 6.8 \times 10^{-5}$).

Figure 4.1: Density curves for aspiration duration in [aTa] stimuli for Experiments 2 and 3

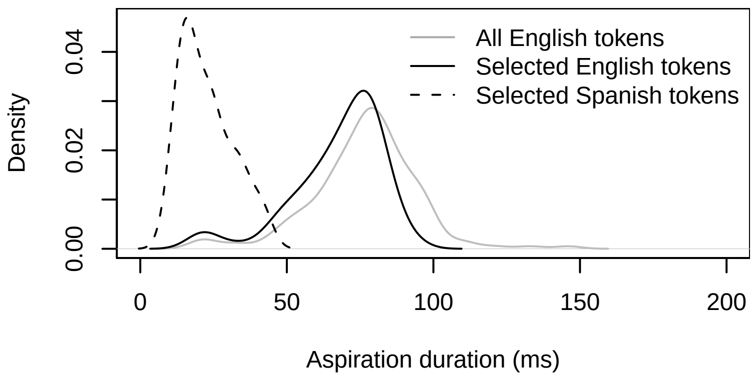


Figure 4.2: Length of first vowel in stimuli for Experiments 2 and 3. Stars mark within-language differences that are significant at $\alpha = .05$

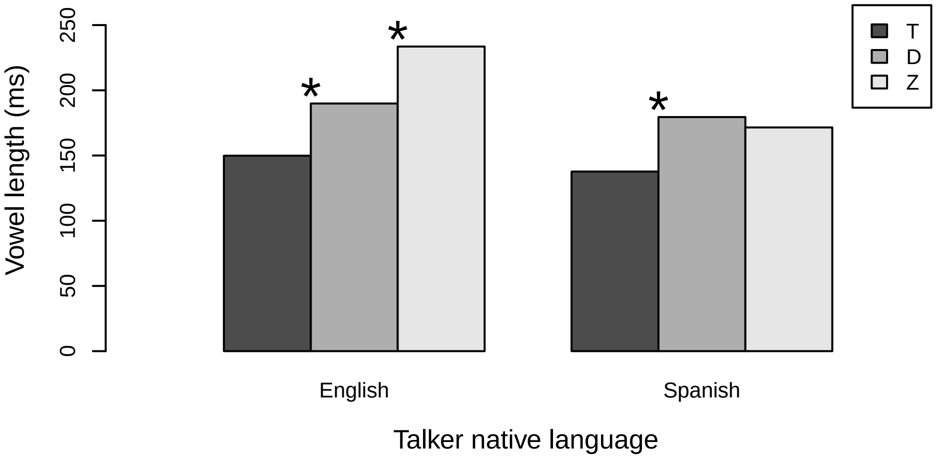
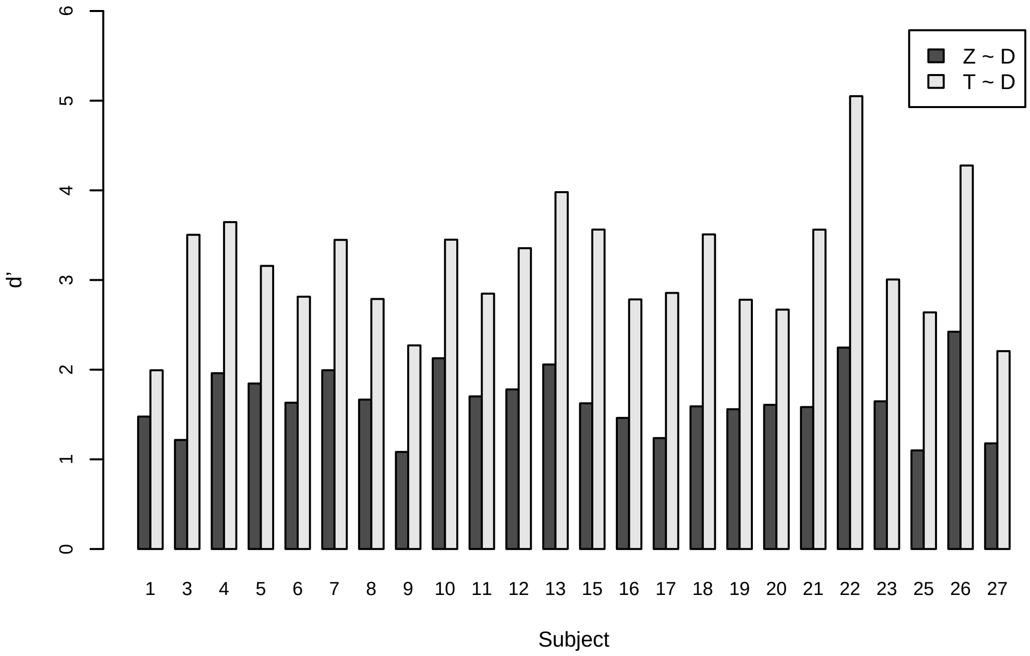


Figure 4.2 shows the duration of the first vowel in the stimuli used in the experiment. Vowel length is shown separately for English and Spanish talkers. As expected, for both groups of talkers, the vowel is significantly longer in [aDa]

Figure 4.3: Sensitivity to Z $\sim$ D vs. T $\sim$ D by subject in Experiment 2. All within-subject differences are significant at $\alpha = .05$. For each d', $600 \leq n \leq 633$



stimuli than in [aTa] stimuli ($p = 4.0 \times 10^{-8}$ for English talkers; $p = 4.6 \times 10^{-6}$ for Spanish talkers). The vowel is also longer in [aZa] stimuli than in [aTa] stimuli ($p = 5.8 \times 10^{-30}$ for English; $p = 1.0 \times 10^{-4}$ for Spanish). In addition, for English but not for Spanish talkers, the vowel is longer in the [aZa] stimuli than in the [aDa] stimuli ($p = 3.2 \times 10^{-9}$; the difference between [aZa] and [aDa] is not significant for Spanish talkers).

‘Same’-‘Different’ Responses

Figure 4.3 shows sensitivity to the Z $\sim$ D and T $\sim$ D differences, broken down by subject. Sensitivity is measured by d', calculated from subjects’ ‘same’-

‘different’ responses, using the Independent-Observation Model.[4] Each subject is significantly more sensitive to the voicing distinction than to the continuancy distinction; when the results for each subject are broken down by place, as shown in figure 4.4, the direction of the effect is the same in every case. Significance for d' was calculated using the G statistic of Gourevitch and Galanter (1967).

Figure 4.5 shows sensitivity by talker and place of articulation. For each combination of talker and place, sensitivity to T $\sim$ D is significantly greater than sensitivity to Z $\sim$ D. Recall that talkers 1 – 3 are the English speakers, and talkers 4 and 5 the Spanish speakers.

Subjects were slightly more sensitive to all differences in later blocks than in the first block; however, the differences in sensitivity are very small.

---

[4] The design of this experiment is neither a pure ‘fixed’ design nor a pure ‘roving’ one. See Kabak and Idsardi (2007, fn. 6) for an argument for the appropriateness of the Independent-Observation Model for this kind of experiment.

Figure 4.4: Sensitivity to Z $\sim$ D vs. T $\sim$ D by subject and place of articulation in Experiment 2. Stars mark within-subject and within-place differences that are significant at $\alpha = .05$. For each d', $194 \leq n \leq 213$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'

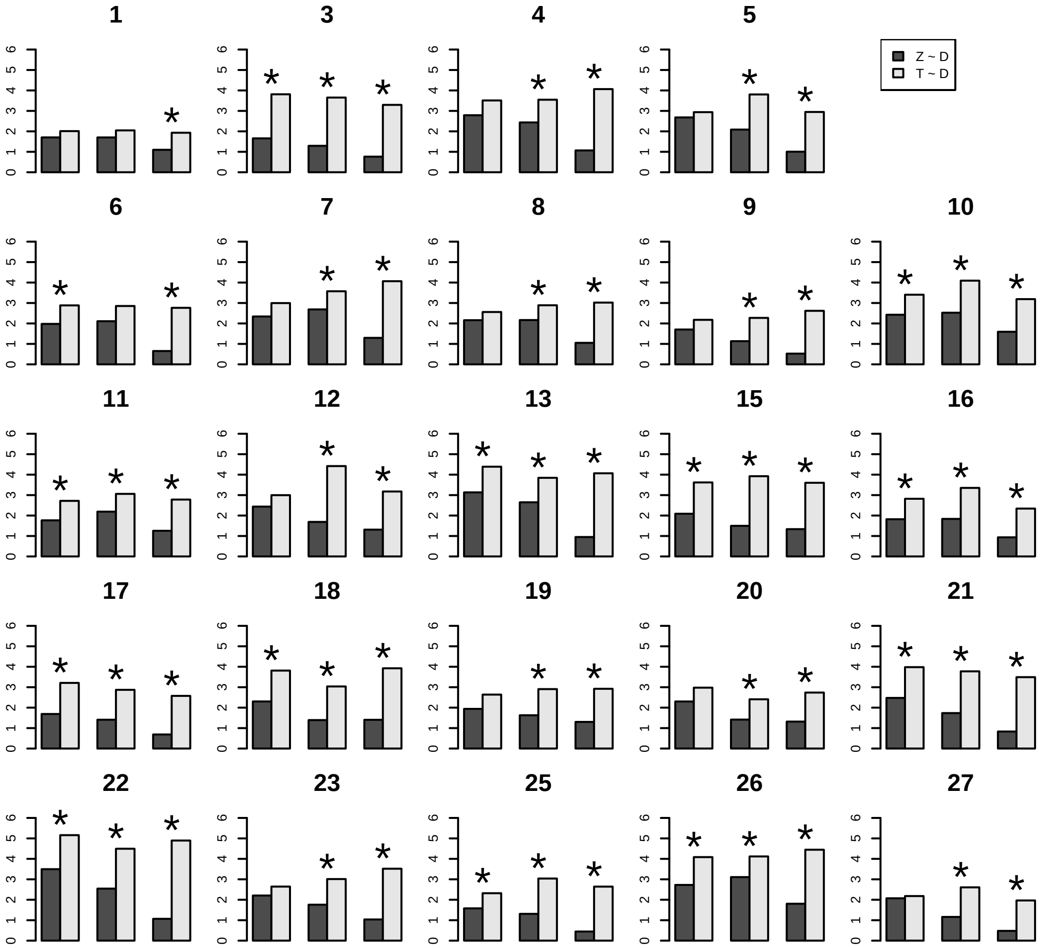
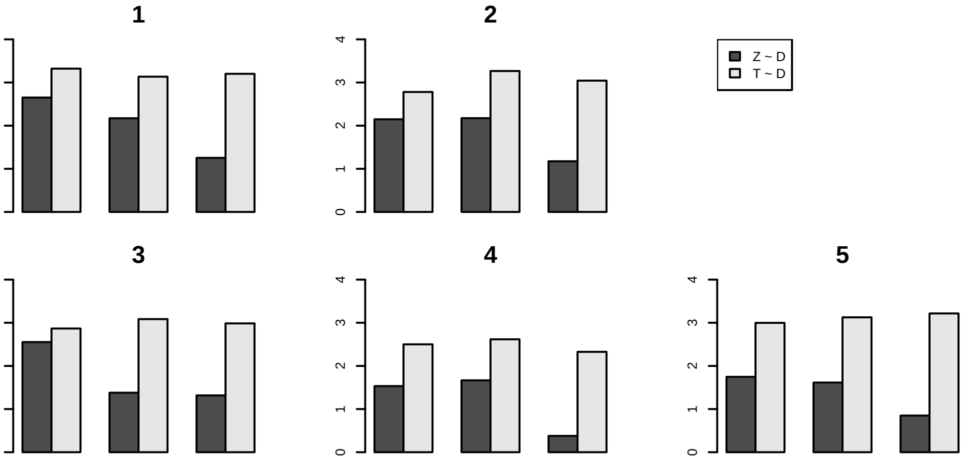


Figure 4.5: Sensitivity to Z $\sim$ D vs. T $\sim$ D by talker and place of articulation for Experiment 2. All within-talker and within-place differences are significant at $\alpha = .05$. For each d', $864 \leq n \leq 1203$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'



Reaction Times

Reaction times were analyzed with a linear mixed-effects model, which predicted log reaction time from the factors Comparison (Z $\sim$ D vs. T $\sim$ D), Place, Trial (the number of the trial for each subject), and their two-way interactions as fixed effects and the factors Talker and Subject as random effects. The model also included by-Subject random effects of Trial. Significance of each fixed effect was estimated from its *t*-statistic with degrees of freedom equal to the number of observations minus the number of fixed-effects parameters (Baayen 2008, 248). Only reaction times for 'different' trials to which subjects responded correctly were included. In general, responses were faster in the same conditions in which subjects exhibited greater sensitivity, as detailed below.

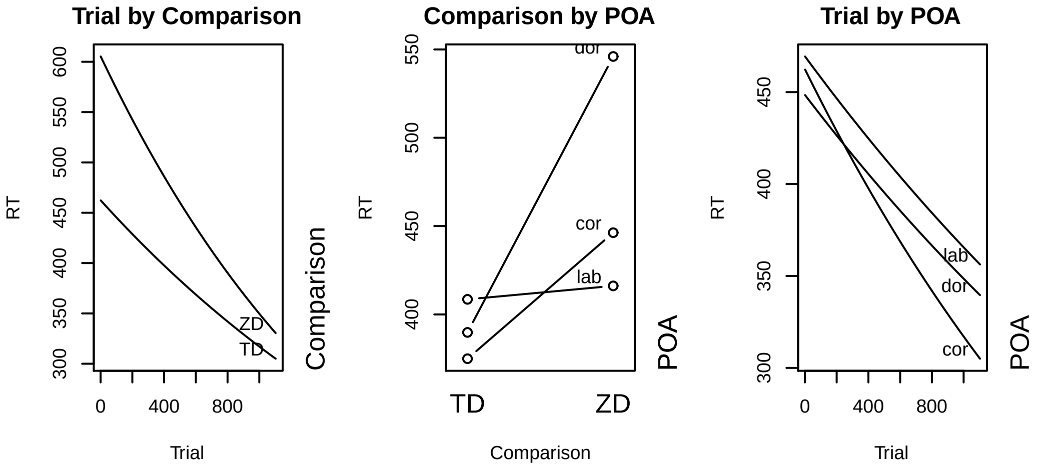
A model of this type with multiple categorical factors, such as Comparison and Place, does not produce information about the main effect of one factor independently of the other. For example, the main effect of Comparison in this model is actually the effect of Comparison within the level of Place that happens to be chosen as the baseline (that is, among either labials, coronals, or dorsals, but not across the three places). For the crucial effect of Comparison, I built three separate models, one with each place of articulation as the baseline, in order to determine whether the effect of Comparison was significant for each place. For other effects, I report the results from the model with coronals as the baseline; unless otherwise noted, all significant effects in this model were also significant in the two other models.

The effect of Comparison was significant for all three places of articulation: labials ($p = 1.6 \times 10^{-3}$), coronals ($p = 2.7 \times 10^{-12}$), and dorsals ($p = 1.2 \times 10^{-21}$). For all three places of articulation, subjects responded more slowly to Z $\sim$ D trials than they did to T $\sim$ D trials. The middle graph in figure 4.6 shows the partial effects of the interaction between Comparison and Place.

There was a significant effect of Trial ($p = 4.7 \times 10^{-7}$) such that subjects got faster over the course of the experiment. There was also a significant interaction between Comparison and Trial ($p = 3.1 \times 10^{-3}$); subjects improved more quickly on Z $\sim$ D trials than they did on T $\sim$ D trials. Finally, there was a marginal interaction between Place and Trial: subjects improved more quickly for coronals than they did for either labials ($p = .088$) or dorsals ($p = .034$); the difference between labials and dorsals was not significant ($p = .97$). Figure 4.6 shows the partial effects of the interactions between Trial and Comparison (left) and between Trial and Place (right).

To determine how robust the effect of Comparison is, this model was compared

Figure 4.6: Partial effects of the interactions between Trial and Comparison, Comparison and Place, and Trial and Place in Experiment 2



to one that included an interaction between Subject and Comparison as a random effect, and one that included an interaction between Talker and Comparison as a random effect. Either interaction, if significant, would suggest that the effect of Comparison seen above is particular to this group of subjects or talkers. The likelihood ratio test described in Baayen (2008, 253) showed that the improvement in the model that added an interaction between Subject and Comparison approached significance ($p = .061$), while the model that added an interaction between Talker and Comparison was significantly improved ($p = 1.7\times10^{-3}$). However, none of the by-subject or by-talker interactions reversed the direction of the effect: reaction times were slower for Z $\sim$ D comparisons for each individual subject and talker. Thus, difficulty with Z $\sim$ D trials is robust across subjects and talkers.

Finally, recall that reaction times were measured from the end of the stimulus. An alternative method would be to measure reaction time from the end of the consonant, since it is the consonants (not the vowels) that consitute the comparisons of interest. The choice between these options depends on two factors:

1. Does the length of the second vowel vary systematically by consonant (or talker)?
2. Did subjects use cues from the second vowel in the perceptual task?

If the length of the second vowel varies systematically among the different stimuli, then measuring reaction times from a point after the end of the vowel may artificially increase or decrease reaction times for certain stimuli. On the other hand, if subjects used cues from the second vowel in the discrimination task, then measuring from the end of the vowel may be a more accurate representation of the point at which subjects had enough information to make a decision.

The answer to question 1 is 'yes'. For example, for every talker except talker 4, the second vowel in the [aZa] stimuli is significantly longer than the second vowel in the [aDa] stimuli. What is not certain, however, is whether subjects used this difference in duration – or other cues within the second vowel – to distinguish the two types of stimuli. In the absence of a clear answer, I built a second set of models based on the reaction times as measured from the end of the consonant. The resulting models are nearly identical to those based on reaction times measured from the end of the stimulus. All of the effects reported above as significant are also significant in this second set of models, except for the result of adding by-subject and by-talker interactions with Comparison. The latter models are not significantly improved by adding such an interaction, either for subjects ($p = .38$) or for talkers ($p = .39$).

4.1.3 Discussion

Overall Results

Overall, the results of the experiment are consistent with a perceptual account of the broad strokes of lenition: listeners are less sensitive to the distinction between voiced stops and spirants intervocalically than they are to the distinction between voiced and voiceless stops. This difference was manifested both in subjects' 'same'-'different' responses (reflected in the d' scores) and in their reaction times: subjects were quicker to identify the more salient voicing distinction than the continuancy distinction.

This effect is highly robust and seems unlikely to be due to the particulars of the experimental design. The difference in d' scores was seen for every combination of talker and place of articulation; this fact indicates that the sensitivity difference is probably not a peculiarity of a few talkers' pronunciations. Nor is the effect likely to depend on particular characteristics of the experimental subjects: the difference in d' scores was seen for every combination of subject and place of articulation, and even with a random effect for the interaction of Subject and Comparison, every subject responded more slowly on Z $\sim$ D trials than on T $\sim$ D trials.

Effect of Acoustic Characteristics of Stimuli

Recall that there are at least two ways in which the English talkers' tokens might have influenced the main result that the voicing distinction is more salient than the continuancy distinction: the use of aspiration to cue voiceless stops, and the difference in length between the initial vowels in the [aDa] and [aZa] stimuli. The aspiration in the [aTa] stimuli might have artificially enhanced the T ∼ D distinction, while the long vowels in the [aZa] stimuli might have artificially enhanced the Z ∼ D distinction.

It is unlikely that the English talkers' aspiration is driving the entire difference in perceptibility between the Z ∼ D and T ∼ D comparisons. The overall effect holds not only for the English talkers, who produced a significant amount of aspiration, but also for the Spanish talkers, who produced far less. Indeed, if anything, we might expect the subjects' sensitivity to the T ∼ D contrast for the Spanish talkers to be artificially *low* – as English speakers, the subjects would expect to be able to rely on the aspiration cue that was much less pronounced in the Spanish tokens.

As for vowel length, the effect of the long vowels in the [aZa] tokens should, if anything, encourage the opposite of the effect found here: by using vowel length as a cue to the Z ∼ D distinction, subjects should have been better able to distinguish those stimuli than they otherwise would have been able to. The T ∼ D distinction was nevertheless more salient, suggesting that this is a robust result.

Effect of the English Consonant Inventory

Finally, it is important to consider whether the main result that the T ∼ D distinction is more salient than the Z ∼ D distinction is simply an artifact of the consonant inventory of the English-speaking subjects. For example, since English has phonemic [k] and [g] but lacks [ɣ], it is only to be expected that English speakers are better able to tell the difference between phonemic sounds in their language ([g] and [k]) than between a phonemic and non-phonemic sound ([g] and [ɣ]) (Boomershine et al. 2008), especially if the unfamiliar [ɣ] was assimilated to the native category [g] (Best et al. 1988; Kuhl and Iverson 1992). A similar explanation could be put forward for the labials: English has [p] and [b], but not [β].[5]

However, it is unlikely that the English segment inventory alone is responsible for the main result that the voicing distinction is more salient than the continuancy distinction. The result holds for coronals as well: even though all three consonants tested in the coronal conditions are phonemic in English ([ð], [d], and [t]), English speakers were less sensitive to one distinction than the other. Indeed, the interaction between Comparison and Place illustrated in figure 4.6 suggests that coronals suffered *more* in the Z ∼ D condition than the labials did, despite their phonemic advantage.

In addition, there is reason to believe that the results for labials and dorsals are just as relevant to understanding spirantization as those for coronals. As discussed in chapter 2, the typological survey of Gurevich (2004) shows that spirantization of voiced stops, like other lenition processes, is usually non-neutralizing. In other words, if a language spirantizes a voiced stop, the resulting voiced spirant exists in the language only as an allophone of the stop; the two are not phonemically contrastive, just as English [g] and [ɣ] do not contrast. On the other hand, many of these languages *do* have voiceless stops that contrast with the voiced stops, just as English [k] and [g] constrast.

There are at least two possible explanations for this state of affairs. One is that once intervocalic spirantization becomes part of a language's phonology,

---

[5] It is possible that subjects assimilated these [β]s to the native categories [v] or [w]. To the extent that subjects were successful in assimilating [β] to a native segment other than [b] (an "opposing-category assimilation" in the terminology of Best et al. (1988, 347)), the same argument applies here that is discussed below for coronals.

that language is likely to lose the contrast between stops and spirants in other environments as well. Another possibility, however, is that spirantization is more likely to enter a language in the first place if that language does not have a contrast between voiced stops and spirants. If the latter is true, then the (non-)phonemic status of [k] ∼ [g] ∼ [ɣ] for English speakers is reflective of the situation for languages that actually acquire lenition; in that case, the results for dorsals and labials may represent the very conditions that encourage spirantization in the first place.

Effect of Place of Articulation

The results of Experiment 2 clearly show that the voicing distinction is perceptually more salient than the continuancy distinction for the relevant segments and environments. Interestingly, the interaction between Comparison and Place illustrated in figure 4.6 suggests that this difference in perceptual salience is not equally large at every place of articulation. Experiment 2 tests whether there are indeed differences by place, or whether this result was simply an artifact of the English-speaking subjects' native consonant inventories.

4.2 Experiment 3: Effect of Place of Articulation

As noted above, Spanish is one language that has a variant of the lenition process of interest: the voiced stops [b], [d], and [g] become the spirants [β], [ð], and [ɣ] between vowels (among other environments, depending on the dialect). In addition, Spanish contains all three voiceless stops investigated here ([p], [t], and [k]), which contrast phonemically with the voiced stops/spirants. Thus, native Spanish speakers cannot provide data on whether the T $\sim$ D or Z $\sim$ D distinction is more salient: since allophonic distinctions are perceived more poorly than phonemic ones (Boomershine et al. 2008), we expect Spanish-speaking subjects to be more sensitive to the T $\sim$ D distinction purely by reason of their native consonant inventory.

However, the effect of the native segment inventory of Spanish should be the same at each place of articulation: for each place, the T $\sim$ D distinction is phonemic and the Z $\sim$ D distinction is allophonic. Therefore, unlike English speakers, Spanish speakers provide the perfect opportunity to study those perceptual differences by place that are inventory-independent.

4.2.1 Design

11 native speakers of Spanish participated in the experiment; all were undergraduate students at the University of California, Santa Cruz. Subjects received either monetary compensation or extra credit, depending on which courses they were recruited from. The procedure was exactly the same as that in Experiment 2, except that all of the instructions were in Spanish. The Spanish instructions were intended to encourage the subjects to draw on their Spanish resources when perceiving the stimuli, especially for those subjects who were native bilingual speakers of both Spanish and English (subjects 2, 3, 6, 7, and 11). Excluding these subjects from the analyses below does not substantially affect the results. In the discussion of reaction times below, I report the results for an analysis of

all 11 subjects; unless otherwise noted, all significant effects are also significant in an analysis of only the subjects who are not also native speakers of English. No subject was a native speaker of any language besides Spanish or English, and no subject had participated in the previous experiment.

4.2.2 Results

Acoustic Properties of the Stimuli

Figure 4.7: Duration of aspiration by place of articulation in [aTa] stimuli for Experiments 2 and 3. All pairwise within-language differences are significant at $\alpha = .05$

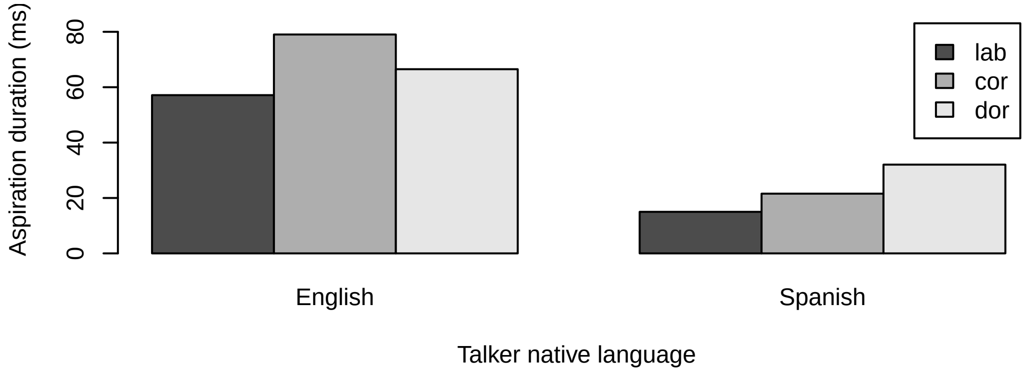
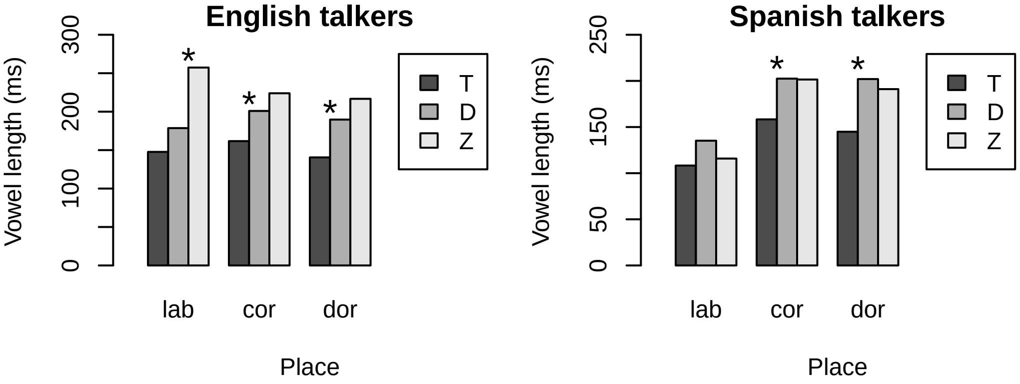


Figure 4.7 shows the duration of aspiration in the [aTa] stimuli, broken down by place and talker's native language. For both groups of talkers, the [apa] tokens had the shortest aspiration. Among the English speakers, the [ata] tokens had the longest aspiration, while among the Spanish speakers, the [aka] tokens had the longest.

Figure 4.8 shows the duration of the first vowel, broken down by place and talker's native language. The patterns shown in figure 4.2, where length was

Figure 4.8: Length of first vowel by place of articulation in stimuli for Experiments 2 and 3. Stars mark within-language and within-place differences that are significant at $\alpha = .05$



pooled across place, are largely preserved here, with some loss of significance. For both groups of talkers, the difference in vowel length between [aTa] and [aDa] stimuli is significant only for coronals and dorsals. Among English speakers, the unusually long vowel durations in the [aZa] stimuli are significant only for tokens of [aβa].

‘Same’-‘Different’ Responses

Figure 4.9 shows subjects’ sensitivity to the Z $\sim$ D and T $\sim$ D differences by place of articulation, pooled across subjects. The difference is largest for dorsals, and comparable for labials and coronals. This pattern appears to hold for most (but not all) of the individual subjects, as shown in figure 4.10. Note that having an extremely large difference for dorsals does *not* seem to be correlated with whether the subject is also a native speaker of English: subjects 6, 7, and 11 all have especially large differences among the dorsals, but subjects 2 and 3 do not.

Figure 4.11 shows sensitivity by place of articulation and talker. The trend

Figure 4.9: Sensitivity to Z $\sim$ D vs. T $\sim$ D by place of articulation in Experiment 3. All within-place differences are significant at $\alpha = .05$. For each d', $5355 \leq n \leq 5727$

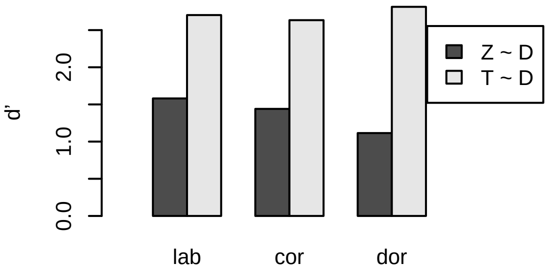
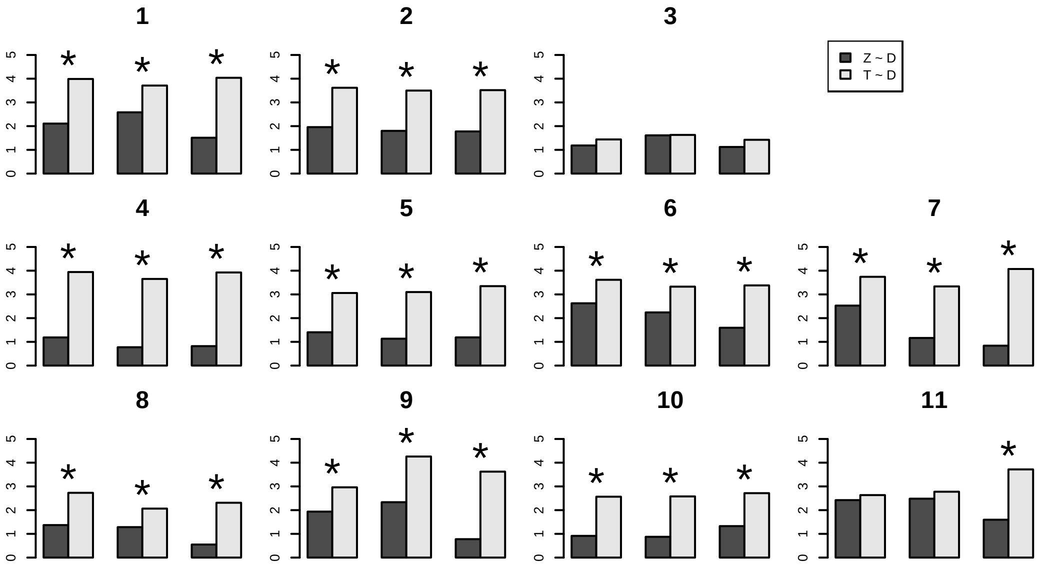
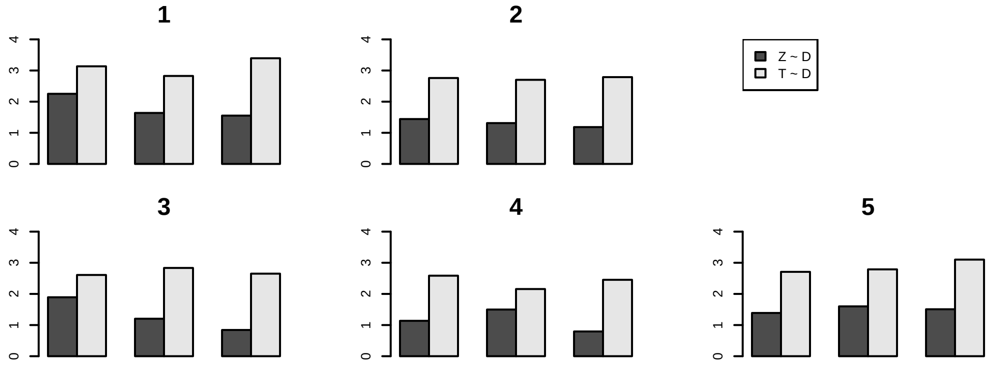


Figure 4.10: Sensitivity to Z $\sim$ D vs. T $\sim$ D by subject and place of articulation in Experiment 3. Stars mark within-subject and within-place differences that are significant at $\alpha = .05$. For each d', $198 \leq n \leq 216$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'



towards larger differences among dorsals is especially pronounced for talkers 1, 3, and 4.

Figure 4.11: Sensitivity to Z $\sim$ D vs. T $\sim$ D by talker and place of articulation for Experiment 3. All within-talker and within-place differences are significant at $\alpha = .05$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'



Reaction Times

Reaction time data was analyzed in exactly the same way as described for Experiment 2. Adding interactions between Comparison and Subject and between Comparison and Talker resulted in significant improvements to the model of reaction times for Experiment 3; therefore, those interactions were included in the final model. (For the subset of subjects who are not native speakers of English, only the addition of an interaction between Comparison and Subject was a significant improvement.)

For T $\sim$ D trials, subjects responded more quickly to coronals than to either labials ($p = 7.7 \times 10^{-4}$) or dorsals ($p = 4.7 \times 10^{-3}$); the difference between labials and dorsals was not significant ($p = .24$).[6] For Z $\sim$ D trials, subjects responded more slowly to dorsals than to either coronals ($p = 2.8 \times 10^{-8}$) or

---

[6] For the subset of subjects who are not native speakers of English, the difference between coronals and dorsals was not significant ($p = .13$), but the difference between labials and dorsals was ($p = 5.7 \times 10^{-3}$): subjects responded more quickly to dorsals than to labials.

labials ($p = 1.7 \times 10^{-6}$); the difference between labials and coronals was not significant ($p = .32$).[7] From these results, we can draw three conclusions about the size of the effect of Comparison at different places of articulation:

- The difference between the Z $\sim$ D and T $\sim$ D comparisons is greater among coronals than among labials: subjects responded more quickly to coronals than to labials on T $\sim$ D trials, but for Z $\sim$ D trials, the two places were approximately the same.

- The difference between the two comparisons is greater among dorsals than among labials: although subjects responded at about the same rate to both places on T $\sim$ D trials, subjects responded more slowly to dorsals than to labials on Z $\sim$ D trials. For the subset of non-English-speakers, we can draw the same conclusion in a slightly different way: as with the difference between coronals and labials, subjects responded more quickly to dorsals than to labials on T $\sim$ D trials; for Z $\sim$ D trials, however, the advantage of dorsals disappeared, and subjects responded to the two places at about the same rate.

- Possibly, the difference between the two comparisons is greater among dorsals than among coronals: for the subset of non-English-speakers only, responses were about the same in the T $\sim$ D trials for coronals and dorsals; for the Z $\sim$ D trials, on the other hand, responses were slower for dorsals.

Analyses of reaction times from the end of the consonant produce exactly the same pattern of significant and non-significant effects, except that responses were

---

[7] For the subset of non-English-speakers, the difference between dorsals and labials was not significant ($p = .18$).

significantly faster to coronals than to labials in Z $\sim$ D trials ($p = .021$).

4.2.3 Discussion

Effect of Place of Articulation

The Spanish-speaking subjects, like the English-speaking subjects, were more sensitive to (and responded faster to) the voicing distinction than the continuancy distinction. This result was expected on the basis of the Spanish consonant inventory alone: the voicing distinction, but not the continuancy distinction, is phonemic for Spanish speakers for these segments.

The new result of Experiment 3 is that the effect of comparison type depends on place of articulation. The differences in sensitivity by place, and the interactions between place and comparison type in the reaction times, suggest that the size of the effect is greatest for dorsals and smallest for labials. Thus, the continuancy distinction may be difficult to perceive in general, but it is *especially* difficult for dorsals. This effect is *not* predicted on the basis of the Spanish consonant inventory; thus, the results of Experiment 3 provide information about the perceptibility of these different places of articulation independent of segment inventory.

Effect of Acoustic Properties of the Stimuli

To be confident that the distinction between [aɣa] and [aga] is especially poorly perceived, we must ask whether the differences that were observed for each place of articulation could be the result of particular acoustic properties of these stimuli. Recall that two cues might be influencing the relative perceptibility of the T $\sim$ D and Z ∼ D distinctions: aspiration in [aTa] tokens and length of the first vowel.

As discussed above, the Spanish speakers produced more aspiration in the [aka] tokens than at any other place of articulation. If subjects picked up on this cue, then their discrimination of [aka] and [aga] may have been artificially inflated; in turn, it is possible that the difference between the [aɣa] ∼ [aga] and [aka] ∼ [aga] distinctions is no greater than it is at other places of articulation (although both dorsal distinctions are lower than the corresponding labial or coronal distinctions). However, an explanation along these lines cannot account for the fact that the subjects in Experiment 3 were just as sensitive overall to the [aka] ∼ [aga] distinction for the English-speaking talkers ($d' = 2.88$) as for the Spanish-speaking talkers ($d' = 2.73$; the difference is not significant, $p = .36$).

As for vowel length, figure 4.8 and the associated discussion show that there are two primary ways in which the overall effects of consonant type on vowel length are modulated by place: English speakers' long vowels in [aZa] stimuli are significantly different from those in [aDa] stimuli only for labials, and neither group of speakers exhibits a significant difference in vowel length between tokens of [apa] and [aba]. Either of these effects might have reduced the difference between the [aβa] ∼ [aba] and [apa] ∼ [aba] comparisons: the first by increasing the perceptibility of the [aβa] ∼ [aba] distinction by providing an extra cue for [aβa] stimuli, and the second by decreasing the perceptibility of the [apa] ∼ [aba] distinction by eliminating a cue to stop voicing.

First, to assess the effect of the English-speaking talkers' long vowels in the [aβa] tokens, we can examine subjects' sensitivity to just the Spanish-speaking talkers, for whom the difference in vowel length between [aβa] and [aba] is not significant (and, indeed, the [aβa] tokens have *shorter* vowels than the [aba] tokens). Unsurprisingly, subjects were more sensitive to the [aβa] ∼ [aba] distinction for the English talkers ($d' = 1.82$) than for the Spanish talkers ($d' = 1.28$; the difference is significant, $p = 1.1 \times 10^{-5}$). However, even for tokens produced by the Spanish talkers, the effect of comparison type is greater for dorsals than it is for labials: a mixed-effects model of reaction times identical to the one described above but including only Spanish talkers reveals a significant interaction between Comparison and Place such that subjects responded faster to [aka] ∼ [aga] trials than to [apa] ∼ [aba] trials, but faster to [aβa] ∼ [apa] trials than to [aɣa] ∼ [aga] trials.

The impact of the lack of a significant difference in vowel duration between [apa] and [aba] tokens is more difficult to assess, since both groups of talkers have this property. A closer examination of the individual talkers, however, reveals that three of the five talkers do in fact exhibit a significant difference: talkers 2 ($p = 6.4 \times 10^{-3}$), 3 ($p = 3.2 \times 10^{-2}$), and 5 ($p = 2.2 \times 10^{-5}$). But subjects were no more sensitive to the [apa] ∼ [aba] distinction for these three talkers (joint $d' = 2.68$) than for the two remaining talkers who do not exhibit a difference in vowel length (joint $d' = 2.81$; the difference is not significant, $p = .41$). In addition, the only pair of individual talkers who had significantly different d's for the [apa] ∼ [aba] distinction was talkers 1 and 3: for talker 1, who does not have a significant difference in vowel length, subjects were *more* sensitive ($d' = 3.13$) than for talker 3, who does ($d' = 2.61$; $p = 3.8 \times 10^{-2}$). Thus, it seems unlikely that subjects suffered from insufficient cues to the voicing distinction in the labial stimuli for some talkers.

4.3 Experiment 4: Relative Perceptibility of Voicing and Spirantization for Voiceless Stops

Experiment 4 was designed to test the relative perceptibility of two logically possible (and attested) repairs for intervocalic voiceless stops: voicing and spirantization. The structure of the experiment is the same as that of Experiment 2, except that different stimuli were used.

4.3.1 Design

Recording of Stimuli

The stimuli for Experiment 4 included the 9 consonants given in table 4.3; as in Experiments 2 and 3, the stimuli had the form [a__a], with initial stress and with some consonant from table 4.3 intervocalically.

Table 4.3: Perceptibility comparisons in Experiment 4

	[−voi] Spirants		**[−voi] Stops**		**[+voi] Stops**
Labials	f	∼	p	∼	b
Coronals	θ	∼	t	∼	d
Dorsals	x	∼	k	∼	g
Cover Symbol	**S**	∼	**T**	∼	**D**

Three talkers were recorded for Experiment 4. Talker 1 was a male native speaker of Bulgarian and second-language speaker of English; talker 2 was a female native speaker of German and second-language speaker of English; talker 3 was a male native speaker of English and second-language speaker of German. Thus, for talkers 1 and 2, [x] was a native phoneme and [θ] a non-native phoneme with which

Table 4.4: Elicitation of stimuli for Experiment 4

Stimulus	**Orthography**
[aba]	aba
[ada]	ada
[aga]	aga
[apa]	apa
[ata]	ata
[aka]	aka
[afa]	afa
[aθa]	aθa
[axa]	axa

they had some familiarity (as second-language speakers of English); for talker 3, [θ] was a native phoneme and [x] was non-native (but familiar from German). Talker 1 also had linguistic training. All three talkers were naïve to the purposes of the experiment. Stimuli were presented to the talkers in a single block in IPA (see table 4.4); before recording, the talkers and experimenter discussed the intended pronunciation of each type of stimulus.

Selection of Stimuli

As was the case for Experiments 2 and 3, it is possible that the different native segment inventories of the talkers who produced the stimuli for Experiment 4 introduce bias that might influence the perceptual results. At least three possibilities come to mind:

1. As in Experiments 2 and 3, the English- (and German-)speaking talkers make use of a cue to the voicing distinction among stops that the Bulgarian-speaking talker lacks: aspiration. Thus, the D $\sim$ T distinction might be

easier to discriminate for the English- and German-speaking talkers.

2. All of the consonants that were non-native for at least some talkers were spirants ([θ] and [x]). It is possible that the center of gravity or the intensity of the frication noise in these segments differed by talker in ways that might influence the perceptibility of the S $\sim$ T distinction.
3. As in Experiments 2 and 3, the vowels in the non-native stimuli may have been distorted, thus introducing artificial cues to the relevant distinctions.

For each talker, 10 tokens of each stimulus were selected (except for talker 3's productions of [aθa], where only 8 tokens were viable). Tokens were selected as follows:

1. All tokens with any obvious abnormality were excluded.
2. For the English and German speakers, the tokens of the [aTa] stimuli were selected that had the shortest period of aspiration after the stop.
3. The F1 and F2 values of the remaining vowels were measured. For each talker and place of articulation, the [aDa] and [aSa] tokens were chosen that were closest to the [aTa] tokens for that speaker and place of articulation (as measured by the formants of the first vowel), such that the first vowel for the [aTa] tokens was not systematically closer to either those of the [aDa] tokens or those of the [aSa] tokens.

Tokens were trimmed and amplitude-normalized as in Experiments 2 and 3. Further acoustic characteristics of the stimuli are discussed below.

Participants

18 native speakers of English participated in the experiment; all were students in undergraduate linguistics courses at the University of California, Santa Cruz. Subjects received either monetary compensation or extra credit, depending on the course. Six subjects were native speakers of a second language besides English: subjects 2, 9, and 11 were native speakers of Spanish; subject 5 of Japanese; subject 15 of Mandarin; and subject 17 of Cantonese. Separate analyses were performed excluding these bilingual subjects; where the results for the subset of monolingual subjects differs from the group as a whole, this is noted below.

Procedure

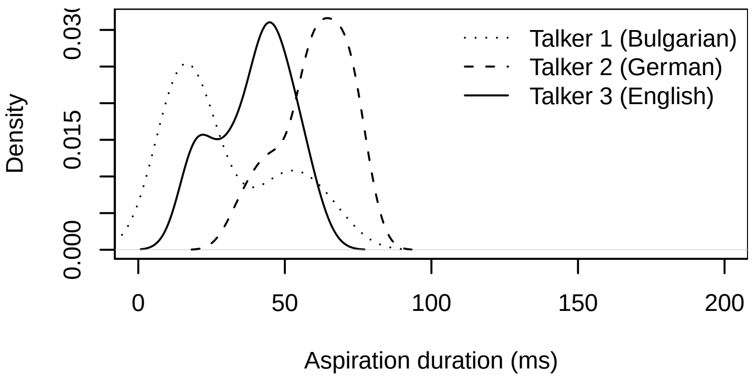
The experiment was run with E-Prime version 2.0 and a PST serial response box. The instructions and procedure were the same as in Experiment 2. There were three types of 'same' trials ($D \sim D$, $T \sim T$, and $S \sim S$) and four types of 'different' trials ($D \sim T$, $T \sim D$, $S \sim T$, and $T \sim S$). The practice session in Experiment 4 lasted 3 minutes for all subjects, rather than terminating after criterion had been reached as in Experiments 2 and 3. The entire experiment lasted about 45 minutes.

4.3.2 Results

Acoustic Properties of the Stimuli

Unsurprisingly, as was the case for Experiments 2 and 3, there was a certain amount of inter-talker variation in the stimuli used in Experiment 4. The following discussion documents some differences among talkers that may have affected

Figure 4.12: Density curves for aspiration duration in [aTa] stimuli for Experiment 4



subjects' performance on the perceptual task.

Figure 4.12 shows the duration of aspiration in the [aTa] stimuli. The tokens produced by the German speaker had significantly more aspiration than those produced by the English speaker ($p = 1.4 \times 10^{-6}$); the tokens produced by the English speaker, in turn, had significantly more aspiration than those produced by the Bulgarian speaker($p = .013$).

Figure 4.13 shows the duration of the voiceless fricative by talker and place of articulation. Interestingly, after Holm correction for multiple comparisons, only one within-place difference turns out to be significant: talker 1 (Bulgarian) produced significantly longer [θ]s than talker 2 (German, $p = .016$). The difference between talkers 1 and 3 for [afa] approached significance ($p = .076$).

Figure 4.14 shows the center of gravity of the voiceless fricative by talker and place of articulation. For [θ], the native English speaker (talker 3) had a lower center of gravity than talkers 1 and 2 ($p = 2.7 \times 10^{-16}$ and 1.3×10^{-9}, respectively). For [x], the native speaker of Bulgarian (talker 1) had a higher center of gravity

Figure 4.13: Duration of consonant in [aSa] stimuli for Experiment 4. Stars mark within-place differences that are significant at $\alpha = .05$

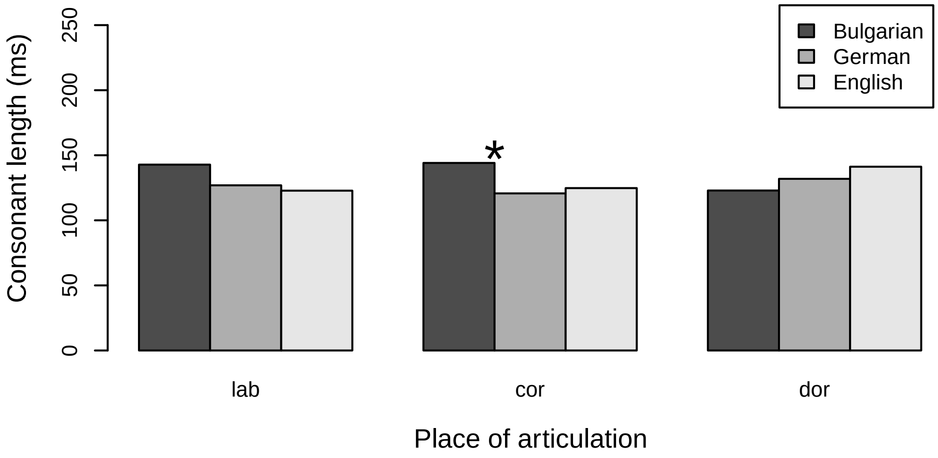
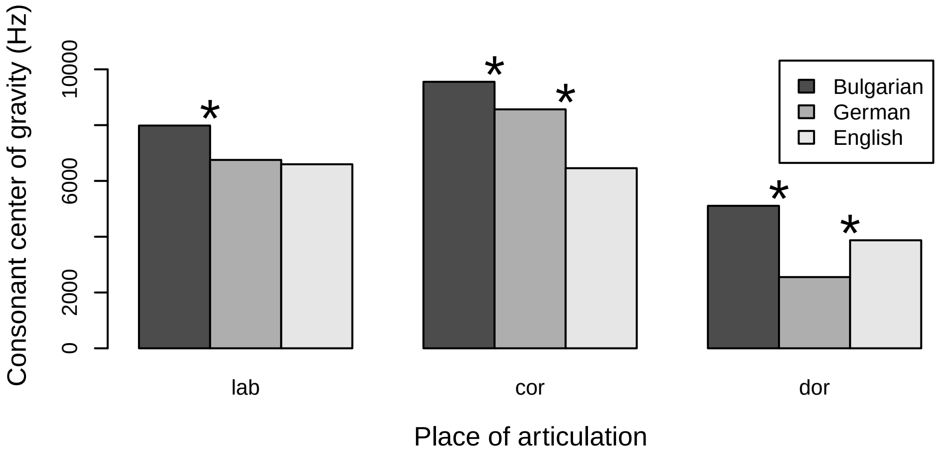
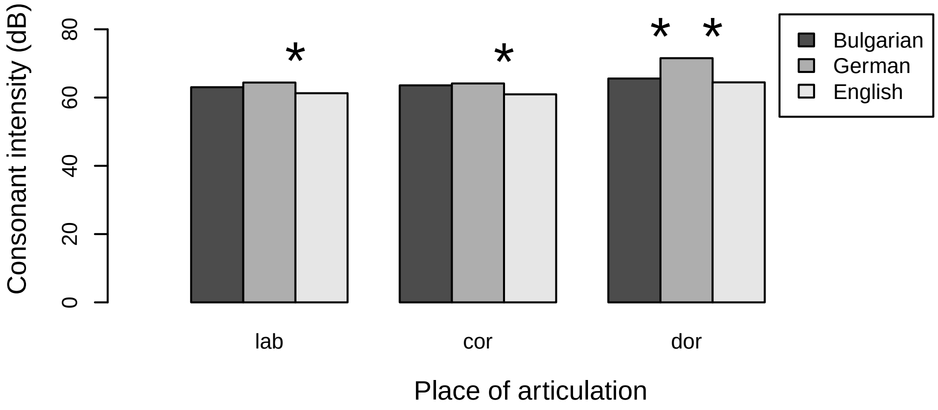


Figure 4.14: Center of gravity of consonant in [aSa] stimuli for Experiment 4. Stars mark within-place differences that are significant at $\alpha = .05$



than the native speaker of German (talker 2, $p = 3.7 \times 10^{-13}$); the center of gravity for talker 3's [x]s was between the two and significantly different from both ($p = 1.5 \times 10^{-4}$ and 6.1×10^{-5}, respectively). There are also differences for

Figure 4.15: Intensity of consonant in [aSa] stimuli for Experiment 4. Stars mark within-place differences that are significant at $\alpha = .05$

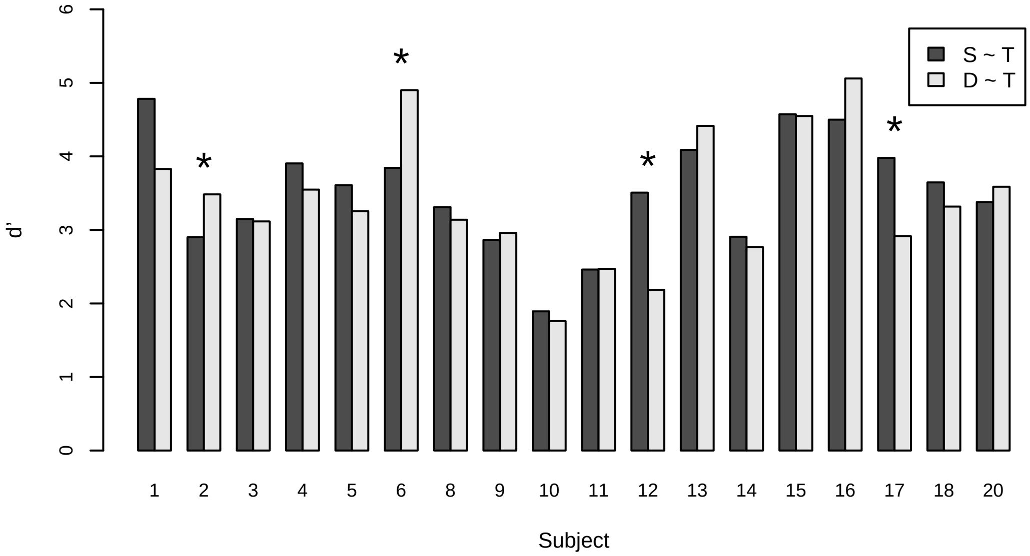


[f]: talker 1 produced tokens with a higher center of gravity than talkers 2 and 3 ($p = 1.5 \times 10^{-4}$ and 2.5×10^{-5}, respectively).

Figure 4.15 shows the intensity of the voiceless fricative by talker and place of articulation. The [θ]s produced by the native English speaker (talker 3) were significantly less intense than those of the native German speaker (talker 2, $p = 6.9 \times 10^{-3}$); the difference between talkers 3 and 1 for [θ] approached significance ($p = .051$). For [x], the native German speaker (talker 2) produced more intense tokens than talkers 1 and 3 ($p = 7.1 \times 10^{-9}$ and 1.5×10^{-11}). In addition, the native German speaker (talker 2) produced significantly more intense [f]s than the native English speaker (talker 3, $p = 6.4 \times 10^{-3}$).

While any of these by-talker differences has the potential to affect subjects' performance on the perceptual task, it is also possible that some or all of the differences could be factored out by subjects in the way that listeners ordinarily adjust to individual differences among talkers, such as differences in F_0. At any

Figure 4.16: Sensitivity to S $\sim$ T vs. D $\sim$ T by subject in Experiment 4. Stars mark within-subject differences that are significant at $\alpha = .05$. For each d', $n = 540$

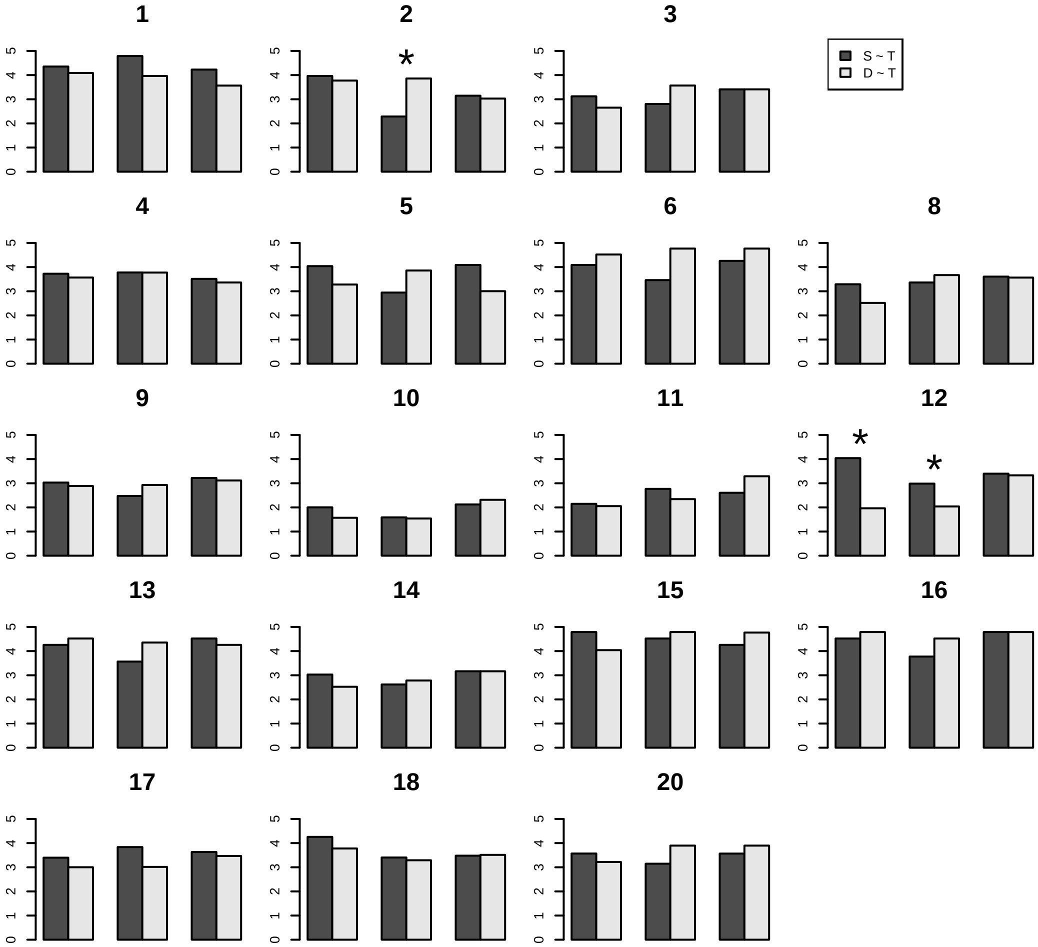


rate, section 4.3.3 below shows that inter-talker differences did not appear to have a large effect on subjects' performance.

'Same'-'Different' Responses

Figure 4.16 shows sensitivity to the S $\sim$ T and D $\sim$ T differences, broken down by subject. In contrast to the results for Experiment 2, neither difference is clearly more perceptible than the other. 11 subjects have a larger d' for S $\sim$ T than for D $\sim$ T; the difference is significant for two of them, subjects 12 and 17 (the latter also a native speaker of Cantonese). 7 subjects have a larger d' for D $\sim$ T than for S $\sim$ T; the difference is significant for two of them, subjects 2 and 6 (the former also a native speaker of Spanish). When the results are further broken down by place of articulation (shown in figure 4.17), we see a similar lack

Figure 4.17: Sensitivity to S $\sim$ T vs. D $\sim$ T by subject and place of articulation in Experiment 4. Stars mark within-subject and within-place differences that are significant at $\alpha = .05$. For each d', $194 \leq n \leq 213$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'



of an overall pattern, and only three differences are significant. However, it is interesting to note that for labials, the S $\sim$ T distinction is more salient than the D $\sim$ T distinction for the majority of subjects, although it is significant only for subject 12.

Figure 4.18: Sensitivity to S $\sim$ T vs. D $\sim$ T by talker and place of articulation for Experiment 4. Stars mark within-talker and within-place differences that are significant at $\alpha = .05$. For each d', $n = 1080$. Within each plot, the three pairs of bars are for labials, coronals, and dorsals (left to right), and the y-axis measures d'

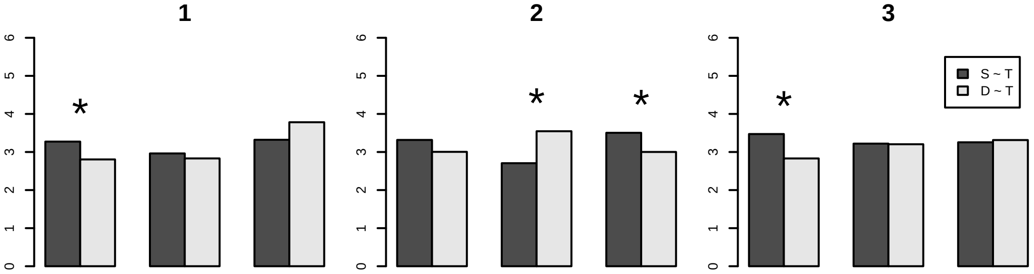
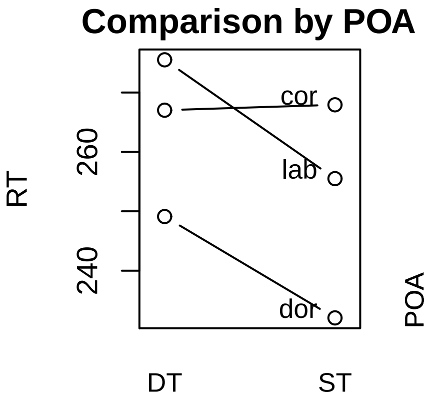


Figure 4.18 shows sensitivity by talker and place of articulation. Trends in both directions (S $\sim$ T more or less perceptible than D $\sim$ T) are seen for each talker; only some of the differences are significant. Note, however, that for both talkers 1 and 3, the difference is significant for labials; talker 2 has a non-significant trend in the same direction. These results, combined with the d' results in figure 4.17, constitute somewhat weak evidence that the [afa] $\sim$ [apa] distinction may be more salient than the [aba] $\sim$ [apa] distinction.

Reaction Times

Reaction times were analyzed in the same way as for Experiments 2 and 3. The effect of Comparison was not significant for coronals ($p = .31$) but approached significance for dorsals ($p = .093$) and labials ($p = .054$). For labials and dorsals, subjects responded more slowly to D $\sim$ T trials than to S $\sim$ T trials. Figure 4.19 shows the partial effects of the interaction between Comparison and Place. When bilingual subjects are excluded, the effect of Comparison is not significant at any

Figure 4.19: Partial effects of the interaction between Comparison and Place in Experiment 4



place ($p = .19$ for labials, .40 for coronals, and .16 for dorsals).

For reasons that are unclear to me, a large proportion of the responses in Experiment 4 (about 17%) were registered before the end of the second stimulus, in contrast to the less than 1% of responses in Experiment 2. Thus, many responses are excluded from the model described above because they involve negative reaction times. When reaction time is calculated from the end of the consonant, only just over 1% of responses are excluded in this way. The results of this model are comparable: responses are significantly slower to D $\sim$ T trials than to S $\sim$ T trials for labials ($p = .011$) and dorsals ($p = .023$), but not for coronals ($p = .21$). (When bilingual subjects are excluded, none of the differences are significant, but the difference for dorsals is nearly so; $p = .054$.)

4.3.3 Discussion

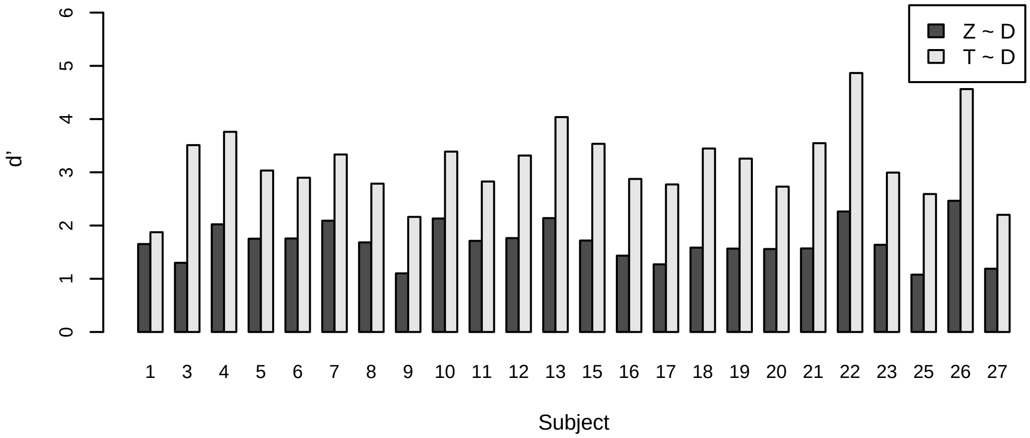
Overall Results

Unlike the results of Experiment 2, the results of Experiment 4 do not provide strong evidence for a large difference in perceptibility by comparison type. Subjects' sensitivity as measured by d' does not reveal a consistent difference between the D $\sim$ T comparison and the S $\sim$ T comparison; about half of the subjects have larger d's for each comparison, and the differences are significant for only a handful of subjects (some in each direction). However, there is some evidence that the S $\sim$ T comparison is more salient than the D $\sim$ T comparison for labials; the difference is significant for two of the three talkers, and 15 of the 18 subjects have trends in this direction.

The reaction time data suggests that the D $\sim$ T distinction may have been a bit more difficult for subjects to perceive than the S $\sim$ T distinction, especially for labials and dorsals; reaction times are slightly slower to the former trials than to the latter. However, the effect is much less robust than the difference by comparison type for Experiment 2; the differences are sometimes significant and sometimes not, depending on how reaction time is measured and which subjects are analyzed – factors that did *not* affect the main result for Experiment 2.

This constellation of results suggests that although there is a difference in perceptibility between the Z $\sim$ D and T $\sim$ D contrasts, there is no difference of a comparable magnitude between the S $\sim$ T and D $\sim$ T contrasts. However, it is possible that the lack of a robust result in Experiment 4 is due to its smaller sample size: for example, the d's in figure 4.16 are calculated from 540 observations each, while those in figure 4.3 are calculated from at least 600. To determine whether

Figure 4.20: Sensitivity to Z $\sim$ D vs. T $\sim$ D by subject in a subset of data from Experiment 2. Within-subject differences are non-significant at $\alpha = .05$ only for subject 1. For each d', $474 \leq n \leq 488$



Experiment 4 simply lacked the power to detect a difference in perceptibility as large as the one found in Experiment 2, I analyzed a random subset of the data from Experiment 2. For each subject and trial type (Z $\sim$ Z, D $\sim$ D, T $\sim$ T, Z $\sim$ D, and T $\sim$ D), I randomly selected 180 trials from Experiment 2 to include in the subset (the same as the number of trials for each subject and trial type in Experiment 4[8]). Thus, this subset of data represents what the results of Experiment 2 might have looked like if it had had the same power as Experiment 4. Figure 4.20 shows subjects' sensitivity in this 'smaller' Experiment 2; for every subject except subject 1, the difference between the Z $\sim$ D and T $\sim$ D comparison types is still significant. I conclude that Experiment 4 had ample power to detect a difference in perceptibility as large as the one found in Experiment 2.

---

[8] In many cases, there were fewer than 180 trials for a given subject and trial type in Experiment 2; in those cases, I included all of the relevant trials in the subset.

Effect of Acoustic Characteristics of Stimuli

In §4.3.2, I described four between-talker differences among the stimuli that might have influenced subjects' perception of the relevant contrasts: aspiration in the [aTa] stimuli; and consonant duration, center of gravity, and intensity in the [aSa] stimuli.

Greater aspiration in the [aTa] stimuli provides an extra cue to the voicing distinction and therefore may increase the perceptibility of the D $\sim$ T contrast. As shown in figure 4.12, the German- and English-speaking talkers produced more aspiration than the Bulgarian-speaking talker. However, subjects do not differ in sensitivity to the D $\sim$ T contrast for any of the three talkers ($d' = 3.0$ for talker 1, 3.2 for talker 2, and 3.1 for talker 3; none of the differences are significant). Alternatively, aspiration in the [aTa] stimuli might *decrease* the perceptibility of the S $\sim$ T contrast if subjects are likely to misperceive aspiration as frication noise or vice versa. However, the three talkers do not differ in sensitivity to the S $\sim$ T contrast either ($d' = 3.2$ for talker 1, 3.1 for talker 2, and 3.3 for talker 3; none of the differences are significant). Thus, it is unlikely that aspiration (or lack thereof) among some talkers is masking an effect of comparison type in Experiment 4.

As shown in figures 4.13 – 4.15, there are a number of differences among talkers at each place of articulation in the duration, center of gravity, and intensity of the voiceless fricatives; any of these differences has the potential to influence subjects' ability to perceive the S $\sim$ T contrast. Among all three talkers and all three places of articulation, there are only two cases in which two talkers differ significantly in subjects' sensitivity to the S $\sim$ T contrast at some place of articulation: subjects are significantly less sensitive to the [aθa] $\sim$ [ata] contrast for talker 2 ($d' = 2.6$) than for either talker 1 ($d' = 3.0$; $p = .046$) or talker 3 ($d' = 3.2$; $p = 2.0 \times 10^{-3}$). This difference might be attributable to talker 2's particularly short and intense [θ]s, although we cannot be certain.

It is possible, then, that the significant difference in sensitivity between [aθa] $\sim$ [ata] and [ada] $\sim$ [ata] that subjects exhibited for talker 2 is an anomaly. If so, then the only remaining significant effects in figure 4.18 would be cases where sensitivity to S $\sim$ T is smaller than sensitivity to D $\sim$ T, suggesting that the latter truly is the more salient distinction. However, even eliminating this significant effect would not eliminate all of the variability seen in the d' results, nor would it render the difference (if real) by comparison type large enough to be of comparable magnitude to the one found in Experiment 2. Thus, talker-particular characteristics of the voiceless fricatives in the [aSa] stimuli do not seem to be masking a large overall effect of comparison type in Experiment 4.

Effect of English Consonant Inventory

On the basis of the English consonant inventory, we would expect there to be an effect of comparison type among dorsals: subjects should be more sensitive to the [aga] $\sim$ [aka] contrast (phonemic in English) than to the [axa] $\sim$ [aka] contrast (since English lacks [x]). Interestingly, the results broken down by place in figures 4.17 and 4.18 show no such effect. This lack of a result suggests that the experimental task may have influenced subjects to tap into more fine-grained phonetic differences and to be less influenced by their native inventory.[9] Alterna-

---

[9] It is also possible that Experiment 4 lacked the power to detect a difference that was actually present among dorsals. As discussed above, this experiment had ample power to detect differences of the magnitude found in Experiment 2; thus, any difference among dorsals that did

tively, subjects may have assimilated [x] to a native category other than [k] (Best et al. 1988).

For labials and coronals, however, both comparison types are on equal footing: [p], [b], [f], [t], [d], and [θ] are all phonemic in English. Any difference by comparison type for labials or coronals *must* have been due to inherent differences in perceptibility between the continuancy and voicing contrasts (modulo talker-specfic effects). No such consistent differences were found.

4.4 Discussion and Implications

The results of Experiment 2 show that the distinction between voiced and voiceless stops intervocalically is perceptually more salient than the distinction between voiced stops and spirants in the same environment. This finding suggests that a perceptual account of the direction of lenition along the lines of the P-map is viable: we could hypothesize that intervocalic devoicing is unattested because intervocalic spirantization is perceptually a better (less salient) option.

The results of Experiment 4 show that there is no difference of a comparable size in perceptibility between voiced and voiceless stops vs. voiceless stops and fricatives intervocalically. This finding, too, is compatible with the P-map: voiceless stops may undergo lenition *either* by voicing *or* by spirantization, because neither alternation is perceptually superior to the other. (See §5.1 for a discussion of the implications of a possible small difference in comparison type among labials.)

The results of Experiment 3 show that for voiced stops, the difference in per-

---

not show up in the present results must be relatively small.

ceptibility between the continuancy and voicing distinctions is modulated by place of articulation: the effect of comparison type is greater for dorsals, for example, than it is for labials. If the implications of this perceptual fact for phonological theory are analogous to those for the facts shown by Experiment 2, then we should predict that spirantization of labials (a more salient change) implies spirantization of dorsals (a less salient change, and therefore to be preferred). However, as discussed in §2.2, this prediction is not borne out: if anything, spirantization of labials is *more* common than spirantization of dorsals. Table 2.8 shows that spirantization of labials does not imply spirantization of dorsals; there are ten languages in the database that spirantize labials only (Apatani, Assamese, Bashkir, Dahalo, Cardiff English, Kagate, Nepali, Nkore-Kiga, Ayt Seghrouchen Tamazight Berber, and Chitwan Tharu).

Thus, Experiments 2 – 4 make some correct predictions with respect to phonological typology and some incorrect ones. Chapter 5 discusses the implications of these results for phonological theory.