PERCEPTUAL SIMILARITY IN NON-LOCAL LARYNGEAL RESTRICTIONS[*][#]
Gillian Gallagher (NYU)
November, 2010

0   Introduction

This paper addresses the treatment of non-local restrictions on laryngeal features, focusing on the role of perceptual similarity and perceptual grounding in understanding non-local phonotactic patterns. The phenomena in question are instances in which the laryngeal features of one segment in a root are restricted based on the laryngeal features of some other non-adjacent segment. An example of such a restriction is a language where roots may have one ejective but not two, ✓ *k'apa* *k'ap'a*.

I will make two claims about this type of restriction. First, I will argue that long-distance restrictions on laryngeal features are restrictions on contrast. Non-local phonological restrictions mirror long-distance conditions on the perceptual strength of a laryngeal contrast. For example, a language may disallow roots with two ejectives because roots of this type are judged too perceptually similar to a root with only one ejective. I present results from two perception experiments and show that the perceptual strength of a contrast in ejection or aspiration correlates with the typology of long-distance restrictions on these segments.

The second claim is that laryngeal cooccurrence restrictions involve restrictions on auditory properties, as opposed to abstract phonological features or articulatory categories (though articulation may also play a role). The cooccurrence restrictions on ejectives and aspirates in Quechua provide evidence for grouping these two types of segments based on their shared property of long VOT. A third experiment supports the idea that long VOT is relevant in explaining the interaction of ejectives and aspirates, thus supporting the proposal that auditory dimensions of contrast are relevant to phonology (Flemming 1995; Steriade 1997).

1   The contrast based approach to long-distance assimilation and dissimilation

Long-distance restrictions on laryngeal features loosely fall into one of the two patterns in (1). In a language with the pattern I refer to as 'dissimilation', stops in a root may not share the same laryngeal feature, (1a). In a language with the pattern I refer to as 'assimilation', stops in a root *must* share the same laryngeal feature; stops with different laryngeal features may not cooccur, (1b).

(1) a. <u>Dissimilation</u>   ✓ K'-T   *K'-T'
    b. <u>Assimilation</u>   *K'-T   ✓ K'-T'

Long-distance restrictions commonly apply to ejectives, aspirates and implosives. These two patterns are exemplified for ejectives by Secwepemctsín (formerly known as Shuswap; Kuipers 1974; MacEachern 1999) and Chaha (Rose and Walker 2004; Rose and King 2007). In Secwepemctsín, as can be seen in (2), roots may have one ejective and one plain stop, but no root may have two ejectives.[1]

(2) Dissimilation – Secwepemctsín [p, t, ts, k, kʷ, q, qʷ; p', t', ts', k', kʷ', q', qʷ']
      s-**k'**le**p**   'coyote'   *s-**k'**le**p'**
      qʷi**ts'**   'to wash'   ***qʷ'**i**ts'**

Dissimilatory restrictions on laryngeal features are also found in Chol (Mayan) (Gallagher and Coon 2009), as well as the languages surveyed by MacEachern (1999): Hausa (Afro-Asiatic), Tz'utujil (Mayan), Ofo (Siouan), Sanskrit (Indo-Aryan), Old Georgian (Caucasian), Gojri (Indo-Aryan), Souletin Basque (isolate), Cuzco Quechua (Quechuan), Bolivian Aymara (Aymaran), and Peruvian Aymara (Aymaran).

    In Chaha, alveolar and velar stops that cooccur in a root must both be voiceless unaspirated, or both ejective; there are no roots with one ejective and one voiceless unaspirated stop.

(3) Assimilation – Chaha [t, k; t', k']
      jɨ-**k**əf**t**   'he opens'   *jɨ-**k'**əf**t**
      jɨ-**t'**əβ**k'**   'it is tight'   *jɨ-**t**əβ**k'**

Assimilatory restrictions on laryngeal features are also found in Amharic (Semitic) (Rose and King 2007), Kalabari Ijo (Ijoid) (Hansson 2001; Jenewari 1989), Bumo Izon (Ijoid) (Hansson 2001; Efere 2001), Zulu, Ndebele, Xhosa and Swati (Bantu) (Hansson 2001; Khumalo 1987), and Gitksan (Tsimshianic) (Brown 2008). Dissimilation seems to be more common than assimilation. Whether this is in fact the case, and what the explanation for such asymmetry would be, is left to future research.

    These two types of patterns raise two broad questions for phonological theory. First, why do restrictions on laryngeal features operate long-distance? Many phonological phenomena are local, meaning that string-adjacent segments interact with one another. While long-distance phonological phenomena are also well attested, not all features show the same range of restrictions as laryngeal features. Major place, continuancy, sonorancy and nasality all show a much more restricted typology of long-distance interactions, or no long-distance interactions at all (see Hansson 2001 for an extensive survey of non-local consonant assimilation). Additionally, not all local phonological phenomena have non-local counterparts. For example, post-nasal voicing is extremely common between a nasal and an immediately following stop, but is unattested between a nasal and a non-adjacent stop.

---

[1] Secwepemctsín also has a series of glottalized sonorants, which freely cooccur with ejectives, e.g. ***t'uxʷnˀ*** 'scouring rush'. Roots with pairs of glottalized sonorants are independently restricted due to the general absence of glottalized sonorants in initial position. Glottalization on sonorants and obstruents thus patterns differently (cf. Steriade 1997).

(4)   a. local post-nasal voicing
            Zoque (Wonderly 1951): /m+pama/ → [mbama]   'my clothing'
        b. unattested non-local post-nasal voicing
            /ma+paka/ → [mabaka]

These asymmetries show that locality is not a fully independent parameter in the phonological grammar, but rather that the locality conditions on a phonological pattern are an integral part of that pattern.

    The second major question raised by the patterns of dissimilation and assimilation is why languages exhibit two seemingly contradictory cooccurrence patterns. Dissimilatory and assimilatory restrictions are opposites; what is required in one type of pattern is disallowed in the other. The contradictory nature of these restrictions is a challenge for a standard theory of markedness, since it is impossible to claim that one of the given laryngeal configurations is both more and less marked than the other.

    The argument here is that the answers to both of the above questions lie in the domain of contrast strength and similarity. The idea is that laryngeal features like ejection and aspiration are subject to non-local restrictions because they interact non-locally in perception. The strength of a contrast in ejection or aspiration is affected by the non-local context. Specifically, the idea is that a laryngeal contrast is degraded if there is another laryngeally specified stop in the same root. Thus, an ejective-plain contrast is stronger in a pair of roots like [**k'**api, **k**api], where there are only plain stops in the contrasting roots, than in a pair like [**k'**ap'i, **k**ap'i], where there is another ejective in each of the contrasting roots. This hypothesis is schematized in (5). The hierarchy in (5a) states that a contrast between one and two laryngeally marked stops in a root is perceptually weaker than a contrast between one and zero laryngeally marked stops in a root. The hierarchy in (5b) is equivalent, but states the asymmetry in somewhat different terms, following Steriade (2001). Here, the perceptual distance (indicated by "Δ") between a laryngeally marked and a plain stop is weaker in a root with another laryngeally marked stop than in a root with another plain stop.

(5) a.   1 vs. 2 (e.g. [k'ap'i, kap'i])   <   1 vs. 0 (e.g. [k'api, kapi])
    b. Δ[K'-K] / [...K'...]$_{RT}$ < Δ[K'-K] / [...K...]$_{RT}$

Given the hierarchies in (5), the existence of non-local phonological restrictions on laryngeal features is unsurprising, as the non-local nature of the phonological pattern correlates with the hypothesized non-local nature of the perceptibility facts. Additionally, the existence of both dissimilatory and assimilatory patterns is unsurprising. If a contrast between roots with one and two laryngeally marked stops is perceptually challenging, languages may disprefer having both types of roots. In dissimilatory languages, forms with two instances of a laryngeal feature are disallowed because they are not sufficiently perceptually distinct from forms with a single instance of that feature. Similarly, in assimilatory languages, forms with one instance of a laryngeal feature are disallowed because they are not sufficiently perceptually distinct from forms with two instances of that feature. The same grammatical pressure against the contrast in one vs. two laryngeal features can be satisfied by allowing only forms with one laryngeally marked stop or only forms with two laryngeally marked stops. This understanding of dissimilatory and assimilatory restrictions is schematized in (6). Given the four possible laryngeal configurations for a root with two stops, there is a contrast between forms with two,

one and zero laryngeally marked stops. From this set of potential forms, languages with assimilatory restrictions select the forms with two and zero laryngeally marked stops, while languages with dissimilatory restrictions select the forms with one and zero laryngeally marked stops. In each type of language *either* a form with one *or* a form with two laryngeally marked stops is grammatical, never both.

(6)　　　　K'-T'　　　K'-T　　　K-T'　　　K-T　　　　　2 vs. 1 vs. 0
　　　　　　　|　　　　　　　　　　　　　　　　　　|
　　　　　　[K'-T', K-T]　　　　　　[K'-T, K-T', K-T]
　　　　　2 vs. 0, assimilatory　　1 vs. 0, dissimilatory

By considering laryngeal cooccurrence restrictions as restrictions on *contrasts* between forms, the apparent contradiction in dissimilatory and assimilatory patterns is resolved. The next section provides experimental support for the asymmetries schematized in (5).

2   Experimental support

The general hypothesis is that the contrast between forms with one and forms with two instances of a laryngeal feature is perceptually confusable, and thus neutralized. This hypothesis is tested and supported by two discrimination experiments, testing the perceptual strength of ejective and aspirate contrasts. The ejective experiment is also presented and discussed in Gallagher (2010b).

　　The two experiments test the perception of ejective-plain and aspirate-plain contrasts. These experiments compare three contrasts in laryngeal configurations between stimuli: 1 vs. 0 (*k'api-kapi*), 2 vs. 0 (*k'ap'i-kapi*) and 1 vs. 2 (*k'ap'i-kap'i*) for both ejectives and aspirates. The results show that for both ejective and aspirate contrasts, the 1 vs. 2 contrast is more difficult than the 1 vs. 0 contrast, which is in turn more difficult than the 2 vs. 0 contrast.

2.1 The hypotheses

The first hypothesis is that the strength of a laryngeal contrast varies depending on the non-local context. This hypothesis is stated in (7).

(7)　Hypothesis 1　　Pairs of roots that contrast 1 vs. 2 ejectives/aspirates are less distinct than pairs of roots that contrast either 1 vs. 0 or 2 vs. 0 ejectives/aspirates.
　　　　　　　　　　$\Delta([T'…K']-[T'…K])\ <\ \Delta([T'…K]-[T…K]), \Delta([T'…K']-[T…K])$

The second hypothesis is that a contrast in two laryngeal features is more perceptible than a contrast in just one laryngeal feature.

(8)　Hypothesis 2　　Pairs of roots that contrast 1 vs. 0 ejectives/aspirates are less distinct than pairs of roots that contrast 2 vs. 0 ejectives/aspirates.
　　　　　　　　　　$\Delta([T'…K]-[T…K])\ <\ \Delta([T'…K']-[T…K])$

These two hypotheses are summarized in (9a), which can be restated as (9b).

(9) a.   1 vs. 2   <   1 vs. 0   <   2 vs. 0
         k'api-k'ap'i   *k'api-kapi*   *k'ap'i-kapi*

    b. $\Delta([T'\ldots K']\text{-}[T'\ldots K])$ $<$ $\Delta([T'\ldots K]\text{-}[T\ldots K])$ $<$ $\Delta([T'\ldots K']\text{-}[T\ldots K])$

For some contrast category A to be "weaker" than some contrast category B, subjects should have more difficulty discriminating pairs of forms that fall into category A than B (i.e. incorrectly think that pairs of different words are the same). The experiments test these hypotheses by presenting subjects with pairs of CVCV nonce words that differ only in whether the consonants are ejective or plain. Subjects are then asked to decide whether the words they hear are the same or different from one another (an AX discrimination task).

2.2 The stimuli

The stimuli for this experiment are pairs of CVCV disyllables manually spliced together from recordings of South Bolivian Quechua. Recordings of a middle-aged female speaker of South Bolivian Quechua were made in Cochabamba, Bolivia by the author, using a Marantz PMD660 solid-state recorder and Audio Technica 831b microphone. The speaker was asked to read phonotactically legal nonce roots from a computer screen, embedded in the carrier phrase *Noqa X simita qellqani* 'I wrote the word X'. The stops in pairs of stimuli differ only as to whether the consonants are plain or ejective (experiment 1) or aspirate (experiment 2).

    The target stimuli have one of four laryngeal configurations: CVCV (0 ejectives), C'VCV (1 ejective, initial position), CVC'V (1 ejective, medial position), or C'VC'V (2 ejectives). Three places of articulation (labial, alveolar and velar) and two vowel patterns (a-i and i-u) were used. Not all four of these desired laryngeal configurations are grammatical in Quechua, however, so it was not possible to record all four types of words. Quechua is a dissimilating language that does not allow pairs of ejectives in a word. Additionally, Quechua does not allow non-initial ejectives in words with two stops (e.g. *sap'a* 'a kind of basket', but **tap'a*). Consequently, nonce words were constructed which conformed to the phonotactics of Quechua for the purposes of recording, and the experimental stimuli were made by splicing together individual syllables. The recorded words took one of three grammatical laryngeal configurations in Quechua: CVCV (two plain stops), C'VCV (initial ejective/aspirate, medial plain stop) or sVC'V (initial [s], medial ejective/aspirate). Stress consistently fell on the initial syllable of the nonce root.

    The experimental stimuli were made by splicing together CV sequences from the original recording, using the Praat software for speech analysis (Boersma and Weenink 2010). Stimuli were spliced together during the closure of the second C, keeping VC transitions intact and resulting in natural sounding stimuli. A stimulus like *k'ap'i*, for example, was made by splicing the first CV and VC transition of *k'api* with the second CV of *sap'i*. Initial and final syllables in the stimuli were always spliced from original initial and final syllables in the recorded speech.

    Pairs of stimuli were either "same" or "different", and fell into one of the following categories (only examples with ejectives are shown; the same categories were used for aspirates.

(10)	same		different	
	0 0	e.g. *pitu-pitu*	0 1	e.g. *pitu-**p'**itu* or *pitu-pit'u*
	1 1	e.g. *p'itu-p'itu* or *pit'u-pit'u*	0 2	e.g. *pitu-p'it'u*
	2 2	e.g. *p'it'u-p'it'u*	1 2	e.g. *p'itu-p'it'u* or *pit'u-**p'**it'u*

5

The 'different' pairs were presented in both possible orders (1–0 and 0–1, 2–0 and 0–2, 1–2 and 2–1). In stimuli pairs where one stimulus has only one ejective, the ejective could be in either initial or medial position (C'VCV or CVC'V). The total number of stimuli in each of the 6 categories in (10) was 48, resulting in a total of 288 pairs presented to each subject for each experiment.

Acoustic analysis of the stimuli revealed that the main correlates of the ejective-plain contrast are VOT (avg: 131 ms in ejectives, 27 ms in plain stops) and burst amplitude, measured in arbitrary units as the difference between the highest and lowest points in the waveform (avg: .7 in ejectives (about 70 dB), .3 in plain stops (about 64 dB)). Ejectives have a much louder burst than plain stops, and are followed by a long period of silent VOT before the modal voicing of the following vowel begins. The aspirate-plain contrast is similarly cued by longer (noisy) VOT in the aspirate (avg:155 ms), and breathy phonation at the onset of the following vowel, measured as H1-H2 of the first 30 ms of the vowel (Ladefoged 2003) (avg: 1.7 db in aspirates, -.7 db in plain stops).[2] More detailed acoustic analysis of laryngeal contrasts in Quechua is available in Gallagher (2010a).

2.3 Procedure

Subjects were presented auditorily with pairs of nonce roots and asked to decide whether the roots they heard were the same or different from one another. The experiment was presented using the Psyscope X software. Subjects indicated their response by pressing one of two clearly marked keys on a standard USB keyboard, the position of "same" and "different" keys were the same for all subjects. Subjects listened to the stimuli through a pair of high-quality headphones, while looking at a computer screen. The stimuli were presented while the computer screen was black, with a 300 ms inter-stimulus interval. After the second stimulus, a green line appeared on the screen. Subjects were told to indicate their response as quickly as possible once the green line appeared, but not before. All subjects thus listened to both stimuli in their entirety before indicating their choice of "same" or "different". Response time was limited to 1500 ms. If subjects did not respond within this time, they automatically went on to the next trial.

The subjects were 19 speakers of American English with no exposure to a language with ejectives. English speakers were chosen as subjects instead of native Quechua speakers to avoid a phonological bias in perception. The English subjects were told that the sounds they were hearing were not from English, and would sound foreign. Each subject listened to five trial pairs to accustom themselves to the stimuli and task. The experiment took about 20 minutes, including a break halfway through, and subjects were paid for their participation.

2.4 Results

Subjects' performance supports the two hypotheses. The 1 vs. 2 contrast is indeed weaker than other laryngeal contrasts, for both ejective and aspirate contrasts. Additionally, the 1 vs. 0 contrast is significantly weaker than the 2 vs. 0 contrast for both ejectives and aspirates.

A binomial Linear Mixed Model was run, using the "lmer" function in the R software, with "correct" as the dependent variable, a fixed effect of contrast category (1 vs. 2, 1 vs. 0, 2 vs. 0), and a random effect of subject. For ejectives, the model shows a significant difference between

---

[2] H1-H2 was measured off of a spectrum of the recorded sound, no correction was done for the radiation factor or to reverse the effects of formants.

the 1 vs. 0 and 1 vs. 2 contrast categories (β=-1.24, St. Err.=.12, t=-10.6, p < .0001) and the 1 vs. 0 and 2 vs. 0 contrast categories (β=1.19, St. Err.=.17, t=7.19, p < .0001). For aspirates, the same result was found, 1 vs. 0 differs from 1 vs. 2 (β=-0.6, St. Err.=.11, t=-5.71, p < .0001) and 1 vs. 0 differs from 2 vs. 0 (β=1.04, St. Err.=.13, t=8.17, p < .0001).

For discussion of Mixed Models and why this is the appropriate model for binary forced choice tasks, see Jaeger (2008). The statistical model only tested for differences between subjects' performance on the different categories. Performance on same trials was very high across the board, showing both that subjects were able to accurately perceive ejectives and that subjects do not generally misperceive pairs of same tokens as different. The results on the same trials are taken into account in computing the d-prime statistic of sensitivity, discussed below. Performance on the different trials for ejectives is shown in Figure 1, and aspirates in Figure 2.

It should be noted that the two comparisons between contrast categories here have a somewhat different character from one another. In the 1 vs. 2 and 1 vs. 0 categories, the contrasting stimuli differ in the laryngeal features of one segment. In the 2 vs. 0 category, however, the contrasting stimuli differ in the laryngeal features of two segments. The supposed greater perceptibility of the 2 vs. 0 contrast is likely deducible from general psychological or cognitive principles, and is not necessarily specific to a linguistic task or to laryngeal contrasts.
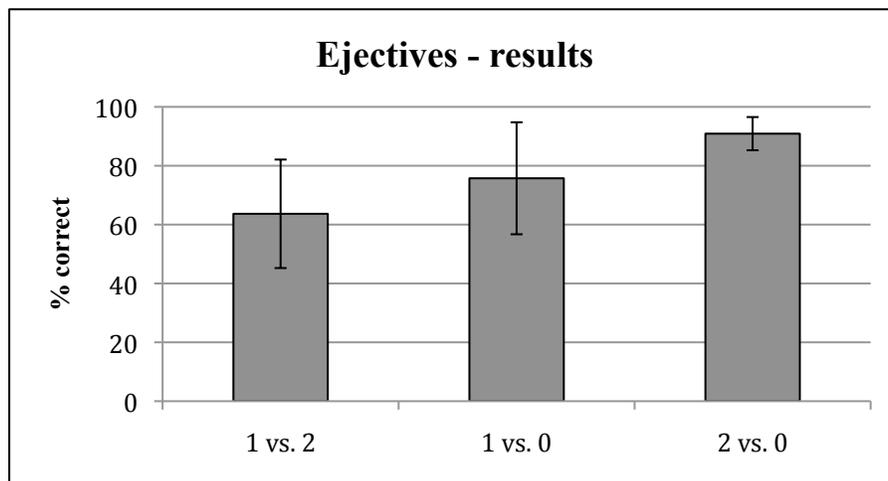
Figure 1:   Percent correct by contrast category, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.
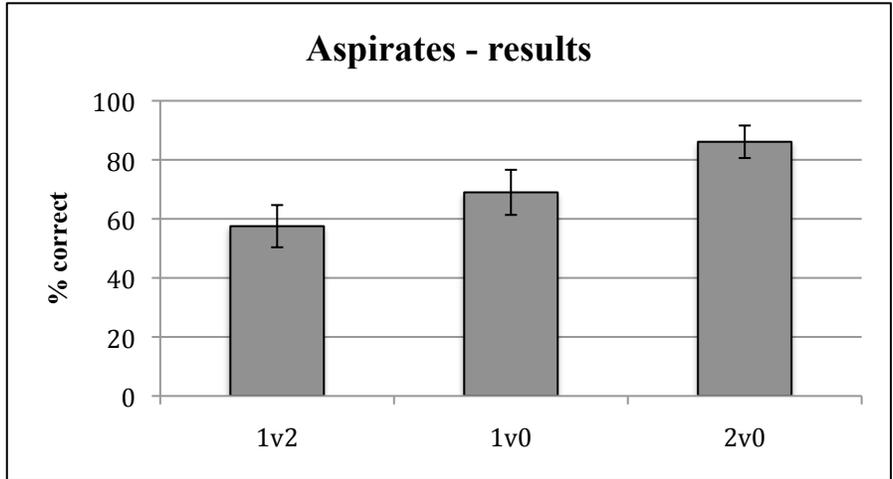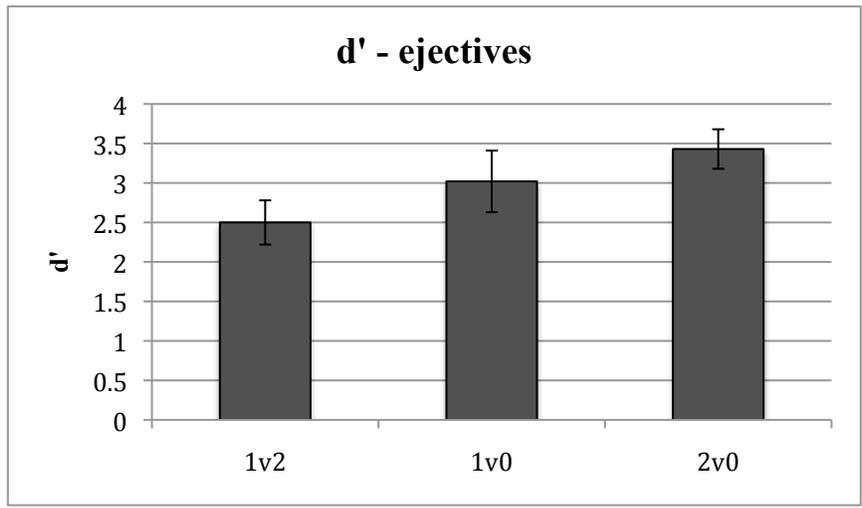
Aspirates - results

Figure 2:   Percent correct by contrast category, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.

The d-prime statistic (MacMillan and Creelman 2004), a measure of sensitivity, was also calculated. Figures 3 and 4 show average d-primes for each of the contrast categories for ejectives and aspirates. The d-prime for a given category is computed based on the hit rate for the relevant different category (1 vs. 2, 1 vs. 0, 2 vs. 0) and the false alarm rate for two of the same categories (1 vs. 1 and 2 vs. 2, 1 vs. 1 and 0 vs. 0, 2 vs. 2 and 0 vs. 0, respectively). A one way ANOVA shows a significant effect of contrast category in determining d-prime (ejectives: $F_{(2,53)}=8.3$, $p < .001$; aspirates: $F_{(2,50)}=3.0$, $p < .03$); two ad-hoc two tailed t-tests confirm that 1 vs. 2 differs from 1 vs. 0 (ejectives: $p < .01$; aspirates: $p < .01$) and 1 vs. 0 differs from 2 vs. 0 (ejectives: $p < .0001$; aspirates: $p < .001$).

d' - ejectives

Figure 3:   D-prime by contrast category, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.
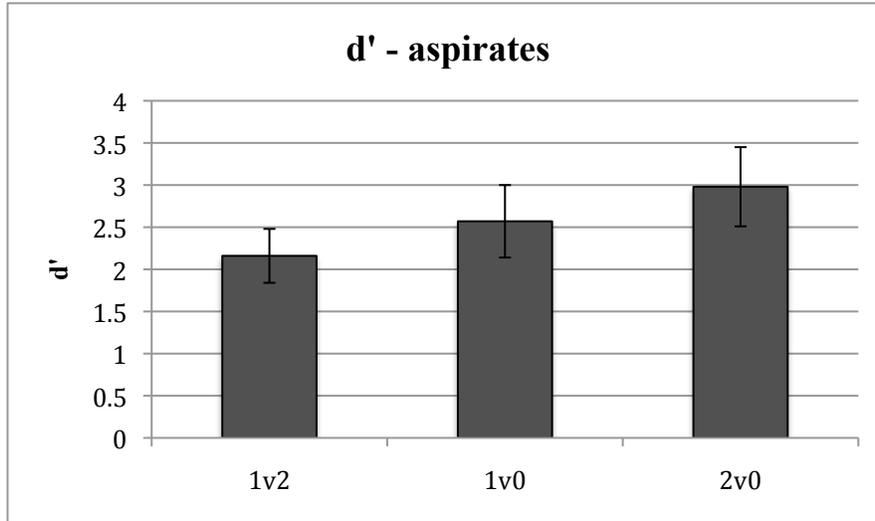
Figure 4: D-prime by contrast category, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.

There are two factors other than the category of laryngeal contrast that could have influenced performance. First, it is possible that subjects' discrimination of the ejective-plain and aspirate-plain contrasts varies depending on the place of articulation of the stop in question (e.g. *k'api-kapi* vs. *t'api-tapi*). Second, the position of the contrast may have an effect (e.g. *k'api-kapi* vs. *kap'i-kapi*). A second Linear Mixed Model was run with fixed effects of contrast category (1 vs. 0 and 1 vs. 2), position of contrast (initial or medial) and place of articulation of contrast (labial, coronal or velar). Only the 1 vs. 0 and 1 vs. 2 contrast category were analyzed in this model because only in these conditions is there a single locus of contrast that can be evaluated for place or position of contrast. The results of interest are given in Table 1 and Table 2. The baseline category is 1 vs. 0, baseline position is initial and baseline place is velar.

fixed effects	estimate	st. error	t value	p value
category	-1.6	0.47	-3.4	< 0.001
position	-2.83	0.44	-6.42	< 0.0001
place(labial)	.45	0.65	0.69	p > 0.05
place(coronal)	-1.7	0.46	-3.7	p < 0.01
category x position	0.06	0.53	0.11	p > 0.05
category x place(labial)	-0.8	0.73	-1.1	p > 0.05
category x place(coronal)	0.48	0.55	0.87	p > 0.05

Table 1: Ejectives – LMM with 'correct' as the dependent variable and fixed effects of category, position and place.

fixed effects	estimate	st. error	t value	p value
category	-1.14	0.36	-3.15	< 0.01
position	0.4	0.29	1.4	p > 0.05
place(labial)	0.77	0.29	2.65	p < 0.01
place(coronal)	1.04	0.32	3.24	p < 0.01
category x position	0.78	0.42	1.87	p > 0.05
category x place(labial)	-0.38	0.43	-0.89	p > 0.05
category x place(coronal)	-0.4	0.41	-0.99	p > 0.05

Table 2:   Aspirates – LMM with 'correct' as the dependent variable and fixed effects of category, position and place.

The above results show a significant effect for position in ejectives, and a significant effect of place for both ejectives and aspirates, but no significant interactions.

Figures 5 and 6 show performance on the 1 vs. 2 and 1 vs. 0 categories, broken down by place of articulation. Performance on contrasts at the labial, coronal and velar places of articulation vary somewhat, but within each place of articulation 1 vs. 2 is always worse than 1 vs. 0, and performance on 1 vs. 2 at any place of articulation is worse than performance on 1 vs. 0 at any place of articulation.

Figures 7 and 8 show performance on the 1 vs. 2 and 1 vs. 0 categories, broken down by position. For the ejectives in Figure 7, it can be seen that performance is better when a contrast is in initial position than when it is in medial position. Performance on 1 vs. 2 in initial position is slightly better than performance on 1 vs. 0 in medial position, reversing the main effect of contrast category, but this difference is not significant.
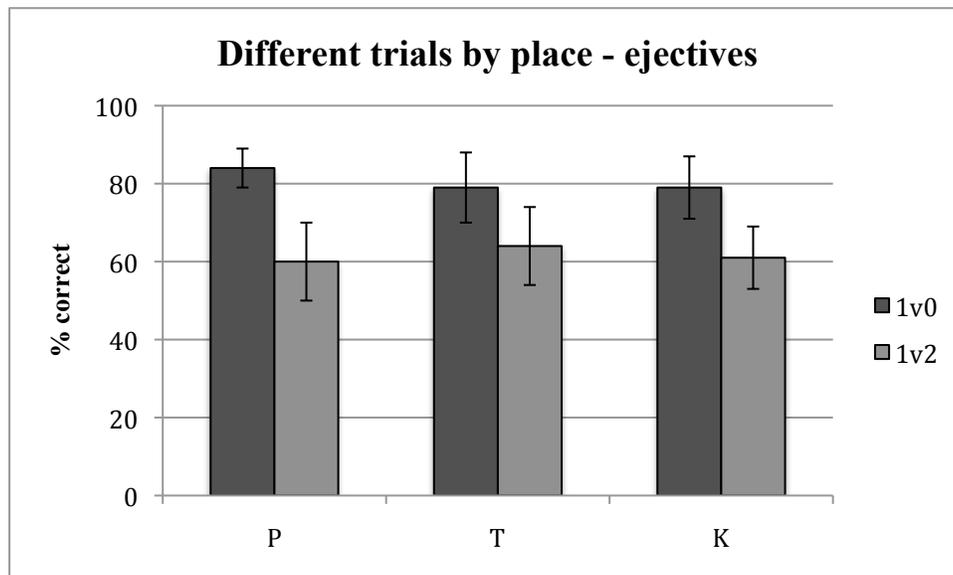

Figure 3:   Ejectives – percent correct for 1 vs. 0 and 1 vs. 2 contrast categories by place of articulation, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.
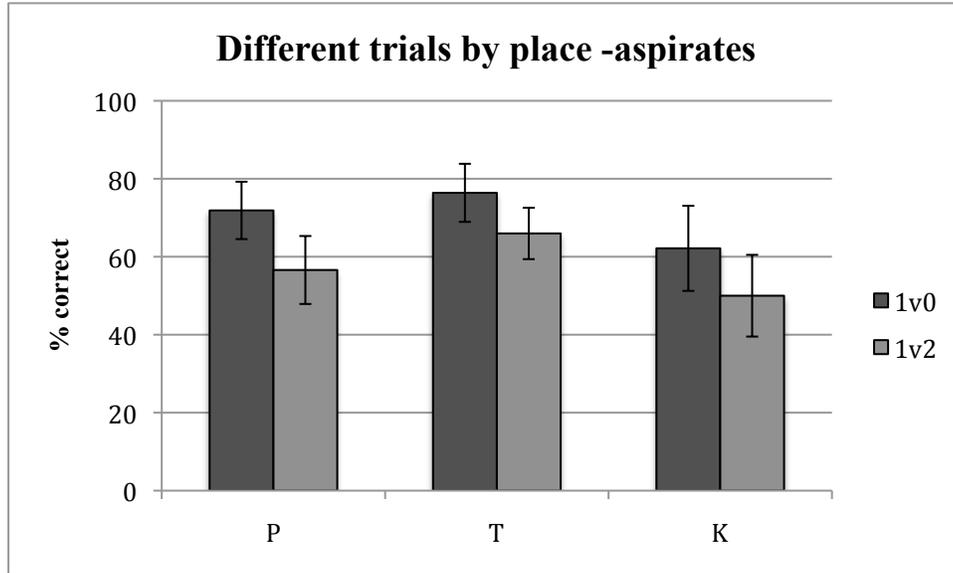
Figure 4:   Aspirates - Percent correct for 1 vs. 0 and 1 vs. 2 contrast categories by place of articulation, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.
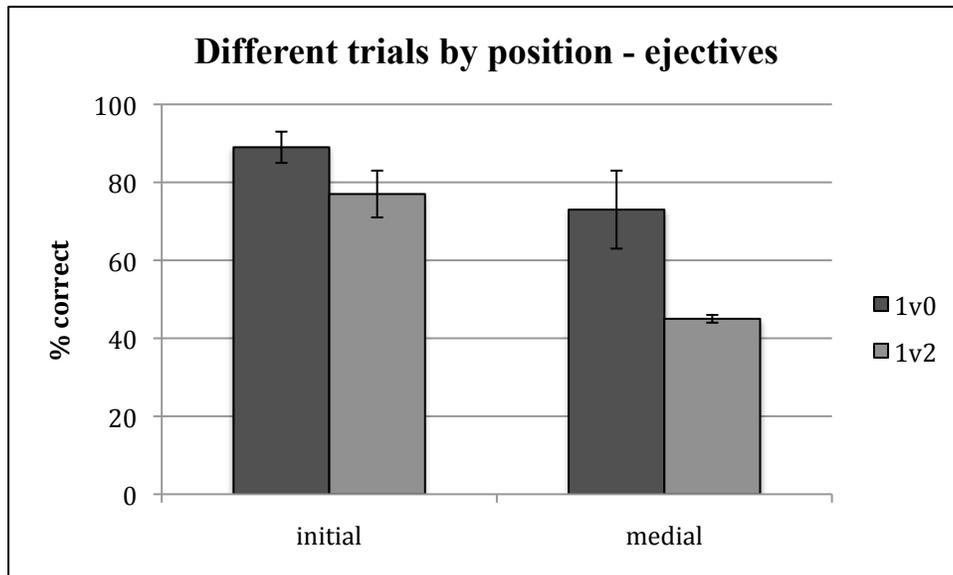

Figure 5:   Ejectives - Percent correct for 1 vs. 0 and 1 vs. 2 contrast categories by contrast position, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.
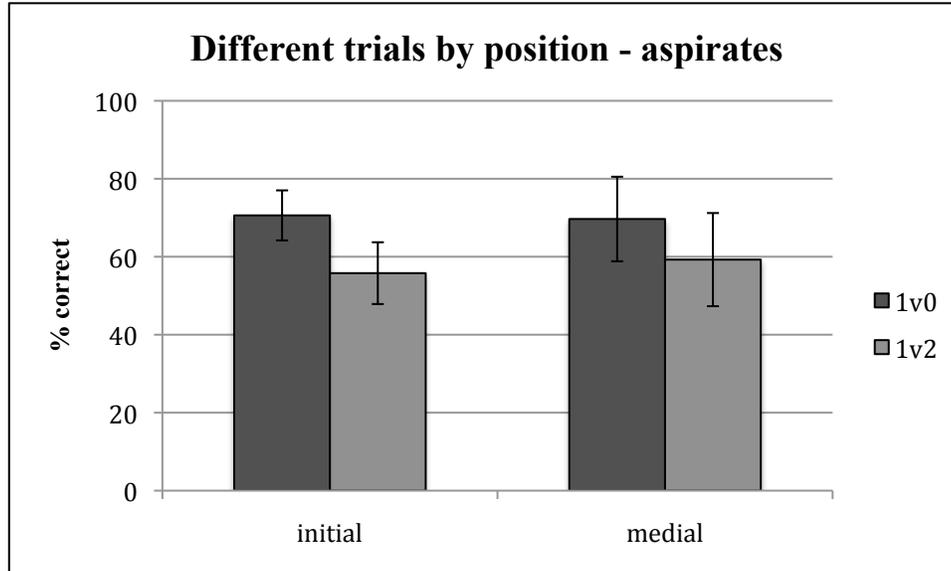
Figure 6:   Aspirates - Percent correct for 1 vs. 0 and 1 vs. 2 contrast categories by contrast position, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.

2.5 Discussion

The perceptual experiments reported here show that long-distance phonological restrictions on laryngeal features have a parallel in the perceptual strength of laryngeal contrasts between roots. It was found that the perception of both ejectives and aspirates, two classes of segments that are subject to non-local phonological restrictions, is degraded if there is another ejective or aspirate in the word.

The key result of these two experiments is that the perceptibility of a laryngeal contrast varies depending on other, non-adjacent segments. Pairs of roots in both the 1 vs. 0 and 1 vs. 2 conditions differ in whether one segment is laryngeally marked or not, e.g. the pair *k'api-kapi* (1 vs. 0) and the pair *kap'i-k'ap'i* (1 vs. 2) both differ in whether the initial consonant is [k] or [k']. Despite the fact that the difference between the roots in these two pairs is acoustically identical, the pair showing a 1 vs. 0 contrast (*k'api-kapi*) is more perceptually distinct than the pair showing a 1 vs. 2 contrast (*kap'i-k'ap'i*). Thus, it is found that the strength of the contrast between an ejective and a plain stop or an aspirate and a plain stop is degraded when there is another ejective or aspirate in the root. Dissimilation and assimilation can then both be understood as the result of a grammatical pressure to neutralize the perceptually indistinct 1 vs. 2 contrast, *{K'-T, K'-T'}.

Under the proposed account, long-distance restrictions are phonetically grounded in perceptual asymmetries, as has been proposed for many local phonological neutralizations (Steriade 1997; Flemming 2004). While local neutralizations are often shown to correlate with the availability of acoustic cues, the experimental findings here show that perceptual asymmetries may also arise in the absence of acoustic ambiguity (i.e. when all the cues to a contrast are available). Both local and long-distance phonological restrictions then reflect conditions on contrast strength, though perceptual asymmetries in contrast strength cannot always be reduced to the availability of locally identifiable cues.

The correlation between cue availability and contrast strength has been an integral part of phonological analyses based on contrast markedness. Consider local laryngeal neutralization as an example. In Ancient Greek, voiceless aspirated and unaspirated stops contrast in pre-sonorant position (11a), but only voiceless unaspirated stops are found in pre-obstruent position (11b) (Steriade 1997).

(11) a. dei**k**-nu:-mi   'I show'   b. dei**k**-teos   'to be shown'
       e**kʰ**-o   'I have'   he**k**-teos   'to be had'

In (11), the root *deik* is invariant, while the laryngeal features of the final stop in *ekʰ* vary depending on context. This root appears with an aspirated stop when followed by a vowel in (11a), but as a voiceless unaspirated stop when followed by an obstruent (11b).

The insight of Steriade's analysis of Ancient Greek (and many other languages that show similar patterns) is that the environment for aspiration neutralization correlates with the availability of acoustic cues to aspiration. The difference between voiceless aspirated and unaspirated consonants is primarily one of VOT; the time between the release of the closure and the onset of voicing is long in an aspirated stop and short in an unaspirated stop. This cue is available only in pre-sonorant position. Before an obstruent or in final position, there is no onset of modal voicing following the stop and thus no VOT. The neutralization pattern for aspiration contrasts mirrors the distribution of phonetic cues to this contrast. Aspiration contrasts are more perceptible in pre-sonorant position than in pre-obstruent or final position, as shown schematically in (12).

(12)   {Kʰ, K} / ___ {[-son], #}   <   {Kʰ, K} / ___ [+son]

The analysis of local laryngeal neutralization outlined above makes crucial reference to phonetic cues. Contrasts in aspiration are dependent on VOT cues, and are prone to neutralization when these cues are absent. In long-distance laryngeal restrictions, however, local acoustic cues are not at issue. The inability of two ejectives to cooccur in a root cannot be reduced to their local context, as the presence of one ejective in a root does not alter the acoustic cues to another ejective elsewhere in the root.

Long-distance laryngeal neutralization initially seems like a very different phenomenon from local neutralization, since the cues to a laryngeal contrast are unaffected by the non-local environment. The perception experiments presented here show that perceptual asymmetries exist independent of acoustic ambiguity and the presence of cues. The 1 vs. 2 contrast in laryngeal features is more difficult than the 1 vs. 0 contrast, despite the fact that the difference between forms in these two contrast pairs is identical. A pair of roots like {k'a**p'**i, k'a**p**i} is more difficult to distinguish than a pair like {ka**p'**i, ka**p**i}, even though the acoustic difference between [p'] and [p] is identical in the two pairs. These experiments thus document an effect of higher level, long-distance perceptual interference, independent of ambiguity in the acoustic signal.

A key result of the experiments and their application to cooccurrence restrictions is to show that long-distance and local laryngeal neutralization both correlate with perceptual asymmetries, though they differ as to whether these perceptual asymmetries can be reduced to the availability of acoustic cues. Long-distance restrictions do not contradict the hypothesis that phonological neutralizations occur to optimize the perceptual distance between contrasting forms. Rather, both local and non-local phonological phenomena show a preference for more perceptible contrasts.

A final point about the relation between the experimental results and the typology of laryngeal restrictions is needed here. While the experiments tested interactions between consonants separated by a single vowel, long-distance phonological restrictions are attested across more intervening segments. For example, in the Secwepemctsín word [s-**k'**le**p**] 'coyote', the interacting consonants are separated by both a consonant and a vowel. Further perception experiments must test the interaction of consonants across more than a single vowel.

3 Further applications – ejective-aspirate interactions in Quechua

The preceding experiments showed that both ejectives and aspirates interact across an intervening vowel in perception, a fact that provides considerable insight into the nature of long-distance phonological restrictions on these features. An outstanding issue is what it is about ejectives and aspirates that makes them subject to these long-distance perceptual asymmetries. The terms "ejective" and "aspirate" cover a range of auditory or articulatory properties. This section proposes that is the auditory property of long VOT, shared by both ejectives and aspirates, that is relevant. There is both typological and experimental evidence in favor of this proposal.

3.1 Typological evidence – Quechua

Quechua (Quechuan) has a three-way contrast among ejective, aspirated and voiceless unaspirated (plain) stops and affricates, as shown in (13).

(13)	labial	alveolar	palatoalveolar	velar	uvular
ejective	p'	t'	tʃ'	k'	q'
aspirate	pʰ	tʰ	tʃʰ	kʰ	qʰ
plain	p	t	tʃ	k	q

Cooccurrence restrictions in Quechua apply to both ejectives and aspirates. No root in Quechua may contain two ejectives, two aspirates or one ejective and one aspirate (MacEachern 1999). The examples in (14), from the Ajacopa et al. dictionary (2007), illustrate the restriction. Roots in Quechua are CV(C)CV, and ejectives and aspirates may only appear in onset position. The optional coda in a root is most often a non-stop. In (14a), examples are given of attested roots with one ejective and one plain stop, one aspirate and one plain stop, or two plain stops. The theoretically possible combinations of two ejectives, two aspirates, or ejective-aspirate pairs given in (14b) are all unattested.

(14) a. **k'**in**t**i   'a pair'   ✓ K'-T
           kʰas**t**uy   'to chew'   ✓ Kʰ-T
           kin**t**u   'a bunch'   ✓ K-T

      b. *__k'__in**t'**i   *K'-T'
           *__kʰ__as**tʰ**uy   *Kʰ-Tʰ
           *__k'__in**tʰ**i / *__kʰ__in**t'**i   *K'-Tʰ

14

The Quechua pattern doesn't follow from two independent restrictions on "ejectives" and "aspirates". The absence of forms with two ejectives or two aspirates follows from the hypothesis that Quechua has an independent restriction on each of these segment types, but the absence of forms with one ejective and one aspirate does not. Hypothetical roots with ejective-aspirate pairs do not fall under the purview of any constraint referring to only ejectives or only aspirates. The restriction against ejective-aspirate pairs in Quechua necessitates reference to ejectives and aspirates as a class, to the exclusion of voiceless unaspirated stops.

Ejectives and aspirates are articulatorily distinct: ejectives involve a constriction of the glottis while aspirates involve spreading of the vocal folds. What these two types of segments have in common is long VOT. The release of an aspirated stop is characterized by a period of aspiration noise before the onset of the following vowel; the release of an ejective is followed by a period of silence, corresponding to glottal closure, before the onset of the following vowel. The interaction of ejectives and aspirates in the cooccurrence restriction of Quechua can be understood as a restriction on long VOT segments. Stating the restriction in auditory terms is natural, considering the perceptually based explanation for cooccurrence restrictions proposed in the preceding sections.

It should be noted, however, that the auditory unity of ejectives and aspirates only holds for those ejectives that are produced with a VOT of comparable length to an aspirate. The production of ejectives has been shown to vary considerably (both between languages and between speakers of the same language, and even for a single speaker), falling into at least two categories (Kingston 1985, Lindau 1984, Warner 1996). 'Strong' ejectives are produced with a loud burst amplitude, long VOT and modal voicing in the following vowel; 'weak' ejectives have a weaker burst amplitude, short VOT and creaky voicing in the following vowel. Ejectives in Quechua fall into the 'strong' category, and form a class with aspirates with respect to long VOT, as can be seen from the acoustic measurements in §2.2 above. Under the current line of analysis, weak ejectives are not predicted to interact with aspirates phonologically. As far as I know, this prediction is correct. The only cases of ejective-aspirate interactions that have been documented in the literature are in Quechua and the Peruvian variety of Aymara, both languages spoken in the Andes.

3.2 Experimental evidence

The proposal is that ejectives and aspirates are subject to long-distance perceptual interactions, and thus long-distance phonological restrictions, because they are both characterized by long VOT. It is the 1 vs. 2 contrast in long VOT that is perceptually difficult. Ejectives and aspirates pattern together in Quechua because they share long VOT, and thus interact in perception. This section presents a third perception experiment that tests the perceptual interaction of ejectives and aspirates. The hypothesis tested in the third experiment is stated in (15).

(15)   Hypothesis: A contrast in long VOT is weaker in a root with another long VOT segment.

If the hypothesis in (15) is true, then ejectives should interfere with the processing of a contrast in aspiration, and aspirates should interfere with the processing of a contrast in ejection. Thus, we predict asymmetries as in (16). In (16a), an aspirate contrast is predicted to be more difficult if there is another ejective in the root than if there are only other plain stops. Similarly, in (16b),

an ejective contrast is predicted to be more difficult if there is another aspirate in the root than if there are only other plain stops.

(16) a. *k'ap^{h}i-k'api*   <   *kap^{h}i-kapi*
      b. *k^{h}ap'i-k^{h}api*   <   *kap'i-kapi*

The stimuli for the third experiment were spliced together in the same method described for the ejective and aspirate experiments in §2. The method was also the same. There were 18 subjects, who were presented with the following "different" pairs.

(17)   1 vs. 2 ejective   e.g. *k^{h}ap'i-k^{h}api*   1 vs. 0 ejective   e.g. *kap'i-kapi*
                                    k'ap^{h}i-kap^{h}i   ***k'api-k**api*
       1 vs. 2 aspirate   e.g. *k'ap^{h}i-k'api*   1 vs. 0 aspirate   e.g. *kap^{h}i-kapi*
                                    k^{h}ap'i-kap'i   ***k^{h}api-k**api*

The duration of VOT in the stimuli was comparable for ejectives and aspirates. Ejectives had an average VOT of 131 ms. and aspirates had an average of 155 ms. of VOT. These VOT values contrast starkly with the average VOT of plain stops, 30 ms.

Subjects' performance supports the hypotheses. The 1 vs. 2 contrast in VOT is weaker than the 1 vs. 0 contrast. The result holds for both contrasts in ejection and aspiration. A LMM was run, with "correct" as the dependent variable, fixed effects of contrast category (1 vs. 0 or 1. vs 2), type of contrast (ejective or aspirate) and position of contrast (initial or medial), and random effect of subject. The results of the model are summarized in table 3. The baseline category is 1 vs. 0, the baseline contrast type is ejective and the baseline position is initial.

fixed effects	estimate	st. error	t value	p value
category	-0.54	0.16	-3.3	< 0.001
position	-1.34	0.16	-8.57	$p < 0.0001$
type	-1.29	0.16	-8.23	$p < 0.0001$
category x position	0.45	0.21	2.11	$p < 0.04$
category x type	0.23	0.21	1.06	$p > 0.05$

Table 3:   Long VOT – LMM with 'correct' as the dependent variable and fixed effects of category, position and place.

The results of the LLM show a main effect of contrast category, position and type, as well as an interaction between category and position. Figure 9 shows that performance is better on 1 vs. 0 contrasts than on 1 vs. 2 contrasts. The average d-prime for 1 vs. 0 condition is 2.14, and for the 1 vs. 2 condition -1.65. A two tailed t-test shows that this difference is significant ($p < .0001$).

Long VOT - results
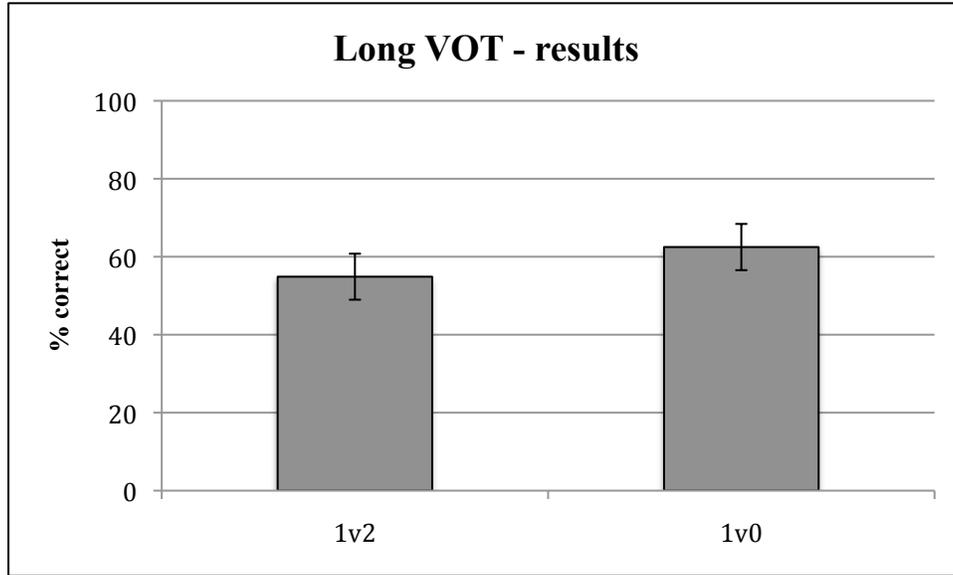
Figure 9:   Percent correct by contrast category, averaged across all subjects.
             Error bars indicate 95% confidence value, showing variation
             between subjects.

Figure 10 shows that performance is better on initial contrasts than on medial contrasts. Moreover, performance on a 1 vs. 2 contrast in initial position is better than performance on a 1 vs. 0 contrast in medial position, reversing the main effect and Figure 11 shows that performance is better on ejective contrasts than on aspirate contrasts.

Different trials by position - long VOT
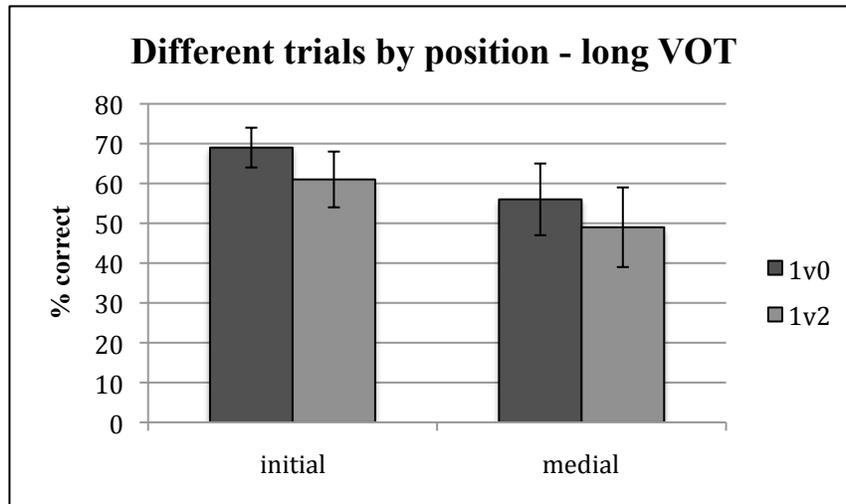
Figure 10:   Percent correct by position and contrast category, averaged across
             all subjects. Error bars indicate 95% confidence value, showing
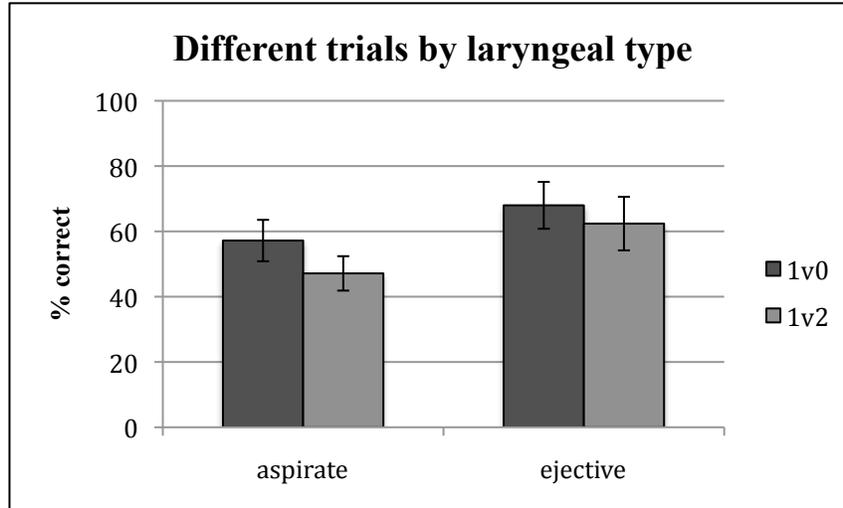   variation between subjects.

Figure 11: Percent correct by laryngeal category and contrast category, averaged across all subjects. Error bars indicate 95% confidence value, showing variation between subjects.

These results support the hypothesis that it is the auditory property of long VOT that is subject to a long-distance perceptual interaction.

4 Conclusion

The proposal developed here is that long-distance cooccurrence restrictions on laryngeal features are restrictions on contrasts, specifically contrasts in certain auditory properties. This line of analysis presents some clear directions for future work and lays the groundwork for answers to the two overarching questions raised by laryngeal cooccurrence restrictions that were introduced in §1.

First, laryngeal cooccurrence restrictions have been shown to correlate with long-distance perceptual interactions. The hypothesis is that only segments with those auditory properties that exhibit long-distance perceptual interactions should be subject to long-distance phonological restrictions. Thus, asymmetries between features in long-distance restrictions are predicted. For example, continuancy is not known to show either long-distance dissimilation or assimilation, and it is thus predicted that continuancy contrasts should not be sensitive to the non-local context. Specifically, the following asymmetries in perceptual strength should *not* hold.

(18) Perceptual asymmetries in continuancy not predicted:
1 vs. 2 continuant, e.g. {fa**k**i, fa**x**i} < 1 vs. 0 continuant, e.g.{pa**k**i, pa**x**i}

Additionally, it is predicted that the typology of local and non-local segmental interactions is different. It was pointed out earlier than there is no non-local counterpart of post-nasal voicing. Nasals induce voicing on an immediately following stop, but never on a non-adjacent stop. This follows from the hypothesis that both local and non-local phonological patterns are perceptually grounded. While a nasal interferes with the perception of a voicing contrast on an immediately following stop (Hayes and Stivers 1995), it presumably has no effect on a non-adjacent voicing contrast.

18

There are two clear paths for following up on and extending the results reported here. The first is to investigate what features and segments show the long-distance perceptual interactions found for ejectives and aspirates, and to see if the typology of perceptual interactions does indeed correlate with the typology of phonological interactions. The second open question is the source of the long-distance perceptual interaction. The perceptual results here are somewhat surprising precisely because they arise in the absence of any acoustic ambiguity. Future work must investigate the explanation for, or source of, non-local perceptual interference.

References

Ajacopa, Teofila Laime, et al. 2007. *Diccionario Bilingüe, Iskay Simipi Yuyayk'ancha: Quechua–Castellano Castellano–Quechua*. La Paz, Bolivia.

Boersma, Paul and David Weenink. 2010. Praat: doing phonetics by computer. Version 5.1.30, available at www.praat.org.

Brown, Jason. 2008. Theoretical aspects of Gitksan phonology. Doctoral dissertation, University of British Columbia.

Efere, Emmanuel Efereala. 2001. The pitch system of the Bumo dialect of Izon. In Suzanne Gessner, Sunyoung Oh and Kayono Shiobara (eds.), *Current Research on African languages and linguistics* (*UBC Working Papers in Linguistics* 4): 115-259.

Flemming, Edward. 1995. Auditory representations in phonology. Doctoral Dissertation, University of California, Los Angeles.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, Robert Kirchner and Donca Steriade eds., *Phonetics in Phonology*, Cambridge: Cambridge University Press.

Gallagher, Gillian. 2010a. The perceptual basis of long-distance laryngeal restrictions. Doctoral Dissertation, Massachusetts Institute of Technology.

Gallagher, Gillian. 2010b. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology* 27:435-480.

Gallagher, Gillian and Jessica Coon. 2009. Distinguishing total and partial identity: evidence from Chol. *Natural Language and Linguistic Theory* 27:545-582.

Hayes, Bruce and Tanya Stivers. 1995. The phonetics of post-nasal voicing. Ms. University of California, Los Angeles.

Hansson, Gunnar Olafur. 2001. Theoretical and typological issues in consonant harmony. Doctoral Dissertation, University of California, Berkeley.

Jaeger, Florian. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* (special issue on *Emerging Data Analaysis*) 59:434-446.

Jenewari, Charles E.W. 1989. Ijoid. In John Bendor-Samuel ed., *The Niger-Congo languages*, 105-118. Lanham: University of Press America.

Khumalo, James Steven Mzilikazi. 1987. An autosegmental account of Zulu phonology. Doctoral Dissertation, University of Witwatersrand.

Kingston, John. 1985. The phonetics and phonology of the timing of oral and glottal events. Doctoral Dissertation, University of California, Berkeley.

Kuipers, Aert. 1974. *The Shuswap Language*. The Hague: Mouton.

Ladefoged, Peter. 2003. *Phonetic Data Analysis*. Oxford: Blackwell Publishing.

Lindau, Mona. 1984. Phonetic differences in glottalic consonants. *Journal of Phonetics* 12: 147-155.

MacEachern, Margaret. 1999. *Laryngeal co-occurrence restrictions*. New York: Garland.

Macmillan, Neil and Douglas Creelman. 2004. *Detection Theory: A user's guide*. 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Rose, Sharon and Lisa King. 2007. Speech error elicitation and cooccurrence restrictions in two Ethiopian Semitic languages. *Language and Speech 50*:451-504.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475-531.

Steriade, Donca. 1997. Phonetics in phonology: the case of laryngeal neutralization. Ms. UCLA.

Warner, Natasha. 1996. Acoustic characteristics of ejectives in Ingush. *Proceedings of the International Conference on Spoken Language Processing*, 1525-1528.

Wonderly, W. 1951. Zoque: phonemics and morphology. *International Journal of American Linguistics* 17:1-4.