

## VCCV Perception: Putting Place in its Place

Steve Winters  
swinters@ling.ohio-state.edu

### Abstract

Jun (1995) and Hume (1998) incorporate perception into analysis of cross-linguistic trends in place assimilation and metathesis by claiming that the perceptual salience of specific segments motivates the ranking of relevant OT constraints. This study investigates the specific claims Jun and Hume make concerning the perceptual salience of cues for stop place of articulation to determine whether their salience actually could motivate the proposed OT rankings. Since both Jun and Hume based their proposals on a consideration of cues for stop place of articulation in the appropriate (VCCV) context for place assimilation and metathesis, this study only tested the salience of stops in this context. Listeners heard unreleased stops of three places of articulation (labial, coronal, dorsal) and two manners (oral, nasal) in two stress patterns *preceding* pre-vocalic oral stops of three other places of articulation. The perceptual salience (as measured in d') of stops in this context did not always bear out the predictions made by Jun and Hume. Interestingly, labials were generally the most salient place of articulation while dorsals were the worst. Nasal stops also turned out to be more salient than oral stops. Less surprisingly, pre-vocalic stops were more salient than post-vocalic stops, and place salience was highest for stops preceding coronals in pre-vocalic position. The variable success of Jun's and Hume's proposed hierarchies of place salience underscores the need to test the empirical validity of hypotheses concerning the interaction of phonology and perception.

## 1. Introduction

The role of perception in phonology has a long but largely unsung history, dating back to at least the early 1970's work of Björn Lindblom and his theories of adaptive dispersion in vowel spaces. The more recent influence of Optimality Theory in linguistic circles provides some new perspectives on how perception might influence phonology. Studies such as Jun (1995) and Hume (1998), for example, attempt to account for phonological processes such as place assimilation and metathesis by appealing to the perceptual salience of specific cues for place of articulation in stop consonants. Though the specifics of their accounts differ, both Jun and Hume propose that differences in perceptual salience can lead to different rankings of phonological constraints (which are encoded in terms of the articulatory intentions which define a speaker's grammar). Note that this subtle gap between perceptual salience and articulatory intentions implies that perception is not literally a part of phonology, but rather has an indirect influence on grammatical possibilities. However, the fact that different constraint rankings constitute different grammars in Optimality Theory has an interesting implication: if the relative perceptual salience of various kinds of sounds is universal, their corresponding constraint rankings would provide some limitation on the kinds of grammars that could possibly exist.

That Jun and Hume both draw phonological conclusions based on cues for place of articulation in stop consonants has interesting implications for the study of stop place perception. Universal or context-invariant acoustic information in the cues for specific places of articulation has been notoriously difficult to find (though note the work of Stevens & Blumstein 1978). This contrasts sharply with the acoustic characteristics of vowels, each of which has a comparatively uniform and easily identifiable pattern of formant frequencies to which it might be assumed the human perceptual mechanism directly responds. The apparent lack of invariant acoustic information for the place of stop consonants has led some to conclude that the perception of these sounds takes place not so much on the basis of a reaction to something that is "out there" but rather as the result of complex and highly specialized perceptual processing in the human brain (Lieberman & Mattingly, 1985).

Nevertheless, researchers such as Miller and Nicely (1955), Wang and Bilger (1973) and Winters (2000) have attempted to identify universal patterns in the perception of various stop places on the basis of merely what a listener can hear coming in from "outside." Perhaps unsurprisingly, studies of this nature have yielded conflicting results concerning the relative perceptual "salience" for different stop places of articulation. With respect to the places labial, coronal and dorsal in pre-vocalic position, for instance, Miller and Nicely (1955) found that salience was highest for coronals, but not substantially different between dorsals and labials. Wang and Bilger (1973), in turn, found that salience was equally high for labials and coronals but lower for dorsals. Not to be outdone, Winters (2000) concluded that salience was highest for labials and dorsals but lower for coronal stops.

What to make of such empirical confusion? One difficulty in comparing results across these experiments is that each study used different methods, which may conceal underlying commonalities in the results. Another problem is that taking such broad swipes at determining the universal "salience" of various places of articulation may ignore the troublesome context-based variance in the acoustic cues which signal stop place. Some variance may disappear when looking at individual contexts, or--even more specifically--at particular "packages" of acoustic cues for stop place. A listener may not necessarily generalize perceptual information across such contexts and packages in developing perceptually-based constraints for their Optimality Theoretic phonologies.

## 2. Theoretical Proposals

Interestingly, the perceptually-based OT constraints proposed by Jun and Hume only address the salience of various cues for stop place in specific contexts. Jun, for instance, accounts for cross-linguistic patterns in place assimilation by only considering what cues for place are present in the appropriate VCCV context for this process. If the first consonant in the CC sequence is a stop, the release burst of the first consonant is commonly dropped, thereby making the vowel-to-consonant transition the only cue for the first stop's place of articulation. From this observation, Jun concludes that the salience of stop place in post-vocalic position depends entirely on the relatively uniform acoustic characteristics of transition cues for the different places of articulation. In coronals, for instance, "Tongue tip gestures are rapid; thus, they have rapid transition cues. In contrast, tongue dorsum and lip gestures are more sluggish; thus, they have long transitions. Consequently, noncoronals have more robust perceptual cues than coronals." Jun's reasoning here seems to be based on the not unintuitive idea that extending the duration of acoustic information will increase the salience of that acoustic cue.

Jun's reasoning in comparing the relative salience of dorsal and labial cues, however, is slightly more complex.

"Unlike labials and coronals, velars have an acoustic attribute, i.e., compactness (Jakobson, Fant and Halle, 1963). Velars can be characterized by a noticeable convergence of F2 and F3 of a neighboring vowel. These two formants can form a prominence in the midfrequency range. As argued and discussed by Stevens (1989), such a midfrequency prominence of velars can form a robust cue for place of articulation...Based on Stevens' claim, we assume that velars have an additional acoustic cue, i.e., compactness, for place of articulation, compared to coronals and labials."

By virtue of this reasoning, then, Jun claims that post-vocalic unreleased dorsal stops are more salient than labials, which are, in turn, more salient than coronals.

If these claims are true, they do a neat job of accounting for certain cross-linguistic tendencies to assimilate such unreleased, post-vocalic stops. In the spirit of Mohanan (1993), Jun performed a cross-linguistic survey of assimilation processes and noted a number of intriguing implicational relationships. For one, Jun notes that neither dorsals nor labials are targets of place assimilation unless coronals are as well. Secondly, he notes that dorsals do not assimilate unless labials do so, too. The pattern seems clear: a more salient place of articulation will not assimilate unless a less salient place already does so.<sup>1</sup>

Jun assumes that such patterns are assimilated into a speaker's grammar under the rubric of "preservation" constraints, which he defines as:

(1) "Pres(X(Y)): Preserve perceptual cues for X (place or manner of articulation) of Y (a segmental class).

Universal ranking: Pres(M(N)) >> Pres(M(R)),  
where N's acoustic cues for M are stronger than R's cues for M."

The appropriate ranking of preservation constraints for place in unreleased stops, then, would be:

(2) Pres(pl(dor<sup>ˀ</sup>)) >> Pres(pl(lab<sup>ˀ</sup>)) >> Pres(pl(cor<sup>ˀ</sup>))

Jun does not stop there; he also looks at patterns in place assimilation with regard to manner, syllabic position and trigger place Jun proposes the following constraint rankings for the relevant groups of sounds:

(3) Manner: Pres(pl([stop]C)) >> Pres(pl[nasal]C))

(4) Position: Pres(pl(onset)) >> Pres(pl(coda))

(5) Trigger: Pres(pl(\_\_cor)) >> Pres(pl(\_\_noncor))

These rankings are also based on Jun's analysis of the relative salience for each sound group's context-dependent acoustic cues. Since his analysis of perceptual salience is based on speculation and not experimentation, however, it seems fair enough to ask if these conclusions are really valid. Is this really an example of perception influencing phonology or are these patterns the result of some other cross-linguistic influence?

Such questions seem even more relevant when looking at Hume's (1998) analysis of consonant/consonant metathesis. Hume proposes that this process may often be driven by perceptual factors; specifically, she claims that,

---

<sup>1</sup> Though see Tserdanelis and Hume (2000) for potential counterevidence to these assimilation patterns.

"...by metathesis, a perceptibly vulnerable consonant shifts to a context in which the phonetic cues to the sound's identification are more robust, thereby enhancing the consonant's auditory prominence and, in turn, strengthening syntagmatic and paradigmatic contrast among sounds in a given language. By perceptibly vulnerable, I refer to a consonant with comparatively weak segment internal and/or contextual cues to, e.g., place and/or manner of articulation." (295-296)

The proposed role that the salience of acoustic cues plays in shaping phonological structures is slightly different here; instead of weakly-cued segments being eliminated (as in Jun), they are shifted into a context in which they would be more salient. The formal mechanism whereby such perceptual optimization is implemented is a family of "AVOID" constraints, which Hume defines as:

(6) AVOID C/X: Avoid positioning a consonant (C) in a context (X) in which it is perceptually weak.

Whether or not perception influences phonology through a strategy of "avoidance" or "preservation"--or even some other strategy--is an interesting (and difficult) research question in its own right. But in this case a more tractable question is whether or not it really is the relative perceptual salience of different cues for stop place that is influencing cross-linguistic patterns in metathesis and place assimilation. Hume proposes that labials have relatively low salience in certain contexts, which can motivate their metathesis into a more salient context. To wit, Hume notes: "...labials can be considered particularly vulnerable given inherently short vowel transitions and relatively weak bursts, as compared to coronals and velars." This analysis can account for stop/stop metathesis in a language like Kui, where labials only metathesize when preceded by a dorsal in a stressed coda position. "The shift of the labial stop from an unstressed to stressed position at the expense of a velar in Kui is therefore not surprising, given that prosodic prominence in the language results in greater duration of transitions into the labial" (296).

However, Hume's claims about the "vulnerability" of labials seems to be at odds with Jun's proposal that labials are *not* the least salient place of articulation (in precisely the same context!). Part of the confusion here may stem from the fact that both researchers are *speculating* about what place cues are more or less salient; the rest of the confusion only follows from the lack of empirical data on which places (and cues) for stops are more or less salient.

Assertions about the relative "strength" or "weakness" of various acoustic cues for place beg the question of how, exactly, we might know whether an acoustic cue is "weak" or "strong". Hume and Jun base their claims on spectrographic analyses of typical labial, coronal or dorsal productions, but listeners who are actually in the business of acquiring and using phonologies have no such electro-mechanical luxury. These listeners have to

base their categorical decisions of place salience upon their own auditory experiences--whatever their evaluative mechanism might be. As a matter of operational fact, then, the "strength" or "weakness" of acoustic cues could only affect phonological structure inasmuch as they are reflected in listeners' perceived experiences of acoustic reality.

We need not speculate blindly about such experiences; listeners themselves can let us know what they are (within certain limits). So, given the proper interpretation of such experiences within some experimental paradigm, it should be possible to establish empirically what the relative strengths and weaknesses of various acoustic cues are. It should be possible, for instance, to investigate hypotheses such as those of Hume and Jun and determine whether their rankings of salience might genuinely serve as the motivation behind the phonological patterns they have found.

### 3. Methods

This study tested Jun's and Hume's claims about the perceptual salience of cues for stop place by investigating listeners' perception of vowel-stop-stop-vowel sequences. The utterances used to create the stimuli in this study were borrowed from Winters' (2000) study of audio and visual cues for place of articulation. All the original stimuli were of the form CVhVC, where the initial and final consonants were always identical, and both vowels were always [a]. These consonants could be either voiced oral or nasal voiced stops and could have either labial, coronal or dorsal places of articulation. There was also stress on either the first or second syllable of the nonsense CahaC word. Varying all of these factors made it possible to test Jun's and Hume's combined claims about salience in different syllabic positions, for different places and manners of articulation, and in stressed and unstressed syllables.

Two speakers, one male and one female, produced all of the relevant CahaC tokens while being videotaped in a sound-proof booth (for recording details, see Winters (2000)). For the original study, the videorecording of these production tokens was then digitized and edited into audio-visual and audio-only VC or CV tokens; the current study simply appropriated the audio-only tokens and digitally spliced them together to form the desired VCCV stimuli. Crucially, this study also eliminated stop bursts from coda position, since Jun's original proposals only considered the salience of unreleased coda stops. Practically speaking, this meant that the VC portion of the to-be-spliced-together VCCV stimuli contained the entire VC articulation up until a release burst (if any) or the offset of any noticeable closure voicing in the waveform. The CV tokens were then spliced directly after these edited VC tokens. The interval between the first vowel's offset and the second vowel's onset was a uniform 150 ms; in certain cases silence had to be inserted after the stop closure to augment the intervocalic duration (see Figure 1a). This particular time interval of 150 ms was chosen after it was found that shorter intervals generally induced a percept of only one consonant between the two vowels. Tokens with nasals in the coda position included nasal murmur during some of the 150 ms of intervocalic closure (see Figure 1b).

The resultant VCCV tokens could vary in place for both consonants, and could have stress on either the first or second syllable. Manner only varied in the first consonant, though, since not all nasals can appear in onset position in English. This meant that there were 3 (C1 place) x 3 (C2 place) x 2 (nasal/oral) x 2 (stressed/unstressed) = 36 token types; since these were produced by two different speakers, this amounted to 72 basic stimuli. These stimuli were randomized by computer and each presented twice to listeners in a sound-proof booth over headphones. After hearing each stimulus, a computer presented them with the question "What did the speaker say?" and gave them nine different responses to choose from (see Figure 2). These alternatives were written as if they were two words (e.g., 'ab da') and differed only in place of articulation for the coda and onset consonants. In order to reduce the listener's task to this point, the stimuli were presented to the listeners in separate blocks with nasal stops and oral stops.

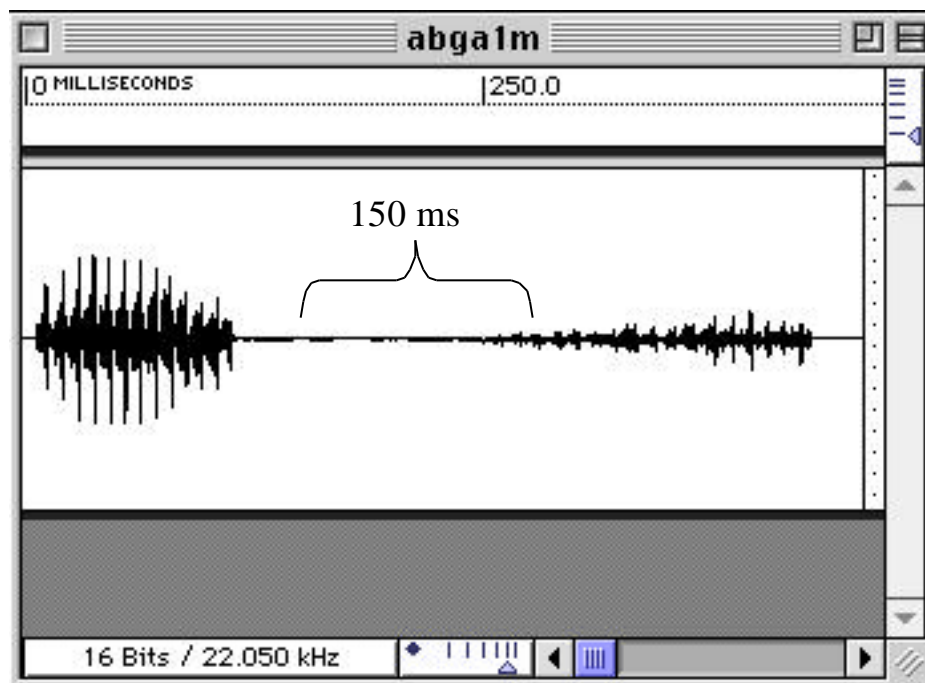


Figure 1a: Waveform for male production of “abga,” with stress on first syllable.

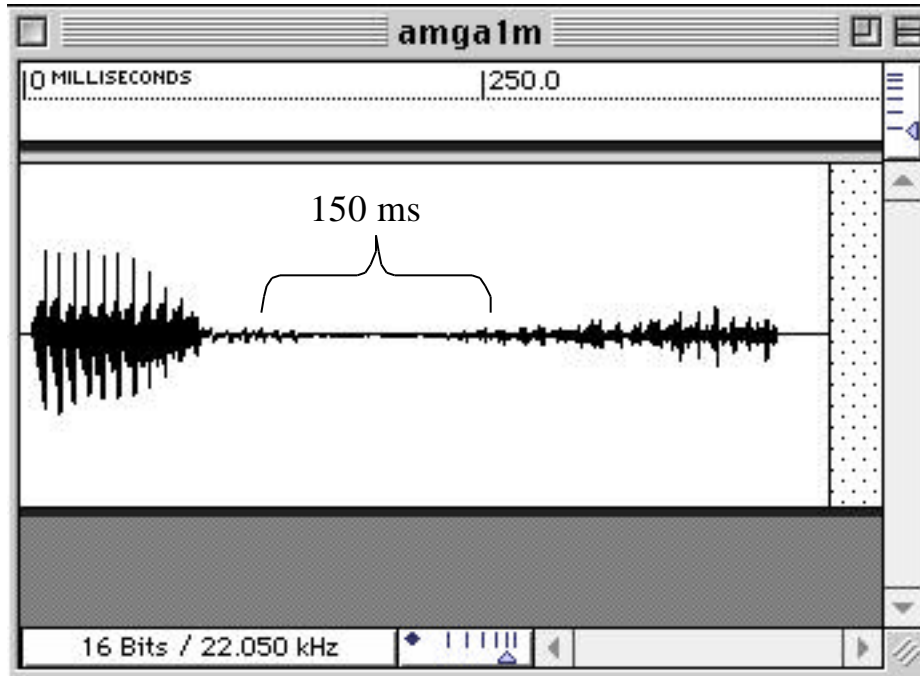


Figure 1b: Waveform for male production of “amga,” with stress on first syllable.

What did the speaker say?

ab ba	ad ba	ag ba
ab da	ad da	ag da
ab ga	ad ga	ag ga
Exit		

Figure 2: Presentation of experimental response alternatives

Twenty-four listeners worked through these blocks of stimuli twice: once at their speech reception threshold and once again at a comfortable listening level. A listener's speech reception threshold was determined with the same pre-test used in Winters (2000); the listener first completed this pre-test and then worked through the experiment at their speech reception threshold before they heard the stimuli again at a comfortable listening level. Listeners heard these tokens at two different sound levels to elicit a comparison between the two conditions--the assumption being that the "salience" of some acoustic cue for place should be directly related to its robustness in resisting a decrease in sensitivity when its amplitude is significantly diminished. However, there were little (if any) interactions between volume level and the factors tested in this experiment, so the effects of volume will be ignored in the discussion of the experimental results.

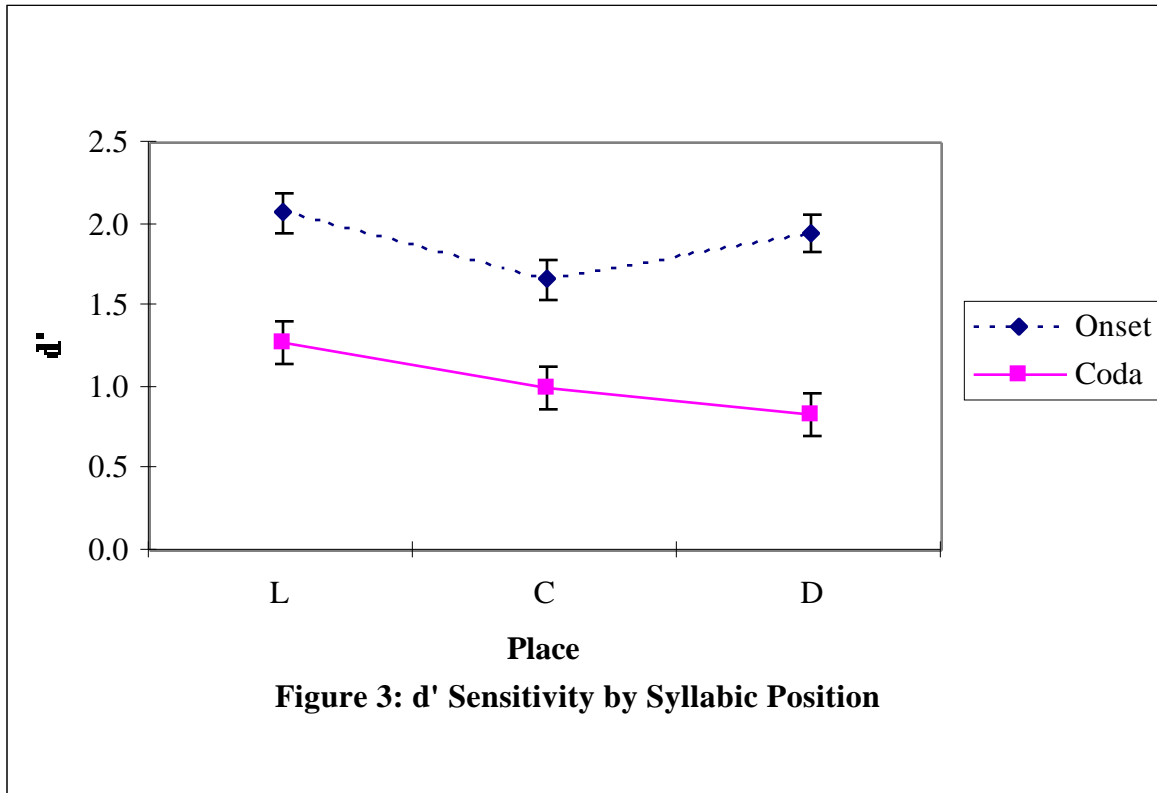
#### 4. Results

Working on the assumption that measures of sensitivity most accurately reflect the "salience" of a particular sound, the results of this experiment were converted into scores of  $d'$ , a standard measure of sensitivity in signal detection theory (MacMillan and Creelman, 1991), for each token type. Calculating  $d'$  involves eliminating listener *bias* in the experimental response options. Every time a listener gives a particular response (e.g., 'ab'), that response was either a *hit* (i.e., an 'ab' stimulus) or a *false alarm* (i.e., not an 'ab' stimulus). The proportion of false alarms for a particular response option reflects a listener's *bias* towards that response category, since it reflects a listener's tendency to respond with that option without receiving any evidence for it.  $D'$  is calculated by first converting the raw proportions of hits and false alarms into z-scores (i.e., the distance from the mean of a standard normal distribution) and then subtracting the z-score of the false alarms from the z-score of the hits. This step essentially eliminates the bias from the proportion of hits and results in a  $d'$  score that represents a listener's sensitivity (measured in units of perceptual distance) to a particular category of sounds.

Listeners only heard each token type four times (twice each for male and female productions), so most of the resultant confusion matrices contained zeros or fours for some response categories. It is impossible to calculate the z-score of a zero or one response ratio, so these ratios had to be converted into effective minima and maxima of .125 and .875 ( $1/2*n$  and  $1-1/2*n$ , following Macmillan and Creelman (1991)). In order to calculate values of  $d'$ , hit rates were calculated for each response category, and false alarms from both competing categories were lumped into one "false alarms" category.  $D'$  therefore reflected the distinctiveness between one sound category and all other response alternatives in the experiment.

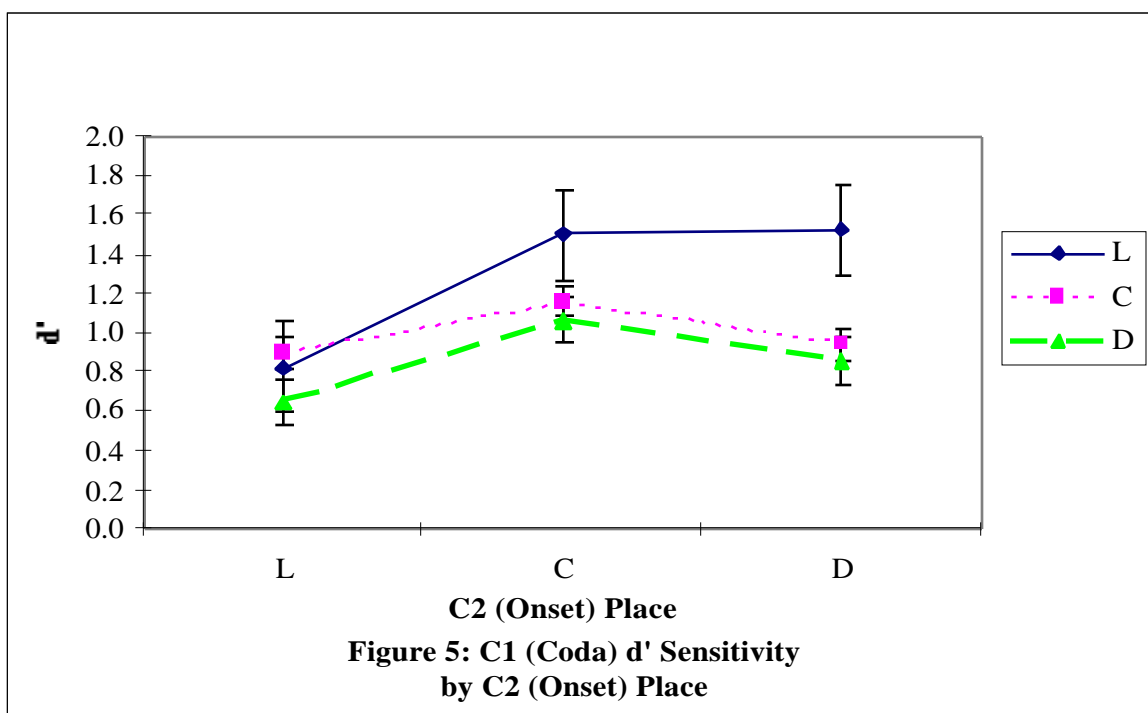
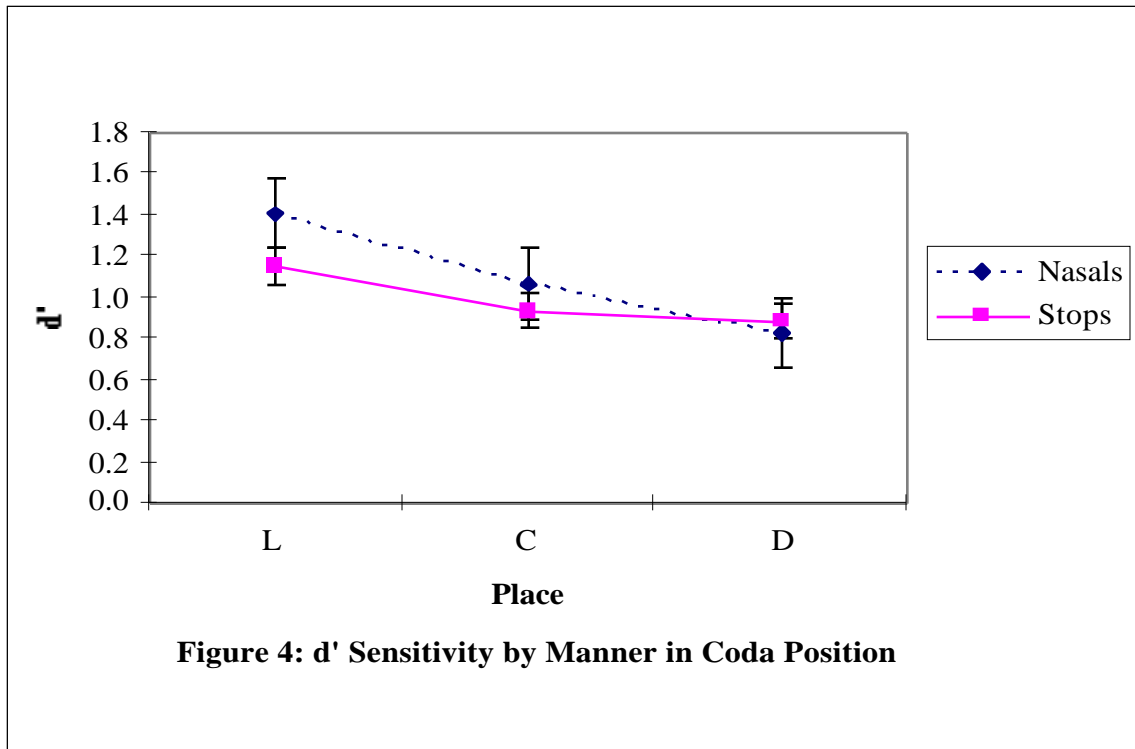
As in the Winters (2000) study, the data yielded conflicting results concerning Jun's and Hume's proposals about the relative salience of different places of articulation. Appendix I gives raw confusion matrices for listener responses in all conditions, while the following figures show average  $d'$  values across listeners for the theoretically relevant

conditions. Figure 3, for instance, shows listener sensitivity in  $d'$  to labial, coronal and dorsal places of articulation in both post-vocalic (coda) and pre-vocalic (onset) positions. Unsurprisingly, sensitivity to stop place was significantly higher in onset position, thereby verifying Jun's least controversial hypothesis (4). (See Appendix II for a description of statistical methods and specific results). However, the relative sensitivity of unreleased place in coda position contradicted Jun's assumptions in (2)--labial was the most salient place in this condition, followed by coronal, and then dorsal.



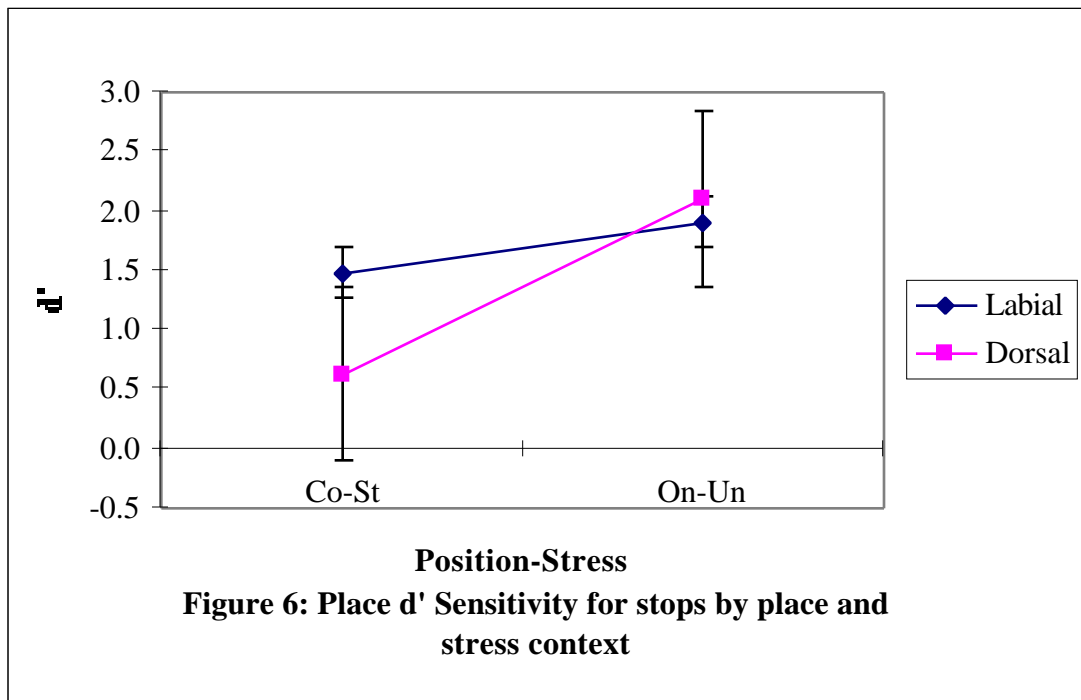
The results also failed to bear out Jun's claims about the comparatively higher salience of oral stops over nasal stops in coda position (3). In fact, it re-confirmed the surprising result of Winters' (2000) study that, if anything, nasals are more salient than stops in this position. Measured in  $d'$  (Figure 4), there is no significant difference between sensitivity for nasal stops and oral stops; nasal stops just enjoy a slight sensitivity advantage. (Superimposed on these results is the same labial > coronal > dorsal pattern in sensitivity that was seen in Figure 3. However, coronals are significantly more salient than dorsals only in nasal stops.) For some reason not apparent in the acoustic signal, listeners actually seem to be more sensitive to place information in nasals than in oral stops in coda position.

One of Jun's more interesting claims was that place sensitivity in coda position was itself sensitive to the place of a following stop consonant (5); Jun assumed salience would be higher before coronal stops than non-coronals, due (again) to the rapidity of coronal gestures and the corresponding lack of articulatory overlap in comparison to non-coronal gestures. For a  $d'$  analysis (Figure 5), the results generally supported this hypothesis; Onset Place was a significant ANOVA factor ( $F = 39.802$ ;  $df = 2,22$ ;  $p < .001$ ).



Post-hoc tests also showed that sensitivity was almost always significantly higher before coronal stops than before dorsals or labials; the only exception here were labial stops, which were not significantly more salient before coronals than before dorsals.

Though some of Jun's hypotheses seem to be supported by these results, the numbers do not bode well for Hume's hypothesis of labial stop "vulnerability". Labial salience seems to be particularly strong in any context, thereby seemingly invalidating any motivation to metathesize these segments into some more salient position. Looking at the specific context for dorsal-labial metathesis in Kui, however--a stop in a stressed coda followed by a stop in an unstressed onset--seems to show that the perceptual optimization of the *dorsal* stop may motivate this process. Figure 6 shows that (oral) dorsal stops in stressed codas have remarkably low salience in comparison to (oral) labial stops in the same position--a fact which is, of course, consistent with the results from Figure 3. In unstressed onset position, however, dorsal salience increases significantly while labial salience does not change drastically. The overall salience of a labial-dorsal stop sequence in this prosodic context would therefore be significantly higher than the overall salience of a dorsal-labial sequence--and it is precisely the more salient sequence that the speakers of Kui choose to produce. Although the labial stop vulnerability hypothesis may be incorrect, Hume's analysis of *why* this process occurs may be appropriate--Kui may be avoiding the production of dorsals in the weak (coda) context.



## 5. Discussion

The fact that communication is language's primary function no doubt plays a role in the kinds of phonological patterns we find in languages throughout the world. It is not unreasonable to suggest that the drive for communicative ease may spawn phonological processes that seem to be articulatory simplifications or acoustic enhancements. Nor is it unreasonable to expect that sound inventories will more commonly include articulatorily simple segments or vowels that are maximally dispersed throughout acoustic space (as in, e.g., Liljencrants and Lindblom 1972). These tendencies do not, of course, preclude the formal possibility for more complex articulations or vowels with unlikely formant patterns--but this is no reason to deny such tendencies any place in the theoretical analysis of language. Explaining grammatical patterns in language on the basis of their communicative function is no less valid (or interesting) than explaining them as purely formal entities. All that is really crucial--in *both* approaches--is establishing the empirical validity of the proposed explanation.

This is where functional analysis can run into trouble. The functional accounts proffered by Jun and Hume for cross-linguistic patterns in metathesis and place assimilation are easy enough to accept on an intuitive basis--who, for example, would not believe that cues for nasal stops are less salient than cues for oral stops? Without the empirical justification provided by studies such as this one, however, such assumptions may just as likely be untrue. Understanding that most language use takes the form of communication provides the linguistic imagination with a wealth of hypotheses about why we find the patterns in phonology that we do--but this is only the first step towards establishing a functional *explanation* for the same phenomena.

The paradigm used in this study was intended to provide one objective means of establishing such explanations, but it does, of course, have its limitations. The results are far from yielding conclusive information about the *universal* salience of stop place cues in these various contexts, since the study only tested English listeners and also used only one vowel ([a]) context. However, it did conclusively show that some of the hypotheses Jun and Hume propose about the relative salience of places of articulation do not necessarily hold in *all* languages. Whether or not some unique aspect of English speakers' experience or environment is concealing a more universal pattern of place salience is a question that is left to future research.

The fruitfulness of restricting the context for specific place cues is also left open to question. There is no *a priori* reason to presume that a listener will generalize across the acoustic manifestations of a particular sound in various contexts; the amount of perceptual detail that is available to a listener in constructing a constraint-based phonology is potentially limited only by the listener's psychophysical capabilities. Discovering what connections there may be between the psychophysical input in speech communication and the formal structures a listener develops in constructing a grammar is the exciting possibility offered by this line of speech perception research. Finding out

what limits there might be to these connections--and thereby addressing the issue of *granularity* (Pierrehumbert, 1999)--is the further knowledge that this research may reveal for the study of cognition.

### Acknowledgements

I would like to thank Keith Johnson, Beth Hume, Mary Beckman, Ilse Lehiste, Pauline Welby, Kiyoko Yoneyama, Jenny Vannest, Giorgos Tserdanelis, Jeff Mielke, Misun Seo, Matt Makashay, and all of Ohio State's PiPsters for their input on this work. I would also like to thank the participants of the S5 conference for their helpful comments on the first presentation of this paper. Part of this material is based on work supported under a National Science Foundation Graduate Fellowship.

### References

- Hume, Elizabeth (1998). The role of perceptibility in consonant/consonant metathesis. In K. Shahin et al. (eds.), *Proceedings of the West Coast Conference of Formal Linguistics* 17: 293-307.
- Jun, Jongho (1995). Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings of the West Coast Conference of Formal Linguistics* 14: 221-237.
- Liberman, Alvin and Mattingly, Ignatius (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- Liljencrants, Johan and Lindblom, Björn (1972). Numerical simulation of vowel quality systems. *Language*, **48** (4), 839-862.
- MacMillan, Neil and Creelman, Douglas (1991). *Detection theory: a user's guide*. Cambridge: Cambridge University Press.
- Miller, G.A. and Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338-352.
- Pierrehumbert, Janet (1999). Formalizing Functionalism. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan and K. Wheatley (eds). *Formalism and Functionalism in Linguistics*, John Benjamins, Amsterdam. Vol. I, 287-305.
- Stevens, Kenneth and Blumstein, Sheila (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.
- Tserdanelis, Georgios and Hume, Elizabeth (2000) Nasal place assimilation in Sri Lankan Portugese Creole: implications for markedness theory. Paper presented at MOT Workshop in Phonology, York University, Toronto.
- Wang, M.D. and Bilger, R.C. (1973). Consonant confusion in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, **54**, 1248-1266.
- Winters, Stephen (2000). Turning phonology inside out: testing the relative salience of audio and visual cues for place of articulation. In Levine, R., A. Miller-Ockhuizen, T. Gonsalvez (eds.), *Ohio State Working Papers in Linguistics* **53**: 168-199.