Steriade, D. (1995). *Positional neutralization.* Unpublished manuscript, University of California, Los Angeles.

Steriade, D. (1999a). Alternatives to the syllabic analysis of consonantal phonotactics. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of linguistics and phonetics 1998* (pp. 205–245). Prague: Charles University Press.

Steriade, D. (1999b). Lexical conservatism in French adjectival liaison. In B. Bullock, M. Authier, & L. Reed (Eds.), *Formal perspectives in Romance linguistics* (pp. 243–270). Philadelphia: John Benjamins.

Steriade, D. (2000). P-map effects on constraint organization. Unpublished manuscript, University of California, Los Angeles.

Stevens, K., & Blumstein, S. (1975). Quantal aspects of consonant production and perception. *Journal of Phonetics*, *3*, 215–233.

Street, C. S., & Mollinjin, G. P. (1981). The phonology of Murinbata. In B. Waters (Ed.), *Australian phonologies: Collected papers* (pp. 183–244). Darwin: Summer Institute of Linguistics.

Stroop, J. (1981). Metathesis von s en p. *Spektator*, *11*, 224–248.

Wackernagel, J. (1896). *Altindische grammatik.* Göttingen, Vandenhoeck und Rupricht.

Waters, B. (1979). *A distinctive features approach to Djinang phonology and verb morphology*. Darwin: Summer Institute of Linguistics, Australian Aboriginal Branch.

Whitney, W. D. (1889). *Sanskrit grammar*. Cambridge: Harvard University Press.

Wingstedt, M., & Schulman, R. (1988). Listener's judgments of simplifications of consonant clusters. *Linguistics*, *26*, 105–123

Zwicky, A. (1976). Well, this rock and roll has got to stop, Junior's head is hard as a rock. In S. Mufwene, C. Walker, & S. Steever (Eds.), *Proceedings of the Chicago Linguistics Society, 12* (pp. 676–697). Chicago: Chicago Linguistics Society.

10
CHAPTER

Perceptual Cues in Contrast Maintenance

Richard Wright

Department of Linguistics
University of Washington
Seattle, Washington 98195-4340

- **I. Introduction**
- **II. Defining *Cue* and *Contrast***
- **III. Degradation and Robustness**
- **IV. Contrast Maintenance**
- **V. Experiment 1**
  - A. Methods
  - B. Results
  - C. Discussion
- **VI. Experiment 2**
  - A. Methods
  - B. Results
  - C. Discussion
- **VII. Conclusions**
- **Appendix A: Confusion Matrices for Experiment 1**
- **Appendix B: Confusion Matrices for Experiment 2**
- **Acknowledgments**
- **References**

The Role of Speech Perception in Phonology

I. INTRODUCTION

Constraint-based phonologies, especially Optimality Theory and its offshoots, have revived interest in the role of perception in a variety of phonological phenomena. Particularly fertile ground for this area of phonological research is in the interaction of formal structural constraints with constraints based on economy of effort and with constraints related to contrast maintenance. Indeed, several recent works have investigated the role of perception in contrast maintenance including the perceptual basis of contrasts and features (Flemming, 1995; Gordon, 1999; Kirchner, 1997), perceptual and positional licensing (Steriade, 1995), and the interaction between articulatory and perceptual demands (Byrd, 1996; Jun, 1995; Silverman, 1997; Wright, 1996). As phonologists turn to the perceptual literature in search of the underpinnings of perceptually motivated constraints, they find volumes of perceptual research on speech cues but few experiments designed with phonology and phonological phenomena in mind. This leaves them with two choices: (1) piecing together evidence from a panoply of experiments from different languages using a variety of perceptual tasks, or (2) designing experiments to test phonological hypotheses themselves. With a few exceptions such as Cutler & Clifton (1983), Ohala (1990), or Pierrehumbert (1979), phonologists are often not equipped with the training or facilities to conduct their own perceptual research and opt for the first choice. However, piecing together evidence in this manner results in generalizations that may not be valid because the differences in the questions being asked and experimental methodologies may result in nonapplicable findings. The purpose of this study is to test two hypotheses that are relevant to phonological theory using experiments with a uniform set of stimuli, tasks, and subjects. The first hypothesis is that formant transitions in syllable onset are a more reliable cue to consonant place than offset transitions. This hypothesis is key to understanding the potential role of perception in positional licensing (Beckman, 1997; Steriade, 1995), the more general NO CODA constraint (Prince & Smolensky, 1993), or the role of perception in sonority sequencing as proposed for example in Kawasaki (1982) and Wright (1996). The second hypothesis is that manner of articulation largely determines the recoverability of consonant cues that are not based on formants: fricative noise, consonant release transients, and nasal murmur. The second hypothesis is crucial to understanding the relationship between sonority violations and perception.

II. DEFINING *CUE* AND *CONTRAST*

A chapter about perceptual cues and contrast maintenance should be explicit about what is meant by the terms *contrast* and *cue*. A commonly used though fairly

limited definition for *contrast* is *a distinction that minimally differentiates one word or morpheme from another in the same language*. Phonological contrasts then are those *speech sounds* that minimally distinguish any two morphemes in a language. While this use of contrast has its foundation in distinctive feature theory (Jakobson *et al.*, 1951) and may need refining for use in constraint-based grammars, it is a good starting point for the purposes of this chapter. Implicit in this definition is the perceptual nature of contrast; when we say "to distinguish" we mean to distinguish perceptually, that is, while a phonological contrast might in principle be based on a distinction that did not depend on perceptual *recoverability*, it would in practice be a rather poor contrast and it would be unlikely to be passed on to future generations of language learners. It is also worth noting that phonological contrast is both syntagmatic and paradigmatic — a particular speech sound must be distinguished both from existing neighboring sounds within forms, and also from that pool of all potential sounds in a position that make up a language's sound inventory. Moreover, one word may be distinguished from another by the presence or absence of a particular speech sound — for example, the listener distinguishes "lap" and "lapse" by the presence of /s/ frication noise

In this chapter, the term "cue" is used to mean information in the acoustic signal that allows the listener to apprehend the existence of a phonological contrast. This is a relatively limited definition of *cue*, since much of the information in the signal does not directly concern phonological contrasts. For example, there is information in the spoken signal that identifies the speaker and the rate of speech. However, though this information can affect what cues are salient to a listener by changing the listener's attention and though it has been shown to affect word identification, it is not being investigated in the current set of experiments (Nygaard & Pisoni, 1995). At first blush, these definitions may lead to the mistaken assumption that "cue" and "contrast" may be used interchangeably. However, on closer inspection it should be apparent that phonological contrasts are built on cues; there may be a one-to-one relationship, a many-to-one relationship, or a one-to-many relationship between cues and a phonological contrast, that is, the same aspect of the signal that provides the listener with information about one contrast may simultaneously provide the listener with information about a neighboring contrast. For example, while the second formant transition out of a stop consonant's closure is a cue to the place of articulation of a consonant, it is also a cue to the vowel quality of the vowel following the consonant.

Thinking about cues in the abstract without reference to their role in the transmission of information may lead to some dubious assumptions. One unsupported assumption is that the amount of information that a cue may carry is proportional to its importance to a contrast in phonological processes. This type of assumption has led to a search for a single invariant "primary cue" without regard for robustness in noise or variability across prosodies or syllable positions

(Blumstein & Stevens, 1981). This type of research has also led to the assumption that the perceptual representation of contrasts and the morphemes they distinguish should be underspecified. There are a variety of reasons that have to do with signal transmission for questioning these sorts of assumptions. Some of the relevant points will be discussed below. There is also empirical evidence that listeners depend on an array of cues in the signal, choosing information that is readily available rather than relying on a single cue (see Neary, 1997, for a discussion).

III. DEGRADATION AND ROBUSTNESS

Under optimal laboratory listening circumstances, any one of an array of cues may be sufficient for a listener to recover a phonological contrast. However, while any one cue may in principle suffice, not all cues will be equally effective in conveying their information to the listener in all environments. The inequality of cues comes from the fact that spoken language must rely on transmission through an acoustic medium and on reception of the signal by the listener. Under these conditions, opportunities abound for the introduction of noise into the process. Any event that shapes or distorts the acoustic signal or the signal reception process and bears a random relationship to the information content of the signal may be considered noise. Note that this definition excludes individual speaker variation and variation due to coarticulation, while environmental masking, hearing loss, and listener distractions are included. Perhaps the most ubiquitous type of noise is *environmental masking*. Environmental masking occurs when an event in the environment generates a signal that overwhelms portions of the speech signal. It is rare for speech to occur in the absence of at least some form of environmental masking but the type of masking may vary from moment to moment. What this means for speech is that a robustly encoded phonological contrast is more likely to survive signal degradation or interference in reception. A robust encoding involves cue redundancy, resistance of cues to environmental masking, the ability of cues to survive momentary distractions on the part of the listener, and the auditory impact of cues. These are not mutually exclusive conditions, but rather are largely overlapping.

As has been frequently noted, as more information is brought to bear on the lexical decision, the probability of decision error decreases and the ability of the system to recover from an error increases (Miller *et al.*, 1951). Information redundancy can be equated with a signal-to-noise ratio (e.g., Miller *et al.*, 1951; Sumby & Pollack, 1954), and can occur at a variety of linguistic levels in an utterance: semantic, syntactic, and phonological. At the phonological level, certain syllable structures permit a much greater degree of redundancy than others. It is probably no accident that the typologically most common syllables are those that ensure a robust encoding of information through informational redundancy (see Wright, 1996, for a discussion). As noted by Mattingly (1981), speech gestures that can overlap without destruction of information result in greater speed of signal transmission and an increase in cue redundancy, that is, a preferred syllable structure will allow greater gestural compression resulting in shorter and more information rich syllables. If coarticulation results in changes to the signal that are audible and nondestructive, there is an increase in redundant look-ahead, look-back, and concurrent information about the phonological content of an utterance. The greatest benefit is achieved if constrictions are released in decreasing degree of stricture, and constrictions are made in increasing degree of stricture. The similarity of Mattingly's preferred syllable structure and the Sonority Sequencing Principle should be obvious (e.g., Bell & Hooper, 1978; Selkirk, 1982). Because formants provide the most reliable type of coarticulatory information, phonological forms with alternating consonants and vowels are the optimal organization. The importance of segmental ordering to the information redundancy is illustrated in Figure 10.1, a spectrogram of a VCV sequence and a VCCV sequence. Figure 10.1A illustrates the increased redundancy when closure is made from lesser to greater stricture *and* when aperture is made from greater to lesser stricture; Figure 10.1B illustrates the loss of information when two peak strictures abut.

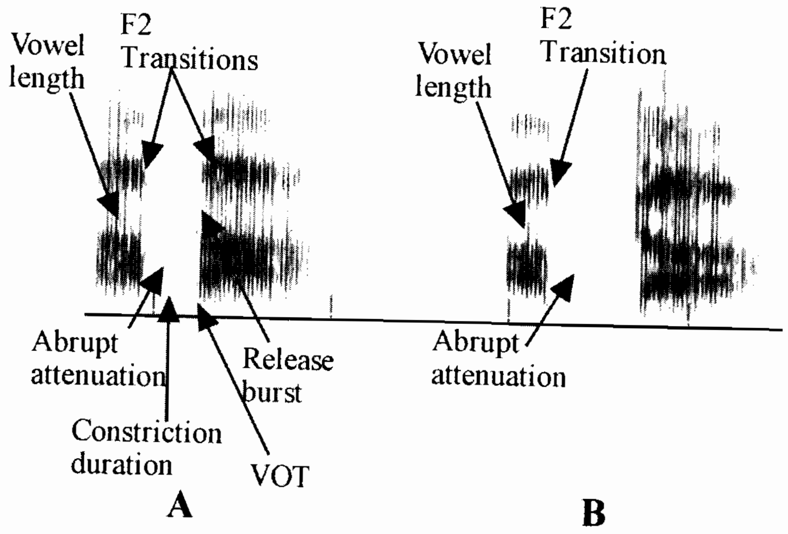

Figure 10.1. Spectrograms of the two nonsense words /aba/ and /abda/ illustrating increased redundancy in V+stop+V (**A**) as opposed to V+stop+stop+V (**B**) sequences. The acoustic cues for the bilabial stop are indicated by the arrows.

A second element of *robustness* of encoding comes from the resistance of cues to environmental masking and other forms of interference. The likelihood that a cue will be lost in transmission and reception is determined in part by its physical characteristics. It is obvious that signals with lower intensity are more likely to be lost to environmental masking (ex: weak /f/ frication vs. strong /s/ frication). Perhaps equally as obvious but less frequently acknowledged is that signals that are transient are more likely to be missed due to listener distraction (ex: release bursts). Moreover, even if a cue is not wholly lost to distraction, the relative importance of cues shift as attentional demands on the listener changes (Gordon *et al.*, 1993). Aperiodic signals are more easily masked by periodic or aperiodic noise than periodic signals are, and this has a direct impact on the type of cues that are most easily masked (Miller & Nicely, 1955). Aperiodic cues such as those contained in isolated release bursts or fricative noise are susceptible to loss or confusion in noise unless they have sufficient intensity to overcome most masking noise (i.e., unless they are sibilant fricatives), whereas periodic cues such as formant transitions are less susceptible to loss or confusion due to noise. Moreover, the more transient a signal, the more likely it is to be masked by abrupt changes in environmental sounds.

The third element of a robust encoding is the *auditory impact* of the signal containing the cue. The importance of auditory factors has been noted for some time (Bladon, 1986; Lindblom, 1990; Greenberg, 1995), but it still remains relatively unaddressed in phonological theories of syllable structure, and, with the exception of Silverman (1997), diachronic and synchronic phonological processes. This section will reiterate and expand on the points of the previous work.

The predominance of onsets over codas in the phonotactics of the world's languages has been frequently noted (Jakobson, 1962). Part of this predominance may be due to the greater inherent cue redundancy for some sounds in onset: while coda stops may be optionally audibly released, onset stops must have a release. Because of the mandatory release, onset stops may have a greater number of cues to place, manner, and voicing contrasts than codas. While redundancy alone may account for the predominance of onsets over codas for stops, other manners with little or no release burst such as nasals, fricatives, affricates, and glides would not gain the same redundancy benefit. A second benefit that onsets have over codas can be found in the way in which our auditory system processes speech sounds and complex sounds in general. Auditory nerve fibers' responses exhibit a dynamic nonlinear response that depends on the environmental context and on the rise-time characteristics of the stimulus signal itself (Kiang *et al.*, 1965; Smith, 1979). An abrupt onset of stimulation (sharp rise-time) in a frequency region that previously had little or no stimulation results in a burst of activity in the firing pattern of the nerve fibers. The initial peak response rate is followed by a very rapid attenuation during the initial 5 ms of the stimulus (rapid adaptation). Rapid adaptation is followed by a slower attenuation in response rate over the next 50 ms (short-term adaptation), settling thereafter into a steady low level of activity as the auditory nerve response becomes saturated (for stimuli at intensity levels typical of speech). After saturation takes place, the nerve fiber becomes much less sensitive to changes in intensity and frequency until after a resting period of approximately 40–50 ms. In the absence of saturation, auditory nerve response rate is equated with signal intensity; thus, the transient boost of activity before rapid adaptation and the elevated level of activity over the next 50 ms is seen as amplifying the onset portion of the stimulus (Delgutte & Kiang, 1984b; Greenberg, 1995). The magnitude of the onset response depends on three factors: (1) the intensity of the stimulus, (2) the abruptness of the onset, and (3) the amount of activity in the frequency region of the nerve fiber immediately preceding the stimulus onset. The greater the intensity and the more abrupt the onset, the greater the response. The less activity in the frequency region of the nerve fiber and the longer the period of inactivity (up to approximately 50 ms), the greater the initial response. There is no equivalent increase in response for signal offsets. This onset asymmetry in the auditory nerve response is mirrored throughout the auditory pathway as high up as the auditory cortex (Greenberg, 1995).

The auditory nerve fiber encoding of a variety of manners of consonants has been studied for both CV (Delgutte & Kiang, 1984a,b; Sinex & Geisler, 1983) and for VC word-final consonants (Sinex, 1995). Overall, the results from such experiments indicate that the onset peak is present for complex speech signals. There is no equivalent boost associated with transitions into a coda closure. Thus, one of the reasons onsets are typologically preferred in the world's languages may be due to low-level auditory processing. Figure 10.2 is a schematic illustration of the onset–offset asymmetry in response to a speech signal.

IV. CONTRAST MAINTENANCE

There may be a variety of ways that contrasts can be maintained. Changes of suboptimal syllables are among the most commonly cited types of phonological repair strategies. One of these is a limitation on the degree of gestural overlap in consonant sequences (see Browman & Goldstein, 1992; Byrd, 1996; Wright, 1999, for discussions). If gestural overlap results in a loss of information in the signal, then a restriction on overlap can preserve critical information about a linguistic contrast that might otherwise be lost. For example, in word-initial stop+stop onsets the release burst may contain the only available cues to the C1 stop; therefore, a limitation on the degree of overlap is the only way of preserving cues to the C1 in the acoustic signal. On the other hand, a word-initial fricative+stop onset, although not as optimal as an alternating CV syllable, may not need the same limitation on

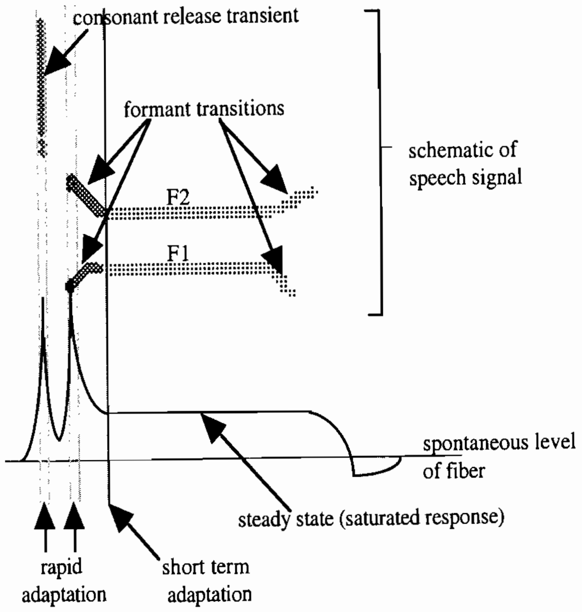

Figure 10.2. Schematic illustration of the onset–offset response asymmetry showing rapid and short-term adaptation. The portions of the signal that benefit from the boost (release transient, onset transitions) are darkened. After Wright (1996).

gestural overlap because the fricative's cues are at the peak of stricture rather than in the release. This type of pattern has been documented in acoustic detail in Tsou, an Austronesian language with a wide array of clusters (Wright, 1999). A related and more commonly cited repair strategy is the insertion of epenthetic vowels between consonants. The epenthetic vowel guarantees not only the release burst for stops, but formant transitions for all consonant types. The insertion of epenthetic consonants to break up sequences of vowels has the effect of increasing the redundancy of cues to vowel contrasts. Hume (1998) analyzes metathesis as a type of repair strategy that is motivated by perceptual considerations as well. Last, a common "repair" strategy is to abandon contrast maintenance altogether and simplify the syllable structure. It is worth noting that cluster simplification favors preserving the consonant with the stronger perceptual cues such as the C2 in word-initial stop+stop clusters (Steriade, 1982).

While repair strategies may be motivated by contrast maintenance, a more fundamental contrast maintenance strategy is for a language to maintain more optimal syllable structure in underlying forms. While there are undoubtedly articulatory factors to consider in defining an optimal syllable, the main proposal of this study is that perceptual factors play a significant role in determining the optimal syllable. The following two experiments test two of the most important claims made in the introduction of this chapter. Experiment 1 tests the claim that formant transitions in the onset provide cues that are more robust than formant transitions in the coda. Experiment 2 tests the claim that periodicity and transience affect the robustness of cues in noise.

V. EXPERIMENT 1

There have been several experiments that have probed the onset advantage using a variety of methodologies in a variety of languages (e.g., Fujimura *et al.*, 1978; Ohala, 1990); however, most of these works have compared the relative strength of codas with following onsets. Other studies have examined confusions in onsets or in codas alone without directly comparing onsets to codas (e.g., Miller & Nicely, 1955; Ohala & Ohala, 1998). Few have made a direct comparison between onsets and codas, and even fewer have looked at the interaction of syllable position with noise. The goal of this experiment is to directly compare the relative strength of cues in onsets and codas under a variety of noise conditions. Generally speaking, place cues are the most vulnerable (Miller & Nicely, 1955); therefore, this experiment will focus on identification of place. The general prediction is that listeners will be more accurate in identifying the place of articulation in onsets than they are in codas. A second prediction is that there will be an interaction of syllable position (onset, coda) with noise conditions such that coda cues will be more affected by masking noise than onset cues.

A. Methods

1. Subjects

Twenty-one unpaid volunteers participated in Experiment 1: 11 male and 10 female. All subjects were graduate and undergraduate students at the University of Washington in Seattle. Some were linguistics majors, but all lacked phonetic training and all were naive to the object of the study. All of the subjects had self reported normal hearing, and all were native speakers of American English.

2. Stimulus Choice

Natural speech tokens were used instead of synthetic tokens to ensure stimulus naturalness and to remove any confounds that might be introduced by synthesis methods. Synthetic stimuli give the experimenter more control over the stimulus attributes and have their place in experiments that require delicate and controlled manipulations (such as manipulating the slope of F2 by equal increments across stimuli). It might be argued that synthesizing a coda that is identical to the mirror image of the onset should ensure that the only variable in the experiment is the syllable position. However, it has been common knowledge for decades that synthetic stimuli are highly impoverished in relation to natural speech (Pisoni, 1981). Moreover, a quick glance at even a small number of spectrograms of the target of investigation reveals that onsets do not have the same dynamic characteristics as codas. If there is a difference in the degree to which the stimuli approximate the target of inquiry, then asymmetry in similarity to the target of the study is confounded with syllable position. If the stimulus is modeled closely after a natural production, the control of variables is lost, but an equivalent gain in naturalness is not achieved. Thus, natural speech was chosen to be sure that the responses to the stimuli were not due to the asymmetrically impoverished nature of synthetic stimuli. This does open up debate as to whether or not any results that show a positional advantage may be due to some difference in the signal dynamics between onsets and coda transitions.

3. Stimulus Recordings

Stimuli were constructed from the nonsense syllables /ba, da, ga, ab, ad, ag/ and read by a male speaker of American English. Prior to the recording session, syllable types were subjected to three randomizations and printed out as three lists for the speaker to read. In the list, each syllable type was printed five times on a separate line. During the recording session, the speaker was instructed to read each line at a moderate rate with a flat intonation on all but the last token of each line. The speaker paused for a breath between each line. The speaker took a 1-minute break between each list. The third repetition of each syllable type from the initial list was used to create stimuli. The same position from the same list for all stimuli was used to ensure the greatest similarity in intonation and original signal-to-noise ratio for all the syllable types. The speaker wore a close-talking head-mounted microphone with a flat response between 50 and 15,000 Hz (Sure SM10A) with a specially modified pickup to increase signal-to-noise ratio without changing frequency response. The recordings were digitized direct to disk at 22.05 kHz (16 bit) using Kay's Computerized Speech Laboratory 4300B.

4. Stimulus Editing

The files were leveled so that they all had the same average rms intensity as measured in the 100 ms surrounding the vowel's midpoint. This was done to ensure that the effect of noise was equivalent across places of articulation and across syllable position. The onset and offset bursts were digitally excised because the experiment was designed to concentrate on the relative impact of the information carried in the formant transitions in onset and coda. To ensure that no transient was introduced in the editing, the beginning and end of the cut-point for the excision was made at a zero crossing. The files were visually examined and played out to three trained phoneticians to ensure that no transient had been introduced. After the burst had been excised, a 50-ms period of silence was appended to the beginning and end of each file. Two levels of white noise were then mixed with copies of the stimuli. One noise level was set to be 2 dB below the rms average of the stimuli, and the other level was set at 2 dB above the rms average of the stimuli. There were six base stimuli with three "noise" conditions (no noise, –2 dB noise, +2 dB noise), resulting in 18 stimuli total.

5. Task

Experiment 1 was a self-paced three-way forced-choice task in which the listeners responded to stimuli by pressing one of three keys labeled "B," "D," and "G." The listeners were told that they would hear syllables that contained one of three consonant sounds — "b," "d," or "g" — occurring either at the beginning or at the end of the syllable. They were instructed to listen carefully to the syllables being played by the computer and to respond as quickly as possible by pressing the key that corresponded to the consonant in each syllable. They were to guess at a response if they were uncertain. There were 10 randomizations of each stimulus set resulting in 180 responses per subject. All listeners heard 10 randomizations, but a different randomization was used for each listener. The stimuli played out binaurally over headphones at a comfortable listening level that was fixed across listeners. Psyscope running on a Macintosh PowerBook G3 was used for randomization play-out and response collection.

B. Results

Perceptual salience of cues to place in onset and coda formant transitions was measured using a non-parametric measure of sensitivity, **A′** (Grier, 1972). **A′** varies from 0 (least sensitivity) to 1 (most accurate). A sensitivity measure was used because it takes into account potential biases in the listeners' responses (place in this case) by calibrating the proportion of correct responses (hits) with the proportion of incorrect uses of a particular response (false alarms). See the appendices for the overall confusion matrices used in Experiments 1 and 2.

The sensitivity data were submitted to a repeated-measures analysis of variance with *sensitivity* as the dependent variable and *consonant type*, *syllable position*, and *noise condition* as independent variables. The results are shown in Table 10.1. With an alpha level of 0.01, the analysis revealed a significant main

TABLE 10.1

Repeated Measures ANOVA Results for Sensitivity (A′) for Experiment 1

	df	F value	*p*	Power
Position	1	7.025	0.009	0.759
Consonant	2	13.846	<0.0001	0.999
Noise condition	2	58.189	<0.0001	1.000
Position * consonant	2	2.901	0.0588	0.548
Noise * position	2	7.742	0.0006	0.961
Noise * consonant	4	5.886	0.0020	0.989
Noise * pos * con	4	3.989	0.0037	0.914

effect for syllable position (coda, onset), consonant type (b, d, g), and noise condition (clear, noise at –2 dB below syllable rms, noise at +2 dB above syllable rms). There was a significant interaction between noise condition and consonant type, and between syllable position and noise condition. The three-way interaction between syllable position, noise condition, and consonant type was also significant. The greatest effect was for noise condition, and consonant type had a greater effect than syllable position.

The effect of syllable position on sensitivity is illustrated in Figure 10.3, a bar chart with sensitivity plotted on the *y*-axis and syllable position plotted on the *x*-axis. The *y*-axis shows a range from 0 to 1 (100% accuracy). The clear and reliable syllable position effect bears out the main hypothesis that place is more reliably recovered from cues in the onset transitions than cues in the coda transitions. Because the release burst was removed, this result is not due to redundancy alone, but rather indicates a more general perceptual onset advantage.

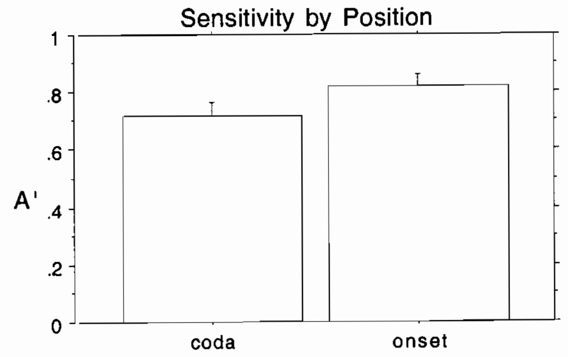

Figure 10.3. Bar chart illustrating the effect of syllable position on sensitivity. Sensitivity is plotted on the *y*-axis, and syllable position on the *x*-axis. The error bars show a 95% confidence interval.

Figure 10.4 illustrates the effect of noise condition across syllable position: sensitivity is plotted on the *y*-axis, and syllable position and noise level are plotted on the *x*-axis.

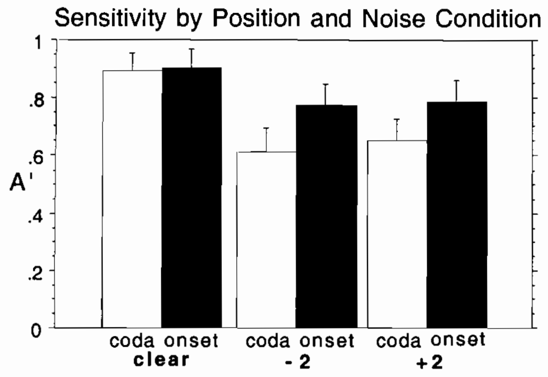

Figure 10.4. Bar chart illustrating the interaction of syllable position and noise level. Sensitivity is plotted on the *y*-axis, and syllable position and noise level on the *x*-axis.

In comparing the clear condition to the two noise conditions, the difference between the onset and the coda increases with the addition of noise in a way that indicates that the coda is more affected by noise than the onset is. However, while there is a noticeable difference between the clear condition and each of the noise conditions, there is very little difference between the two noise conditions. This pattern may indicate that the position-by-noise interaction is an artefact of the test conditions: the lack of difference between the onset and coda in the clear condition may be the result of a ceiling effect that artificially narrows the onset–coda disparity. A further experiment with larger step sizes or more steps in the noise conditions is needed to be sure that the syllable position-by-noise condition interaction is real. Thus, while the position-by-noise interaction appears to bear out the secondary hypothesis that coda transitions are more detrimentally affected by noise than onset transitions are, any conclusions must be tentative.

The reliable effect for consonant type and the reliable interactions between consonant type and noise conditions is also of interest. Although not significant, the trend in consonant type-by-syllable position interactions should also be considered. It bears directly on assumptions, such as those in Jun (1995), that certain places of articulation are less reliably recovered than others. In Figure 10.5, which illustrates these interactions, sensitivity is plotted on the *y*-axis, and place of articulation, syllable position, and noise condition on the *x*-axis. In general, the labial place of articulation is the most reliably recovered across syllable positions and noise conditions. In general, there is little difference between /d/ and /g/ across conditions. The results were submitted to Fisher's PLSD and Scheffé post-hoc tests with an alpha level of 0.01. While the differences between /b/ and /d/, and /b/ and /g/, were significant, the difference between /d/ and /g/ didn't reach significance.

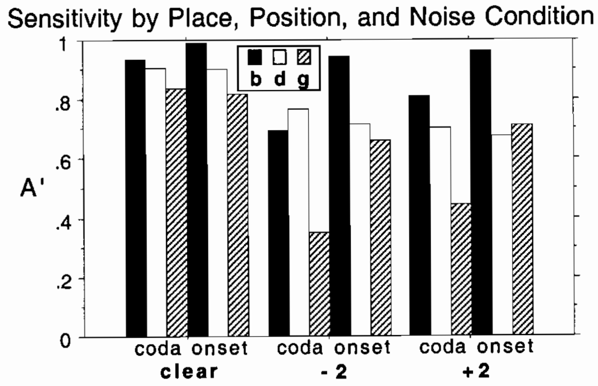

Figure 10.5. Bar chart illustrating the interaction of consonant type, syllable position, and noise level. Sensitivity is plotted on the *y*-axis, and consonant type, syllable position, and noise level on the *x*-axis.

Figure 10.5 illustrates an asymmetry across places of articulation in the noise conditions. Both /b/ and /g/ showed an onset advantage in the noise conditions, but /d/ either showed no onset advantage or an apparent coda advantage. The reason for the asymmetry may be an inherent place of articulation difference. However, it may be that listeners rely more on the release burst cues for alveolar stops in the onset and that the removal of the burst degraded the naturalness of the /d/ stimuli more than the others. If listeners are more reliant on /d/ release bursts, it may be due to their greater intensity and lower variability across vowels than /b/ or /g/ releases (Ohala, 1990). Because syllable final stops are often unreleased, the asymmetry across place of the relative contribution of the release burst may be absent. Further experiments that compare stimuli with and without release bursts are needed to determine the source of the asymmetrical /d/ behavior.

C. Discussion

The results of Experiment 1 are of interest for several reasons. First of all, they show a clear syllable position advantage even in a language that maintains place contrasts in the coda. If this sort of result were found for a language that maintained no coda contrasts, or even a reduced set of contrasts, the results might be interpreted as a language specific attentional factor. For example, speakers of a language such as Japanese might be expected to perform worse on this task than English speakers, because Japanese contains no place contrasts for the coda that are not part of an intervocalic geminate (which would per force have an onset portion). This sort of language-specific attentional factor has been reported for Japanese versus Dutch speakers in response to voicing contrasts in the coda (van Wieringen, 1995). The assumption here is that, because English has stop place contrasts in the coda, English speakers are trained to pay attention to information in the coda transitions, yet they still perform more poorly on place identification in the coda. The difference in salience of onset and coda transitions may be a factor that contributes to the crosslinguistic preference for onsets over codas. It may also be seen as motivating phonological constraints that govern such synchronic processes as resyllabification (MAX ONSET), coda deletion (NO CODA), loss of contrasts in the coda (see, e.g., Steriade, 1995), and metathesis (see, e.g., Hume, 1998).

The place of articulation effect is also of interest. First of all, it is interesting that the results of this experiment contradict both the assumptions in Jun (1995) and the findings of Hume *et al.* (1999). In this study, labial place appeared to be the most reliably recovered. If the findings of this study prove replicable, the interaction between place of articulation, syllable position, and noise condition indicates that we should be cautious in concluding too much about the salience of any particular place of articulation without taking into account syllable position and noise effects.

VI. EXPERIMENT 2

As discussed above, periodicity and transience are factors in a signal's ability to resist masking. The implications of this for phonotactics are obvious: in general, syllables that permit all consonants to have formant transitions in adjacent vowels should be crosslinguistically preferred, while those that strand consonants without formant transitions should be dispreferred. There are further predictions about the types of consonants that are more likely to be recoverable in the absence of vowel formant transitions: the less transient and the more intense a signal, the better internal cues it will have. Thus, fricatives, and especially the sibilants, can still be recovered in the absence of a flanking vowel, whereas a stop will be more likely to be lost. It is also important to remember that one of the choices that a listener is faced with in making lexical choices based on an incoming signal is not only about place, manner, and voicing, but also about whether or not a speech sound occurred at all. Thus, although nasal place information is poorly encoded in the signal, the presence of a nasal is well encoded because of the nasal's periodicity and relatively long duration (relative to release transients). In examining the types of syllable onsets that regularly violate the Sonority Sequencing Principle (and, by extension, Mattingly's preferred syllable types), fricative+stop clusters are the most common, nasal+stop are second most common, and stop+stop are the rarest (Greenberg, 1978). Experiment 2 examines the robustness of what can be called *internal cues*, cues that are found at the peak of stricture and at the onset of release of stricture. The prediction is that fricatives will be more reliably recovered than nasals or stops because they contain the most robust place cues. A second prediction is that there will be an interaction between noise levels and identification such that nasals, while poorly distinguished from each other over all, will be more resistant to masking noise than fricatives or stops, and that stops will be the most affected by noise.

A. Methods

1. Subjects

Experiment 2 had the same 21 unpaid volunteers who had participated in Experiment 1.

2. Stimulus Choice

Natural speech tokens were used: release burst noise, fricative noise, and nasal murmur excised from nonsense VC syllables produced in isolation.

3. Stimulus Recordings

Stimuli were constructed from the nonsense syllables /af, as, at, ap, am, an/ read by a male speaker of American English. "Word-final" consonants were used to ensure that the release bursts were unaccompanied by vowel formants or voicing. The recording methods were identical to those in Experiment 1, except that the speaker was instructed to use slow and careful pronunciation. Careful pronunciation was used to ensure that the word-final stops would be released.

4. Stimulus Editing

For each of the chosen syllables, the vowel portion was digitally removed, leaving behind only the release transient /p, t/, fricative noise /f, s/, or nasal murmur /m, n/. After isolating the stimuli, each pair was leveled. For the fricative pair and the nasal pair, stimuli were leveled to a common rms average intensity. The stop transients were leveled so that their peak intensity was equal to the rms average of the other stimuli. The stimuli were leveled so that the addition of noise was equivalent across stimuli. Because the hypothesis being tested has to do with between-manner and not within-manner differences, this was seen as the only way of being able to test the hypothesis while minimizing the effect of other correlated but untested variables. It must be noted here that leveling potentially decreases the naturalness of the stimuli, and it may degrade within-manner differences that listeners may use in perceiving place of articulation. The most dramatic effect of leveling is in the fricative case, where the amplitude of /s/ was dropped to match the rms value for /f/. The leveling may also artificially increase the sensitivity scores for the stop release bursts, especially /p/, because they become equivalent to the intensity of /s/. After leveling, 50 ms of silence was appended to the beginning and end of stimuli. The stimuli were mixed with the same noise levels as in Experiment 1 (clear, +2 dB, –2 dB).

5. Task

For consonants that are stranded without any flanking vowels (for nasals and obstruents), a listener must distinguish not only one speech sound from another but also whether a speech sound has occurred at all. Therefore, in this task the listeners were presented with three stimulus types: one type for each stimulus pair, and one file that contained no information (only silence or masker noise). Experiment 2 was a self-paced three-way forced-choice task that was blocked for stimulus pairs. For the stop block there were nine stimulus files total: /p/ transient at three noise conditions (no noise, –2 dB of noise, +2 dB of noise), /t/ transient at three noise conditions, and an empty file of the same length at three noise levels (silence, masking for the –2 dB condition, and the masking noise for the +2 dB condition). Similarly, there were nine stimuli for the fricative block, and nine stimuli for the nasal block. The listeners responded to stimuli by pressing one of three keys labeled "P,T,None," "F,S,None," or "M,N,None." In between blocks, the listeners took a break while they received a new set of instructions. The listeners were told at the beginning of the stop block that they would hear syllables that contained one of three types of sounds: "p," "t," or "nothing." Similar instructions were given at the beginning of the other two blocks. They were instructed to listen carefully to the sounds being played by the computer and to respond as quickly as possible by pressing the key that corresponded to the appropriate sound. They were instructed to guess at a response if they were uncertain. There were 10 randomizations of each stimulus block in 90 responses per block and 270 responses over all. All listeners heard 10 randomizations and 3 blocks, but a different randomization and a randomly generated block order was used for each listener. The same playout and button press collection was used as in Experiment 1.

B. Results

Perceptual salience of cues to place in onset and coda formant transitions was measured using the same non-parametric measure of sensitivity, with the empty stimulus treated as a "place." The sensitivity data were submitted to a repeated-measures analysis of variance with *sensitivity* as the dependent variable and *manner* and *noise condition* as independent variables. The results are shown in Table 10.2. With an alpha level of 0.01, the analysis revealed a significant main effect for manner (fricative, stop, nasal), and for noise condition (clear, noise at –2 dB below syllable rms, noise at +2 dB above syllable rms). There was also a significant interaction between noise condition and manner. The greatest effect was for noise condition.

TABLE 10.2

Repeated Measures ANOVA Results for Sensitivity (A′) for Experiment 2

	df	F-value	*p*	Power
Manner	2	70.669	<0.0001	1.000
Noise condition	2	60.093	<0.0001	1.000
Manner * noise	4	19.041	<0.0001	1.000

The results were also submitted to Fisher's PLSD and Scheffé post-hoc tests with an alpha level set at 0.01. Across manner, the differences between 0 and –2 dB was significant, as was the difference between the –2 dB and +2 dB conditions. The differences between nasal and stop, and between nasal and fricative, were significant; however, the difference between stop and fricative were not. Figure 10.6 illustrates the effect of noise condition across manner: sensitivity is plotted on the *y*-axis, and manner and noise level are plotted on the *x*-axis.

One unsurprising aspect of the results is the overall poor identification of nasals. Nasals are in general hard to distinguish from each other in the absence of formant transitions. On the other hand, as predicted, nasals were much less affected by noise than the stops. Somewhat surprisingly, when there is no masking noise, listeners were able to identify the stop place of articulation with more accuracy than the fricative stimuli. Although the stop release advantage may be due to stimulus leveling, it is more likely due to the information-rich nature of stop releases (Blumstein & Stevens, 1981). When the transients are heard without masking noise and when the listener's attention is fully on the task of identification, there is little penalty for transience. However, there is a sharp drop off in identification accuracy with the addition of noise, which supports the prediction that transient signals will be more affected by noise than signals with longer durations. At the highest noise level, listener accuracy for stop place identification drops below even the nasal place identification level. The lack of significance between stop and fricative manners in the post-hoc tests may be due to the dramatic effect of noise on the stop release; they range from the highest sensitivity score in the 0 noise condition, to a level equivalent to the stops in the –2 dB noise condition, to having a lower score than the nasals in the +2 dB noise condition.

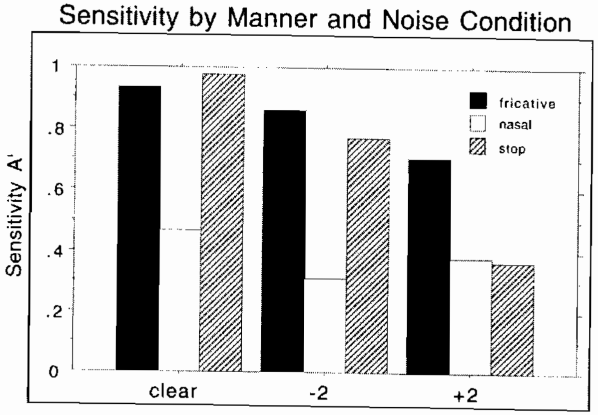

Figure 10.6. Bar chart illustrating the interaction of consonant manner and noise level. Sensitivity is plotted on the *y*-axis, and manner and noise level on the *x*-axis.

To examine the change in sensitivity across noise conditions of the individual manners, a second measure was used: the difference between the clear condition and the noisiest condition (+2 dB). Figure 10.7 is a box plot illustrating the effect of noise on the three manner types; the difference in sensitivity is plotted on the *y*-axis against consonant manner on the *x*-axis.

From the plot, two aspects of the data are evident. The first is that the difference in sensitivity between the clear condition and the noise condition is greatest for stops. Moreover, there is a slight difference between the means of the fricatives and the nasals. The second is that there is more variability, seen as the height of the box, in the data for both the fricatives and the nasals than there is for

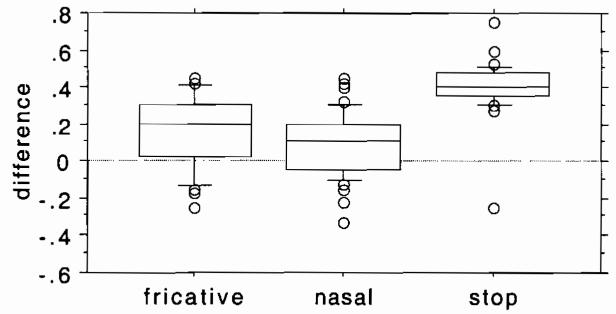

Figure 10.7. Box plot of the difference in sensitivity between the "clear" condition and the "noise" (+2 dB) condition. Difference in sensitivity (A′) is plotted on the *y*-axis against consonant manner on the *x*-axis.

the stops. The differences were submitted to an unpaired *t*-test with an alpha of 0.01. As might be expected from the plot in Figure 10.7, there is a reliable difference in the stop–fricative comparison and in the stop–nasal comparison, but the fricative–nasal difference is not reliable.

C. Discussion

The results of Experiment 2 are of interest because they point to the importance of considering the asymmetrical degradation of information across manners of articulation. Both nasal murmur and frication are signals with relatively long durations, giving them an inherent advantage over stop releases when a listener is not devoting all of her attentional resources to the task and when there is low-level background noise, as is the normal case for spoken language. It should be noted that even the "high-noise" condition is not greater than is frequently observed for a wide variety of speaking conditions. It is interesting that, despite being aperiodic, the fricatives still resisted the noise condition as well as the nasals. This may be an indication that at moderate noise levels the periodicity of the signal is of less importance than transience in determining the robustness of information. Without further tests such as phoneme monitoring or other detection tasks, the relative weighting of transience over periodicity remains a speculation and not a real finding.

The differences across manner should be of particular interest to those considering phonetic motivation for constraints on segmental ordering within the syllable or word. This type of finding can motivate a revision of the Sonority Sequencing Principle: instead of segments being ordered from greatest stricture to least stricture going from the onset to the syllable nucleus, and from least stricture to greatest stricture from the nucleus to the coda, segmental ordering can refer to the relative recoverability of information, or to cue robustness. With this type of perceptually motivated sequencing, the most common sonority "violation" patterns are expected to occur. For example, fricative+stop clusters, the most common sonority reversal, are by far more common than stop+stop clusters. One nice side effect of recasting the Sonority Sequencing Principle as perceptually motivated is that this unifies several phonotactic constraints into a single perceptual sequencing constraint: segments should be ordered so that transitions from one into the next provide sufficient information for the lexical item to be recovered under normal listening conditions (see Wright, 1996, for discussion).

VII. CONCLUSIONS

The results from both experiments illustrate the importance of taking into account environmental factors when considering the role that perception may play in phonological processes and in typological patterns. Experiment 1 demonstrates an onset advantage that may be one of the factors in determining the typological preference for onsets over codas, and may be the foundation for the NO CODA constraint in OT. It is worth noting that the effect was not symmetric across places of articulation; while listeners responses to bilabial stops showed a clear advantage for onsets over codas, and a less extreme but similar pattern for velar stops, responses to alveolar stops show no clear advantage for onsets or codas. Also of interest was the difference across place: on the whole, bilabial place was identified with more accuracy than either alveolar or velar place. It may be that with more vowel conditions the velar–alveolar difference would tease apart, or that the accuracy ranking would change altogether (see Hume *et al.*, 1999). This set of findings should sound a note of caution for those phonologists who use the findings of previous experiments to motivate phonological constraints: not all cues can be treated as equivalent. A cue that was established through work on CV syllables with no noise conditions may not have the same perceptual strength when it occurs in different syllable positions or in noise. The interaction of noise with syllable position in Experiment 1 and with consonant manner in Experiment 2 clearly indicates that cues should not be considered abstract entities; rather, they should be considered under conditions that are more like those of spoken language: with noise and distractions. Although enticing, this set of experiments is but a first step in that direction. Other variables (such as distractions or cognitive loads) should be considered before we can make categorical statements about the role of a cue or cues in a phonological process.

APPENDIX A:
CONFUSION MATRICES FOR EXPERIMENT 1

Onsets

Confusion matrix for onsets no noise

	b	d	g
b	203	3	4
d	1	180	29
g	1	44	165

Confusion matrix for onsets -2 dB noise

	b	d	g
b	167	30	13
d	4	127	79
g	2	122	86

Confusion matrix for onsets +2 dB noise

	b	d	g
b	180	16	14
d	2	90	118
g	11	189	10

Codas

Confusion matrix for codas no noise

	b	d	g
b	193	14	3
d	7	194	9
g	12	53	145

Confusion matrix for codas -2 dB noise

	b	d	g
b	123	47	40
d	9	192	9
g	86	83	41

Confusion matrix for codas +2 dB noise

	b	d	g
b	151	31	28
d	143	39	28
g	46	105	59

APPENDIX B:
CONFUSION MATRICES FOR EXPERIMENT 2

Frication noise

Confusion matrix for fricatives with no noise

	f	s	none
f	182	3	25
s	7	169	34

Confusion matrix for fricatives with -2 dB noise

	f	s	none
f	142	17	51
s	25	146	39

Confusion matrix for fricatives with +2 dB noise

	f	s	none
f	106	39	65
s	30	120	60

Nasal murmur

Confusion matrix for nasals with no noise

	m	n	none
m	139	71	0
n	133	77	0

Confusion matrix for nasals with -2 dB noise

	m	n	none
m	100	87	23
n	126	68	16

Confusion matrix for nasals with +2 dB noise

	m	n	none
m	113	86	11
n	132	65	13

Stop release

Confusion matrix for stops with no noise

	p	t	none
p	189	18	3
t	0	208	2

Confusion matrix for stops with -2 dB noise

	p	t	none
p	104	71	35
t	14	177	19

Confusion matrix for stops with +2 dB noise

	p	t	none
p	66	60	84
t	80	58	72

ACKNOWLEDGMENTS

The author would like to thank Susie Levi for assistance in data collection and tabulation, the organizers of the satellite meeting, and the reviewers and members of the audience for helpful comments.

REFERENCES

Bell, A., & Hooper, J. B. (Eds.) (1978). *Syllables and segments*. Amsterdam: North-Holland.

Bladon, R. A. W. (1986). Phonetics for hearers. In G. McGregor (Ed.), *Language for hearers* (pp. 1–24). Oxford: Pergamon.

Beckman, J. (1997). *Positional faithfulness*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Blumstein, S. A., & Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, *10*, 25–32.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 263–282.

Cutler, A., & Clifton, C. (1983). The use of prosodic information in word recognition. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performances, X: Control of language processes* (pp. 183–196). New York: Lawrence Erlbaum.

Delgutte, B., & Kiang, N. Y. S. (1984a). Speech coding in the auditory nerve, III: Voiceless fricative consonants. *Journal of the Acoustical Society of America*, *75*, 887–896.

Delgutte, B., & Kiang, N. Y. S. (1984b). Speech coding in the auditory nerve, IV: Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America*, *75*(3), 897–907.

Flemming, E. (1995). *Auditory representations in phonology*. Unpublished doctoral dissertation, University of California, Los Angeles.

Fujimura, O., Macchi, M. J., & Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, *21*, 337–346.

Gordon, M. (1999). *Syllable weight: Phonetics, phonology and typology*. Unpublished doctoral dissertation, University of California, Los Angeles.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*, 1–42.

Greenberg, J. H. (1978). Some generalizations concerning initial and final consonant clusters. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 243–379). Stanford: Stanford University Press.

Greenberg, S. (1995). Auditory processing of speech. In N. J. Lass (Ed.), *Principles of experimental phonetics*. St. Louis: Mosby.

Grier, J. B. (1972). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.

Jakobson, R. (1962). *Selected writings, I: Phonological studies*. The Hague: Mouton.

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis*. Cambridge: MIT Press.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. Unpublished doctoral dissertation, University of California, Los Angeles.

Hume, E. (1998). The role of perceptibility in consonant/consonant metathesis. In S. Blake, E.-S. Kim, & K. Shahin (Eds.), *WCCFL XVII Proceedings* (pp. 293–307). Stanford: CSLI.

Hume, E., Johnson, K., Seo, M., & Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 2069–2072.

Kawasaki, H. (1982). *An acoustic basis for universal constraints on sound sequences*. Unpublished doctoral dissertation, University of California, Berkeley.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., & Clark, L. F. (1965). *Discharge patterns of single fibers in the cat's auditory nerve*. MIT Research Monograph No. 35. Cambridge: MIT Press.

Kirchner, R. (1997). Contrastiveness and faithfulness. *Phonology*, *14*, 83–111.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.

Mattingly, I. G. (1981). Phonetic representations and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representations of speech* (pp. 415–420). Amsterdam: North Holland.

Miller, G. A., & Nicely, P. (1955). An analysis of perceptual confusions among English consonants. *Journal of the Acoustical Society of America*, *27*, 338–352.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The Intelligibility of speech as a function of the context of test materials. *Journal of Experimental Psychology*, *41*(5), 329–335.

Neary, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*(6), 3241–3254.

Nygaard, L. C., & Pisoni, D. B. (1995). Speech Perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 63–96). San Diego: Academic Press.

Ohala, J. J. (1990). Phonetics and phonology of aspects of assimilation. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology, I: Between the grammar and the physics of speech* (pp. 258–275). Cambridge: Cambridge University Press.

Ohala, M., & Ohala, J. (1998). *Correlations between Consonantal VC transitions and degree of perceptual confusion of place contrast in Hindi. Proceedings of the 5th International Conference on Spoken Language Processing, Sydney*, pp. 2795–2798.

Pierrehumbert, J. B. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363–369.

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *Journal of the Acoustical Society of America*, *70*, S98.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Unpublished manuscript, Rutgers University, New Brunswick, NJ, and the University of Colorado, Boulder.

Selkirk, E. O. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *Features, segmental structure and harmony processes*, Part II (pp. 337–383). Dordrecht: Foris.

Silverman, D. (1997). *Phasing and recoverability*. New York: Garland.

Sinex, D. G. (1995). Auditory nerve fiber representation of cues to voicing in syllable-final stop contexts. *Journal of the Acoustical Society of America*, *90*, 2441–2449.

Sinex, D. G., & Geisler, C. D. (1983). Responses of auditory-nerve fibers to consonant–vowel syllables. *Journal of the Acoustical Society of America*, *73*(2), 602–615.

Smith, R. L. (1979). Adaptation, saturation and physiological masking in single auditory-nerve fibers. *Journal of the Acoustical Society of America*, *65*(1), 166–178.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Steriade, D. (1995). *Neutralization and the expression of contrast*. Unpublished manuscript, University of California, Los Angeles.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215.

van Wierengen, A. (1995). *Perceiving dynamic speechlike sounds*. Unpublished doctoral dissertation, University of Amsterdam.

Wright, R. (1996). *Consonant clusters and cue preservation in Tsou*. Unpublished doctoral dissertation, University of California, Los Angeles.

Wright, R. (1999). Tsou consonant clusters and auditory cue preservation. In E. Zeitoun & P. Li (Eds.), *Selected papers from the eighth international conference on Austronesian linguistics* (Vol. 1, pp. 227–312). Taipei: Academia Sinica, Institute of Linguistics.“