# On The Perception of /s/ and /ʃ/: Considering the Effects of Phonotactics

Honors Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation "with
Honors Research Distinction in Linguistics" in the Undergraduate
Colleges of The Ohio State University

By Kristen Scudieri

The Ohio State University
February 2012

Project Advisors: Professor Elizabeth Hume, Department of Linguistics
Assistant Professor Cynthia Clopper, Department of Linguistics

## *1. Background*

The current study explores the interaction between phonological relationships and speech perception. While phonemic status has been thoroughly considered in the speech perception literature, much less research has been devoted to the perception of phones of different phonological status, such as those exhibiting allophony or partial contrast. The goal of the current study was to extend our understanding of partial contrast by considering partial contrast due to phonotactic constraints. Specifically, the research questions were: to what extent do gaps in phonemic distribution affect the perceptual distinctiveness of otherwise contrastive pairs? Within a single language and in a range of environments where a pair of segments differs in degree of contrast, does the degree of perceptual similarity differ, or is perceptual similarity constant no matter the environment?

Recent literature has begun to consider this new area of study through a look at the effects of partial contrast and allophony on perception. In 2003, Hume and Johnson explored partial contrast due to neutralization in their study of Chinese tone. Specifically, they examined the low-falling-rising tone (214) and the mid-rising tone (35) of Mandarin Chinese, which in nearly all environments are contrastive, but whose contrast is neutralized after a low-falling-rising tone. The results showed that perception of the tones by Mandarin speakers in both neutralized and non-neutralized contexts nearly mimicked perception of non-contrastive tones in a speeded discrimination task. American English speakers perceived these same tones with some difficulty, rating four distinct Mandarin tones closely to one another. These results suggest that, at least for some prosodic features, partial contrast plays a role in perception by reducing perceptual distinctiveness of otherwise contrastive segments.

More recently, Boomershine et al. (2008) studied the perception of coronals in allophonic versus phonemic relationships in Spanish and English. Consistent across two tasks, they found that allophonic coronals were perceived as more similar than those enjoying phonemic status. Since the examined consonants contrast differently in the two languages, ([d]/[ɾ] being allophonic and [ð] being phonemic in English; [ð]/[d] being allophonic and [ɾ] being phonemic in Spanish), the pairs perceived as more similar changed based on native language of the listener. This finding further emphasized the suggestion that allophony influences speech perception, and specifically, that allophonic pairs are perceived as more similar than phonemic pairs.

The findings of Hume and Johnson (2003) and Boomershine et al. (2008) suggest that partial contrast and allophony individually affect speech perception, specifically decreasing perceptual distinctiveness of segments. Most recently, however, Johnson and Babel (2010) compared the perception of allophonic and partially contrastive pairs within a single study. The authors compared the perception of /s/ and /ʃ/, among other fricatives, in Dutch and English, classifying these phones as allophonic in Dutch, aside from their use in non-adapted loanwords, and fully contrastive in English. Still, given the phones' spotty distribution over phonotactic gaps (see Section 2.1.) and morpho-phonological alternations in English (Johnson and Babel 2010;129), this study is truly a comparison of allophony in Dutch verses partial contrast in English.

In a perceptual similarity rating task, Johnson and Babel found that listeners perceived /s/ and /ʃ/ as more similar when they were in an allophonic relationship (Dutch) than when their relationship was partially contrastive (English). In a speeded AX discrimination task, on the other hand, Johnson and Babel (2010) found no significant difference in responses between languages.

Both Hume and Johnson (2003) and Johnson and Babel (2010) compared partial contrast in one language with allophony or non-contrastiveness in another. Based on the responses of their Mandarin listeners, Hume and Johnson (2003) found that a partially contrastive pair's neutralization in one environment caused its perceptual distinctiveness to be reduced in all environments. Johnson and Babel (2010), on the other hand, looked at their partially contrastive fricatives in three intervocalic environments, none of which exhibited the neutralization of contrast shown in certain English environments. English listeners' perception varied significantly across each of these intervocalic environments, but the cause is unknown because of the absence of neutralized contexts. Nonetheless, across all environments, the partially contrastive pair (English /s/ and /ʃ/) was perceived as more distinct than the non-contrastive/allophonic pair (Dutch /s/ and /ʃ/) in a perceptual similarity rating task[1].

These two studies offer two different findings on whether partial contrast reduces the perceptual distinctiveness of segments in different environments. Naturally, the studies had different aims and used different experimental designs. Additionally, Hume & Johnson (2003) considered partial contrast only with respect to prosodic tones and created by contextual neutralization, while Johnson and Babel (2010) looked at partial contrast in fricatives. Although any of these factors might account for the different

---

[1] See Section 1.1. on perceptual similarity rating tasks verses speeded discrimination tasks.

findings on partial contrast, the differing results spark the need for further study into the nature of partial contrast and its influence on perception.

The present study considered one aspect of partial contrast through a close examination of the English fricatives /s/ and /ʃ/. These specific phones are of special interest because of their phonemic status in English, yet their variation in contrast depending on environment. In many contexts, /s/ and /ʃ/ are fully contrastive (e.g. ‘seep’/sip/ ~’sheep’ /ʃip/, ‘fasten’ /fæsən/ ~‘fashion’/fæʃən/, ‘lease’/lis/ ~ ‘leash’ /liʃ/). In other environments, however, they can be non-contrastive (e.g. ‘shrimp’/ʃɹɪmp/ ~ *‘srimp’ /sɹɪmp/, ‘asleep’/əslip/ ~ * ‘ashleep’/əʃlip/) or even allophonic across speakers (‘street’/stɹit/ ~ ‘street’ /ʃtɹit/). What is unclear from previous literature is if a pair of sounds is perceived as equally similar across all environments or whether the pairs’ perception is influenced by the individual sounds’ distribution.

### *2.1. Distribution of English /s/ and /ʃ/*

In English, /s/ and /ʃ/ are sibilant fricatives that differ only by place of articulation; while /s/ is classified as alveolar, /ʃ/ is generally described as palato-alveolar or post-alveolar (Rutter 2011). The two are distinct phonemes as they are contrastive in all major word positions (word-initial, word-medial, and word final). Again, we find minimal pairs in each of these word contexts: ‘seep’/sip/ ~ ‘sheep’ /ʃip/, ‘fasten’ /fæsən/ ~‘fashion’/fæʃən/, and ‘lease’/lis/ ~ ‘leash’ /liʃ, respectively. Despite this phonemic contrast, however, Johnson and Babel (2010) point out that /s/ and /ʃ/ do occasionally alternate with one another allophonically as in *miss* [mɪs] ~ *miss you* [mɪʃu] and morpho-phonologically as in *confess* [kənfɛs] ~ *confession* [kənfɛʃən] (129).

Additionally, while this fricative pair is contrastive in each major environment in English and consistently in pre-vocalic position, it is non-contrastive in a few clusters. In these instances, consonant sequences including an initial /s/ or /ʃ/ are non-occurring because of phonotactic gaps. These phonotactic gaps make the distribution of the fricatives far from even across English.

In cluster-initial position, the permissibility of /s/ and /ʃ/ may vary to different degrees. In specific environments, phones within the pair /s/ and /ʃ/: may neither be permissible, one phone may be marginally permissible, one may be disallowed, or the use of the two phones may be allophonic. The first case is exemplified by the fact that neither /s/ nor /ʃ/ may appear before /g/ among other phones. In the second case, the phone /ʃ/ rarely, but occasionally, appears in a CC cluster before /l/, while the phone /s/ is often found in this position (Giegerich 1992; Hammond 1999). Although the /ʃl/ form is not completely disallowed, its occurrence is marginal, only appearing in a few English words such as ‘schlep’, ‘schlub’, and ‘schlok,’ all of which are borrowings from Yiddish. The third case is shown by the fact that the phone /ʃ/ but not /s/ may appear in a CC cluster before /ɹ/.

The fourth case is a slightly more complicated issue. Traditionally, only /s/ could act as the first constituent in an English CCC cluster (Giegerich, Hammond); however, recent studies have shown that in this specific context, pockets of speakers

*Table 1. Distribution and Relationship of English /s/ and /ʃ/.*

| **Fricative Pair** | **Phonemic Relationship** | **Environment** | | **Example** | **Type Frequency** |
|---|---|---|---|---|---|
| [sɑ]/[ʃɑ] | fully contrastive | **Word-Initial** | sV | ‘seep’ /sip/ | 1029 |
| | | | ʃV | ‘sheep’ /ʃip/ | 157 |
| | | **Intervocalic** | VsV | ‘fasten’ /fæsən/ | 909 |
| | | | VʃV | ‘fashion’ /fæʃən/ | 1007 |
| [slɑ]/[ʃlɑ] | marginally contrastive | **Word-Initial** | slV | ‘sleep’ /sip/ | 79 |
| | | | ʃlV | ‘shlep’ /ʃlɛp/ | 0 |
| | | **Intervocalic** | VslV | ‘asleep’ /ʌslip/ | 40 |
| | | | VʃlV | ‘ashlar’ /æʃləɹ/ | 1 |
| [sɹɑ]/[ʃɹɑ] | non-contrastive | **Word-Initial** | sɹV | ø | 0 |
| | | | ʃɹV | ‘shrimp’ /ʃɹɪmp/ | 22 |
| | | **Intervocalic** | VsɹV | ‘crossroad’ /kɹɑsɹod/ | 10 (all heteromorphemic) |
| | | | VʃɹV | ‘mushroom’ /mʌʃɹum/ | 2 |
| [stɹɑ]/[ʃtɹɑ] | non-contrastive/allophonic | **Word-Initial** | stɹV | ‘string’ /stɹɪŋ/ | 82 |
| | | | ʃtɹV | ‘string’ (alt. pronunciation) /ʃtɹɪŋ/ | 0 |
| | | **Intervocalic** | VstɹV | ‘astrology’ /əstɹɑlədʒi/ | 106 |
| | | | VʃtɹV | ‘astrology’ (alt. pronunciation) /əʃtɹɑlədʒi/ | 0 |

around the country are now using an innovative form. Specifically, they are producing orthographic ‘str’ as [ʃtɹ], [ʃtʃɹ] or as a form shown by intermediate spectral energy and spectral peak to be transitional between [stɹ] and [ʃtɹ] (Durian 2007; Rutter 2011). While the factors conditioning the use of this innovation are not fully known, findings from Durian (2007) seem to suggest that the innovation is not equal across speakers, for use of the innovative form increases as age and socio-economic class decrease and as “urban affiliation” increases[2].

This phenomenon has now been strongly attested in production, yet the current literature provides no information on how /s/ and /ʃ/ are *perceived* in the environment [_tɹ]. Although slight variability exists within the productions of a single speaker, Rutter (2011) proposed that each speaker tends to prefer one form to the others. Thus, the variability in phonotactic well formedness of /ʃ/ in this environment seems be observed across speakers rather than within them as in the [_l] environment.

In summary, although /s/ and /ʃ/ share phonemic status in English, there are certain environments where they are contrastive, non-contrastive, marginally contrastive or allophonic because of phonotactics. More precisely, in word-initial position, [sɹ] is an illicit cluster, while the permissibility of [ʃl] is marginal and that of [ʃtɹ] is illicit or allophonic with [stɹ]. All other clusters are unconstrained by phonotactics.

---

[2] Several of these factors are questionable as Durian used his own perceptions during rapid anonymous surveys in forming his conclusions. He did not record or analyze any of the data from this experiment but instead simply trusted his ear. Rutter points out that in his own experiment, several of the sounds initially perceived as [s] were found under spectrographic scrutiny to more closely resemble [ʃ] (34). We can expect Durian made similar mistakes and thus, might expect more instances of [ʃ] under close scrutiny.

Table 1 summarizes the permissibility of /s/ and /ʃ/ in each of the aforementioned contexts and shows their level of contrast in each relationship. Type frequency counts are included so as to attest to the phonological distribution described above. These are not real-life counts, but rather, are taken from the Hoosier Mental Lexicon, a database of some 20,000 words. For sequences that are present in the lexicon but are not theoretically possible because of phonotactic constraints, it is important to mention whether they are heteromorphemic or monomorphemic. The reasoning behind this is explained in Section 2.3., but for now it is simply noted next to the frequency counts.

### *2.2. Acoustic cues of /s/ and /ʃ/*

In addition to the phonological distinctions between /s/ and /ʃ/ in English, certain acoustic distinctions between the two play a role in their perception. In her article on perception, Steriade (1997) almost intuitively suggests that the better a set of cues for phone identification in each environment, the more likely that contrast will be maintained. In her suggestion, a better cue package consists of more cues and cues of better quality (Steriade 1997). For fricatives, four factors have generally been viewed as cues to identity: spectral qualities of the frication, amplitude of the frication, duration of the frication, and transition from the frication to the following vowel (Jongman, Wayland, and Wong 2000). Of these factors, those qualities within the fricative itself have been seen as primary in previous research.

This is supported by the fact that formant transitions (in F2 and sometimes F3) have been shown as poor markers of fricative identity. Nittrouer and Miller (1997) assert that while formant transitions between fricatives and the following vowels do affect fricative identification, this affect varies across age groups. Therefore, they say, this is especially true among children (Nittrouer and Miller 1997). As they grow older, though, adults use these same transitions more as identifiers of the following vowel quality (Nittrouer and Miller 1997). This is consistent with the findings of Soli (1998) who says that in environments where the fricative precedes a stop or a vowel, anticipatory formant coarticulation is useful only in the identification of the following phone, not in the identification of the fricative (Soli 1981).

Fricative cues within the frication itself have also been shown as extremely important in the phones' identification. Spectral peak is consistently measured higher for alveolar fricatives [s,z] than for palatal fricatives [ʃ,ʒ] (Soli 1981). Previous research has shown that this and other cues determining the place of articulation within sibilant fricatives are found primarily within the frication itself, rather than in the surrounding vowel transitions (Harris 1958, Martin & Peperkamp 2011). Even lone spectral slices of frication seem to provide enough information for correct identification of sibilant fricatives (Hughes and Halle 1956). Because these cues are so strong and those within the following vowel transitions are questionable, the frication itself appears to be the primary means of conveying fricative identity.

### *2.3. Syllabic considerations*

Before moving on to the current study, it is necessary to make one final consideration about syllable and morpheme boundaries, as alluded to in Table 1. First, the issue of whether phonotactic constraints hold across syllable boundaries needs to be considered.

While some authors assert that phonotactics are only relevant within the syllable (Giegerich 1992), others propose the syllable is only one of several phonotactic domains, and describe phonotactics as likewise active in consonant clusters across a syllable boundary (Harris 1994; Hammond 1999). Practical experience and cluster restrictions for multisyllabic medial clusters listed in language grammars most definitely suggest the latter position (see Hammond 1999).

Scholars assert that phonotactic restrictions do not hold across *morpheme* boundaries, however, saying that no cluster is forbidden in heteromorphemic compounds that could be broken up into an allowable coda and onset (Hammond 1999). This position fits with the data in Table 1 taken from the Hoosier Mental Lexicon. No tokens were found in the corpus of illicit /s/ and /ʃ/ clusters occurring within a single morpheme, and only one example exists of a marginal cluster [ʃl] occurring within a single morpheme (i.e. 'ashlar'). Further, experience with the English language confirms that several of the present study's illicit or marginal stimulus sequences do occur medially in English across morpheme boundaries (e.g. 'crossroad' [kɹɑs+ɹod], 'marshland' [mɑɹʃ+lænd], etc.). Taking this evidence into consideration, we expect that morpheme boundaries matter significantly in the perception of phontactically illicit clusters whereas syllable boundaries have no bearing.

This consideration is especially important because in addition to the word-initial environments listed above, the present study also considers /s/ and /ʃ/ in parallel intervocalic sequences. Thus, for each word-initial pair, there is one matching pair differing only by the addition of an initial /ɑ / (i.e. for [sɑ]/[ʃɑ], we have [ɑsɑ]/[ɑʃɑ], for [sɹɑ]/[ʃɹɑ], [ɑsɹɑ]/[ɑʃɹɑ] and so on). In summary, the present study positions /s/ and /ʃ/ in eight different environments: word initially ([_ɑ], [_ɹɑ], [_lɑ], [ _tɹɑ]) and intervocalically ([ɑ_ɑ], [ɑ_ɹɑ], [ɑ_lɑ], [ɑ_tɹɑ]).

## 3. *Current Study*

### *3.1. Tasks for paired comparison*

Following the model of previous perception studies in phonology (e.g. Johnson and Babel 2010), the present study used two paired comparison tasks to help illuminate the perceptual differences between /s/ and /ʃ/ in specific environments. These two paradigms are a perceptual similarity rating task (PSRT), which asks participants to rate the similarity of two sounds without time constraints, and an AX task, which asks participants to quickly judge whether pairs of sounds were the same or different. In this second task, faster reaction time is taken as a measure of greater perceptual distinctiveness because it is easier to judge to distinct sounds as different than it is to judge two less distinct sounds (See Shepherd, Kilpatric and Cunningham 1975, Nosofsky 1992). Traditionally, a PSRT has been assumed to reach a deeper, phonological level of speech processing and therefore show language-specific effects on perception (See Boomershine 2008, Johnson & Babel 2010, Huang 2004). An AX test, on the other hand, has been assumed to reach a lower level of processing based on acoustics (See Fox 1984; Strange and Dittman 1984; Werker and Logan 1985, Johnson and Babel 2010).

Although AX tests are designed to show little to no language-specific effects, several recent studies suggest that ridding experimental results of linguistic experience may not be entirely possible (Boomershine et al. 2008, Huang 2004, and Krishnan et al. 2005). In their study of Spanish and English coronals, Boomershine et al.'s (2008) PSR and AX tasks both showed effects of native phonology. Similarly, Huang's (2004) study of Mandarin tone neutralization considered participants with differing knowledge of tone—American English speakers and Mandarin speakers. Across 3 discrimination tasks, listeners with knowledge of Mandarin tone distinguished between tones, at least in part, based on native language neutralization patterns.

These results suggest that the role of language background can never be totally eliminated, yet prior to these studies, most discrimination tasks using the same methods have apparently removed linguistic experience successfully. As neither of the interpretations has emerged as the clear front runner, we are left to question the use of AX tasks as a successful means to understand raw acoustics. The current study used both this AX task and a PSRT for two purposes. First, the PSRT was used to understand the effects of a partially contrastive relationship in a pair of sounds whose distribution differs in different positions in English. Second, an AX test was included to contribute to the debate on linguists' ability to completely remove language experience as a factor in perception.

### *3.2. Predictions*

The aims of this study are two-fold: first, to compare the perception of fricative sequences between our two word positions, word-initially and intervocalically; Second, to compare the fricatives within the same word position but within different local contexts or consonant sequences. Although our study tests pairs of different types (e.g. [sɑ]/[ ʃɑ]) and pairs of the same type (e.g. [sɑ]/[sɑ]), our focus is on the different pairs as the same trials were added primarily to break up large blocks of different trials. Perception of these same pairs may provide some interesting results, but as they are tangential to our goals, they will not be discussed here.

### *3.2.1. Rating task predictions*

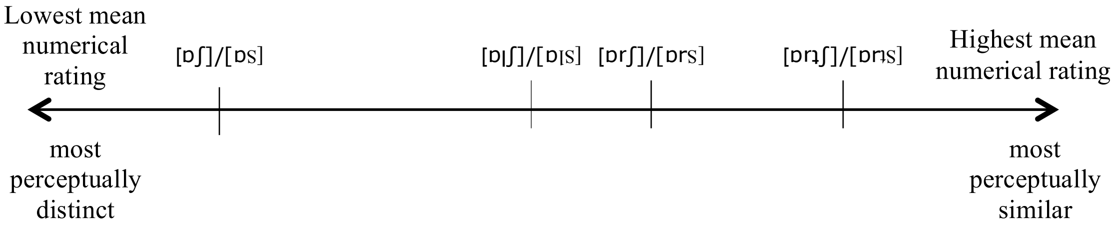
To start, I will consider the perceptual similarity rating task in light of my first aim: comparing the perception of word-initial consonant sequences with the perception of intervocalic sequences. Considering acoustic properties of the fricatives alone, I expect that the stimuli will be similarly easy to identify regardless of environment. This is because the major cues for fricative identification are all found within the frication itself (i.e. peak, amplitude, and duration).

Still, we must consider more than acoustic properties for a thorough prediction of perception in different word positions. We must also consider if listeners will choose to segment the illicit or marginal intervocalic sequences (VʃlV, VsɹV, VʃtɹV) through morpheme boundaries. Because the syllables in the intervocalic data do not resemble any morphemes in the listeners' linguistic experience, I expect participants to perceive the sequences as monomorphemic. This choice would preserve the phonotactic limitations on the stimuli, making each sequence as licit or illicit as its word-initial counterpart. Both the findings on phonotactics across morpheme boundaries and the findings on fricative cues support the prediction that word environment will not affect perception of the stimulus data. Therefore, I expect the intervocalic sequences (ɑsɑ]/[ɑʃɑ], [ɑsɹɑ]/[ɑʃɹɑ], [ɑslɑ]/[ɑʃlɑ], [ɑstɹɑ]/[ɑʃtɹɑ]) to behave like their word initial counterparts.

This prediction leads us to question how the word-initial sequences will themselves behave or, similarly, how the different local contexts [_ɑ], [_ɹɑ], [_lɑ], [ _tɹɑ] affect perception. In experiment 1, I expect neutralization due to phonotactic restrictions to play an extensive role. More specifically, I expect listeners' judgment of perceptual similarity to increase in intervals at each point where pairs move from contrastive, to marginally contrastive, to non-contrastive, to allophonic.

Recalling the phonotactic distribution of /s/ and /ʃ/ as mentioned in section 2.1., we can see that [sɑ] and [ʃɑ] are the only fully contrastive pair. Therefore, I expect this pair to be rated as the most perceptually distinct. As the remaining pairs contain an illicit or marginal member, I expect a large jump in perceptual similarity from [sɑ]/[ʃɑ] to the other pairs. I expect [slɑ]/[ʃlɑ] (marginally contrastive) to be the next most distinct, followed by [sɹɑ]/[ʃɹɑ] (non-contrastive). Lastly, I predict [stɹɑ]/[ʃtɹɑ] (non-contrastive or allophonic) to be perceived as most similar of all. These predictions can be spatially represented on the following continuum:

*Figure 2. Continuum of predicted perceptual similarity (PSRT).*



*3.2.2. Discrimination task predictions*

As previously mentioned, speeded discrimination tasks in linguistic studies have produced differing results. Therefore, because the ability of discrimination tasks to circumvent the phonological system is unclear, I offer two sets of predictions for the present study. If, as Boomershine et al. (2008), Huang (2004), and Krishnan et al. (2005) have found, native phonology cannot be bypassed, Experiment 2 should produce much the same results as Experiment 1. The intervocalic sequences will be perceived similarly to the corresponding word-initial sequences. In terms of consonant sequence, [sɑ] and [ʃɑ] will be perceived as the most distinct because of their full contrast, followed by [slɑ]/[ʃlɑ] because of their marginal contrast, [sɹɑ]/[ʃɹɑ] because of their lack of contrast, and [stɹɑ]/[ʃtɹɑ] because of their non-contrastive or allophonic nature.

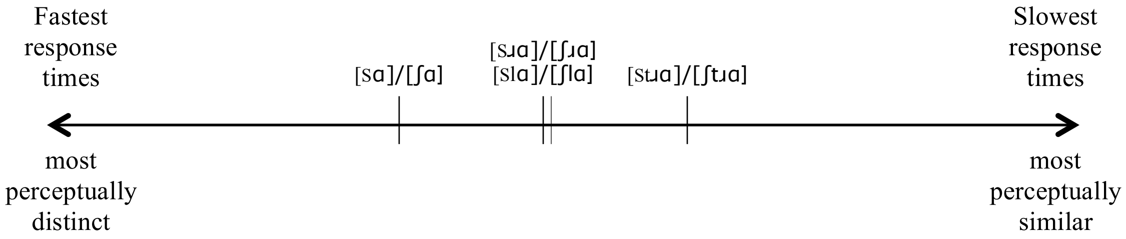
Conversely, the results of the present study will differ if the AX task successfully reaches a lower level of speech processing. Because participants are making quick judgments, I would not expect them, under this view, to identify or classify each speech sound before rating. We might then expect less influence from the listener's phonological system and a greater reliance on the raw, acoustic properties of the sounds to complete the task. In this estimation, I expect the perception of intervocalic and word-initial sequences to be insignificant, but for slightly different reasons. Because listeners do not access linguistic knowledge before comparing, I expect they will not process any information about morpheme boundaries. Taking this factor out of the equation, participants will most likely compare fricatives based on acoustic properties as mentioned in Section 2.2. Again, because the fricative noise itself encodes for place of articulation, I expect that the word position will not matter significantly.

Predictions for how consonant sequences will compare in different local contexts are more complex. If, again, we consider the AX task a means to show raw, acoustic similarity, we can expect level of contrast to play little to no role in the perception of the stimuli. Instead, we would expect only acoustic differences to determine perception. This, too, however, is difficult to determine. Acoustic differences between pairs of consonant sequences are difficult to be measured and often are tested through listener perception, which itself, may be biased towards native language. This leaves researchers in a catch-

22 because listener response determines acoustic similarity and acoustic similarity determines listener response.

The few clues we have to the outcome, therefore, leave a nonspecific picture of how the sounds will be perceived. As contrast would play no role in measuring acoustic similarity, we would expect all pairs to be perceived as more similar to one another than in Experiment 1. The fact that they differ from the PSRT alone would be a good indication that at least some of the phonological constraints were eliminated. Moreover, Rutter (2011) suggests that acoustic characteristics of consonants in a multi-item cluster are less important in distinguishing the utterance than when a single consonant is used. Referring to Blevins (2005), he says "sound changes are more likely to occur in items where acoustic detail is less important for successful retrieval from the lexicon." (30). He specifically refers to the [stɹɑ]/[ʃtɹɑ] merger, but we can extend this logic to our other stimuli as well. We would therefore expect that the stimuli with longer consonant clusters to rely less on the acoustic distinctiveness of the fricative and those with shorter consonant clusters to rely more on the fricatives' precise acoustic properties. Under this assumption, [sɑ] and [ʃɑ] will hold the most acoustic cues in their fricatives and thus be most distinct, while [stɹɑ] and [ʃtɹɑ] will hold the fewest acoustic cues in their fricatives and will thus be the most perceptually similar. Between these two pairs would be [slɑ]/[ʃlɑ] and [sɹɑ]/[ʃɹɑ], with two-consonant clusters. Still, it is unclear which of the two is more acoustically similar. Spatially, these predictions are shown in Figure 3.

*Figure 3. Continuum of predicted perceptual similarity in the AX task.*



## *4. Experiment 1: Perceptual Similarity Rating Task*

The first experiment was a perceptual similarity rating task used to understand the language specific effects of phonology on speech perception (see e.g. Boomershine et al. 2008).

### *4.1. Method*

#### *4.1.1. Participants*

In total, 28 people participated in Experiment 1. Because of an error in the experimental design, 9 participants heard and responded to only the intervocalic stimulus data. They, therefore, did not complete the entire experiment, and their results were excluded from analysis. In addition, those who reported to be native speakers of a language other than English and those who had been bilingual from age five or under were excluded from analysis. Lastly, one participant reported a speech disorder and was also omitted. Considering these exclusions, 15 participants were left, 4 males and 11 females.

Participants for Experiment 1 were recruited in two ways. 18 participants were from the Linguistics Outside of the Classroom (LOC) subject pool at the Ohio State University and were undergraduate students at Ohio State taking an introductory linguistics course.

They were given partial course credit for their involvement in the experiment. The other 10 participants were recruited via email through the friend of a friend approach. These participants received no monetary compensation for their involvement, but were instead compensated with homemade baked goods.

### *4.1.2. Stimuli*

Sixteen distinct stimuli were presented in this study. Each stimulus used one of eight environments and one of two relevant fricatives, /s/ and /ʃ/, creating eight pairs distinguished only by the place of articulation of the fricative. Once again, four of these pairs placed the fricative as word-initial in environments preceding /ɑ/, /ɹɑ/, /lɑ/, or /tɹɑ/. The second four pairs were identical to these except they included a word-initial /ɑ/, creating disyllabic, intervocalic environments instead of monosyllabic, word-initial environments. The full list of stimuli is given in Table 4.

*Table 4. Stimulus data by pair and environment.*

| Word-Initial | Intervocalic |
|---|---|
| [sɑ]/[ʃɑ] | [ɑsɑ]/[ɑʃɑ] |
| [sɹɑ]/[ʃɹɑ] | [ɑsɹɑ]/[ɑʃɹɑ] |
| [slɑ]/[ʃlɑ] | [ɑslɑ]/[ɑʃlɑ] |
| [stɹɑ]/[ʃtɹɑ] | [ɑstɹɑ]/[ɑʃtɹɑ] |

All stimuli were produced by a graduate student in the linguistics department at the Ohio State University, who is a native English speaker and trained in the production of non-native consonant clusters. Several repetitions were recorded, and in each environment, the stimuli most closely matching for intonation, presence or absence of creaky voice, presence or absence of a vowel-initial glottal stop, and vowel quality were chosen. The intensity of all stimuli was normalized and the sampling rate was adjusted to 22050 Hz.

### *4.1.3. Procedure*

Each trial consisted of two word-initial or intervocalic fricative tokens, separated by an interstimulus interval of 1000 milliseconds. Members of each pair either differed in their choice of fricative (*different*) or were two tokens of the same type utterance (*same*). In each environment, there were 8 unique ordered *different* pairs and 4 unique ordered *same* pairs. The specific orders follow and use subscripts to indicate different token choices. The same pairs were repeated twice.

Different pairs: [ɑsɹɑ]$_a$/[ɑʃɹɑ]$_c$, [ɑsɹɑ]$_a$/[ɑʃɹɑ]$_d$, [ɑsɹɑ]$_b$/[ɑʃɹɑ]$_c$, [ɑsɹɑ]$_b$/[ɑʃɹɑ]$_d$, [ɑʃɹɑ]$_c$/ [ɑsɹɑ]$_a$, [ɑʃɹɑ]$_d$/ [ɑsɹɑ]$_a$, [ɑʃɹɑ]$_c$/ [ɑsɹɑ]$_b$, [ɑʃɹɑ]$_d$/ [ɑsɹɑ]$_b$

Same pairs: [ɑsɹɑ]$_a$/[ɑsɹɑ]$_b$, [ɑsɹɑ]$_b$/[ɑsɹɑ]$_a$, [ɑʃɹɑ]$_c$/[ɑʃɹɑ]$_d$, [ɑʃɹɑ]$_d$/[ɑʃɹɑ]$_c$

Participants heard the stimuli through headphones while seated at individual computers in a lab with up to four other subjects. Their environment was semi-private as each station was partially enclosed. After hearing each pair, participants were directed to rate how similarly they perceived the two stimuli using a scale of 1 to 7. They typed their responses on a keyboard that only accepted numbers 1-7 as possible answers.

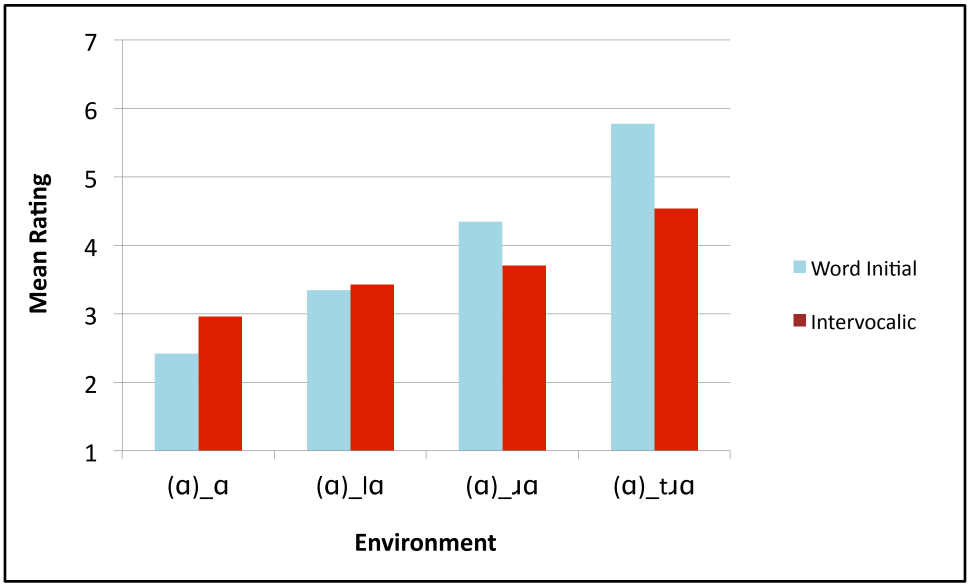
Subjects were not directed to rate different tokens of the same stimuli as different. Rather, oral and written directions informed them that a rating of 1 indicated the two stimuli were very similar to identical, while a rating of 7 indicated they were very dissimilar to not similar at all. Subjects were not instructed on how quickly they should respond so as to allow them ample time to process each pair. Further, they were given no feedback throughout the trials.

Trials were blocked into two groups—the intervocalic sequences and the word initial sequences. Each block contained 64 trials played twice for a total of 128 trials in each block and 256 trials in the entire task. Within each of these blocks the trials were randomized separately for each subject using E-prime software. Additionally, the blocks themselves were presented in two orders. Half of participants (N=8) heard the block of word-initial trials followed by the block of intervocalic trials. The remaining half (N=7) heard these blocks in reverse order.

### *4.2.1. Results*

The data from Experiment 1 were analyzed using a repeated measures ANOVA in the statistical program R. The within-subjects variables tested were consonant sequence ([ _], [ _l],[ _ɹ], [_tɹ]) and word position (intervocalic or word-initial). The results showed a significant main effect of consonant sequence (F(3,42)=78.49, p<.001), word position (F(1,14)=5.29, p=.037) and an interaction between consonant sequence and word position (F(3,42)=16.97, p<.001). Figure 5 shows this consonant sequence by word position interaction and its effect on mean rating.

*Figure 5. Mean rating of /s/ and /ʃ/ in different consonant sequences and word environments*



Within each word position, the mean difference in numerical responses between all consonant sequences was calculated using several paired t-tests. In word-initial position, the [sɑ]/[ʃɑ] pair was rated significantly more different than the [slɑ]/[ʃlɑ] pair (t(14)=-5.60, p<.001), the [sɹɑ]/[ʃɹɑ] pair (t(14)=-9.82, p<.001), and the [stɹɑ]/[ʃtɹɑ] pair (t(14)=-13.05, p<.001). Additionally [slɑ]/[ʃlɑ] was rated as significantly more different than [sɹɑ]/[ʃɹɑ] (t(14)=-5.02, p<.001) and [stɹɑ]/[ʃtɹɑ] (t(14)=-9.06, p<.001) and [sɹɑ]/[ʃɹɑ] was rated significantly more different than [stɹɑ]/[ʃtɹɑ] (t(14)=-7.14, p<.001). In brief, all contexts were rated as significantly different from one another in word-initial position.

The results in intervocalic position proved very similar to those in word-initial position. Here, [ɑsɑ]/[ɑʃɑ] was rated significantly more different than [ɑslɑ]/[ɑʃlɑ] (t(14)=-3.98, p=.002), [ɑsɹɑ]/[ɑʃɹɑ] (t(14)=-3.68, p=.003), and [ɑstɹɑ]/[ɑʃtɹɑ] (t(14)=-5.35, p<.001). Moreover, [ɑslɑ]/[ɑʃlɑ] was rated significantly more different than [ɑstɹɑ]/[ɑʃtɹɑ] (t(14)=-4.48, p<.001), but not significantly more different than [ɑsɹɑ]/[ɑʃɹɑ]. Finally, [ɑsɹɑ]/[ɑʃɹɑ] was rated significantly more different than [ɑstɹɑ]/[ɑʃtɹɑ] (t(14)=-4.50, p<.001). Similar to perception of sequences in word-initial position, the perception of /s/ and /ʃ/ was significantly different in nearly all contexts. Unlike in word-initial position, though, the pairs [ɑslɑ]/[ɑʃlɑ] and [ɑsɹɑ]/[ɑʃɹɑ] were not significantly different in intervocalic context.

Across word positions, there were also some significant differences in perception. The pair [sɑ]/[ʃɑ] was rated significantly more similar in intervocalic position than in word-initial position (t(14)=2.47, p=.027). Moreover, the pairs [sɹɑ]/[ʃɹɑ] and [stɹɑ]/[ʃtɹɑ] were rated significantly more different in intervocalic position than in word initial position (t(14)=-3.35, p=.005 and t(14)=-5.93, p<.001 respectively). The pair [slɑ]/[ʃlɑ] was not found to be significantly more different in either environment.

### *4.2.2. Discussion*

Experiment 1 seems to suggest that phonotactic constraints affect perception of otherwise contrastive pairs because the results fit with the phonemic distribution of /s/ and /ʃ/ in English. As predicted, ease of perception was not constant across environments. Rather, it appears to be related to degree of contrast. Across both word-initial and intervocalic positions, the greatest perceptual distance was perceived in the environment where /s/ and /ʃ/ exhibit full contrast (i.e. [(ɑ)_ɑ]). As contrast decreases between the fricatives from marginal [(ɑ)_lɑ] to non-existent [(ɑ)_ɹɑ] to allophonic [(ɑ)_stɹɑ], perceptual salience also decreases in significant steps. The only exception was between the pairs [ɑslɑ]/[ɑʃlɑ] and [ɑsɹɑ]/[ɑʃɹɑ] which showed no significant difference in perception in intervocalic position.

In addition to the impact of consonant sequence, the influence of word position was also found to be significant in this experiment. While word position did not affect the order of perceptual similarity between fricative pairs, it did affect degree of perceptual similarity across contexts within pairs. Specifically, Experiment 1 found that [sɑ] and [ʃɑ] were easier to distinguish in word-initial position, while [sɹɑ]/[ʃɹɑ] and [stɹɑ]/[ʃtɹɑ] were easier to distinguish in intervocalic position. This result can be interpreted in two ways.

As the two clusters that are illicit in word-initial position were easier to perceive intervocalically, we could assume that listeners chose to interpret them as bimorphemic.

In this view, the morpheme boundary makes the phonotactic constraints inactive. This interpretation, however, leaves us to wonder why [sɑ]/[ʃɑ] were easier to distinguish in word-initial position than in intervocalic position.

A second interpretation of why the intervocalic sequences were perceived differently despite the same local context may offer an explanation that is simpler and applies more broadly to both the perception patterns. This interpretation suggests an equalized view of the intervocalic sequences. It should be noted that the intervals between pairs in intervocalic position were generally much smaller than in word-initial position. Even more, the range of mean responses to the intervocalic stimuli was from 2.94 to 4.54 (a difference of 1.60), while the range of mean responses to the word-initial stimuli was from 2.42 to 5.78 (a difference of 3.36). Because the response range was so small in the intervocalic sequence, it is possible that listeners had more trouble distinguishing the phonological differences in this environment, perceiving them all as equally similar and thus rating them towards the middle of the scale. This could signal that the perceptual cues in intervocalic environment are less apparent.

## *5. Experiment 2: AX discrimination test*

Our second experiment was an AX discrimination task that asked listeners to quickly determine whether two stimuli of a pair were the same or different. Once again, this experiment was designed to measure the phonetic similarities of the relevant fricatives, but is questioned in its effectiveness. Unlike the PSRT, which is assumed to force participants to classify speech before comparison, this task hypothetically forces participants to compare speech sounds without classification.

### *5.1 Method*

#### *5.1.1. Participants*

15 subjects participated in Experiment 2. Participants were recruited from the same group of LOC students used in Experiment 1 under the same precepts. Again, any participant who self-reported as a non-native English speaker or a bilingual speaker from age 5 or younger was excluded from analysis. None of the participants reported a known speech or hearing disorder.

#### *5.1.2. Stimuli*

Experiment 2 utilized the same stimuli as Experiment 1.

#### *5.1.3. Procedure*

The trials for Experiment 2 were once again composed of two word-initial or two intervocalic sequences, using the fricatives /s/ and /ʃ/. The A and X stimuli in each trial were either two tokens of the same type or two tokens of different types using the same environment, but different fricatives. Moreover, the interstimulus interval remained at 1000 milliseconds. As with Experiment 1, trials were randomized and blocked with half of participants (N=8) hearing the blocks in word-initial, intervocalic order, and half (N=7) hearing the reverse.

In this task, participants were instructed to judge each pair in a trial as the same or different. Each participant heard the stimuli through headphones and indicated his or her response on a 5-key button box hooked up to his or her computer station. Two of the buttons were marked with black dots to indicate the two response choices, SAME or DIFFERENT, and their meaning was reinforced throughout the experiment on the computer screen. Subjects were not given feedback during the course of the experiment.

As with Experiment 1, they were not instructed how to handle two tokens of the same type.

An important aspect of this task is a quick response time. Therefore, participants were carefully instructed to respond as quickly as possible without sacrificing accuracy. Further, they were directed to use two fingers when responding (either both forefingers or both thumbs) in order to prevent the lag of reaching across the button box.

### *5.2.1. Analysis*

As with Experiment 1, I chose again to only focus on the DIFFERENT stimulus pairs of Experiment 2. In order to correctly represent the perceptual difference within stimulus pairs, I chose to exclude responses over 2000 ms. These did not reflect *speeded* responses and thus, gave speakers the time to individually classify each member of the pair before comparison. Additionally, I felt that some of these delayed response times reflected missed hearing of the stimuli or a distraction that kept participants from careful comparison.

Further, I excluded all responses that incorrectly identified the two stimuli in a trial as SAME or DIFFERENT. Johnson & Babel's study found that participants achieved near perfect responses in speeded discrimination, stating that across all participants and pairs, subjects achieved roughly 95% accuracy (133). Surprisingly, we found that mean accuracy across all *different* pairs was much lower in our study, at 79.17%. Nearly all sequences had a mean correct response rate in the 80th or 90th percentile; however, one pair, both in initial and intervocalic position skewed the overall results. This was the [stɹɑ]/[ʃtɹɑ] pair, which was identified with only 41.88% accuracy in word-initial and intervocalic position. While some speakers were able to distinguish this pair at as high as 95.31% overall, others had a mean accuracy of 0% across all pairs.

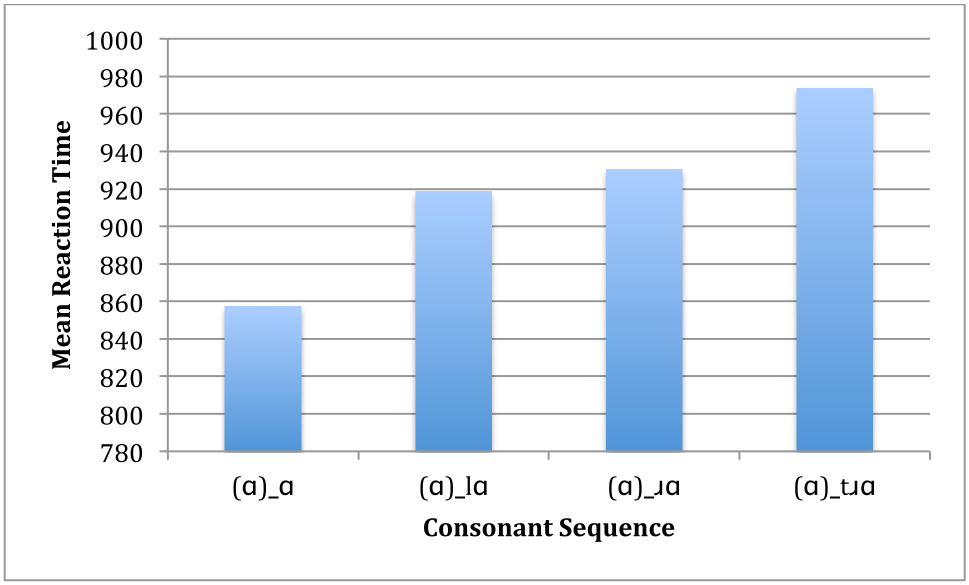
This inability to identify [stɹɑ] and [ʃtɹɑ] as different was particularly obvious for two participants. Subject 4 had a 0% accuracy across both intervocalic and word-initial pairs. Subject 7 incorrectly identified all [stɹɑ]/[ʃtɹɑ] pairs in word-initial position, but not in intervocalic position. Because word position proved not to be a significant factor in the analysis, word-initial and intervocalic responses were averaged across subjects. This left Subject 7 with some correct responses (37.5%), but Subject 4 with no correct responses. As the analysis only considered correct responses, his responses were excluded.

The results of the two aforementioned subjects as well as the overall low accuracy for correctly identifying the difference between [stɹɑ] and [ʃtɹɑ] were especially interesting with respect to the claims of Durian (2007) and Rutter (2011). Although the authors attested that certain speakers were beginning to produce the innovative form [ʃtɹɑ], this is the first indication that a perceptual merger between [stɹɑ] and [ʃtɹɑ] is occurring for some speakers. Further, the fact that these sounds are indistinguishable to some, but accurately distinguishable to others supports their claims that this sound change is quite actively occurring across speakers. Unfortunately, the present study has no way of telling whether those listeners with a low accuracy were the producers of the innovative or the standard form. Future studies would do well to test both the production and perception [stɹɑ] and [ʃtɹɑ] within speakers to flesh this out.

### *5.2.2. Results*

The results for Experiment 2 were again evaluated using a repeated measures analysis of variance with within-subjects variables of consonant sequence and word position. As with Experiment 1, the analysis showed a significant main effect of consonant sequence (F(3,38)=4.89, p=.006). However, unlike the previous experiment, this experiment's results showed no significant effects of word position or consonant sequence by word position. Figure 6 shows the main effect of consonant sequence on mean reaction time across participants and word environments.

*Figure 6. Mean reaction time to /s/ and /ʃ/ in various consonant sequences.*



Again, paired t-tests were used to show how reaction times compare between each of the consonant sequences. These tests showed that reaction times for [sɑ] /[ʃɑ] differentiation were significantly faster than reaction times for the [slɑ]/[ʃlɑ] pair (t(13)=-4.65, p<.001), the [sɹɑ]/[ʃɹɑ] pair (t(13)=-5.13, p<.001), and the [stɹɑ]/[ʃtɹɑ] pair (t(13)=-3.00, p=.010). The [slɑ]/[ʃlɑ] pair proved not to be significantly faster than the [sɹɑ]/[ʃɹɑ] or [stɹɑ]/[ʃtɹɑ] pairs. Lastly, [sɹɑ]/[ʃɹɑ] showed no significant difference in reaction times in relation to [stɹɑ]/[ʃtɹɑ].

### *5.2.3. Discussion*

Recalling the predictions form Section 3.2.2., we can take these results as an indication that Experiment 2 was able to reach a lower level of processing that reflects the raw acoustic properties of the stimuli. If, as found by Boomershine et al. (2008), Huang (2004) and Krishnan et al. (2005), the results had mirrored the results from the PSRT, we would have taken them as stemming from native language phonology. As the significant factors were found to be considerably different, though, the results can be considered to have at least in part bypassed the native phonology of the listeners.

Unlike in the PSRT, where word environment and consonant sequence both played a significant role, the AX task found that these were not factors in listeners' perception. This suggests that, as predicted, participants did not individually classify and segment the disyllabic stimuli they heard. Instead, the evidence suggests they quickly compared the sounds without categorization. Relying only, therefore, on the fricative cues, they rated stimuli in intervocalic and word-initial position as not significantly different.

These results also provide some clues as to the raw acoustic properties of the stimuli. The predictions supposed that those stimuli with fewer consonants in their clusters would be acoustically more different because, in these sequences, there is less other information from which to take cues. This held true in that participants reacted to [sɑ] and [ʃɑ] pairs much more quickly, thus showing they were easier to perceive. On the other hand, all the remaining pairs ([slɑ]/[ʃlɑ] , [sɹɑ]/[ʃɹɑ], [stɹɑ]/[ʃtɹɑ]) were not significantly different from one another. This could suggest that the difference between a two-constituent consonant cluster and a three-constituent consonant cluster make little impact on perception. Alternately, we can take this as a sign that the length of clusters play little to no role in perception and that the significant difference between [sɑ]/ [ʃɑ] and the other pairs was caused by another factor.

This unknown factor should be fleshed out in future studies by using listeners of two or more languages, where /s/ and /ʃ/ contrast differently or are distributed differently by phonotactic constraints. If participants respond similarly in the AX task despite their native language, we can take this as a measure of raw acoustics and gain a better understanding of the purely auditory similarities among the pairs.

### *6. Conclusions*

The present study contributes to the larger effort by researchers to examine the nuances of phonology's effects on perception. The results fit with previous research that considers how contrast is perceived and provide a new basis for understanding how cases of partial contrast due to phonotactic constraints affect perception. Moreover, the study supports the idea that partially contrastive phones are not perceived equally in different word environments and consonant clusters.

In the perceptual similarity rating task, participants rated stimulus pairs as more similar or different according to their contrastive relationship. As phonological contrast decreased from one context to the next, listeners found the fricatives to be more perceptually similar. This finding suggests that phonemes are not perceived as equally salient in all contexts, but that the level of contrast within a specific environment affects the perception of otherwise contrastive phones.

Additionally, this experiment found that word environment (word-initial v. intervocalic position) mattered significantly in rating. As the [sɑ]/[ʃɑ] pair was more distinct in word-initial position and the pairs [sɹɑ]/[ʃɹɑ], [stɹɑ]/[ʃtɹɑ] were more distinct intervocalically, we might suspect that listeners inserted morpheme boundaries into the disyllabic stimuli; however, most likely, these results stem from the overall medial ratings of the intervocalic pairs.

The results from the speeded discrimination task were quite different from the results of the PSRT. Specifically, the task was able to reach a lower, acoustic level of processing consistent with the traditional view and the findings of Johnson and Babel (2010), Fox (1984), Strange and Dittman (1984), and Werker and Logan (1985). While word position did not matter in this task, consonant sequence was significant in determining perceptual distinctiveness. All stimuli were judged as more similar to one another than they were in the PSRT, although [sɑ]/[ʃɑ] stood apart from the rest as being more acoustically distinct.

As suggested, future studies might repeat this examination of /s/ and /ʃ/ in various languages and compare the results. The more consistent the PSRT findings with the phonological relationships of the fricatives, the stronger the claim that partial contrast and phonotactic gaps have a strong effect on perception. Further, a comparison across languages would help provide a better understanding of the acoustic relationship between sounds in the AX task as a clear pattern should develop across languages. This study could also be expanded upon by looking at different segment types that face neutralization due to gaps in the phonological system. Ultimately, partial contrast due to phonotactic gaps deserves more study in order to understand its full effect within languages.

**References**

Blevins, Juliette (2005). The role of phonological predictability in sound change: Privileged reduction in Oceanic reduplicated substrings. *Oceanic Linguistics* 44, 455–464.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher, & K. Rice (Eds.), Phonological contrast: Perception and acquisition (pp. 146–172). New York: Mouton de Gruyter.

Durian, David (2007). Getting [ʃ]tronger every day?: More on urbanization and the socio-geographic diffusion of (str) in Columbus, OH. *University of Pennsylvania Working Papers in Linguistics,* 13(2), 65-79.

Fox, Robert (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human perception and performance, 10(2),* 526-540.

Giegerich, Heinz (1992). *English Phonology: An Introduction.* Cambridge, UK: Cambridge University Press.

Hammond, Michael (1999). *The Phonology of English.* Oxford: Oxford University Press.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1, 1-7.

Harris, John (1994). *English Sound Structure*. Cambridge, MA: Blackwell Publishers.

Huang, T. (2004). Language specificity in auditory perception of Chinese tones. Ph.D.

dissertation, Ohio State University.

Hughes, G. W. & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America* 28, 303-310.

Johnson, K. & Babel, M. (2010). On the perceptual basis of distinctive features: evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics,* 38, 127-136.

Jongman, A,, Wayland R. & Wong S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108,* 1252-1263.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25, 161–168.

Nittrouer, S. and Miller, M. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 102 (1).* 572-580.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25–53.

Peperkamp, S. & Martin, A. (2011) Speech Perception and Phonology. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (eds.) *Companion to Phonology.* Hoboken, N.J. : Wiley-Blackwell.

Rutter, Ben (2011). Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *Journal of the International Phonetic Association*, 41, 27-40.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, 7, 82–138.

Soli, S. D. (1981) Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation**.** *Journal of the Acoustical Society of America, 70 (4).* 976-984.

Steriade, D. (1997) Phonetics in phonology: the case of laryngeal neutralization. *MIT Linguistics.* Retrieved April 17, 2012, from http://web.mit.edu/linguistics/people/faculty/steriade/publications.html.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36, 131–145.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63.