Universal Principles and
Native Language Influences
on the Perception of Minimally Distinct Uvular Sounds

Avital Kaplan

A thesis submitted in partial fulfillment of the Honors requirements for the degree of
Bachelor of Arts in the field of Linguistics
at the College of Arts and Science of New York University

Under the advisement of Professor Gillian Gallagher

April, 2014
I. INTRODUCTION

One of the most difficult aspects of functioning within a society of one’s second language, even for quite advanced learners of an L2, is actually understanding native speakers’ speech. It is often much easier to understand the written word than the spoken one. Words are perceived as unenunciated and indistinct, and sounds are mistaken for others.

The human vocal tract is capable of producing a suite of sounds, yet no natural language employs all those sounds in its phonology. When babies are born, they are capable of properly perceiving all the sounds of the world’s languages. However, by the end of the first year of life, the infants’ auditory mechanisms specialize to the phonology of the ambient language, and proper perception of non-native sounds dull. Such is the state of the mature listener: markedly poor perception of non-native sounds. This is the reason why it is so difficult to perceive even a long-studied L2: our minds are attuned to a given phonology, to the detriment of all others (Werker and Tees, 1984; Stager and Werker, 1987).

An additional difficulty, perhaps more challenging than learning to perceive novel sounds, is the fact that various sounds can be categorized differently in different languages. Two sounds that differ, “contrast,” in only one aspect of their production, may be categorized as variants of the same sound in a given language, and are realized in differing, predictable environments. However, other languages may retain the distinction between those two sounds. For speakers of the former category of languages, it can be exceptionally difficult to perceive the distinction between the two sounds in the latter set of languages.

In other words, our native phonologies influence our perception of non-native sound and of non-native sound distinctions.

But how exactly does this occur? What rules, if any, dictate which non-native sounds, and to what degree, will prove challenging to a listener? Through what mechanisms does our native language effect the perception of sounds that exist in differing categories in a foreign language?
To test these questions, I examine two uvular consonants: the voiced, fricated rhotic, /ʁ/, and the voiceless fricative, /χ/. These two sounds exist as separate phonemes in Hebrew, and are quite distinct to speakers of that language. The rhotic exists as a phoneme in French, however, it is devoiced when produced in a cluster after a voiceless consonant. In that environment, the realized production of the rhotic sounds more like the fricative. In that case, French employs both sounds, but does not distinguish between them. It should be significantly difficult for a native French speaker to distinguish between two Hebrew words that differ only by those two sounds, for example, /χam/, “hot,” and /ʁam/, “esteemed.”

English doesn’t contain either of these sounds in its phonology. It contrasts voicing and utilizes both rhotics and fricatives, but it does not employ the uvula as a place of articulation. It should indeed prove difficult for native English speakers to perceive these Hebrew sounds.

A discrimination test of Hebrew nonsense words containing these two sounds was performed by French speakers as well as by American-English speakers. The results from each population were compared to the ceiling-level discrimination of Hebrew speakers. Additionally, the two populations’ results were compared to one another to determine who was more challenged by the test. The purpose of these tests was to determine not just the degree of difficulty, but the nature of the difficulty experienced in the perception of non-native sounds by speakers of languages with differing relationships to those sounds.

**II. BACKGROUND**

The concept of Contrastive Analysis (CA) was originally articulated in Robert Lado’s Contrastive Analysis Hypothesis (CAH), whose basic claim was that by comparing a native language and target language, a correct prediction could be made of where a learner will have acquisitional difficulty. Lado outlined three levels of the language-specific sound system-phonemic, phonetic and distributional- and posited that differences between sounds on these levels
in the L1 and L2 would lead to difficulty with perception of sounds in the target language, and from there, difficulty in production of those sounds. For an example of a cross-linguistic phonemic difference, the American English (AE) rhotic, /ɹ/, is extremely rare among the worlds’ languages. When tutoring a native Hebrew speaking 8 year old boy in English, I found that his closest production to this phone was realized somewhat like a /w/, because the AE rhotic simply does not exist on any level in Hebrew, and the child had no understanding of the sound or of its proper articulation. A phonetic difference is that both Spanish and English employ the voiced coronal obstruents /d/ and /ð/, however in English these are two separate phonemes, while they are allophonic in Spanish. When Spanish speakers learn English, it is common for them to pronounce what should be a /d/ as its spirantized counterpart /ð/ in the environments that would condition that variation in Spanish (Lado, 1957). Finally, distributionally, English employs the nasal consonant /ŋ/, but never in word-initial position. When studying Sherpa, I had a very hard time distinguishing it in word-initial position, where it often occurs, from the nasal /m/ sound.

Contrastive Analysis has been criticized by many for yielding theoretical findings about Second Language Acquisition (SLA) which make incorrect predictions about L2 phone production difficulties (Wardhaugh, 1974; Lehtonen and Sajavaara, 1980). Among the criticisms of CA is that, rather than capturing rule or constraint-based generalizations about a given language’s sound system, it demands an exhaustive list of both L1 and L2 phonology. That is to say, rather than accounting for the features of the phonemes and their organization within a language in a systemized and predictable fashion, the CAH compares the entire suite of sounds of different languages (Wardaugh, 1974). Additionally, it proposes to make these comparisons in a systematic form, wherein this could be accomplished by way of written grammars; without studying actual speakers in talk situations, much critical, realized speech information is overlooked (Lehtonen and Sajavaara, 1980).
After the insufficiencies of the theory were observed, a weak form of it was promulgated which dispensed with the attempt at prediction and instead focused on retroactive analysis of acquisitional shortcomings of L2 phonology. This analysis was based on a comparison of phonetic features between allophones of mutual categories, allowing space for consideration of various interferences and avoiding confusion over identically labeled yet differently realized phonemes across the languages in question (Wardhaugh, 1974).

While the strong form has proven to be flawed in its methodology, its conceptual foundation was never dismissed. There are many revisions and emendations to the Contrastive Analysis Hypothesis’ foundational claim that attempt to direct its theory into an applicably accurate predictive tool in the field of SLA, specifically as it concerns phonology. (Eckman, 1977; Lehtonen and Sajavaara, 1980; Best et. al. 2001; Hallé et. al. 2003; Best, 1995; Hallé et. al. 1999; Fowler et. al. 2008; Flege, 1995; Kuhl, 1991; Kuhl and Iverson, 1995). A major theory that attempts to rectify the shortcomings of the strong form of the hypothesis is the Markedness Differential Hypothesis (MDH). It posits that the discrepancy between the Contrastive Analysis Hypothesis’ (CAH) predictions and tested results is an outcome of CAH not taking relative learning difficulty of the specific phonological difference into account. The MDH employs a generative perspective in remedying the problem: it compares the L1 and L2 phonological systems for the given feature/s in question in order to determine which is more typologically marked. The level of markedness is determined based on the level of dependence of a certain phonological phenomenon on another: if a phenomenon can occur on its own, it is unmarked, and if its presence requires that another phenomenon also occur, then it is marked. For example, a language’s velars can palatalize without anything else doing so; velar palatalization is unmarked. However, if its dentals palatalize, its velars must do so as well. Dentals cannot palatalize without velars doing the same, so dental palatalization would be considered, in this system, typologically marked (Chen, 1973). If the L1
system is marked compared to the L2, there should be no difficulty learning, whereas acquisitional difficulty should arise if the L2 is more marked than is the L1 (Eckman, 1977).

Other theories have arisen with the same goal of understanding the difficulty of L2 perception and production that don’t see themselves as permutations of the CAH, but as autonomous systems that share the goal of CAH. One theory that attempts to explain the relationship between native and second languages in phonological acquisition is the perceptual assimilation model (PAM). Emphasizing gestural phonology, PAM claims that a non-native phone will assimilate to the native phone that is articulatorily closest to it (Best et. al. 2003; Best et. al. 2001; Best, 1995; Hallé et. al. 1999). In other words, non-native phones will be organized into phoneme categories not based on descriptive featural similarity, but according to articulatory closeness. It understands phonological organization as shaping itself around the gestures of phone production. That is to say, it claims that when we hear a speech sound, we very quickly deduce and map onto it the details of the gestures involved in producing that sound in the vocal tract. Therefore, a non-native contrast would be heard in the framework of the relationship between the two native sounds under which they are gesturally categorized (Hallé et. al. 2003).

This method leads to three general possibilities:

1) Categorized: A given non-native sound is assimilated clearly to a native phoneme, whether it has excellent or poor goodness-of-fit to the phoneme;

2) Uncategorized: The non-native sound falls somewhere in between native phonemes, meaning that it is similar to two different phonemes yet not quite like either;

3) Nonassimilable: The sound bears no salient similarity to any native phoneme, as if it were a non-speech sound.

Taking these possibilities and applying them to non-native contrast identification, six potential relations come about:
1) Two Category (TC) assimilation is when the two sounds are phonetically similar to two separate native phonemes. This mapping to different native phonemes allows the non-native contrast to be heard as two distinct, easily discernible sounds.

2) Single Category (SC) assimilation yields difficult differentiation, as the two sounds assimilate equally well to a single native phoneme. Due to this lack of distinction in assimilation, listeners hear the contrast as the same sound, and have a difficult time discerning the two.

3) Category Goodness (CG) mitigates the latter situation, as the two sounds assimilate to the same native phoneme, but with differing levels of goodness-of-fit. One sound is heard as a better token of the given native phoneme than is the other, and therefore it is relatively easy to discriminate between the two, though not as easy as in TC assimilation.

4) If only one of the non-native sounds is assimilable to a native phoneme, while the other is Uncategorized, an Uncategorized-Categorized (UC) relation arises. Because one is heard as a recognized speech sound and the other sound is not, it should be simple to distinguish them from each other.

5) Both phones may be uncategorized, an “Uncategorized-Uncategorized” (UU) relation. These will be discriminated on a spectrum from fair to good, depending on the perceived closeness of the two sounds and the closeness of the native phonemes to which they are similar.

6) Finally, the two sounds may both be so distant from the phoneme set of the native language that they are both nonassimilable (NA). It is PAM’s claim that NA sounds are not subject to the same auditory processing mechanisms as speech sounds, and therefore should be easy to discriminate.

When the contrasting sounds are assimilated to separate native phonological categories, or when only one of the phones is assimilated to a native category, they should be easier to differentiate. When they fall into the same native phonological categories, the task becomes significantly more difficult, depending on the variance of goodness-of-fit between the two sounds. PAM claims that a listener is capable of detecting phonological, phonetic and non-linguistic/
acoustic cues in speech sounds. Presumably, TC contrasts are detected at the phonological level, CG ones at the phonetic level and SC ones, when there are no other higher-order cues detectable by the listener, at the non-linguistic/acoustic level (Best et. al. 2001). In Hebrew, the rhotic and the fricative differ phonemically, and constitute a TC contrast. The degree to which the French perceive the fricative as the rhotic, meaning if the fricative functions for them as an allophone to the rhotic, will determine if the contrast is a TC one or a CG one. If, as predicted, the fricative does function for them as an allophone to the rhotic, their discrimination will be somewhat poor as in a CG contrast. The performance of the AE speakers on the discrimination task will help to determine the inherent acoustic similarity between those two sounds, and the sort of contrast they constitute when not bound by language-specific phonological categories.

The native language magnet (NLM) model is based on the findings of the Perceptual Magnet Effect which proposes that exposure to an ambient language alters a person’s phonetic perception of the place of phones within the vocal tract, at any age of exposure (Kuhl, 1991; Kuhl, 1994; Kuhl and Iverson, 1995). Simply put, “...language experience alters the mechanisms underlying speech perception, and thus, the mind of the listener.” (Kuhl and Iverson, 1995, p.121) During the first year of life, infants can hear and properly identify all sounds of the world’s languages, but by the end of that year, their minds have already specialized to the ambient language, leading to poor phoneme identification for sounds that are not part of that language (Werker and Tees, 1984; Stager and Werker, 1987; Kuhl 1994; Maye et. al., 2002). Phonemes of the native language, according to the theory, are stored in the mind as “prototypes.” The prototype functions much like a magnet, pulling similar, non-prototypical sounds towards it, thereby creating the distortion of the perceptual space. Multilingual adults will then have a different mental representation of the perceptual space than will monolingual speakers of any of the given languages. Non-native speech sounds will be assimilated to the phoneme prototype of the native language in whose perceptual space the non-native sound falls; crucially, the model claims that no
matter how much experience is gained in an L2, prototypes for that language will not form (Kuhl, 1991; Kuhl, 1994; Kuhl and Iverson, 1995).

The speech learning model (SLM) was formulated to investigate the cause behind inaccuracies in L2 speech production (Flege, 1995). It states that the main, though not singular, factor behind L2 production errors is perceptually based, and that without “perceptual ‘targets’ to guide the sensorimotor learning of L2 sounds, production of the L2 sounds will be inaccurate” (Flege, 1995 p. 238). It is predicated on the following four postulates:

1) The mechanisms involved in L1 acquisition do not disappear, but can be employed at any point in life to learn an L2.

2) Phonetic categories are part of the long-term memory, and represent language-specific speech sounds.

3) Phonetic categories evolve throughout life to reflect proper categorization of L1 and L2 speech sounds that are identified as belonging in each category.

4) Although the speech sounds of all languages one speaks exist in a single phonological space, speakers attempt to retain a contrast between L1 and L2 phonetic categories (Flege, 1995). Studies of non-native bilinguals who learned their L2 at a variety of ages-of-learning (AOL) showed compromises to characteristics of sounds that occur in both languages (in line with postulate 3) yet still retained different pronunciations of the sound in each language (in line with postulate 4). For example, native French speakers who were proficient in English were shown, on the one hand, to produce a word-initial /t/ in French with a longer VOT value than would be expected, reflecting English influence. On the other hand, it did not reach the VOT length of native English (Flege, 1995).

The SLM posits that L2 production becomes more inaccurate with greater AOL. It also claims that an L2 sound without an L1 counterpart will be produced more accurately than will be an L2 sound that does have an L1 counterpart. This notion is based on Trubetzkoy’s analogy (1939)
that a mature L1 phonological system functions like a sieve, allowing only information about the L2 that is important to L1 contrasts to pass through. In some cases, this could lead to underdifferentiation, producing the wrong phoneme because the crucial contrastive feature in the L2 is not contrastive in the speaker’s L1. Alternatively, such a phenomenon could be due to misperception of acoustic cues that are relevant to the L2, also due to the L1 sieve effect (Trubetzkoy, 1939). Research suggests that allophones may not be grouped into phonemic categories until a child 5 or 6 years old, meaning that the mind is initially quite sensitive to L1 acoustic properties, but that the sensitivity dulls over time (Flege, 1995; Jusczyk, 1985).

PAM, NLM and SLM all make claims about contrastive analysis and predictions about the ease of perception of non-native contrasts. The basic assumption underlying all of them is that the ability to perceive a non-native contrast is predicated on the process of native language acquisition and on the relationship between native categories and the sounds in question; the theories differ in their understanding of the acquisitional process and auditory mechanisms. While the NLM and the SLM are not identical- the former promulgating a more axiomatic claim about the effects of native phonemes on non-native ones and the latter leaving more room for variations based on the two languages’ respective phonological systems- they both make the basic claim that L2 similarity to an L1 phone will distort the ability of the listener to perceive, and therefore accurately produce, the L2 phone. On the other hand, PAM’s approach predicts the opposite: L1 experience with a category of sounds means that a listener is more attuned to what he is hearing, and better able to discriminate lower-order phonetic and acoustic differences.

Both approaches are in line with the basic principle that production of a sound is based on the perception of it; an L2 speaker doesn’t mispronounce a sound simply because he doesn’t know how to properly articulate it; he doesn’t know how to properly articulate it because he misperceives it (Trubetzkoy, 1939). That is to say, the L1 sieve effect would explain SLM’s prediction of difficulty in discerning the variations between similar sounds. NLM’s foundational magnet effect
principle essentially serves the same function as the sieve effect. Meanwhile, it would be the lack of familiarity with certain gestural combinations in the L1 that would lead to PAM’s prediction about difficulty in perceiving non-inventory phones.

In further testing the theories set forth in these three models, it would be fruitful to test a contrast on native speakers of two separate languages whose respective inventories would presumably categorize the given sounds differently. To that end, native French speakers and native American English (AE) speakers would be expected to discriminate the Hebrew uvular rhotic, /ʁ/, and uvular voiceless fricative, /χ/, in vastly different manners. Hebrew, a West-Semitic AfroAsiatic language, contrasts these two sounds phonologically, as exemplified in the minimal pair, /χʌtam/, 3SG pst “sign,” and /ʁʌtam/, 3SG pst “bridle (an animal)” (Morag, 2009). In French, a Romance language, the uvular rhotic exists phonemically; it is posited that the allophone of that phoneme which exists in environments that condition devoicing is realized as this voiceless uvular fricative. For example, in the French trouver, /tʁuve/ “to find,” the rhotic [ʁ] is often devoiced and bears greater resemblance to a fricative /χ/ than to the uvular rhotic; this as opposed to the word drôle, [dʁɔl], “funny,” where the rhotic is realized as such (Casagrande, 1984). Finally, while English, a West-Germanic language, employs both rhotics and fricatives on the phonemic level, it does not utilize them at the uvula. While the uvula is still part of the phonological space of English, it is not a primary place of articulation, and neither the [ʁ] nor the [χ] exists phonemically or allophonically in English. (Chomsky and Halle, 1968).

Bearing these cross-linguistic distinctions in mind, the various theories discussed above would predict different outcomes for French and AE speakers in discriminating these sounds. The NLM and SLM would posit that the French speakers would have great difficulty with discrimination, as the allophonic relation between the two sounds would render them too similar to be differentiable in the phonological space. On the other hand, English speakers wouldn’t be hindered by any intra-phonemic equivalence classification or phonological space distortion. PAM,
on the other hand, would predict that the French speakers would discriminate the sounds as a Category Goodness contrast, which would result in above chance discrimination. The English speakers, however, would have a more difficult time, as they don’t have similar phonemes to aid in discrimination; they would probably have to rely on lower-order, “last resort” nonlinguistic/acoustic cues in order to discriminate, such as native voicing contrasts in fricatives at other places of articulation.

III. EXPERIMENTS

Two questions are set as the focus of this paper. First, I seek to know if French speakers can discriminate its rhotic, $\mathcal{R}$, from the voiceless, uvular fricative $\chi$. In other words, I posit that in the environment following a voiceless consonant, the allophone of the French rhotic /$\mathcal{R}$/ is not simply a phonetic variant of that phone, but is actually realized as the uvular fricative $\chi$. To test this, native French speakers participated in an ABX test of nonsense words, recorded by a native Hebrew speaker trained in Phonetics and Phonology, as both /$\chi$/ and /$\mathcal{R}$/ are phonemic in Hebrew. The nonsense words were all bisyllabic, and tested the contrastive phones in various environments: word-initially, following a voiced consonant, and following a voiceless consonant. Were the French speakers able to discriminate the sounds across all environments, I would conclude that in fact, the French rhotic allophone that appears after a voiceless consonant is simply a phonetic variant of the /$\mathcal{R}$/ phoneme, rather than an approximation of the Hebrew fricative, /$\chi$/.

However, should the French speakers show significant difficulty discriminating the sounds, especially in the voiceless consonant clusters where the French rhotic allophone tends to be realized, it would be concluded that the realization of the rhotic’s allophone in French is devoiced to a degree that it bears significant similarity to the Hebrew fricative. I expect the latter to be the case.

The second aim is to see if the /$\mathcal{R}$/ and /$\chi$/ phones are inherently acoustically dissimilar, and therefore easy to distinguish universally, or if they are inherently close in the phonological space,
and speakers of languages such as Hebrew which discriminate them phonemically must learn the contrast in their early years of life. To test this, the same ABX test that was performed by native French speakers was also performed by native American English (AE) speakers. Because English does not employ either of these sounds phonemically or phonetically, it can be assumed that the way AE speakers hear the two sounds reflects the sounds’ natural relationship with one another. If the AE speakers results are similar to those of the Hebrew speakers, then it would seem right to claim that the two sounds are in fact inherently dissimilar and universally distinguishable. If, on the other hand, as is expected, the AE speakers results pattern more like those of the French speakers, we could assume that they are in fact similar, and speakers of languages that contrast them, like Hebrew, would have to learn the contrast.

IV. METHODS

A. Stimuli

The nonsense words that were used as stimuli were recorded by a male native speaker of Hebrew, a phonologist, in a sound-attenuated booth in the Phonetics and Experimental Phonology (PEP) lab at the department of Linguistics at NYU. Each token was recorded twice through a hands-free Sony headset microphone, and analyzed in Praat. Any token that did not fit the standard stress pattern of Hebrew employed in the pronunciation, mil’ra, word-final stress, was discarded. The selected tokens were analyzed through their waveforms and spectrograms in Praat, to ensure that all the phones retained their intended articulation, with no assimilation or voicing harmony affecting the quality of the sounds. All selected tokens were run through an intensity scaler, so that no tokens would be distinguishable based on amplitude.

In the word-initial environments, the nonsense words fit a CVCV pattern: the first C was either our rhotic or our fricative (2) X the first vowel was, in turn, [i,e,a,o,u] (5) X the second consonant, which was one of the stops, [p,t,k] (3) X the second vowel, which was [o,a] (2) = 60
nonsense words, or 30 minimal nonsense pairs. One such pair is [χupa]-[ʁupa]. I expect the French speakers to have a relatively easy time discriminating the sounds word-initially, as that is an environment where the rhotic rather than the fricative definitively appears in their phonology. Meanwhile, AE speakers should have a relatively more difficult time, as their phonology provides no framework for these sounds. However, no preference for one sound over the other is expected to be shown in the results of the AE speakers, as neither sound is more familiar to them than the other.

In both the voiceless and the voiced consonant clusters, the nonsense words fit a VCCV pattern: the first vowel was [o] (1) X the first consonant was one of the stops,[p,b,t,d,k,g] (6) X either our rhotic or our fricative (2) X the second vowel, [i,e,a,o,u] (5) = 60 nonsense words, or 30 minimal nonsense pairs. In total, then, there were 120 nonsense words and 60 minimal pairs. An example of a voiced-cluster minimal pair is [ogχo]-[ogʁo]; an example of a voiceless-cluster pair is [opχu]-[opʁu]. I expect the French speakers to have an easy time discriminating the sounds in the voiced clusters, as in that environment, the allophone of the rhotic is the underlying form. Voiceless clusters are where the French speakers are expected to have the most difficulty. AE speakers are expected to have an equally difficult or easy time with both clusters, again because their phonology provides no framework for these sounds.

Examples of stimuli in each context are in Figures 1, 2 and 3. It is clear from those waveforms and spectrograms that none of the stimuli were misarticulated, or underwent any phonetic processes such as voicing assimilation that would distort their character as good exemplars of the respective phonemes. The possibility that any results would be the effect of mis-articulation of the stimuli must therefore be ruled out. Any difference in the levels of accuracy in the perception of these non-native segments must be related to the nature of the segments themselves.
Figure 1: Spectrograms of /ʁeka/ (above) and /χeka/ (below). The long duration and thick background striations of the latter point to the conclusion that the target sound was, in fact, articulated as a pure fricative while the target sound of the former, with its shorter duration, voicing bar, and cleaner striations confirm it as a rhotic.
Figure 2: Spectrograms of /obși/ (above) and /obți/ (below). It is clear from the a-periodicity, background striations and long duration that the target sound in the latter retained its character as a pure fricative.
Figure 3: Spectrograms of /ɔ̃ʁa/ (above) and /ɔ̃χa/ (below). Characteristically, the voiceless fricative has a significantly longer duration than the voiced rhotic. There is also a characteristic dip in F3 during the transition out of the rhotic, and the rhotic’s spectrogram shows significantly less fricative striations.

Each potential ABX set was tested: ABA, ABB, BAA, BAB, leading to 240 sets. The complete list of nonsense words are listed in the Appendix. The trial was done through the E-Prime 2.0 program. It randomized the trials for each participant and automatically recorded the results. The trials were programed so that there was a pause of 400 ms between nonsense words of the trial, and the response time between trials was 2,000 ms. If a participant answered before that time was up, the program automatically moved on to the next trial.

B. Procedure

Each participant filled out a brief questionnaire about their non-native language experiences and any hearing/auditory processing deficits they might have. Two native Hebrew speakers took the ABX test as a sort of control, to make sure that the contrast, as Hebrew sounds, was distinct. Their non-native linguistic experience was not important. Data from the native French speakers could not be used if they had any experience or significant exposure to Hebrew, as that would skew them to have an easy time discriminating the contrast. Data from the native AE speakers could not be used if the participants had any experience or significant exposure to either Hebrew or to French, as that would mean they were familiar with the contrastive sounds in one form or another.

Each participant was told briefly that in the test, they would hear sets of fake words, and that each set consisted of three of those fake words. The first two words in the set would always differ from each other somehow, and the third word would always be the same as either the first or the second. It was their job to decide to which of the two preceding words the third was a match, and press a button corresponding to that word to record their answer. They were warned that the words would be played in quick succession, that they should give their initial, gut-reaction, and that they should do their best to answer all trials. They were then taken into the testing room, shown the
button box, and were encouraged to adjust the volume on the over-the-ear headphones to a level that was comfortable for them. Instructions were displayed on the screen and when they felt ready, the participant began the 6 trial practice. Only once they felt comfortable did they commence the test.

V. RESULTS

A. Native Hebrew Speakers

Two native speakers of Hebrew (1 female, 1 male) sat the study. Their average age was 27. One was phonetically trained and one was not. Both participants were at ceiling: one answered only 2 out of 240 trials incorrectly and left none out, while the other left 2 out but answered all the rest correctly. From this data it was clear that the stimuli were accurate and that the minimal pairs were in fact perceptibly contrastive for Hebrew speakers.

B. Native French Speakers

Seven native speakers of French (4 female, 3 male) participated in the study. Their average age was 31 (r:17-62), and none had any experience, formal or informal, with Hebrew. The results for each environment, including the mean of incorrect answers and Standard Deviation, are given in Table 1. Dividing between trials where X was the same as A versus when it was the same as B, a noticeable recency effect emerges in 4 out of 6 of the environments. Additionally, in every single environment, participants incorrectly chose B more than double the amount of times they incorrectly chose A. This phenomenon can be seen in Figure 4 and Table 1. No effect was shown for unanswered trials.
Figure 4: French speakers’ percent incorrect (Y axis), divided by context of the target sound (X axis) and further by the position of the target sound in the set. + and - refer to the voicing value of the consonant preceding the target in the clusters.

<table>
<thead>
<tr>
<th>Context</th>
<th>Correct</th>
<th>Incorrect</th>
<th>Mean Incorrect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+v singletons, A</td>
<td>73.33%</td>
<td>16.67%</td>
<td>5.00</td>
<td>4.47</td>
</tr>
<tr>
<td>+v singletons, B</td>
<td>85.24%</td>
<td>8.10%</td>
<td>2.43</td>
<td>2.15</td>
</tr>
<tr>
<td>-v singletons, A</td>
<td>70%</td>
<td>23.33%</td>
<td>7.00</td>
<td>4.62</td>
</tr>
<tr>
<td>-v singletons, B</td>
<td>75.24%</td>
<td>8.10%</td>
<td>2.43</td>
<td>2.15</td>
</tr>
<tr>
<td>+v+v clusters, A</td>
<td>73.33%</td>
<td>15.24%</td>
<td>2.29</td>
<td>2.29</td>
</tr>
<tr>
<td>+v+v clusters, B</td>
<td>88.57%</td>
<td>6.67%</td>
<td>1.00</td>
<td>1.15</td>
</tr>
<tr>
<td>-v+v clusters, A</td>
<td>66.67%</td>
<td>26.67%</td>
<td>4.00</td>
<td>2.77</td>
</tr>
<tr>
<td>-v+v clusters, B</td>
<td>80.00%</td>
<td>11.43%</td>
<td>1.71</td>
<td>1.80</td>
</tr>
<tr>
<td>-v-v clusters, A</td>
<td>66.66%</td>
<td>19.05%</td>
<td>2.86</td>
<td>2.41</td>
</tr>
<tr>
<td>-v-v clusters, B</td>
<td>82.86%</td>
<td>7.62%</td>
<td>1.14</td>
<td>1.35</td>
</tr>
<tr>
<td>+v-v clusters, A</td>
<td>66.67%</td>
<td>21.90%</td>
<td>3.29</td>
<td>0.95</td>
</tr>
<tr>
<td>+v-v clusters, B</td>
<td>84.76%</td>
<td>7.62%</td>
<td>1.14</td>
<td>1.35</td>
</tr>
<tr>
<td>Total</td>
<td>76.07%</td>
<td>14.29%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Overall Results for French Speakers by Context. Contexts are based on the position in the word and voicing on the target phoneme. In the clusters, the first voicing value is for the consonant that immediately precedes our phoneme; the second value is that of the phoneme itself. Each context was also broken up into instances in which target X was identical to A and those in which it was identical to B.

Participants had a more difficult time with the fricatives than with the rhotics in two of the three environments, as illustrated in Figure 5. Word-initially and in voiced clusters, the French speakers scored between 2% and 3% more incorrect for the fricatives. However, in clusters following a voiceless consonant, participants actually scored 5.7% more wrong with the rhotic than with the fricative targets. As I have already ruled out the possibility of misarticulation, these results must be taken as reflective of French speakers’ perception. This, however, fits well with my expectations: voiceless environments are typically where the fricative-like French allophone is realized, and it would be in these environments that French speakers’ L1 phonotactics affect their perception of non-native sounds to assimilate the rhotic to a fricative. Because it is following a voiceless consonant that the French rhotic is realized with some devoicing, the voiceless fricative appears to be more easily perceptible and assimilable for French speakers in that environment than is the rhotic. It is a natively permissible phonotactic structure for them, whereas the fricative in other environments is not.
Looking at the rhotics in the three different environments, the same pattern holds. In word-initial and voiced cluster environments, participants scored 79.29% and 80.95% correct, respectively. They scored 12.38% incorrect word-initially and 10.95% incorrect in voiced clusters. Meanwhile in voiceless clusters, participants went down to 73.33% correct, and went up to 19.05% incorrect.

The results for the fricatives in the different environments only varied with regard to the number incorrect. In voiced clusters, participants scored 75.71% correct, with 14.76% incorrect. In voiceless clusters they scored 74.76% correct and 13.33% incorrect. Word initially, the score correct was 72.62%, and the number incorrect went drastically up, to 24.76%.

C. Native American-English Speakers

Four native speakers of American English (3 female, 1 male) participated in the study. They had an average age of 22.25 and none had any knowledge of Hebrew, nor of French. A fifth subject participated, but his data was not used as it was significantly outlying in all environments, to the degree that it seemed as though he might have misunderstood the instructions. Table 2 shows the results for each environment, including mean of incorrect responses and Standard Deviation, broken up further by whether the target X was the same as A or the same as B. English speakers showed recency effects in all environments, as can be seen in Figure 6.
The largest difference was in voiceless clusters. When the target was a fricative, participants scored 77.33% correct, and 19.33% incorrect. With the rhotics, participants scored 83.33% correct and 14.67% incorrect. For all other environments, the difference between rhotics and fricatives- for correct and incorrect answers- was no more than 3 percentages.

The rhotics were fairly stable for American-English speakers. They scored 84.67% correct when word-initial, 86.00% correct when in a voiced cluster and 83.33% correct when in a voiceless cluster. There was a slight range for fricatives. Word initially, participants scored 83.00% correct and 85.33% correct in voiced clusters, but 77.33% in voiceless clusters. Participants’ percent correct were on par for the rhotic and the fricative when word-initial and in voiced clusters. The rhotic targets answered incorrectly in word-initial position came to 12.33%; in voiced clusters, 13.33%; in voiceless clusters, 14.67%. For incorrectly answered fricative targets in word-initial position, the score came to 13.00%; in voiced clusters, 11.33%; in voiceless clusters, 19.33%. This is illustrated in Figure 7.
Figure 7: American-English speakers’ percent incorrect for each of the targets in each environment

<table>
<thead>
<tr>
<th>Context</th>
<th>Correct</th>
<th>Incorrect</th>
<th>Mean Incorrect</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+v singletons, A</td>
<td>83.33%</td>
<td>14.00%</td>
<td>5.25</td>
<td>5.44</td>
</tr>
<tr>
<td>+v singletons, B</td>
<td>86.00%</td>
<td>10.67%</td>
<td>4.00</td>
<td>2.94</td>
</tr>
<tr>
<td>-v singletons, A</td>
<td>76.67%</td>
<td>18.67%</td>
<td>7.00</td>
<td>5.94</td>
</tr>
<tr>
<td>-v singletons, B</td>
<td>89.33%</td>
<td>7.33%</td>
<td>2.75</td>
<td>3.59</td>
</tr>
<tr>
<td>+v+v clusters, A</td>
<td>80.00%</td>
<td>20.00%</td>
<td>3.75</td>
<td>3.86</td>
</tr>
<tr>
<td>+v+v clusters, B</td>
<td>92.00%</td>
<td>6.67%</td>
<td>1.25</td>
<td>0.96</td>
</tr>
<tr>
<td>-v+v clusters, A</td>
<td>76.00%</td>
<td>21.33%</td>
<td>4.00</td>
<td>3.16</td>
</tr>
<tr>
<td>-v+v clusters, B</td>
<td>90.67%</td>
<td>8.00%</td>
<td>1.50</td>
<td>1.91</td>
</tr>
<tr>
<td>-v-v clusters, A</td>
<td>66.67%</td>
<td>30.67%</td>
<td>5.75</td>
<td>2.50</td>
</tr>
<tr>
<td>-v-v clusters, B</td>
<td>88.00%</td>
<td>8.00%</td>
<td>1.50</td>
<td>1.29</td>
</tr>
<tr>
<td>+v-v clusters, A</td>
<td>80.00%</td>
<td>17.33%</td>
<td>3.25</td>
<td>2.63</td>
</tr>
<tr>
<td>+v-v clusters, B</td>
<td>90.67%</td>
<td>5.33%</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Total</td>
<td>83.42%</td>
<td>13.67%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D. Results: Inter-Language Comparison

Hebrew speakers were at ceiling for both targets across all contexts, which is expected as a language that contrasts the two sounds phonemically. French speakers performed the lowest of all populations in each context. The target that was performed better for them in any given context never matched the worse performed target for AE speakers in the same context. Both word-initially and in voiced clusters, they performed slightly better when tasked with recognizing the rhotic than with recognizing the fricative. In voiceless clusters, their recognition of fricatives surpassed their recognition of rhotics, if just slightly. This was due to their performance on rhotics falling slightly below their performance in the other two contexts; their performance for the fricatives was nearly identical across contexts. No target in any context was an outlier compared to the rest; all the values hovered around the 75% accuracy mark.
American-English speakers performed somewhat better than the French speakers, though their results did not approximate those of the Hebrew speakers. They always had stronger results for rhotic recognition than for fricative recognition, though the results for the two targets are nearly identical in word-initial and voiced cluster contexts. In the voiceless clusters, their performance on fricatives drops a bit, and the difference between that and their performance on the rhotics approaches 5 percentage points.

VI. DISCUSSION

The data collected herein were meant to address two questions. The first concerned the status of the voiceless uvular fricative, /χ/, in the phonology of native French speakers. Posited to be an allophone of the voiced uvular rhotic, /ʁ/, it was found that participants indeed performed poorest when their target was preceded by a voiceless consonant, the environment in which French speakers realize the /ʁ/ with a measure of devoicing. That is to say, in the environment where the allophone of the rhotic that is being investigated is typically realized, French speakers had a difficult time perceiving the phonemic rhotic. I conclude that the voiceless fricative, /χ/, is at least sufficiently similar, if not identical, to the allophone of /ʁ/ that is produced immediately following a voiceless consonant in a cluster.

The second point of inquiry focused on the cross-linguistic status of the two sounds. They differ in voicing, which is contrastive in many languages. They also differ somewhat in manner, as the /ʁ/ is a fricated rhotic, with the characteristic rhotic property of a dip in F3 before transitioning into a vowel (see Figure 3, above) while the /χ/ is a pure fricative. Yet, despite two points of divergence, they exist in some languages, typified herein by French, in complementary distribution. This being so, is there some inherent acoustic similarity between the two sounds that overrides their featural differences? The purpose of running the ABX test on AE speakers was to investigate this
question: when untouched by the specific phonology of a given language, how do these sounds behave?

English is a particularly good language to use as a “neutral” tester, as it also contrasts by voicing, and employs rhotics and fricatives as two distinct manners (Chomsky and Halle, 1968). The uvula is contained within the phonological space of English, so even though it doesn’t function as a primary place of articulation, sounds produced there should be heard as possible speech sounds (Best et al. 2001). Therefore, the characteristics of our two sounds are equally represented in English— including manner and voicing, and not including place— and so should theoretically be perceived with the same amount of ease or difficulty by English speakers. Therefore, any phenomena or patterning that were to arise in the results of the AE speakers’ testing would demand some explanation.

In fact, a pattern was found. AE speakers performed better in all contexts when the target phone was the rhotic rather than the fricative. In voiced clusters, the difference was slight. Word-initially, the accuracy differed by less than 2 percentage points, but in voiceless clusters, the difference was greater, approaching 5 percentage points. As seen in Figures 1, 2 and 3, above, all stimuli were produced accurately, without any articulatory processes such as voicing assimilation distorting the nature of either the rhotic or the fricative.

In all trials, participants heard both the rhotic and the fricative, and then one of the two a second time. Due to the strong recency effects exhibited by the AE-speaking participants, I choose to focus only on trials structured as ABA, where the target sound was that which was contained in the first nonsense word of the set (e.g: [χupa]-[ʁupa]-[χupa]).

In such cases, when participants chose the correct target, they recognized, at the very least, that the second sound was not the same as the target. When participants chose incorrectly, they misperceived either the target sound or the sound in B position, or both, to a high-enough degree that the two sounded identical. American-English speakers, then, mistook /χ/ for /ʁ/ when preceded
by the latter more often than they mistook /ʁ/ for /χ/. The most likely explanation to account for this is that rhotics are easier to identify as such than fricatives are as fricatives. Therefore, when the AE participants heard the non-native rhotic in B position, they assimilated it to their native rhotic category, and that familiar category subsumed the following sound as well. However, participants did not have an automatic category to which to assimilate the fricative, as English contains several varied ones, and none particularly close to the uvula. Therefore, when that was in B position, it did not signal a native category, and so could not so easily distort the categorization of the rhotic in the following target word.

There is an important caveat to this approach. It must either assume that universally, rhotics are easier to perceive than are fricatives, or it must be considered specific to languages like English that employ only one rhotic and would therefore easily categorize non-native rhotics under the umbrella of their native one. To differentiate between these two possibilities, it would be necessary to run the test on speakers of languages that employ more than one rhotic as well as more than one fricative. Additionally, ABX tests that contrast two different rhotics could be performed by speakers of any of the languages dealt with herein to see if they are capable of discriminating the two.

The results still give insight into the relationship between these two sounds for English speakers, if not universally. Across all contexts and for both targets, they performed markedly better than did the French speakers, but in no way approximated the ceiling-level performance of the Hebrew speakers. The rhotic proved easier to perceive. One potential explanation for this is that the rhotic was assimilated somewhat easily, though not perfectly, to a native category. That category would likely be the alveolar rhotic, /ɹ/. The fricative must not have been so easily assimilated to a native category. In the Perception Assimilation Theory (PAM) terms, our rhotic behaved as a categorized sound. On the other hand, our fricative did not so easily fall into a native category. The ease with which the rhotic was assimilated to a native category is likely because its
distinctive property of rhoticity appears in only one English phoneme. The fricative, however, does not have any properties that are unique to only a single English phone, and could therefore not be so instantaneously categorized. However because the uvula is in the phonological space of English, the /χ/ must have been heard as a possible speech sound, and functioned in PAM terms as an un categorized sound. It would not have been assimilated to /k/, the closest native English phoneme to the fricative, despite the /k/ functioning as a sort of cognate to the velar /x/ and uvular /χ/ in foreign loan words, such as the German name, “Bach,” [bax], pronounced in English as [bak]. This “cognate” phenomenon is a result of /k/ being the closest sound to those fricatives that English speakers can easily produce. English speakers would not, however, actually confuse the fricatives for a /k/ because English does contrast between plosives and fricatives.

Although PAM’s terms are convenient here, what predictions does it make about our results? PAM’s claim for Uncategorized-Categorized non-native contrasts is that the difference in the two sounds’ level of similarity to native categories should make discrimination easy. The AE speakers did in fact answer the majority of trials correctly- more than 75%. There was also a non-trivial number that they answered incorrectly, which is where the bias towards the categorized rhotic was highlighted.

If PAM is indeed correct, then that is support for the hypothesis that a given rhotic is more easily identifiable as a member of said category than is a fricative of that category of manner of articulation. However to be sure that the categorizations assigned to each phone here in retroactive analysis are accurate, a categorization trial for each phone would need to be performed by AE speakers.

The Native Language Magnet (NLM) model is slightly more difficult to deal with. Its postulate that non-native speech sounds- and we have indeed established that AE speakers hear both our phones as possible speech sounds- are “pulled in” by the prototype phoneme most similar to it creates difficulties. By this model, the attraction exerted on a non-native phone by a similar native
one allows for the categorization, though imprecise, of the non-native phone in the listener’s phonology. In this test, the rhotic was recognized as itself more easily than was the fricative, suggesting, according to the NLM, that the rhotic fit more naturally into a native category than did the fricative, which was more elusive for the listeners.

Yet, how is one to determine to which prototype each of our sounds is most similar? In PAM, we were able to categorize based on articulatory similarity, which allowed the uvular rhotic to be assimilated to the alveolar one; the NLM demands categorization to a given phoneme based on an account of the specific phonological space. Dealing first with their differing manners of articulation with reference to place of articulation, the /χ/ should have a stronger attraction to a native prototype than should the /ʁ/, as the only English rhotic is alveolar, while there are fricatives produced farther back in the vocal tract and closer to the uvula, such as the palato-alveolar /ʃ/.

When it comes to the difference in voicing of our sounds, English employs a phonological voicing contrast extremely close to the uvula, with the velar stops, /g/ and /k/. While that consideration gives no preference to either sound, one could argue that such prototypes favor the perception of the /χ/ over the /ʁ/ in reference to the velar stops, as fricatives are much closer to stops in the sonority hierarchy than are rhotics. This analysis seems to suggest that the fricative is more likely to be pulled into a native prototype than is the rhotic. Although it is quite possible that rhoticity generates exceptionally strong magnet effects across the phonological space, the NLM does not make an account for this, and so it seems as though this model does not properly predict the AE results.

The last model dealt with is the Speech Learning Model (SLM). Its main predictions are meant for speakers who have significant experience with the L2 target language. Because the AE speakers that were tested had absolutely no knowledge of either Hebrew or French, the SLM’s predictions are not tailored for this study. To fully understand why this approach is necessarily

1 The sonority hierarchy, from least to most sonorous: stops, affricates, fricatives, nasal, liquids, glides, vowels.
untenable here, one must allow for a measure of abstraction. The SLM predicts, based on Trubetzkoy’s foundational L1 sieve analogy (1939), that listeners will have an easier time accurately perceiving, and therefore producing, L2 sounds that are dissimilar from any L1 category; sounds similar to an L1 category would be subsumed under that category, and the listener would not be able to detect the phonetic variations of that sound. These predictions relate specifically to the discrimination of similar but not identical native and non-native categories, and are predicated on the assumption that listeners have significant exposure to the L2. The ABX test, for the AE speakers, examined discrimination between two non-native, unfamiliar categories. Therefore, the SLM could not be manipulated to make predictions about the acumen of the AE speakers’ perception here.

While irrelevant to the test for American-English speakers, SLM’s predictions can explain the results of the French speakers’ ABX test quite well. The rhotic, /ʁ/, is a native phoneme, and the purpose of the ABX test was to determine if the fricative, /χ/ was perceived as similar enough to it to be considered a possible allophone of the rhotic in certain contexts. That the French speakers often heard the fricative as a rhotic suggests that, especially when following a voiceless cluster, the fricative may indeed exist as an allophone to the rhotic.

In fact, all three models’ postulates account for the results obtained from the French speakers. In PAM terms, this is a Category-Goodness contrast: two allophones of the same phoneme, with one, the rhotic, functioning as the “elsewhere” allophone. This results in alright but not ceiling perception, especially not in the environment in which the other allophone is usually realized. (Best et. al. 2001). Because the rhotic is a phoneme in French, it functions as an NLM “prototype,” to which its allophone, the fricative, is pulled. This results in distortions in the voiceless clusters, where French speakers usually realize the rhotic as the fricative; essentially, it is in the environment in which the allophone of the phoneme is realized that the “pull” of the fricative to the rhotic is strongest, and therefore most difficult to discriminate for the French speakers.
VII. CONCLUSION

On the basis of the ABX test results, it is found that the voiceless uvular fricative, /χ/, phonemic in Hebrew, phonetically approximates the allophone of the French voiced uvular rhotic, /ʁ/, when preceded by a voiceless consonant. All three of the non-native contrast models discussed herein,—PAM, NLM, SLM—can account for this. As for English specifically, it seems that PAM is most likely to correctly predict our results. That would mean that for English speakers, non-native rhotics are easier to perceive than are non-native fricatives. As to the question of the universal relationship between the /ʁ/ and the /χ/, it is difficult to say. It is possible that our results cannot be generalized beyond English. One would have to administer the test to speakers of languages that employ multiple rhotics as well as multiple fricatives in order to check the validity of any extrapolation of these results. However, it is predicted that universally, rhotics are more easily recognized as members of that category than fricatives are as fricatives, though the level of ease may differ by language.

Hearkening back to the foundation of the study of non-native contrasts, Lado’s Contrastive Analysis Hypothesis (CAH), though flawed, can be seen as a basic structure around which this study was organized. As mentioned, Lado describes three levels of any language-specific sound system: phonemic, phonetic and distributional (Lado, 1957). The /ʁ/ and the /χ/ are phonemic in Hebrew. Certainly, AE listeners showed a robust and systematic preference for one over the other, though it would be inaccurate to refer to that difference as phonemic, as phonetic or as distributional; the sounds simply are not part of the native phonology. Finally, it was not clear if these two sounds bore any relationship in French. Certainly the rhotic is phonemic, but what of the fricative? Our results suggest that the fricative may be the allophone of the rhotic that appears in the environment of a voiceless consonant, making the French distinction between the two, in Lado’s conception, a distributional one.
IX. APPENDIX

A. Stimuli

χυρα-κυρα χαρα-καρα χερα-κερα χιρα-κιρα χωρα-κορα χωτα-κώτα
χατα-κατα χετα-κετα χίτα-κίτα χότα-κότα κύκα-κυκα κάκα-κακα
χεκα-κέκα χικα-κίκα χοκα-κοκα χεκο-κεκο χικο-κίκο χοκο-κοκο
χιρο-κιρο χορο-κορο χιτο-κίτο χωτο-κώτο χετο-κετο χίτο-κίτο
χωτο-κώτο κύκο-κυκό κάκο-κακό κεκο-κέκο χικο-κίκο χοκο-κοκο
οτχο-οτκο οτχα-οτκα οτχε-οτκε οτχι-οτκи οτχο-οτκο οδχο-οδκο
οδχα-οδκα οδχε-οδκε οδχι-οδκι οδχο-οδκο οκχο-οκκο οκχα-οκκα
οκχε-οκκε οκχι-οκκι οκχο-οκκο ογχο-ογκο ογχα-ογκα ογχε-ογκε
ογχι-ογκι ογχο-ογκο ορχα-ορκα ορχα-ορκα ορχε-ορκε ορχι-ορκι
ορχο-ορκο οβχα-οβκα οβχα-οβκα οβχε-οβκε οβχι-οβκι οβχο-οβκο

References


