EFFECTS OF PREDICTABILITY OF DISTRIBUTION ON WITHIN-LANGUAGE PERCEPTION*

Kathleen Currie Hall, Veronica Letawsky, Alannah Turner,
Claire Allen, and Kevin McMullin
University of British Columbia

1. Introduction

The phonological relationship that holds between two sounds in a language is known to affect perception, with members of a pair of sounds that are contrastive being perceived as more distinct from each other than members of a pair that are allophonic (e.g., Kazanina et al. 2006, Boomershine et al. 2008). Although there are several (perhaps interacting) factors that determine whether a pair of sounds is contrastive or allophonic (see, e.g., Steriade 2007, Hall 2013a), the degree to which two sounds are in complementary distribution is one primary means of doing so (see Hall 2009). Traditionally, this measure has been taken to be, essentially, binary: sounds that are in entirely complementary distribution are taken to be allophonic (unless there is some other reason, such as phonetic distance, that would prevent this interpretation), while the existence of even one environment where the sounds overlap is taken to be a sign of contrast. Hall (2009, 2013b) proposes, however, that the notion of complementarity of distribution can better be modelled probabilistically, such that there are meaningfully different degrees to which distributions overlap. While the algorithm for calculating predictability of distribution (ProD) is relatively well-established (see §2.1), the effect of different degrees of predictability on perception has so far been rather inconclusive. Hall (2009) showed limited effects of ProD on the perception of the German voiceless fricatives [s] and [ʃ], but no clear effects on other phone pairs. Hall and Hume (2014, submitted) show effects of ProD, but only in conjunction with other factors such as phonetic similarity, frequency, and functional load, on the confusion patterns between French vowels. In this paper, we present evidence that ProD does play a role in shaping perception, and furthermore, that this is true even within a single pair of sounds in a language across different contexts.

2. Background: Predictability of Distribution

2.1 Calculating Predictability of Distribution

As mentioned above, Hall (2009, 2013b) proposes an algorithm for quantifying the degree to which a pair of sounds is in complementary distribution, that is, the predictability of their distribution (ProD). This metric is based on the Information-

* We gratefully acknowledge support for this research from the SpeeCon Laboratory at the University of British Columbia and the Social Sciences and Humanities Research Council of Canada, through an Insight Development Grant (#430-2013-0802) to Kathleen Currie Hall.

Actes du congrès annuel de l'Association canadienne de linguistique 2015.
Proceedings of the 2015 annual conference of the Canadian Linguistic Association.

Theoretic measure of entropy (Shannon and Weaver 1949). Entropy is a measure of uncertainty. When applied to phonological relationships, it can be used to measure how much uncertainty there is about which of two sounds occurs in any given phonological environment. If there is no uncertainty (an entropy of 0), then only one of the sounds occurs in that environment. As the entropy increases, there is greater uncertainty about which of the two sounds occurs. This measure can be extended from a single environment to cover all environments and thus provide a general measure of ProD. As will be shown below, the maximum entropy for a single pair of sounds is 1; thus, entropy ranges from 0 to 1, with 0 meaning complete certainty (analogous to perfectly complementary distribution, i.e., allophony) and 1 meaning maximal uncertainty (analogous to perfectly overlapping distribution, i.e., contrast).

The formula for entropy is given below in (1). The formula in (1a) shows the overall entropy formula across all environments, which is a function of the entropy in each environment (1b) and the probability of each environment (1c). The entropy of two sounds in a single environment, as shown in (1b), is calculated by taking the probability of each sound in that environment multiplied by that sound's log probability, and summing the products for the two sounds. The sum is then multiplied by -1 simply to make it a positive number. For example, if [d] and [t] both occur word-initially in some language, but of all the instances of either sound in that position, 30% are [d] and 70% are [t], then the entropy in that environment is $-(0.3 * \log_2(0.3) + 0.7 * \log_2(0.7)) = 0.881$. This formula is then repeated across all environments where at least one of the two sounds occurs, and the entropy of each environment is weighted by the probability of that environment's occurring, calculated as in (1c). Thus, if two sounds are largely in complementary distribution, except in one environment, then that one environment will only affect the overall appearance of contrastiveness if it is itself a highly frequent environment. Both the entropy in individual environments (given by (1b)) and that of the entire system (given by (1a)) range from 0 to 1. The maximum entropy of 1 occurs when each of the two sounds occurs with a frequency of 50%; anything other than this perfect split will result in a smaller entropy.

(1) a. Systemic Entropy = $\sum (H(e) * \mathrm{p}(e))$

b. $\mathrm{H(e)} = - \sum p_i \log_2 p_i$

c. $\mathrm{p(e)} = \mathrm{N}_e / \sum \mathrm{N}_{e \in E}$

2.2 The Effect of ProD on Perception

As mentioned in §1, the phonological relationship that holds between sounds is known to have an effect on their perceived similarity. Boomershine et al. (2008), for example, showed that [d] and [ð], which are contrastive in English but allophonic in Spanish, are perceived as more distinct by English speakers than they are by Spanish speakers, while [d] and [ɾ], which are contrastive in Spanish but allophonic in English, had the opposite pattern of perceived similarity. Furthermore, Hume and Johnson (2003) showed that pairs of sounds (in their case, tones) that are neutralized in at least one context are perceived as being more similar than pairs that are never neutralized, even in the contexts where the sounds are not neutralized. This finding suggests that intermediate degrees of ProD matter: neutralization creates an environment of perfect predictability of distribution, and this isolated lack of contrast affects overall perception.

Further attempts to examine the role of ProD have had limited success. Hall (2009) examined four pairs of sounds with varying degrees of ProD in German, but did not find an overall correlation between perceived similarity and ProD. This lack of a result, however, may have been due to the rather different raw acoustic details of the sounds in question: the sounds may simply have been too different acoustically for finer-grained differences in ProD to have had an effect.

Hall and Hume (2014, submitted) do show a consistent role for ProD in predicting confusability patterns among French vowels, though again, only among pairs of vowels that are phonetically similar enough to each other that any confusion is viable. Among such pairs, ProD has the expected effect: pairs of vowels that are more predictably distributed are more likely to be confused than those that are less predictably distributed.

As a general proposition, phonologists are interested in the systemic entropy, that is, the overall degree of ProD in a language. But note that the calculation of ProD automatically also gives us a measure of the predictability of two sounds in individual environments. This characteristic may allow for a way to circumvent the problems found in Hall (2009) and Hall and Hume (2014, submitted) of acoustic differences overshadowing differences in ProD. Specifically, it is possible to look for differences in perceived similarity within a single pair of sounds (thus keeping the acoustics maximally similar), across different environments in which ProD scores vary. While Hume and Johnson (2003) did find that neutralization in one context transferred to greater perceived similarity in other contexts as well, Hall (2009) found some limited support for differences in ProD across environments affecting the perception of [s]/[ʃ] differentially in those environments. Such a finding had not been the intent of that study, however, and so the pairs and environments were relatively limited. The goal of the current study is to further probe the possibility that ProD affects the perceived similarity of pairs of sounds, by looking within single pairs of sounds across different environments. Figure 1 summarizes the hypothesis being tested: the less overlapping the distribution of a pair of sounds, the more similar the sounds will be perceived.

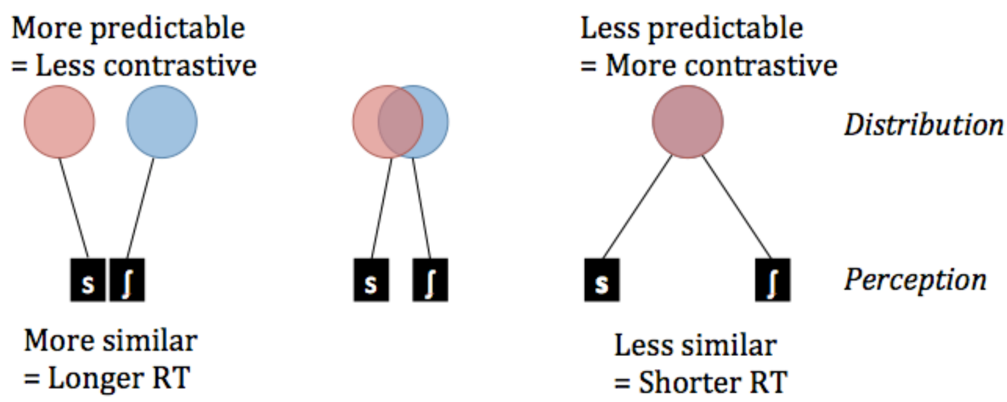

Figure 1: Hypothesized relation between predictability of distribution and perceived similarity

3. Methods

The overall strategy for testing the role of ProD on perceived similarity is as follows. We first ran an AX discrimination experiment on speakers of English to test the perceived similarity of pairs of voiceless fricatives across a variety of contexts. We then correlated the results of the experiment with measures of ProD in those same contexts from a corpus of English. Voiceless fricatives were chosen as the target sounds to minimise the effect of the contexts themselves on the similarity. For instance, if voiceless stops had been chosen, we would expect *a priori* that the sounds in pre-vocalic positions would be perceived as more distinct than those in post-vocalic positions, simply because of the greater availability of acoustic cues in the transition between consonant and vowel than between vowel and consonant. While fricatives are of course also influenced by their contexts, they contain many of the cues to their identity within themselves, thus minimising contextual effects. Voiceless fricatives in particular were chosen because of their cross-linguistic frequency; our intention is to compare the results found here with results of speakers of other languages where the same fricatives occur but with different patterns of ProD. Only results from English-speakers are presented in the current paper, however.

3.1 Experimental Methods

An AX discrimination task was used to probe perceived similarity. Following Boomershine et al. (2008), reaction time (RT) was used as the dependent measure, with longer RTs being assumed to indicate greater perceived similarity.

3.1.1 Stimuli

All stimuli contained one of the voiceless fricatives [f], [h], [s], or [ʃ]. In addition to the reasons for using voiceless fricatives mentioned above, no additional sounds were used as fillers, so as to maximize the acoustic similarity across tokens and encourage participants to make use of phonological differences instead. These fricatives were embedded in VC, CV, and VCV contexts, where the V was one of {[i], [ɑ], [u]}. We will refer to each of these combinations as a "vowel / context block"; there are thus nine vowel / context blocks (namely, VC-[i], VC-[ɑ], VC-[u], CV-[i], CV-[ɑ], CV-[u], VCV-[i], VCV-[ɑ], and VCV-[u]). Stimuli were recorded by two native speakers of Turkish (1 m, 1 f); Turkish speakers were used because in Turkish, all four fricatives are contrastive in all nine vowel / context blocks.

The speakers produced three tokens of each stimulus type. The two most similar stimuli of each type were selected by having four phonetically trained native English speakers listen to all three tokens and select the two that they perceived as most similar in terms of duration, pitch, and vowel quality. Disagreements among the listeners (generally in cases where all three stimuli were judged to be extremely similar) were resolved by selecting the two stimuli that were calculated to be the most acoustically similar using the MFCC-based acoustic similarity algorithm in *Phonological CorpusTools* (Hall et al. 2015).

All pairwise comparisons of fricatives and tokens were included, though each member of a pair was produced by the same talker, and stimuli were blocked by context and by vowel. Pairing four fricatives results in 6 "different" pairs (i.e., [f]-[h], [f]-[s], [f]-[ʃ], [h]-[s], [h]-[ʃ], [s]-[ʃ]), each of which was represented with two tokens of each of the two fricatives, in two orders, by two speakers (= 6 x 2 x 2 x 2 x 2 = 96 "different" pairs per vowel / context block). These were paired with 96 "same" pairs per vowel / context block, produced by pairing each of the two tokens of each of the four fricatives in two orders for two talkers (4 x 2 x 2 = 16) and repeating these 16 pairs six times (16 x 6 = 96). Thus, for any given vowel / context block, there were 192 pairs, half of which were "same" pairs and half of which were "different" pairs.

Any given participant heard two of the nine vowel / context blocks, with a short break in between the blocks. Each block took approximately 15-20 minutes to complete, so listening to two blocks kept the entire experimental procedure (including consent, instructions, and a background questionnaire) to under one hour for each participant. Participants were divided randomly among nine "groups"; within each group, the two blocks heard were the same, though the order of blocks within a group was randomized. Each group's two vowel / context blocks contained the same vowel, but different syllable structures. The groups were created by doing all three pairwise comparisons of syllable structure within vowel context. That is, group 1 heard the CV-[ɑ] and VC-[ɑ] stimuli; group 2 heard CV-[ɑ] and VCV-[ɑ]; and group 3 heard VC-[ɑ] and VCV-[ɑ]. Groups 4-6 heard the same sets, but with [i] vowels, and groups 7-9 heard the same sets but with the [u] vowels. This set-up was created to ensure that each vowel / context block was heard by two different sets of participants, while minimizing the potential comparative effects of different vowel contexts within a group.

3.1.2 Participants

A total of 192 undergraduate students from the University of British Columbia participated in this study. Because students received course credit for their participation in the study, this sample of 192 students included students of many different language backgrounds. The 95 students for whom English was both their first language and their dominant language were selected for analysis here. These participants had been randomly assigned to one of the nine groups described above. Because of the initial random group assignment and the post-hoc selection of English-dominant participants, the participants were not evenly distributed among the groups. The average number of participants per group was 11, but the actual number ranged from 4 to 19. Recall, however, that every vowel / context block was heard by two groups; each vowel / context block was heard by at least 15 participants (average = 21).

The 95 participants ranged in age from 18 to 31, with a mean age of 20.6 years. There were 76 females, 18 males, and one participant who declined to state their sex.

3.1.3 Procedure

An AX discrimination task was used to probe the perceived similarity of the various fricative pairs in each context. On each trial, two stimuli were presented with an ISI of 100 ms, and listeners were given up to 2 s after the end of the second stimulus to indicate whether the members of the pair were the "same" or "different." Same pairs consisted of acoustically different repetitions of the same type by the same speaker (e.g., [isi]-[isi]); different pairs consisted of two different types of stimuli, uttered by the same speaker (e.g., [isi]-[ifi]). If no response was detected within the 2 s, a "No response detected" feedback message was displayed, followed by the next trial. If a response was detected, the participant was told whether the response was correct or incorrect, how quickly they had responded on that trial, and their overall percent correct. This feedback was provided to both alleviate boredom in the task and to encourage participants to respond as quickly and as accurately as possible. Participants indicated their response by pressing either "1" (same) or "5" (different) on a five-button response box; the buttons were also labeled with the names "same" and "different" to remind participants which was which.

Stimuli were presented using the psychological testing software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Participants were assigned to a particular group based on what day and time they came in to do the experiment, with the experimenters simply rotating through groups on a consistent basis to keep the numbers as even as possible across groups. Within a group, E-Prime randomized both the order of the two blocks and the order of the stimuli within a block.

The two blocks were always preceded by four practice trials to familiarise participants with the task. These trials consisted of stimuli produced by two different, non-Turkish speaking talkers producing pairs of same or different VCV nonsense syllables where the C was one of [d], [ɾ], or [ð]; these were tokens and consonant types not otherwise heard in the experiment.

3.2 Corpus Methods

In order to measure ProD for English-speaking participants, the Irvine Phonotactic Online Dictionary (IPHOD; Vaden et al. 2007) was used. In IPHOD, each entry's pronunciation is the American English pronunciation, taken from the Carnegie-Mellon pronouncing dictionary (Weide 1994).

ProD was measured for each of the six fricative pairs in each of the nine vowel / syllable contexts used in the experiment (i.e., [#_ɑ], [ɑ_#], [ɑ_ɑ], [#_i], [i_#], [i_i], [#_u], [u_#], [u_u]). Note that, of course, the actual stimuli were of shapes such as #CV#, i.e., with two word boundaries, so theoretically, the contexts should have been [#_ɑ#], etc. Of course, this would severely limit the actual occurrence of any of these sounds in the lexicon of English, as most of the stimuli are in fact nonsense words. Hence, we prioritized the word boundary next to the consonant, which were the target sounds of interest; thus, CV stimuli were interpreted as representing the environment [#_V], VC stimuli as the environment [V_#], and VCV stimuli as the environment [V_V].

ProD was calculated using the algorithm defined in Hall (2009) and shown above in (1), using *Phonological CorpusTools* software (Hall et al. 2015). As a reminder, ProD is calculated in terms of entropy; it ranges from 0 (complete predictability, allophony) to 1 (complete unpredictability, full contrast). When discussing the measure used, including in the statistical analysis of the results, we will refer to entropy; when discussing the phonological concept measured by this use of the entropy formula, we will refer to ProD.

ProD can be measured using either type or token frequencies of occurrence. Here, type-frequency measures were used following the results of Hall and Hume (submitted), which consistently showed that type-based measures were better predictors of behavioural results than token-frequency measures.

4. Results

The dependent variable of interest here is the reaction time in the AX discrimination experiment. The RTs were measured from stimulus onset, but the stimuli themselves were of different durations: for example, all the VCV stimuli tended to be longer than all the VC or CV stimuli. Thus, the duration of each stimulus was subtracted from the RT on each trial to give a measure of the RT from stimulus offset. Furthermore, RTs were z-score normalized within participant to correct for overall general differences across participants and ensure that the RTs were directly comparable. Finally, RTs that were greater than two standard deviations away from the mean were removed as outliers; this removed 5% of the trials. Trials where the response was inaccurate (also 5%) were also removed.

Figure 2 shows the overall results of correlating mean RTs with the entropy scores calculated in §3.2, for all "different" pairs of fricatives and across all contexts. The first striking aspect of these results is that there is a large number of pairs with an entropy of 0, and that these pairs are associated with a wide variety of RTs. Recall that an entropy of 0 indicates no uncertainty about which of the two sounds occurs. This occurs when one of the two sounds simply does not occur in that environment. For example, [h] never occurs word-finally in English, so all of the VC# contexts in which one of the members of the pair is [h] have an entropy of 0. The fact that these pairs exhibit a wide variety of RTs indicates that ProD, not surprisingly, is not the only factor affecting perceived similarity. In particular, acoustic differences, especially across pairs, presumably play a considerable role; note, for example, that the two pairs with the longest RTs are both instances of [f]/[h], which, as relatively low-noise fricatives, would be predicted to be most similar to each other.

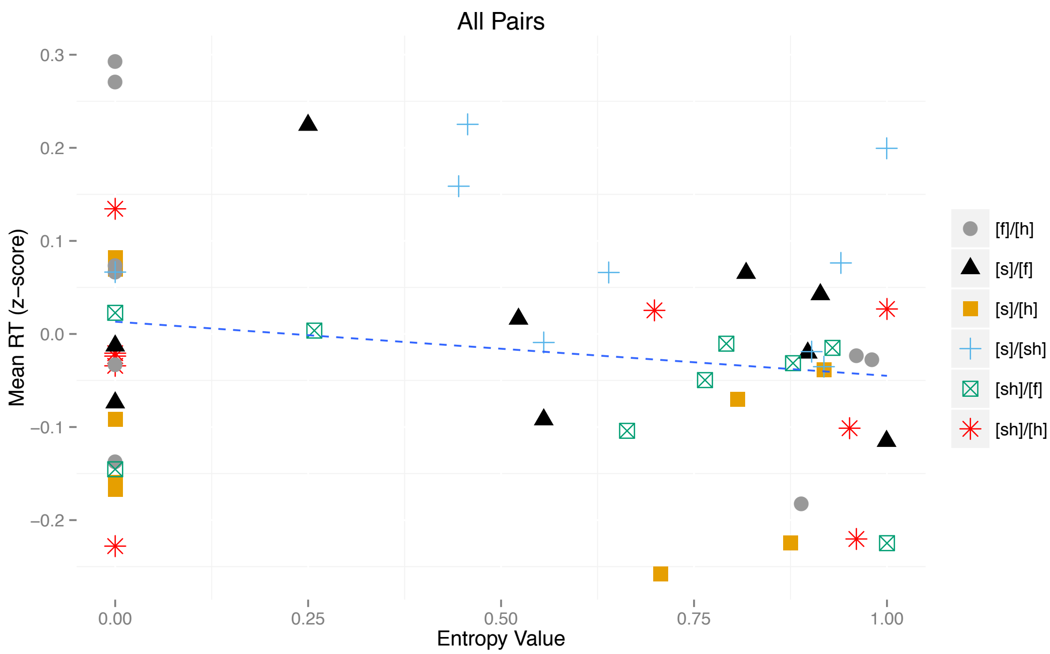

Figure 2: Overall results: RT vs. Entropy, all pairs, all contexts

That said, it is also striking that there is a clear, if relatively weak, negative correlation between RT and entropy. Perhaps more importantly, this negative correlation holds for each individual pair of sounds, as shown in Figure 3. Here, each data point again represents the mean RT for a given pair of fricatives in a given vowel / context environment, but the points are separated out by fricative pair; the shapes of the plotting characters correspond to specific environments. In this plot, error bars representing 95% confidence intervals around the mean of each point are also included. As can be seen, there is a slight but consistently negative correlation between the entropy value and the RT for every pair.

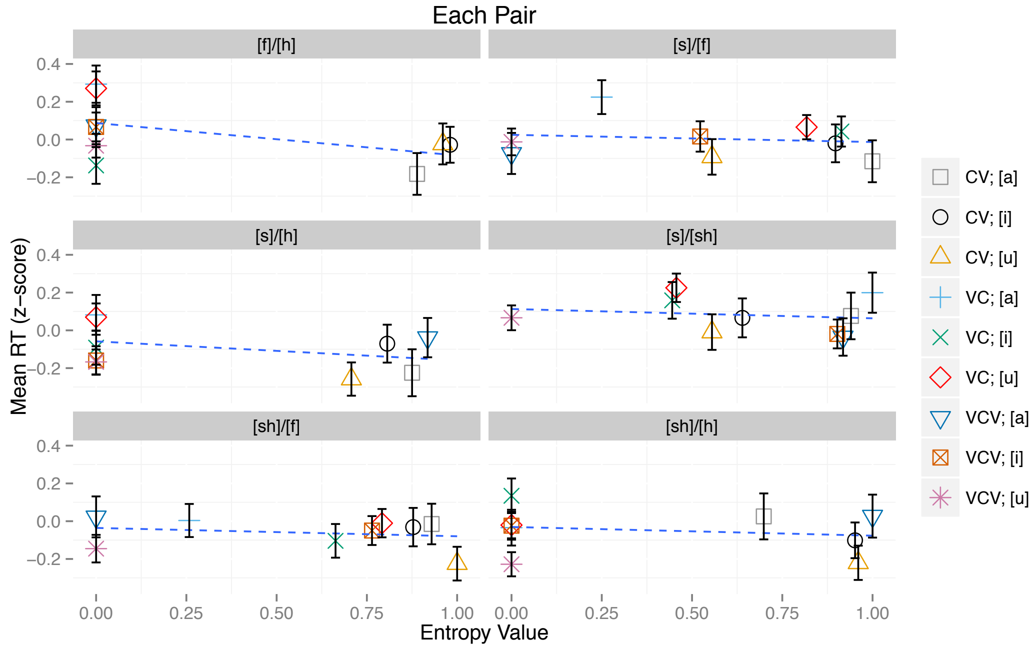

Figure 3: Reaction time as a function of entropy for individual pairs across different contexts

A linear mixed-effects regression model was run using the lme4 package (Bates et al. 2014) in R (R Core Team 2014) to test the statistical significance of the effect of entropy on RT. A model was fitted to predict normalized RTs from the fixed effects of entropy, vowel identity, pair identity, and their interactions, along with random intercepts for participant and talker, and random by-participant and by-talker slopes for the effect of entropy. The appendix shows the full results of this model.

Note that syllabic context (CV vs. VC vs. VCV) was not included in the model. Including all three of vowel, pair, and context would create perfect collinearity with entropy, because given the vowel, pair, and context identity, it is possible to predict the entropy score. Recall that one of the reasons that voiceless fricatives were chosen as the target sound was to minimise the effect of syllabic context on perceived similarity by concentrating acoustic cues in the target sounds themselves. The identity of the pair and the vowel, on the other hand, are more likely to have a direct impact on the perception of the fricatives. For example, the pair [h] / [s] is likely to be easier than [ʃ] / [s], simply because the first involves a greater acoustic difference in terms of amplitude and sibilance than the second. One might also expect the discrimination of a single pair, such as [ʃ] / [s], to be easier before [ɑ] than before [i], given the tendency for coarticulation with [i] to produce more palatal-like sounds. Thus, we felt it was more important to include pair and vowel than to include context in the model, given that all three could not be included.

This model estimated that the effect of entropy itself was negative, as expected: an increase in entropy (towards contrastiveness) results in a decrease in reaction time.

Although mixed-effects regressions don't produce standard p-values, the t-value for the effect of ProD is -4.29, which is rather large; generally speaking, t-values with an absolute value of 2 or greater are considered to be statistically significant. Furthermore, this model can be compared to one with the same random effects structure, but without the fixed effect of entropy; a likelihood ratio test comparing the two models indicates that the one with entropy is statistically significantly different than the one without it ($\chi^2(18) = 63.15$, $p < 0.001$), with the model including entropy being the better one in terms of having a higher log-likelihood score. Thus, it seems fairly clear that entropy has a significant effect in the expected direction.

There are, however, also a number of interaction effects that also appear to be significant in the model. The baseline (intercept) of the model is for the vowel [ɑ] and the pair [f] / [h]. The estimates for both [i] and [u] are also negative, indicating that overall, reaction times are predicted to be faster with these vowels than with [ɑ]. Similarly, the estimates for each of the other pairs were also all negative, indicating that reaction times are also predicted to be faster with these other pairs than for [f] / [h]. Interestingly, all of the two-way interactions between entropy and either vowel or pair identity have positive estimates, indicating that the combined estimate of entropy and vowel or pair (all of which were negative) is too large, and thus when both are working together, the effects are mitigated. In other words, if there is an estimated baseline reaction time for [ɑ] and [f] / [h], then an increase in entropy, a switch to a different vowel, or a switch to a different pair will each tend to decrease the reaction time. But if there is an increase in entropy *and* a switch to a different vowel, then the predicted decrease in reaction time from both individual switches is overestimated, and so the positive interaction effect counteracts it. Of particular note here is that in almost all cases, the pairwise interaction compensation is not large enough to push the overall reaction time to be *greater* than the original baseline; that is, it is still the case that an increase in entropy and a switch to a different vowel (or pair) generally still reduces predicted reaction time.

Furthermore, all but one of the three-way interactions have negative estimates. Thus, if there's an increase in entropy, a switch to a different vowel, *and* a switch to a different pair, then the reaction time is decreased by each of the individual switches, somewhat increased again by the pairwise interaction, and decreased again by the three-way interaction. (There are also pairwise interactions between vowel and pair that have both positive and negative effects.) Thus, the model structure is somewhat complicated, but it does appear that entropy generally has a statistically significant effect in the expected direction.

The effectiveness of the model can be examined more closely in Figure 4, which shows the original mean data points in larger, coloured points, and the mean predicted values of the best-fit model in smaller, black points. As can be seen, the predictions of the model are at least visually relatively close to the actual values, and if anything, tend to minimize differences in RTs. The Pearson correlation score (R) between the means of the predicted values and the means of original values is 0.86. Thus, the model seems to be both relatively accurate and also to crucially rely on a statistically significant influence of entropy, such that higher-entropy stimuli are discriminated more quickly.

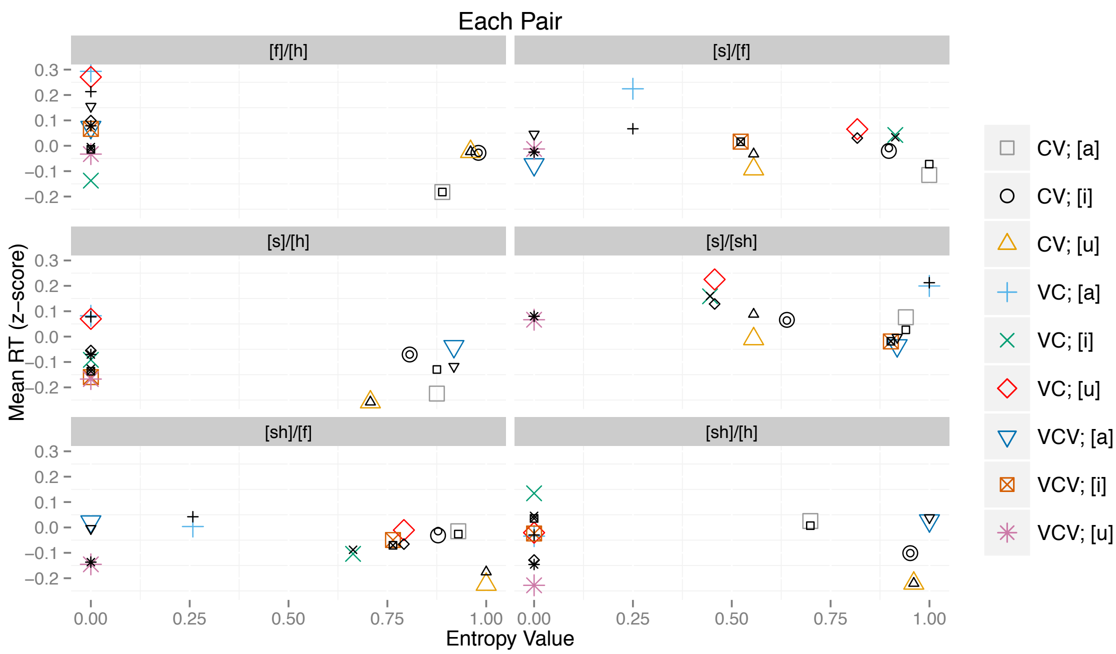

Figure 4: Comparing actual and predicted RTs for each pair of fricatives, by fricative pair, vowel, and context

5. Discussion

The results of this experiment indicate that predictability of distribution does have an effect on the perceived similarity of sounds. Specifically, pairs of sounds are perceived as more similar when they are more predictably distributed.[1] This finding is particularly notable in two regards. First, it means that multiple *degrees* of contrast, at least insofar as predictability of distribution is a measure of contrast, are relevant; language users seem to be sensitive to different amounts of predictability of distribution. Second, this effect is found to hold even *within single pairs of sounds* that as a general proposition are all considered to be categorically contrastive in a language. That is, all of the pairs of fricatives in this experiment consisted of members of the contrastive inventory of English. Yet within each pair, there is a tendency for positions in which the pair of sounds is more overlapping in their distribution to be those in which the discrimination of the sounds is fastest.

These findings can be interpreted in a broader communicative framework such as that outlined in Hume et al. (2014). In that framework, phonological patterns are seen as emergent from the natural forces that affect the communication of meaningful units of

[1] Note that in the original published version of this paper, this conclusion was mis-stated (though was correctly presented throughout the rest of the discussion and conclusion section). It has been corrected as of May 2019.

language (e.g., words). There are competing pressures for such communication to be both as accurate and efficient as possible, which leads to the signal that is used for communication to be manipulated in ways that enhance redundancy in places of greater uncertainty and reduce redundancy in places of greater certainty.

When thinking about the results of the current study, consider the fact that places of greater entropy (lower predictability of distribution; higher contrastiveness) are places where the acoustic signal, by definition, is more important in terms of distinguishing sounds. For instance, in a context like [#_il], it is difficult for a listener to guess whether the initial sound was [s] or [h], given the existence of both the words *seal* and *heal*. Hence, the information conveyed by the acoustic signal itself is what is important for successful message communication. On the other hand, in a context like [li_#], it is quite easy to guess that the final sound must have been [s] and not [h], given the word *lease* and the general lack of any [h]-final words in English. In this context, listeners can rely on information other than the acoustic signal to make a perfect prediction.

Thus, one might imagine that listeners develop the habit of paying more attention to acoustic cues to differentiate signals in positions where such cues are most necessary, and devote fewer resources to this task in positions where they are not. In an experimental task like the current one, then, listeners should be faster at differentiating sounds when they happen to occur in the former type of position, and slower in the latter—which of course is entirely consistent with the current results.

In real life, of course, the situation is much more complicated. In particular, there are at least two aspects of the current experiment that would mitigate any such effects. First, there was no lexical information in the current experiment at all, which presumably means that, to the extent they can, participants would prioritize listening strategies that target the acoustic signal as a whole rather than relying on lexical knowledge of the distribution of sounds. Second, from trial to trial, listeners were being asked to distinguish six different pairs of fricatives. This means, for example, that even if it would be possible to perfectly predict that [s] and not [h] occurs in the context [li_#] (or even [i_#], as the stimulus was here), there are still the competing possibilities of [f] and [ʃ]. Thus, it is of course true that listeners are perfectly capable of listening to the acoustic signal itself and gleaning information from it. But crucially, they are faster at doing so when there are fewer choices overall, because of the phonological distribution of the sounds in their language. It would therefore be fruitful in the future to examine the relation of reaction time to some measure of overall uncertainty about which sound occurs in a given environment, given the entire phonemic inventory and the entire lexicon of a language, instead of just examining it pairwise as presented here.

A third factor that indubitably affects the results seen here is the fact that predictability of distribution is not the only non-acoustic factor that is likely relevant. For example, the functional load of the pairs in question may matter. As explained in Hall (2009), predictability of distribution and functional load are related but not identical measures. Two sounds may occur equally often in some environment (e.g., there are minimal pairs such that for every instance of A in the environment there is also a B), which would give them a low predictability of distribution, but that does not reveal anything about how often the environment itself occurs (e.g., there may be a large or small number of minimal pairs hinging on A and B in the lexicon overall). Thus, another fruitful line of inquiry will be to investigate how the various relevant factors interact with each other to determine perceived similarity (see also Hall and Hume, submitted).

6. Conclusion

In conclusion, this paper has shown clear evidence that the degree of predictability with which two sounds are distributed in a language has an effect on their perceived similarity, with sounds that are more predictably distributed being perceived as more similar than even the same sounds in environments where the distribution is less predictable. These results can be interpreted within a larger framework of communication, in which it is beneficial for listeners to pay more attention to the acoustic signal when that is the only cue to a contrast than when the identity of a sound is derivable from the surrounding phonetic, lexical, syntactic, or social context.

References

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. Available from: http://CRAN.R-project.org/package=lme4.

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, and Keith Johnson. 2008. The influence of allophony vs. contrast on perception: The case of Spanish and English. In *Contrast in phonology: Perception and acquisition*, ed. Peter Avery, B. Elan Dresher and Keren Rice, 145-171. Berlin: Mouton.

Hall, Kathleen Currie. 2009. *A probabilistic model of phonological relationships from contrast to allophony*. Doctoral dissertation, Linguistics, The Ohio State University, Columbus, OH.

Hall, Kathleen Currie. 2013a. A typology of intermediate phonological relationships. *The Linguistic Review* 30:215-275.

Hall, Kathleen Currie. 2013b. Documenting phonological change: A comparison of two Japanese phonemic splits. In *Proceedings of the 2013 Annual Meeting of the Canadian Linguistic Association*, ed. Shan Luo. Toronto: Canadian Linguistic Association.

Hall, Kathleen Currie, Blake Allen, Michael Fry, Scott Mackie, and Michael McAuliffe. (2015). Phonological CorpusTools, Version 1.1.1. [Computer program]. Available from https://github.com/PhonologicalCorpusTools/CorpusTools/releases.

Hall, Kathleen Currie, and Elizabeth Hume. 2014. Predicting perceptual similarity of French vowels: The influence of phonology, phonetics, and frequency. Paper presented at 2nd Annual Meeting on Phonology, 21 September, at MIT.

Hall, Kathleen Currie, and Elizabeth Hume. submitted. Modeling perceived similarity: The influence of phonetics, phonology and frequency on the perception of French vowels. *(Manuscript, submitted August 2015)*.

Hume, Elizabeth, Kathleen Currie Hall, and Andrew Wedel. 2014. Strategic responses to uncertainty: Strong and weak sound patterns. In *Proceedings of the 5th International Conference on Phonology and Morphology*. Korea.

Hume, Elizabeth, and Keith Johnson. 2003. The impact of partial phonological contrast on speech perception. In *Proceedings of the Fifteenth International Congress of Phonetic Sciences*, ed. Maria-Josep Solé, Daniel Recansens and Joaquin Romero, 2385-2388. Barcelona / Australia: Causal Productions.

Kazanina, Nina, Colin Phillips, and William J. Idsardi. 2006. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America* 103:11381-11386.

Psychology Software Tools, Inc. [E-Prime 2.0]. 2012. Available from: http://www.pstnet.com.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical

Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication*. 1998 ed. Urbana-Champaign: University of Illinois Press.

Steriade, Donca. 2007. Contrast. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 139-157. Cambridge: Cambridge University Press.

Vaden, K. I., H. R. Halpin and G. S. Hickok. 2009. Irvine Phonotactic Online Dictionary, Version 2.0. [Data File.] Available from: http://www.iphod.com.

Weide, Robert L. 1994. CMU Pronouncing Dictionary. Available from: http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Appendix: Full Linear Mixed-Effects Regression Model

RT ~ Entropy * Vowel * Pair + (1 + Entropy | Subject) + (1 + Entropy | Talker)

Fixed effects	Estimate	Standard Error	t-value
Intercept (V = [ɑ], pair = [f]/[h])	0.19	0.05	3.51
Entropy	-0.41	0.10	-4.29
vowel = [i]	-0.20	0.07	-2.83
vowel = [u]	-0.09	0.07	-1.39
pair = [s] / [f]	-0.11	0.05	-1.95
pair = [s] / [h]	-0.14	0.06	-2.19
pair = [s] / [ʃ]	-2.83	0.85	-3.32
pair = [ʃ] / [f]	-0.15	0.05	-2.84
pair = [ʃ] / [h]	-0.25	0.06	-4.12
Entropy : [i]	0.42	0.12	3.36
Entropy : [u]	0.36	0.12	3.09
Entropy : [s] / [f]	0.27	0.10	2.68
Entropy : [s] / [h]	0.22	0.10	2.14
Entropy : [s] / [ʃ]	3.27	0.90	3.65
Entropy : [ʃ] / [f]	0.36	0.10	3.49
Entropy : [ʃ] / [h]	0.50	0.10	4.91
[i] : [s] / [f]	0.17	0.13	1.33
[u] : [s] / [f]	0.00	0.07	0.02
[i] : [s] / [h]	0.02	0.08	0.24
[u] : [s] / [h]	-0.01	0.07	-0.17
[i] : [s] / [ʃ]	3.13	0.86	3.65
[u] : [s] / [ʃ]	2.83	0.85	3.32
[i] : [ʃ] / [f]	-0.31	0.28	-1.11
[u] : [ʃ] / [f]	-0.06	0.07	-0.90
[i] : [ʃ] / [h]	0.30	0.08	3.96
[u] : [ʃ] / [h]	0.03	0.07	0.36
Entropy : [i] : [s] / [f]	-0.33	0.19	-1.75
Entropy : [u] : [s] / [f]	-0.21	0.13	-1.61
Entropy : [i] : [s] / [h]	-0.13	0.14	-0.90
Entropy : [u] : [s] / [h]	-0.39	0.13	-2.92
Entropy : [i] : [s] / [ʃ]	-3.60	0.91	-3.96
Entropy : [u] : [s] / [ʃ]	-3.19	0.90	-3.53
Entropy : [i] : [ʃ] / [f]	0.18	0.37	0.49
Entropy : [u] : [ʃ] / [f]	-0.29	0.13	-2.25
Entropy : [i] : [ʃ] / [h]	-0.62	0.13	-4.56
Entropy : [u] : [ʃ] / [h]	-0.46	0.12	-3.71