Gradient Perceptual Facilitation from Phonotactic Knowledge

Hahn Koo and Jennifer Cole
University of Illinois at Urbana-Champaign

1 Introduction

In classical generative phonology, phonotactic constraints define the set of possible sound patterns in a given language by restricting the distribution of sounds, e.g. in certain structural positions or next to certain other sounds. Speakers' knowledge of the phonotactic constraints in their language is manifested in various linguistic tasks they perform. Speakers will judge a sound pattern to be grammatical if it is phonotactically *legal* and ungrammatical if it is phonotactically *illegal*. Their knowledge is also manifested in their performance in on-line speech processing tasks, such as speech perception and production. For example, Brown and Hildum (1956) showed that adult English speakers perceive non-words beginning with phonotactically legal onset clusters (e.g. /pr/) more accurately than non-words beginning with illegal onset clusters (e.g. /zdr/). One obvious source of their sensitivity to phonotactic legality of sound patterns is the distribution of sound patterns in the language. They have encountered one or more examples of phonotactically legal sound patterns but they have not encountered any examples of phonotactically illegal sound patterns in the language.

Under the classical view, the focus was on the categorical distinction between phonotactically legal vs. illegal sound patterns. More recent studies make a finer distinction and focus on the *gradience* of phonotactic constraints. One commonly made argument is that a better description of phonotactic constraints should include not only what are legal and illegal sound patterns in the language but also the frequency or probability with which legal sound patterns occur in the language. Evidence in support of the argument comes from studies that relate gradience in speakers' task performance with lexical statistics. Speakers judge non-words consisting of more frequent sound patterns to be more acceptable than non-words consisting of less frequent sound patterns (Coleman and Pierrehumbert, 1997; Treiman et al., 2000). Speakers perceive non-words with high-frequency syllables more quickly than non-words with low frequency syllables (Vitevitch et al., 1997). Similar to the idea that presence vs. absence of sound patterns in the language results in speakers' sensitivity to the phonotactic legality of sound patterns, the gradience in performance is often attributed to difference in the amount of statistical evidence for different sound patterns. That is, speakers become more familiar with the phonotactic constraint governing a

* This work was supported by NIH Grant HD 44458. We thank Gary Dell, Cynthia Fisher, Jill Warker, and Ye Xu for helpful comments and suggestions.

sound pattern as they encounter more examples and this gradience in familiarity results in gradience in performance.

However, gradience in phonotactic knowledge may also stem from factors related to the formal aspects of the phonotactic constraints or cognitive biases governing learning and generalization. Distribution of some sound patterns may go unnoticed, or take more statistical evidence to learn because they are *unnatural* or formally more *complex* (e.g. Pycha et al., 2003; Wilson, 2003; Newport and Aslin, 2004; Peperkamp et al., 2005). Speakers may be biased to generalize their knowledge to novel sound patterns that are more similar to the examples from which the knowledge was acquired (Wilson, 2006). As a consequence, despite equal amount of statistical evidence, such factors may still cause phonotactic knowledge to be gradient and result in gradience in speakers' performance which manifests their knowledge. For example, Warker and Dell (2006) compared learnability of two types of phonotactic constraints: first-order constraints which restrict syllable position of consonants vs. second-order constraints which restrict syllable position of consonants depending on the identity of the adjacent vowel. The results show that it requires more statistical evidence for the second-order phonotactic constraints than the first-order phonotactic constraints to have a sizable effect on production. Another way to interpret the result is that despite equal amount of statistical evidence, speakers are less familiar with the phonotactic constraint if the constraint requires more statistical evidence to learn, and the effect of phonotactic knowledge on production is weaker.

The aforementioned studies show that gradience in phonotactic knowledge can explain much of gradience that we observe in speakers' linguistic behavior. However, as we have already identified at least three factors (lexical statistics, learnability, and substantive bias in generalization) that result in gradience in phonotactic knowledge, explaining gradience in behavior is a complex task and it requires understanding of what the underlying factors are and how they interact. While not denying the role of gradience in phonotactic knowledge, we propose to extend the range of factors to consider by showing how a task-related factor can result in gradience in behavior. Specifically, we focus on the task of spoken non-word recognition and argue that the degree of perceptual confusion inherent in the task input is another source of gradience in performance. Our argument is elaborated in the perceptual facilitation hypothesis in section 2. Supporting evidence comes from four experiments using the artificial grammar learning paradigm described in section 3. Section 4 suggests how two sources of gradience in performance, gradience in phonotactic knowledge and perceptual confusability, can be integrated with a connectionist model. Section 5 concludes the paper.

2 The perceptual facilitation hypothesis

Studies on the effect of phonotactic knowledge on perception show that speakers' phonotactic knowledge facilitates perception of phonotactically legal sound patterns. Phonotactically legal sound patterns are perceived more accurately

(Brown and Hildum, 1956) and more quickly (Vitevitch et al., 1997) than phonotactically illegal sound patterns. Upon hearing ambiguous speech sounds, listeners perceive them as the sounds which are phonotactically legal in the given context rather than as sounds that are phonotactically illegal (Massaro and Cohen, 1983; Moreton and Amano, 1999). Although the precise mechanism by which phonotactic knowledge affects perception is not well understood, previous works (e.g. Luce et al., 2000) succeed in modeling the effect with computational models for the perception of novel words.

In models of spoken word recognition, both for humans and machines, a phonological unit is selected as the output of the recognition process if it outscores other candidate units given the perceptual input and the knowledge of sound patterns encoded in the model. In connectionist models of human speech perception, the chosen unit is the one with the highest activation (e.g. McClelland and Elman, 1986; Norris, 1994). In statistical models of automatic speech recognition, such as Hidden Markov Models, it is the state or the state sequence with the highest posterior probability (e.g. Rabiner and Juang, 1993; Jelinek, 1997). We can describe the state where there are multiple candidates matching the perceptual input as a state of perceptual confusion. Ideally, the top score should be given only to a single phonological unit among other candidates. Perceptual confusion is reduced by increasing the score of a unit while decreasing the scores of the other candidates. Phonotactic knowledge reduces confusion by adding more scores to units whose combination is phonotactically legal or more frequent and/or deducting scores from units whose combination is illegal or less frequent.

We may consider two recognition scenarios that differ only in the acoustic similarity of competing recognition candidates. In the first scenario, suppose that the acoustic signal corresponding to the current phonological input unit was very distinct to start with (i.e., was not highly similar to acoustic signals for other phonological units), and that the output unit representing a phonotactically legal combinations had a very high score (i.e., was a close match to the input acoustic signal) while the output unit representing a phonotactically illegal combination had a very low score (i.e., was not a close match to the input acoustic unit). In this case, based only on the match between the acoustic signal and the output units, the high score of the legal output unit may be sufficient on its own to select the legal unit as the recognition output. In this scenario, there may be little room for phonotactic knowledge to influence the recognition process through increasing the scores of phonotactically legal units. In other words, an unambiguous acoustic signal may match very well to a single output unit, resulting in a "ceiling effect" where further effects due to phonotactic knowledge have no further impact. In this scenario, the effect of phonotactic knowledge on perception would appear relatively weak.

In the second scenario, suppose the input acoustic signal were ambiguous and its match to an output unit harder to discern, with roughly equal scores assigned to the units of both phonotactically legal and illegal sound patterns. In this case, phonotactic knowledge could have a significant further impact on the recognition

outcome, by increasing the scores of phonotactically legal units and decreasing the scores of phonotactically illegal units. As a consequence, the effect of phonotactic knowledge on perception would appear relatively strong.

The score assigned to all output units based on their match to the input acoustic signal is a measure of the perceptual ambiguity of the input signal given the available phonological units, as defined by the sound inventory of the language. The perceptual facilitation hypothesis states that the size of the effect of phonotactic knowledge on perception depends on this measure of perceptual confusability.

> **Perceptual Facilitation Hypothesis:** The size of the effect of phonotactic knowledge on spoken (non-)word recognition varies as a function of the confusability of the signal as an instance of the phonotactically legal or illegal sound pattern. The greater the confusability of the signal with respect to identification as a legal or illegal sound pattern, the greater the size of the effect of phonotactic knowledge on perception.

As in the aforementioned studies on phonotactic constraints, the effect of phonotactic knowledge is measured in this study by comparing the speakers' performance on phonotactically legal sound patterns with their performance on phonotactically illegal sound patterns in tasks that involve phonological processing. The following section describes how we tested the hypothesis with four experiments using the artificial grammar learning paradigm.

3 Experiments

The experiments presented here test the perceptual facilitation hypothesis in an artificial grammar learning paradigm, by comparing the facilitation effects due to phonotactic learning for the perception of sound patterns under two conditions of confusability. In one condition, the perception stimuli are highly confusable between phonotactically legal and illegal sound patterns, while in the other condition, the stimuli are not very confusable between legal and illegal sound patterns. Subjects learn the phonotactic constraint of the artificial language implicitly, through exposure to the (nonsense) words in the language. The words of the artificial language are well-formed with respect to English phonotactics, and the new phonotactic constraint imposes a further restriction on sound patterns. Subjects are familiarized with a set of words, and then tested to see if their perceptual performance on new words is facilitated by implicitly learning the phonotactic constraint of the language. Some of the new words are legal according to the constraint, and some are illegal by that constraint. The degree of perceptual facilitation is measured by the difference in performance between legal and illegal words.

Subjects are divided into two groups, with one group learning a phonotactic constraint that restricts the distribution of acoustically similar stimuli, and the

other group learning a constraint that restricts the distribution of sounds that are acoustically more distinct. By the perceptual facilitation hypothesis, the constraint on acoustically similar sounds has a greater potential to reduce perceptual confusion. Importantly, both constraints have the same formal structure and the stimuli in both experiments presents the same number of novel nonsense words instantiating the constraint. The prediction is that the degree of perceptual facilitation due to phonotactic learning will be greater in the experiment whose constraint has more potential to reduce confusion.

The constraints adopted in our experiments are all co-occurrence restrictions on non-adjacent sounds that are separated by one intervening sound. The constraint in Experiment 1 restricts co-occurrence of the liquids /l/ and /r/ such that repetition of either /l/ or /r/ is *favored* in the sense that subjects encounter many words instantiating the repetition during the familiarization phase. On the other hand, co-occurrence of /l/ and /r/ is *disfavored* in the sense that subjects never encounter words with the co-occurrence pattern during the familiarization phase. For example, the constraint favors words such as /sa.la.la/ or /sa.ra.ra/, while it disfavors words such as /sa.la.ra/ or /sa.ra.la/. The constraint in Experiment 2 restricts co-occurrence of the two consonants /l/ and /m/ in the same way. For example, the constraint favors words such as /sa.la.la/ or /sa.ma.ma/, while it disfavors words such as /sa.la.ma/ or /sa.ma.la/. Experiments 3 and 4 assume a similar constraint that restricts the co-occurrence of the two high vowels /i/ and /u/ in the same way. For example, the constraint favors words such as /sa.ki.si/ or /sa.ku.su/, while it disfavors words such as /sa.ki.su/ or /sa.ku.si/.

The amount of statistical evidence in support of the experimental constraint depends on how frequently subjects encounter the nonsense words that instantiate the constraint. For example, in Experiment 1, the number of words such as /sa.la.la/ or /sa.ra.ra/ that a subject encounters during the experiment session would be the amount of statistical evidence. We held the amount of statistical evidence fixed across experiments by holding the number of instantiating words constant. The amount of potential confusion that can be reduced by the phonotactic knowledge is measured by how confusable the two constrained phonemes are. For example, the amount of confusion in Experiment 1 would depend on the confusability between /l/ and /r/, whereas the amount of confusion in Experiment 2 would depend on the confusability between /l/ and /m/. The purpose of the experiments is to test the prediction of the perceptual facilitation hypothesis such that the degree of perceptual facilitation will be greater in the experiment whose constraint restricts co-occurrence of a more confusable phoneme pair, despite equal amount of statistical evidence for the constraint in each experiment.

3.1 Phoneme confusability

The constrained phonemes were consonants in Experiments 1 and 2, while the constrained phonemes were vowels in Experiments 3 and 4. The difference in each pair of experiments was in the perceptual confusability of phonemes whose

co-occurrence was restricted; the constrained phonemes were more confusable in one experiment than the other. The constrained consonants were /l/ vs. /r/ in Experiment 1 and /l/ vs. /m/ in Experiment 2. The pair /l/ vs. /r/ is more confusable to each other than /l/ vs. /m/. The constrained vowels were /i/ vs. /u/ in both Experiments 3 and 4. However, test words were presented with white noise in the background with the signal-to-noise-ratio (SNR) at +5dB in Experiment 4, rendering the vowels more confusable.

Our estimation of phoneme confusability was based on the confusion matrices in Luce (1986) and the theoretical measure of phonological similarity proposed in Frisch et al. (1997). The confusion matrices in Luce (1986) summarize how often adult English subjects perceived a phoneme X as a phoneme Y, where a phoneme is one of forty American English phonemes. Subjects listened to either CV or VC syllables with background noise at three different noise levels (SNR=+15dB, +5dB, -5dB) and reported the phonemes they perceived. Frisch et al. (1997) measures the phonological similarity between two phonemes in terms of natural classes. Specifically, the phonological similarity between X and Y is the ratio of the number of natural classes that include both X and Y to the number of natural classes that include either X or Y.

According to the confusion matrix, /l/ and /r/ were confused 8.0% of the time when SNR=+15dB, while /l/ and /m/ were confused 1.7% of the time at the same noise level. According to Frisch et al. (1997), the phonological similarity between /l/ and /r/ is 0.5407 while the similarity between /l/ and /m/ is 0.1579 when the phonemes were specified in terms of the features in the Sound Pattern of English (Chomsky and Halle, 1968) [1] . Therefore, the phonotactically constrained consonants in Experiment 1 were more confusable to each other than the constrained consonants in Experiment 2. According to the confusion matrix, the vowels /i/ and /u/ were confused 0.4% of the time when SNR=+15dB, while they were confused 7.0% of the time when SNR=+5dB[2]. Therefore, the vowels were more confusable in Experiment 4 than in Experiment 3.

3.2 Methods

3.2.1 Subjects

Fifteen adult native speakers of English participated in each of the four experiments. Subjects were students at the University of Illinois at Urbana-Champaign and received course credit for compensation.

---

[1] Similarity was computed using the segmental similarity calculator in Albright (2003).

[2] We manipulated the noise level instead of choosing two different vowel-pairs of different confusability/similarity because no vowel pair in the confusion matrix in Luce (1986) satisfied the following two conditions: (1) both vowels must be allowed to end a syllable in English, and (2) the pair must be significantly more confusable than the /i/-/u/ pair.

3.2.2 Materials

Stimuli were tri-syllabic nonsense words of the form C_1V_1. C_2V_2. C_3V_3 produced by a male native speaker of English. The first syllable was fixed to either /sa/ or /ke/. The constrained positions were C_2 and C_3 for the consonant experiments (Experiments 1 and 2), and V_2 and V_3 for the vowel experiments (Experiments 3 and 4). For the consonant experiments, the constrained positions were filled by {/s/, /k/, /l/, /r/} in Experiment 1 and {/s/, /k/, /l/, /m/} in Experiment 2. The remaining two vowel positions were filled by {/a/, /e/, /i/, /u/}. For the vowel experiments, the constrained positions were filled by {/a/, /e/, /i/, /u/} and the remaining two consonant positions were filled by {/s/, /k/, /l/, /r/} in both Experiments 3 and 4.

From the set of 512 nonsense words that satisfy the above constraint, 92 words of four different types were pseudo-randomly chosen for each experiment session: 16 study, 18 legal, 18 illegal, and 40 filler words. Study words and legal words instantiated the sound patterns favored by the experimental constraint. Study words were presented to subjects multiple times to familiarize the subjects with the constraint, while legal words were presented once to test how familiarization with the constraint facilitates subjects' perception. On the other hand, illegal words were instances of sound patterns disfavored by the constraint. Filler words were distracter items that were neither favored nor disfavored by the constraint. For example, the constraint in Experiment 1 was that repetition of /l/ or /r/ is favored while /l/ and /r/ cannot co-occur. Therefore, words such as /sa.la.la/ or /sa.ra.ra/ would be either study or legal words, whereas words such as /sa.la.ra/ or /sa.ra.la/ would be illegal words. Words such as /sa.la.ka/ or /sa.ka.si/ would be filler words. The words were distributed across five blocks as summarized in Table 1.

	Practice	Block 1	Block 2	Block 3	Block 4	Block 5	Total
Study		16	16	16	16	16	80
Legal				6	6	6	18
Illegal				6	6	6	18
Filler	2	8	8	8	8	8	42
Total	2	24	24	36	36	36	158

Table 1: Distribution of words in an experiment session.

To hold the statistical distribution of sound patterns the same between experiments, words were first chosen for Experiments 1 and 3 and then modified for Experiments 2 and 4, respectively. For Experiment 2, all instances of /r/ in Experiment 1 were replaced by /m/. For example, /sa.la.ra/ in Experiment 1 would be /sa.la.ma/ in Experiment 2. For Experiment 4, the set of words was identical to that in Experiment 3 except that white noise at SNR=+5dB was added to words from blocks 3 through 5.

3.2.3 Procedures

A subject was seated in front of a computer monitor placed in a sound-attenuated booth for each session. For each trial, the subject listened to a stimulus word through headphones and was told to repeat it as quickly and accurately as possible into the microphone placed in front. Response latency for each trial was recorded using a serial response box featuring an integrated voice key. Responses during the session were recorded using a DAT recorder. Stimuli were presented and response latencies were collected using the E-prime software.

3.2.4 Scoring

We measured latency and accuracy of response for each trial. Latency was measured in milliseconds from stimulus offset to response onset. The investigator transcribed each response based on the DAT recordings. A response was marked an error if the transcribed phoneme sequence differed from the phoneme sequence of the stimulus. Latency was averaged per stimulus type (legal vs. illegal) with the following excluded: errors, machine failures, responses that occurred too early, and outliers. Machine failures were responses which the microphone failed to detect in the first attempt. Responses that occurred before the beginning of the final syllable of the stimulus were considered too early. Latencies 2.5 standard deviations away from the mean latency averaged over the remaining responses were considered outliers. Accuracy was averaged per stimulus type with the machine failures excluded.

3.3 Results

The results were consistent with the predictions from the perceptual facilitation hypothesis. In the consonant experiments, subjects perceived the legal words significantly faster than the illegal words when the constrained consonants were more confusable (Experiment 1; $F(1,14)=6.278$, $p=0.025$), whereas the difference in latency between the two failed to reach significance when the constrained consonants were less confusable (Experiment 2; $F(1,14)=0.206$, $p=0.657$). In the vowel experiments, subjects perceived the legal words significantly more accurately[3] than the illegal words when the constrained vowels were more confusable due to background noise (Experiment 4; $F(1,14)=31.381$, $p<0.001$), whereas difference in accuracy failed to reach significance when the constrained vowels were less confusable (Experiment 3; $F(1,14)=0.072$, $p=0.792$). Figure 1 summarizes the mean latencies measured in milliseconds from the two consonant

---

[3] We compared accuracy instead of latency because adding background noise in Experiment 4 caused subjects to make too many errors. Mean error rate was 0.074 in the first two blocks where stimuli were presented without noise, while it was 0.660 in the three test blocks where noise was added. Out of 45 test blocks, three blocks for each of 15 subjects, 23 blocks had an error-rate of 1.0 for either legal (six blocks) or illegal words (17 blocks), which made it meaningless to compare latency between legal and illegal words. In addition, the difference between legal and illegal words failed to reach significance in Experiment 3 even when the performance was measured in latency ($F(1,14)=0.164$, $p=0.691$).

experiments. Figure 2 summarizes the mean accuracies from the two vowel experiments.

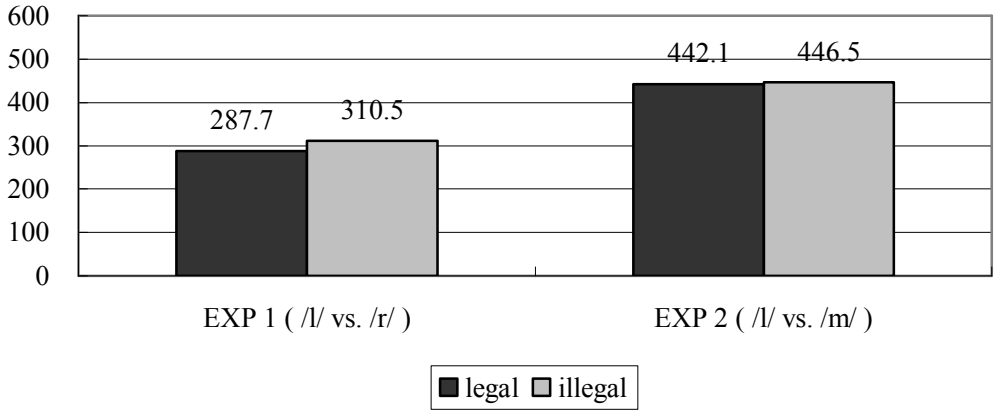

Figure 1: Mean latency per stimulus type in consonant experiments.

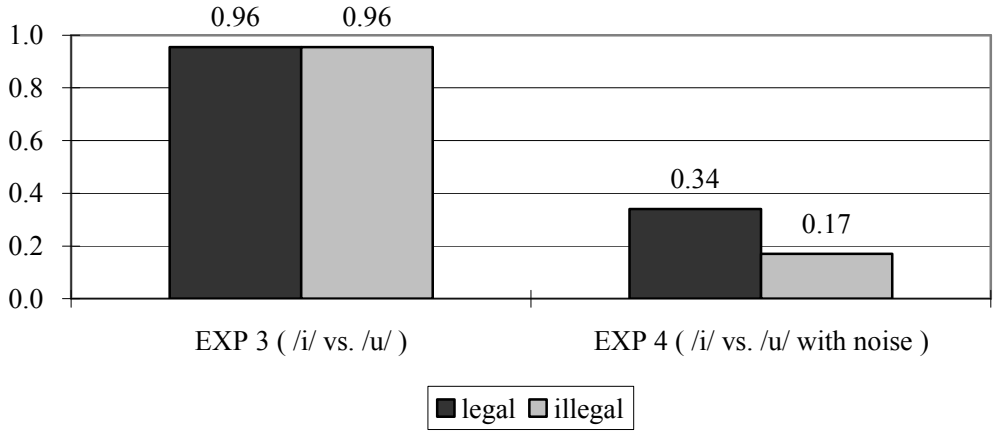

Figure 2: Mean accuracy per stimulus type in vowel experiments.

The results suggest that gradience in performance is not necessarily due to gradience in phonotactic knowledge. Neither difference in statistical evidence in support of phonotactic constraints nor difference in learnability of phonotactic constraints caused gradience in performance in the four experiments. Rather, the gradience in performance observed in our experiments appears to be the consequence of gradience inherent in the nature of the input to the perceptual task which subjects performed. Specifically, when the co-occurrence of a more confusable pair of phonemes was restricted, the effect on perception was greater. The difference was observed while the distribution of types and tokens that pertain to the phonotactic constraint was exactly the same between the compared

experiments (Experiment 1 vs. 2, and Experiment 3 vs. 4), and learnability of phonotactic constraint, which is another source of gradience in the acquired phonotactic knowledge, could not have been different (Experiment 3 vs. 4).

4 Integrating gradience with a connectionist model

We introduce a connectionist model of speech perception that integrates two sources of gradience and simulates the findings of our experiments. The two sources of gradience integrated in our model are gradience in phonotactic knowledge and perceptual confusability inherent in the task input. The task of the model is to predict how "well" an input non-word is perceived by a hypothetical speaker with the knowledge of phonotactic constraints. The effect of phonotactic knowledge on perception is measured by comparing the model's estimate of how well legal words are perceived against its estimate of how well illegal words are perceived. Assuming the acquired knowledge is equally salient between different phonotactic constraints, the model predicts that the effect will be greater if the constrained phonemes are more confusable, as our experiments suggest.

4.1 Structure and processing dynamics

The model consists of two layers of artificial neurons. The two layers are functionally similar to the input level and the pattern level in the PARSYN model of spoken word recognition (Luce et al., 2000). Each neuron in the input layer represents a different phoneme in the given language. For example, if the assumed language is English, there would be forty neurons in this layer, representing the forty phonemes in English. The neurons in the pattern layer are organized into a sequence of blocks. Each block represents a particular segment position and each neuron in the block represents a position-specific phoneme. For example, a neuron labeled /k/ in the third block from the left is relevant to perceiving /k/ as the third phoneme of the input word. The input layer and the pattern layer are fully connected, allowing activation to flow from the input layer to the pattern layer. Each block of the pattern layer is fully connected with the preceding blocks, allowing activation to flow from the preceding blocks[4]. The connections are weighted so that the amount of activation that flows along the connection is modulated. The structure of the model is illustrated in Figure 3.

We assume that the input signal is segmented and presented to the model one phoneme at a time. The input neurons respond to the presented phoneme, say X, by being activated by relatively how often listeners confuse X as the phoneme represented by the input neuron. For example, suppose that the presented phoneme were /k/ and that a confusion matrix of this language suggested that listeners confuse /k/ as /t/ twenty percent of the time whereas they correctly

[4] Here, we follow the standard assumption made in the automatic speech recognition literature that phonemes are recognized from left to right. However, the right-to-left contextual effect can be easily captured by allowing the activation to flow right to left as well.

perceive /k/ sixty percent of the time. When /k/ is presented to the model, the input neuron labeled /k/ will be activated by 0.6, and the one labeled /t/ will be activated by 0.2.

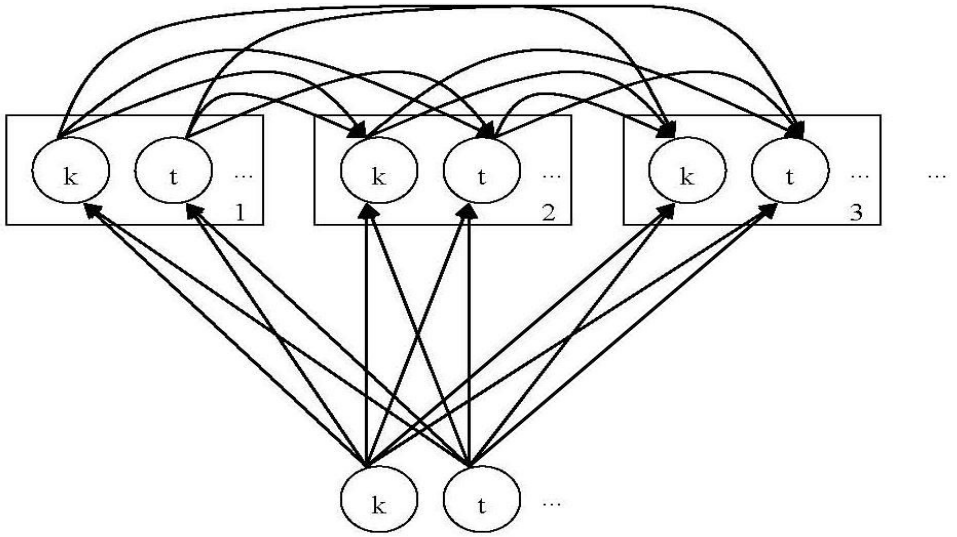

Figure 3: Structure of the connectionist model.

The activation from the input layer then spreads along the connection to activate the pattern neurons in the block pertaining to the given time-frame. This is to capture how listeners perceive the same phoneme in different word positions. At the same time, activations stored in the preceding blocks also spread into the current block. This is to capture the effect of previous phonemes on recognizing the current phoneme. Activation of each neuron in the current block is equal to the result of applying the sigmoid function to the dot product of the activation vector and the weight vector. The activation patterns in the input layer and the preceding blocks constitute the activation vector. The weights of the connections leading from the input layer and the preceding blocks to the current block constitute the weight vector. The resulting activation pattern in the current block is stored for the successive time-frames.

How well the input word would be correctly perceived by a listener, henceforth the *perception score*, can be estimated from the activation pattern in the pattern layer after all phonemes of the input word have been processed. We compute the Luce-ratio (Luce, 1959) of the activation of the neuron representing the correct phoneme in each block and then multiply the Luce-ratios over all blocks. For example, if the input word were /kæt/, the perception score would be computed as follows. We identify by how much the neuron labeled /k/ in the first block is activated, and divide it by the sum of activation over all neurons in the first block. Likewise for /æ/ in the second block and /t/ in the third block. The perception score of /kæt/ is the product of the three Luce-ratios.

4.2 Simulation

4.2.1 Network topology

The model consisted of forty neurons in the input layer and six blocks of forty neurons in the pattern layer. We chose forty neurons because our subjects were native speakers of English and the confusion matrices which we referred to in our simulation included forty American English phonemes. The pattern layer had six blocks because all of our experimental stimuli were six phonemes long.

The most important aspect of the network topology is how we weighted the connections. The perception score depends on the relative amount of activation of the neurons representing the constituent phonemes, and the amount of activation of each pattern neuron depends on how the connection weights the activations from the input layer and the preceding blocks. The model offers two ways to weight the connections: (1) manually specifying the weights as we see fit, and (2) training the model with the study words and fillers so that the model "learns" the proper set of weights[5]. To highlight our assumption that gradience in phonotactic knowledge is held constant across experiments, we manually specify the weights in this paper.

Connection weights were manually specified as follows. As for the connections between the input layer and the pattern layer, if the neurons in the two layers represent the same phoneme, the connection between the two is weighted by +10. On the other hand, if the two neurons represent different phonemes, the connection is weighted by -10. The intuition behind this is that information relevant to identifying the correct phoneme should be valued while any information that could lead to potential confusion should be suppressed. The use of negative connection weights is also similar to the use of inhibitory connections in various connectionist models of speech processing as a way to implement competition. Although the effect of word position on perception could be modeled by varying the magnitude of weights for different word positions, this was not implemented as the positional effect was not the focus of our simulation.

The connections between the blocks in the pattern layer capture the co-occurrence restriction. If two neurons represent the two phonemes whose co-occurrence is favored by the phonotactic constraint, the connection between the two was weighted by +5. On the other hand, if two neurons represent the two phonemes whose co-occurrence is disfavored, the connection was weighted by -5. For example, in Experiment 1, the constraint favored /l/ to repeat as the third and

---

[5] The model can learn the weights on-line by applying the delta rule for each study or filler word it processes. For each word, the ideal activation pattern is defined as having only the neurons representing the constituent phonemes activated in the pattern layer. Error is defined as the deviation of the model's actual activation pattern from the ideal activation pattern. The learning process can be considered as learning to minimize the model's error in recognizing the sequence of constituent phonemes of the input word.

the fifth phoneme of a word, and likewise for /r/. Therefore, the connection between the neuron labeled /l/ in the third block and the neuron labeled /l/ in the fifth block was weighted by +5, and likewise for the two neurons labeled /r/ in the two blocks. On the other hand, the constraint disfavored the co-occurrence where /l/ and /r/ occupied the third and the fifth position in the same word. Therefore, the connection between the neuron labeled /l/ in the third block and the neuron labeled /r/ in the fifth block was weighted by -5, and likewise for the connection between the neuron labeled /r/ in the third block and the neuron labeled /l/ in the fifth block. All other connections in the pattern layer were weighted zero, as they were not the focus of our simulation. Note that by holding the magnitude of connection weights between the restricted phonemes constant across different experiments, we implemented the assumption that there is no gradient difference in the acquired knowledge of experimental constraint[6].

4.2.2 Procedures

The model predicted the perception score for each test word, either legal or illegal, presented to the subjects in the corresponding experiment session. We averaged the mean prediction scores for the two stimulus types and computed the difference between the two means to estimate the size of the effect of phonotactic knowledge on perception.

At the beginning of each word, all neurons in the model had zero activation. The model then received as input each constituent phoneme of the test word one by one. For each phoneme, the input neurons were activated according to the confusion matrices in Luce (1986). For the words which were presented without background noise (all test words in Experiments 1, 2, and 3), we referred to the onset confusion matrix and the vowel confusion matrix for SNR=+15dB. For the words in Experiment 4, we referred to the two matrices for SNR=+5dB. The pattern neurons in the block representing the given word position were activated accordingly and stored until the final phoneme of the word was processed. The perception score for the test word was estimated from the resulting activation pattern in the pattern layer after processing the final phoneme of the word.

4.2.3 Results

In simulating all four experiments, mean perception score was higher for legal words than for illegal words, demonstrating the model's ability to simulate the phonotactic legality effect on perception. More importantly, comparison of the

---

[6]Alternatively, we can represent gradience in phonotactic knowledge by varying the magnitude of connection weights for different phonotactic constraints according to their lexical statistics and learnability. For example, if there were more study words instantiating the constraint in Experiment 1 than Experiment 2, the magnitude of connection weights would be larger for the model simulating Experiment 1 than the model simulating Experiment 2. This would also be the case if the constraint in Experiment 1 were more readily learnable than the constraint in Experiment 2.

simulations between experiments showed that the model's prediction is consistent with the perceptual facilitation hypothesis and the experimental results. The difference in mean perception score between legal and illegal words was greater when the constrained phonemes were more confusable. Between the consonant experiments, the difference was greater in Experiment 1 than Experiment 2 ($t(14)=12.089,\ p<0.001$). Between the vowel experiments, the difference was greater in Experiment 4 than Experiment 3 ($t(14)=20.830,\ p<0.001$). The differences in mean score are summarized in Figure 4.

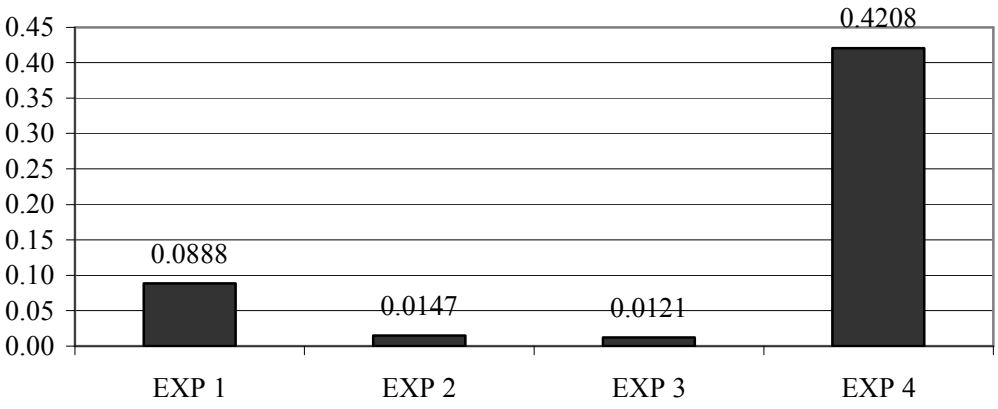

Figure 4: Difference in mean perception score between legal and illegal words.

5 Conclusion

Knowledge of a phonotactic constraint facilitates perception of phonotactically legal sound patterns. Moreover, phonotactic knowledge may be gradient and the gradience in knowledge may lead to gradience in the degree to which perception is facilitated. We proposed the perceptual facilitation hypothesis such that the phonotactic knowledge works to reduce perceptual confusion and that the degree of perceptual facilitation would be greater if there were more room for the phonotactic knowledge to reduce confusion. This suggests that gradience in confusability inherent in the perceptual input could be another source of gradience in perceptual performance. The hypothesis was supported by the results from four experiments where subjects performed fast auditory-repetition tasks to learn one of four artificial co-occurrence restrictions. The effect of phonotactic knowledge on subjects' perception performance was greater when the confusability between the restricted phonemes was greater. The proposed connectionist model could successfully simulate the experimental results and illustrated its potential to integrate the two sources of gradience in perception performance: gradience in phonotactic knowledge and gradience in confusability.

References

Albright, A. 2003. Segmental Similarity Calculator, using the "shared natural classes" method of Frisch, Broe, and Pierrehumbert (1997). http://web.mit.edu/albright/www/. Retrieved 24 Mar 07.

Brown, R. W., & D. C. Hildum. 1956. Expectancy and the perception of syllables. *Language*, *32*, 411-419.

Chomsky, N., & M. Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Coleman, J. S., & J. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational Phonology, Third meeting of the ACL special interest group in computational phonology,* 49-56.

Frisch, S., Broe, M., & J. Pierrehumbert. 1997. Similarity and phonotactics in Arabic. *Rutgers Optimality Archive*, ROA-223-1097. http://www.web-slingerz.com/cgi-bin/oa_list.cgi.

Jelinek, F. 1997. *Statistical Methods for Speech Recognition*. Cambridge: MIT Press.

Luce, P. A. 1986. *Neighborhoods in the Words in the Mental Lexicon*. Ph.D. dissertation, Department of Psychology, Indiana University, Bloomington, Indiana.

Luce, P. A., Goldinger, S. D., Auer, E. T., & M. S. Vitevitch. 2000. Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, *62*, 615-625.

Luce, R. D. 1959. *Individual Choice Behavior*. New York: Wiley.

Massaro, D. W., & M. M. Cohen. 1983. Phonological context in speech perception. *Perception & Psychophysics*, *34*, 338-348.

McClelland, J. L., & J. L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Moreton, E., & S. Amano. 1999. Phonotactics in the perception of Japanese vowel length: evidence for long-distance dependencies. In *Proceedings of the 6th European Conference on Speech Communication and Technology*.

Newport, E. L., & R. N. Aslin. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127-162.

Norris, D. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Peperkamp, S., Skoruppa, K., & E. Dupoux. 2005. The role of phonetic naturalness in phonological rule acquisition. In *Proceedings of the 30th annual Boston University Conference on Language Development*, 464-475.

Pycha, A., Nowak, P., Shin, E., & R. Shosted. 2003. Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of WCCFL 22,* 423-435.

Rabiner, L., & B. -H. Juang. 1993. *Fundamentals of Speech Processing*. New Jersey: Prentice Hall.

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & M. Bowman. 2000. English speaker's sensitivity to phonotactic patterns. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 269-283. Cambridge: Cambridge University Press.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & D. Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47-62.

Warker, J. A., & G. S. Dell. 2006. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 2*, 387-398.

Wilson, C. 2003. Experimental investigation of phonological naturalness. In *Proceedings of WCCFL 22*, 533-546.

Wilson, C. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. Cognitive Science, 30, 945-982.