CHANGE IN THE ADULT PHONOLOGICAL PROCESSING SYSTEM BY LEARNING NON-ADJACENT PHONOTACTIC CONSTRAINTS FROM BRIEF EXPERIENCE: AN EXPERIMENTAL AND COMPUTATIONAL STUDY

BY

HAHN KOO

B.A., Seoul National University, 2002

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2007

Urbana, Illinois

Doctoral Committee:

Professor Richard W. Sproat, Director of Research
Associate Professor Jennifer S. Cole, Co-director of Research
Professor Gary S. Dell
Professor Cynthia L. Fisher

ABSTRACT

Recent studies show that the adult phonological processing system constantly changes as a result of word processing experience; adult speakers can learn new sound patterns from brief experience processing words that exhibit the sound patterns, and how they process words changes as a result of learning. But how malleable is the phonological processing system and what is the mechanism underlying the adaptation of the system to recent processing experience? This dissertation presents experiments and computational models that investigate whether adult speakers can learn non-adjacent phonotactic constraints, and how their perception and grammaticality judgment behavior change as a result of learning.

The experiments show that adults can learn phonotactic constraints that are nonexistent in their language and which restrict co-occurrence of two non-adjacent phonemes with one intervening phoneme. The results further document evidence of the malleability of the adult phonological processing system, and extend the range of learnable sound patterns since non-adjacent phonological dependencies are assumed to be difficult to learn.

As a result of learning, the speakers judge phonotactically legal novel words to be more grammatical than phonotactically illegal novel words. They also perceive the legal ones more quickly and accurately than the illegal ones. In addition, the experiments show that the effect of learning on perception is greater when the learned phonotactic constraint restricts co-occurrence of more confusable phonemes. This subtle effect of learning on perception is expressed as the Perceptual Facilitation Hypothesis, which provides a more detailed account of how the phonotactic knowledge functions in the adult phonological processing system to change its perceptual behavior.

The experimental results are simulated with two computational models that

demonstrate how the adult phonological processing system adapts to recent experience: how it comes to perceive legal sound sequences better than illegal ones after repeatedly processing sequences embodying non-adjacent phonotactic constraints, and how it learns the constraints from observing the perceptual output and computes the probability of the perceived phonological structure in judging its grammaticality. The models suggest possible mechanisms that underlie the adaptation of the adult phonological processing system and guide the direction of future research by providing falsifiable predictions.

ACKNOWLEDGEMENTS

First and foremost, I thank the members of my committee: Richard Sproat, Jennifer Cole, Gary Dell and Cynthia Fisher. This dissertation would not have been possible without their helpful feedback, encouragement, patience and the opportunities they provided. Thanks also go to Gary Linebaugh for recording all the nonsense words I used for my experiments, and Kyle Chambers and Kris Onishi for helping me design the experiments. Finally, I would like to acknowledge that the experiments in this dissertation were a part of the project "The Role of Experience in the Production and Perception of Phonological Sequences" funded by the National Institutes of Health: NIH Grant HD-44458.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The sound patterns of a language are embedded in its words. Speakers learn the sound patterns as they hear and say words, and the acquired knowledge of the sound patterns affects their phonological behavior in turn. For example, speakers can tell whether a novel sound sequence can be a word in their language. Further, in perceiving and producing speech sounds, they are biased towards those that occur in contexts that follow the sound patterns of their language. For example, there is no word in English that starts with /ŋ/, and therefore speakers never hear or say words in English that start with /ŋ/. As a result, native speakers of English do not consider sound sequences that start with /ŋ/ as possible words in English, and they rarely make speech errors that result in sound sequences that start with /ŋ/ (Dell et al., 2000).

The learning of sound patterns and the resulting change in a speaker's behavior is often described as the adaptation of their *phonological processing system* (Chambers, 2003; Whalen and Dell, 2006), defined in Snowling (2004) as one component of the language processing system that is "concerned with how speech sounds are perceived, coded, and produced".[1] But how does it adapt to the speaker's experience in processing words that

[1] The term phonological processing or phonological processing system is often used without clear definition, but some of the explicitly stated definitions include the following. Phonological processing is "the use of phonological information in processing written and spoken language" (Wagner and Torgesen, 1987), or a process that "entails the segmental analysis of words for ordinary speaking and listening, as well as the metaphonological skills required for explicitly analyzing the sound structure of speech into the phonetic

embody the sound patterns of a language? This dissertation focuses on two specific questions related to this matter: how malleable is the phonological processing system, and how does accumulated experience with processing cause the system's phonological processing behavior to change?

This dissertation addresses these two questions by presenting experiments and computational models that investigate whether the adult phonological processing system can learn non-adjacent phonotactic constraints and how its behavior changes as a result of learning. The experiments show that the adult phonological processing system is malleable enough to learn constraints that do not exist in the speaker's language and which restrict the co-occurrence of non-adjacent phonemes. I show that as a result of learning, speakers' perception of novel utterances that follow the acquired phonotactic constraints is facilitated, but the facilitation effect itself is contingent on the perceptual confusability of the phonemes whose co-occurrence is restricted. In addition, speakers judge novel utterances that follow the acquired phonotactic constraints to be more grammatical than those that violate the constraints. The computational models developed here illustrate the process of learning and the resulting change in behavior of the adult phonological processing system, and simulate the experimental results from my human subject experiments.

---

components represented by the alphabet" (Mody et al., 1997; Evans et al., 2002). The phonological processing system is defined as a system "which participates in accessing, assembling, and encoding word forms" in Okada (2005).

1.1. Background

1.1.1. Phonotactic constraints

Phonotactic constraints are restrictions on the distribution of sounds or sound sequences in various prosodic positions or domains of a given language (e.g., Kenstowicz, 1994; Roca and Johnson, 1999). For example, in English, /h/ is restricted in its distribution: it appears in syllable onsets but never in codas, whereas /ŋ/ appears in codas but not in onsets. In generative phonology, phonotactic constraints are traditionally interpreted as absolute restrictions; sound patterns either do or do not occur in a given prosodic domain. However, phonotactic patterns can also be described with stochastic constraints that model the statistical distribution of sound structures in prosodic domains (e.g., Kessler and Treiman, 1997). For example, /ɪʧ/ and /ɑb/ appear in syllable codas more frequently than /ɑʧ/ and /ɪb/ in English (Treiman et al., 2000).

Languages differ in their phonotactic constraints. For example, consonant clusters such as /sm/ or /kr/ appear in syllable onsets in English. However, such consonant clusters do not appear in syllable onsets in Japanese and Korean. In Arabic verbal roots, identical consonants do not repeat in the first two consonant positions (Greenberg, 1950); roots like /dadam/ do not occur. The total Obligatory Contour Principle (total-OCP) is a constraint adopted in McCarthy (1986) to prohibit identical consonants from occupying adjacent consonant positions in the consonantal root in Arabic languages. However, no such constraint is found in English, Japanese, and Korean.

The distributional restrictions effected by phonotactic constraints are more than “artifacts of historical ancestry” (Frisch et al., 1997). There is ample evidence suggesting that speakers are sensitive to the phonotactic constraints of their language in phonological

processing. For example, Japanese and Korean speakers' sensitivity to the constraint against consonant clusters is reflected in loanword adaptation. When loanwords with consonant clusters enter Japanese (e.g., Arakawa, 1977) or Korean (e.g., Kang, 2003), vowels are inserted between the consonants in the cluster. Arabic speakers' sensitivity to the total-OCP constraint is reflected in language games. For example, Bedouin Hijazi Arabic speakers play a game where they freely permute the order of consonants in verbal roots. However, permutations that violate the total-OCP are not allowed (McCarthy, 1986).

Evidence of individuals' sensitivity to the phonotactic constraints of their language also comes from studies on speech production and speech perception. Speakers rarely produce speech errors that violate the phonotactic constraints of their language (e.g., Fromkin, 1971; Stemberger, 1982). English speakers perceive nonsense words consistent with English phonotactics more accurately than those that are inconsistent with English phonotactics (Brown and Hildum, 1956). They also perceive nonsense words that contain relatively common syllables more quickly than those that contain less common syllables (Vitevitch et al., 1997). When a sound that is ambiguous between a phonotactically legal phoneme and an illegal phoneme is presented to a listener, they are more likely to perceive it as the phonotactically legal one (Massaro and Cohen, 1983; Moreton and Amano, 1999).

1.1.2. Malleability of the phonological processing system

We have seen that different languages have different phonotactic constraints, and that individuals are sensitive to the phonotactic constraints of their language, as reflected in their phonological processing behavior. This suggests that phonotactic constraints are learned and the acquired knowledge forms a part of the phonological processing system.

Studies in developmental psychology show that phonotactic learning takes place

during the latter half of the first year of infancy. Nine-month old infants, but not 6-month-olds, prefer to listen to phonemic sequences that are phonotactically legal (Friederici and Wessels, 1993; Jusczyk et al., 1993) and phonotactically more frequent (Jusczyk et al., 1994) in the ambient language. Nine-month-olds can also perceptually segment words from a stream of continuous speech based on how often consonant clusters occur within a word in the ambient language (Mattys et al., 1999; Mattys and Jusczyk, 2001).

The phonological processing system is highly malleable during infancy; it can rapidly learn new sound patterns embedded in the words that are processed, even new sound patterns that are not found in the ambient language (i.e., the language of the caregivers). Eight-month-olds can acquire sensitivity to the frequency with which two CV syllables occur contiguously after only two minutes of exposure to an artificially created stream of syllables (Saffran et al., 1996a). Similarly, nine-month-olds can rapidly learn the syllabic structure and consonant voicing patterns of the words to which they are exposed (Saffran and Thiessen, 2003); in about two minutes of exposure, they become sensitive to whether the structure of a given word is CVCV or CVCCVC, and whether a voiced consonant appears in onset or coda. Similar findings are observed with older infants, too. After brief auditory exposure to artificially created CVC words, 16.5-month-olds can learn whether particular consonants appear in onset or coda (Chambers et al., 2003).

The phonological processing system remains malleable in adulthood. Adults can learn novel phonotactic constraints on consonant positions from brief production experience (Dell et al., 2000; Dell and Warker, 2004; Goldrick, 2004). For example, after repeatedly reciting a list of nonsense CVC words where /f/ was always an onset and /s/ was always a coda, subjects in Dell et al. (2000) showed a tendency to produce speech errors that respect the constraint; /f/ replaced an onset consonant and /s/ replaced a coda consonant. Adults can learn similar constraints on consonant positions from perception experience

alone (Onishi et al., 2002). After repeatedly listening to nonsense CVC words embodying the phonotactic constraints, subjects perceived novel CVC words that followed the constraints more quickly than those that violated the constraints.

1.1.3. Extent of malleability

The adult phonological processing system is not indefinitely malleable. Despite equal amounts of exposure, adults may learn one sound pattern but fail to learn another. Aspects of sound patterns that have been investigated as potential factors related to learnability include the following: formal complexity (Dell et al., 2000; Onishi et al., 2002; Pycha et al., 2003; Warker and Dell, 2006), phonetic or phonological naturalness (Pycha et al., 2003; Wilson, 2003; Morrison, 2004; Peperkamp et al., 2005; Morrison and Kirchner, 2007), and non-adjacency (Gomez, 2002; Newport and Aslin, 2004).

The formal complexity of sound patterns is measured in terms of the number of features invoked in a formal, grammatical characterization of the sound patterns. In general, studies agree that sound patterns that are formally more complex are harder to learn (Pycha et al., 2003; Warker and Dell, 2006). For example, Warker and Dell (2006) compared learnability of two types of phonotactic constraints: first-order constraints that restrict the syllable position of consonants vs. second-order constraints that restrict the syllable position of consonants depending on the identity of the adjacent vowel. The first-order constraint can be interpreted as a function (restriction on the syllable position) of a single feature (identity of the consonant), and the second-order constraint as the same function of two features (identity of the consonant and identity of the adjacent vowel). In this sense, the second-order constraint is formally more complex than the first-order constraint. The results show that it requires more production experience for the second-order constraints

than the first-order constraints to have a sizable effect on speech errors.

The definition of phonetic and/or phonological naturalness is not consistent among studies on learnability of sound patterns. Pycha et al. (2003) considers that a sound pattern is phonetically natural if it could be the result of listener's misinterpretation of the acoustic cues, while Morrison (2004) and Morrison and Kirchner (2007) consider whether the constraint serves functional purposes such as minimizing articulatory effort or avoiding perceptual confusion. The notion of phonetic naturalness in Peperkamp et al. (2005) is in fact more phonological in nature. A phonological alternation is phonetically natural if it satisfies the following conditions: the change involves a small number of distinctive features, the target must either become similar or dissimilar to the trigger, and the *markedness* of the surface form must be reduced. The definition of phonological naturalness is not clear in Wilson (2003); the author considers assimilation and dissimilation to be phonologically natural because speakers seem to have a bias towards learning such alternations over random alternations.

Previous studies reach different conclusions on the relation between naturalness of a sound pattern and its learnability, partly because the notion of naturalness is defined differently across studies. For example, Peperkamp et al. (2005) concludes that natural sound patterns are more readily learned than unnatural sound patterns. The natural patterns in their study involved a morphophonological alternation of intervocalic voicing which maps /f,s,ʃ/ to [v,z,ʒ] and /p,t,k/ to [b,d,g], while the unnatural pattern had the arbitrary alternation mapping /p,g,z/ to [ʒ,f,t] and /ʃ,v,d/ to [b,k,s]. Their subjects learned to generalize the intervocalic voicing rules but failed to generalize the arbitrary alternations.[2]

[2] The same result would be interpreted by Pycha and her colleagues as relating to the relation between formal complexity and learnability, rather than between naturalness and

On the other hand, Pycha et al. (2003) concludes that naturalness of a sound pattern is not related to its learnability. The natural sound pattern in their study was vowel harmony, while the unnatural sound pattern was vowel disharmony. Vowel harmony was considered as natural because it may be the result of listener's misinterpretation of vowel-to-vowel co-articulation. Their subjects learned both vowel harmony and vowel disharmony equally well.[3]

The issue of learnability in relation to adjacency in sound patterns has attracted interest in recent psycholinguistic literature especially since Saffran et al. (1996a) suggested that people track conditional probability between adjacent units to learn linguistic structures. The fact that non-adjacent dependencies do exist in languages means that speakers must also track conditional probabilities between non-adjacent units to learn the dependency. This is a problem because the number of probabilities increases exponentially as a function of distance between non-adjacent units, placing a huge burden on the learner.

Studies show that learning non-adjacent dependencies is harder than adjacent dependencies. Newport and Aslin (2004) shows that adults cannot learn the conditional probability between non-adjacent CV syllables with one intervening CV syllable from a continuous stream of syllables, while Saffran et al. (1996b) shows that adults can learn the

---

learnability. The intervocalic voicing rules can be characterized using just the [voice] feature, while no single feature can characterize the arbitrary alternations. Therefore, the intervocalic voicing rules are formally simpler than the arbitrary alternations.

[3] Taken together with the result that the subjects in Pycha et al. (2003) failed to learn an arbitrary alternation between stem and suffix vowels, Wilson (2003) would interpret the same result as evidence that both vowel harmony and vowel disharmony are natural sound patterns.

conditional probability between adjacent CV syllables to segment words from a stream of syllables. Adults in Gomez (2002) were exposed to three element strings (e.g., *pel-wadim-jic*) and learned the dependency between the non-adjacent elements (e.g., *pel* and *jic*). However, the results suggest that the subjects primarily focused on adjacent dependencies (e.g., *pel-wadim* and *wadim-jic*) and learned the non-adjacent dependency only after the variability of the intervening element (e.g., *wadim*) significantly increased.

Languages, however, do have non-adjacent dependencies between phonemes, such as vowel harmony, and studies show that they can be learned from brief experience. Pycha et al. (2003) shows that adults can learn both palatal harmony and palatal disharmony between CVC stem vowel and VC suffix vowel. Wilson (2003) shows that adults can learn nasal harmony and disharmony applying to the final two consonants of CV.CV.CV words. Adults in Newport and Aslin (2004) learned conditional probability between non-adjacent phonemes with one intervening phoneme from listening to a stream of CV syllables. McLennan et al. (2005) shows that adults can learn which consonants can co-occur in CVC words. Adults in Moreton (2006) learned height harmony between two vowels and voicing harmony between two consonants in CV.CV words. The sound patterns in these studies are all non-adjacent in the sense that there is one intervening phoneme between the mutually dependent phonemes.

1.1.4. Change in the phonological processing behavior

The behavior of the phonological processing system changes in various aspects as a result of learning phonotactic constraints, as introduced in the first two sections of this chapter. Of particular interest to this dissertation is how phonotactic learning affects perception and grammaticality judgment behavior. Speakers perceive phonotactically legal or more

frequent sound patterns more quickly and accurately than phonotactically illegal or less frequent sound patterns. They also judge phonotactically legal or more frequent sound patterns to be more acceptable than phonotactically illegal or less frequent sound patterns. But what is the underlying mechanism? Models and theories differ in various aspects, but one prominent difference lies in assumption on whether the effect of phonotactic learning on phonological processing behavior is either lexical or sub-lexical.[4]

1.1.4.1. Perception

The mechanism that underlies the perceptual facilitation of phonotactically legal words is often explained in terms of models of speech perception or spoken word recognition (McClelland and Elman, 1986; Auer, 1993; Norris et al., 1997; Grossberg et al., 1997; Luce et al., 2000). Existing models typically assume a set of lexical and/or sub-lexical units, which are activated to a certain degree according to the processing dynamics of the model. The model explains the behavior that phonotactically legal words are perceived better by assigning higher activation to lexical or sub-lexical units that are consistent with the phonotactic constraints.

For example, the perceptual facilitation effect is explained in the TRACE model (McClelland and Elman, 1986) as a lexical effect. The model is a connectionist model of speech perception, consisting of three layers of nodes: feature, phoneme, and word. As it processes perceptual input piece by piece from left to right, activation spreads between the layers interactively. The phoneme(s) perceived at each time slice activates words containing

---

[4] The term *sub-lexical*, especially in the speech perception literature, refers to the components such as features, phonemes, and syllables that define the phonological structure of a word.

the perceived phoneme(s), and the activated words in turn activate all of their component phonemes. If there are more words that contain the perceived phonemes, the total amount of activation the constituent phonemes receive from the word layer will be larger. As activation also flows from the phoneme layer back to the word layer, the words that contain these phonemes will be activated even further. As a result, words that contain more common phoneme sequences cumulate more activation over time and end up with higher activation than words with less common phoneme sequences. Notice that this effect is triggered at the word layer; it is the large number of words which contain the perceived phonemes that first causes phonotactically legal sound sequences to receive more activation. If this effect were the result of learning, learning would be no other than simply storing words. In this sense, TRACE models the effect of phonotactic learning on perception as a lexical effect.

On the contrary, the effect of phonotactics on perception is modeled as a sub-lexical effect in PARSYN (Auer, 1993; Luce et al., 2000), in recognition of the finding from previous studies that phonotactic effects are distinct from lexical effects (Pitt and McQueen, 1998; Vitevitch and Luce, 1999; Bailey and Hahn, 2001). Similar to TRACE, the PARSYN model has separate layers for phonemes and words. Figure 1.1 illustrates the structure of PARSYN. However, while phonemes in different word positions are not connected in TRACE, they are connected in PARSYN allowing activation to directly flow from the phonemes in one position to the phonemes in another position. The amount of activation that flows between the two word positions is modulated by the log-frequency-weighted transitional probability between the two phonemes. Simply put, the more frequently the two phonemes co-occur in the language, the higher the activation of the two phonemes; the phonemes in turn facilitate activation of words that contain them, and words that contain phonemes that frequently co-occur are activated more. If this were the effect of

learning, learning would consist of tracking the co-occurrence frequency of phonemes. In this sense, PARSYN models the effect of phonotactic learning on perception as a sub-lexical effect.

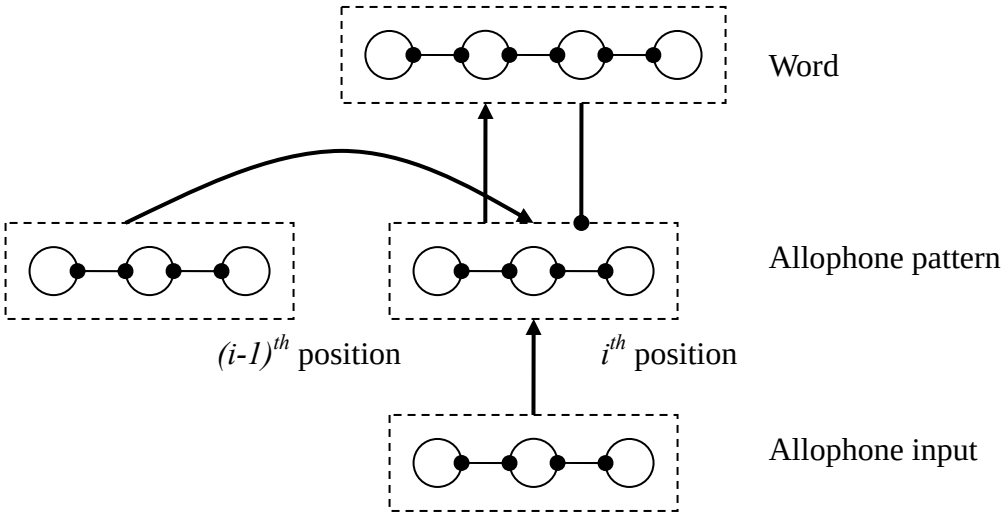

Figure 1.1. A schematic diagram of PARSYN (Luce et al., 2000). The model consists of three layers of nodes: input, pattern, and word. Each node represents a separate word in the word layer, and a separate position-specific allophone in the other two layers. Lines terminating in arrows imply that facilitative activation spreads from one layer (position) to the other layer (position). Lines terminating in circles imply that inhibitory activation spreads from one layer (node) to the other layer (node). For example, the line from the box labeled *(i-1)*th position to the box labeled i^{th} position implies that the allophones in the *(i-1)*th position facilitates the activation of the allophones in the i^{th} position. The line between two nodes in each position implies that the allophones in the same position inhibit the activation of each other.

1.1.4.2. Grammaticality judgment

Grammaticality judgments have played a central role in the development of phonological

theory, especially in generative phonology. The focus of the enterprise has been on explaining why some sound sequences are grammatical while others are not, in terms of a set of abstract rules (e.g., Chomsky and Halle, 1968), well-formedness conditions (e.g., Goldsmith, 1976), or ranked constraints (e.g. Prince and Smolensky, 1993). In such theories, the grammaticality of a sound sequence is determined by testing whether it is consistent with the system of rules or constraints, which are assumed to be the content of speakers' knowledge of sound patterns in their language.

Generative phonology has been successful in explaining the distribution of sound patterns in natural languages using the notion of grammaticality. However, the theory has not been so successful in explaining how speakers actually make grammaticality judgments in experiments. For example, speakers' judgments show gradience on a scale of acceptability and are rarely categorical. In addition, the overall likelihood of the whole sound sequence influences the judgment more than the grammaticality of any specific part of the sequence against a set of abstract rules or constraints (Coleman and Pierrehumbert, 1997).

As summarized in Bailey and Hahn (2001) and Albright (2006), recent studies that model speakers' grammaticality judgment behavior can be classified into two groups[5]: those that adopt a lexical approach and those that adopt a phonotactic approach. The lexical approach claims that speakers judge the grammaticality of a sound sequence based on how similar it is to the words stored in the speakers' mind. For example, Bailey and Hahn (2001) shows that there is a positive correlation between the speakers' wordlikeness judgment of a

[5] Bailey and Hahn (2001) uses the term *wordlikeness* and Albright (2006) uses the term *acceptability* instead of grammaticality. However, I use the term grammaticality in this dissertation for consistency.

novel word and its *neighborhood density*. In its simplest form (Luce, 1986), the neighbors of a word in a language are the words in the language that can be derived by applying a single phoneme edit-operation (substitution, insertion, deletion) on the word. The neighborhood density of a word is often the sum of token frequency of all of its neighbors in the language. For example, the neighbors of *pit* in English would be words such as *pin* (via substitution), *spit* (via insertion), and *it* (via deletion). The neighborhood density of *pit* would be the sum of token frequency of all such words in English. As in the case of TRACE, it is the number of words similar to the novel word that determines the outcome of grammaticality judgment, and the phonotactic learning process that enables the judgment is equivalent to simply storing the words in the language.

On the other hand, the phonotactic approach claims that the speakers' grammaticality judgment of a novel word reflects their knowledge of the *phonotactic probability* of its sub-lexical components, where phonotactic probability refers to the frequency with which sounds occur in certain positions and sequences in a given language (Jusczyk et al., 1994; Vitevitch and Luce, 2004). As the ambiguity of the terms *positions* and *sequences* implies, there are different ways to calculate the phonotactic probability of a word. For example, in Coleman and Pierrehumbert (1997) and Frisch et al. (2000), a position is specified along three dimensions: onset vs. rhyme, word-initial syllable vs. word-final syllable, and stressed vs. unstressed. Vitevitch and Luce (2004) sequentially numbers each segmental slot in a word from left to right and specifies a position in terms of its index. The positional probability of a word is estimated by computing the product (Coleman and Pierrehumbert, 1997) or the sum (Vitevitch and Luce, 2004) over the individual positional probabilities, which is proportional to how often the component sounds occur in each position in the given language.

A sequence is most often specified in terms of *n-grams*, or *n* contiguous phonemes

(Jusczyk et al., 1994; Vitevitch et al., 1997; Vitevitch and Luce, 1998; Bailey and Hahn, 2001; Leigh and Charles-Luce, 2002; Vitevitch et al., 2004; Storkel et al., 2006). In all of the cited studies, the *n* of *n*-grams is two, and the frequency of a *bi-gram*, or a *bi-phone*, is captured in terms of *bi-phone probability*, which is the conditional probability of observing the latter phoneme of a bi-phone given its first phoneme in the given language. The sequential probability of a word is estimated by computing the mean of the component bi-phone probabilities (e.g., Vitevitch et al., 1997; Vitevitch and Luce, 1998; Bailey and Hahn, 2001). As in the case of PARSYN, it is the sub-lexical statistics that determines the outcome of grammaticality judgment, and the phonotactic learning process that enables the judgment is equivalent to learning the statistics such as positional and sequential probabilities.

1.2. Central claims and contribution

The claims and contributions of this dissertation are as follows. Firstly, I show that the adult phonological processing system can adapt to various artificial co-occurrence restrictions on non-adjacent phonemes from brief experience. The experimental results in support of this claim add to the previously reported evidence on the malleability of the adult phonological processing system. Along with the recent studies on learnability of non-adjacent phonological dependencies reviewed in Section 1.1.3, the results extend the range of sound patterns that can be learned by the adult system to dependencies between two non-adjacent phonemes between which one intervening phoneme occurs. Moreover, the results show implicit learning of non-adjacent phonotactic constraints. With the exception of McLennan

et al. (2005),[6] all prior studies on learnability of non-adjacent phonological dependencies present evidence of learning by eliciting speakers' explicit judgment of grammaticality. On the other hand, the results in this dissertation also show that speakers' perception of phonotactically legal novel words becomes more rapid and accurate as a result of implicit learning.

Secondly, I argue that the effect of phonotactic knowledge on perception is subtle and goes beyond simple facilitation. I propose a perceptual facilitation hypothesis that relates the degree of facilitation to the confusability of the constrained phonemes: phonotactic knowledge facilitates perception of legal sound patterns more if the phonotactically constrained phonemes are more confusable to each other. Previous studies show that phonotactic knowledge facilitates the perception of phonotactically legal novel sound patterns. The hypothesis provides a more detailed account of how the phonotactic knowledge functions in the adult phonological processing system to change its perceptual behavior.

Finally, I present two computational models and demonstrate how the adult phonological processing system learns non-adjacent phonotactic constraints and how its perception and grammaticality judgment behavior changes as a result of learning. The models explicitly formulate the mechanisms that underlie the adaptation of the adult phonological processing system and guide the direction of future research by providing falsifiable predictions. In modeling the process, I follow the sub-lexical approach and

---

[6] The non-adjacent phonotactic constraints in this dissertation differ from that of McLennan et al. (2005) in that they restrict co-occurrence of phonemes across syllable boundary while the constraint in McLennan et al. (2005) restrict co-occurrence of consonants within the same syllable.

assume that the effect of phonotactic learning observed in the experiments here reflects the speakers' learning of phoneme co-occurrence patterns.

The effect of phonotactic learning on perception is modeled in a connectionist framework, using a single layered recurrent perceptron. With stipulations similar to PARSYN (Luce et al., 2000), the perceptron demonstrates how perception of phonotactically legal novel words becomes facilitated by repeatedly mapping perceptual input embodying the phonotactic constraints to its corresponding phonological structure. The perceptron also predicts that the degree of facilitation will be larger when the phonotactically constrained phonemes are more confusable, as predicted by the perceptual facilitation hypothesis. The effect of phonotactic learning on the speakers' grammaticality judgment is probabilistically modeled using a Bayesian belief network. Grammaticality of a novel word is assumed to reflect the joint probability of its constituent phonemes computed. Learning of phonotactic constraints is modeled as identifying position-specific phonemes that are conditionally dependent on each other and estimating the probability distribution over the conditionally dependent phonemes.

1.3. Organization of the dissertation

The dissertation is organized as follows. Chapter 2 presents six artificial grammar learning experiments which show that adult speakers of English can learn non-adjacent phonotactic constraints from brief perception and production experience. Evidence of learning is tested by measuring the subjects' performance in auditory repetition tasks in four experiments and grammaticality judgment tasks in two experiments. Evidence of learning in the four auditory repetition experiments shows an interesting asymmetry which leads us to investigate how knowledge of a phonotactic constraint facilitates perception.

Chapter 3 presents the perceptual facilitation hypothesis to account for the asymmetry found in the experiments described in Chapter 2. The hypothesis is that knowledge of a phonotactic constraint facilitates perception more if the constrained phonemes are perceptually more confusable to each other. Two auditory repetition experiments are presented, where stimuli words embody non-adjacent phonotactic constraints on phoneme pairs that are more confusable than the ones in the experiments in Chapter 2. Results from these two experiments are compared with the results from two parallel experiments in Chapter 2. The results from the four experiments collectively serve as evidence supporting the perceptual facilitation hypothesis.

Chapter 4 presents a connectionist model of how phonotactic learning affects perception using a single layered recurrent perceptron. The model estimates how well a word would be perceived by a hypothetical speaker who has been exposed to words instantiating a particular phonotactic constraint. The knowledge of the phonotactic constraint that results from the exposure is encoded among the connection weights of the model. The structure of the model is transparent enough for us to directly encode the phonotactic knowledge by manually specifying the weights. The model can also learn the phonotactic constraint from the instantiating words by adjusting its weights using the delta rule. Simulation studies are presented and show that the model makes predictions consistent with the perceptual facilitation hypothesis and can learn the non-adjacent phonotactic constraints from brief exposure to instantiating words.

Chapter 5 presents a Bayesian belief network that learns phonotactic constraints as conditional dependencies between position-specific phonemes and makes probabilistic grammaticality judgments accordingly. When coupled with the connectionist model in Chapter 4, the belief network correctly predicts the results of the two grammaticality judgment experiments presented in Chapter 2. It also provides a possible explanation as to

why the asymmetry addressed by the perceptual facilitation hypothesis is found in auditory repetition experiments but not in grammaticality judgment experiments.

Chapter 6 summarizes the dissertation and discusses the direction of future research stemming from the experiments and computational models presented in the dissertation.

CHAPTER 2

LEARNABILITY EXPERIMENTS

The six experiments presented in this chapter investigate if the adult phonological processing system can learn artificial non-adjacent phonotactic constraints from brief perception and production experience. The phonotactic constraints are non-adjacent in that it restricts co-occurrence of two phonemes that are one phoneme apart from each other. Because of their non-adjacency, it is questionable whether they will be learned from brief experience from which constraints on adjacent phonemes of the same formal complexity were learned in previous studies. In addition, the phonotactic constraints closely resemble the sound patterns often discussed in phonology, and their natural language counterparts differ in their typological frequency. Considering the conjecture in Newport and Aslin (2004) and Moreton (2006) such that the distribution of sound patterns reflect their learnability, not all constraints may be equally well learned although they are identical in terms of their formal complexity and non-adjacency.

The chapter is organized as follows. Section 2.1 introduces four artificial non-adjacent phonotactic constraints whose learnability is examined by the experiments. The rationale for suspecting learnability of phonotactic constraints to differ with respect to their non-adjacency and typological frequency is also briefly discussed. Section 2.2 introduces the basic idea underlying the two types of experiments presented in this dissertation: auditory repetition experiments and grammaticality judgment experiments. Section 2.3 presents four auditory repetition experiments testing learnability of the four non-adjacent phonotactic constraints. Section 2.4 presents two grammaticality judgment experiments testing learnability of two of the four constraints. Section 2.5 summarizes and discusses the experimental results.

2.1. The non-adjacent phonotactic constraints

The experiments test learnability of four constraints on non-adjacent phonemes that hold in the final two syllables of CV.CV.CV nonsense words: liquid harmony, liquid disharmony, backness harmony, and backness disharmony. Liquid harmony disallows co-occurrence of liquids (/l/ and /r/[7]) of different laterality (/l/ = [+lateral], /r/ = [-lateral]). For example, /sa.la.la/ or /sa.ra.ra/ would be phonotactically "legal", while /sa.la.ra/ or /sa.ra.la/ would be phonotactically "illegal". Backness harmony disallows the co-occurrence of high vowels (/i/ and /u/) of different backness (/i/ = [-back], /u/ = [+back]). For example, /sa.li.ki/ or /sa.lu.ku/ would be phonotactically legal, while /sa.li.ku/ or /sa.lu.ki/ would be phonotactically illegal. Liquid disharmony and backness disharmony are constraints directly opposite to liquid harmony and backness harmony, respectively.

2.1.1. Non-adjacency of the phonotactic constraints

The four phonotactic constraints govern non-adjacent phonemes in that the two constrained positions are one phoneme apart from each other. For liquid harmony and disharmony, the constrained positions are the last two syllable onsets separated by an intervening vowel. For backness harmony and disharmony, the constrained positions are the last two nuclei, or vowels, separated by an intervening consonant.

In terms of the order of formal complexity suggested in studies such as Dell et al. (2000) and Onishi et al. (2002), the four constraints can be described as second order

[7] I use /r/ to denote the alveolar approximant [ɹ] in the dissertation.

phonotactic constraints. In those studies, the order of formal complexity depends on the number of variables invoked in formally characterizing a phonotactic constraint. The first order phonotactic constraint restricts which phonemes can occupy certain syllable positions. For example, /f/ is an onset and /s/ is a coda. This can be interpreted as a function (restriction on the syllable position) of one variable (identity of the consonant). On the other hand, the second order phonotactic constraint restricts which phonemes can occupy certain syllable positions depending on the adjacent vowel. For example, /f/ is an onset and /s/ is a coda when the nucleus is /æ/, while /s/ is an onset and /f/ is a coda when the nucleus is /ɪ/. This can be interpreted as a function (restriction on syllable position) of two variables (identity of the consonant and identity of the adjacent vowel). The non-adjacent phonotactic constraints are second order constraints because they can be interpreted as restrictions on which liquid (high vowel) can occupy the onset (nucleus) in the final syllable depending on which liquid (high vowel) occupied the onset (nucleus) in the previous syllable.

A second order phonotactic constraint dependent on a non-adjacent phoneme is suspected to be more difficult to learn than a second order constraint dependent on an adjacent phoneme as it requires more processing resources to identify the dependency. Identifying the dependency between two adjacent phonemes requires processing at least two phonemes at once. On the other hand, identifying the dependency between two non-adjacent phonemes with one intervening phoneme necessitates processing at least three phonemes at once. In addition, assuming the phoneme inventories are identical in both cases, there are more theoretically possible three phoneme sequences than two phoneme sequences. Therefore, the learner must track more sequences to learn a second order constraint dependent on a non-adjacent phoneme than to learn a second order constraint dependent on an adjacent phoneme.

Evidence in support of this conjecture comes from Gomez and Maye (2005). The

study tested if infants of different ages (12, 15, and 17 months old) could learn a dependency between the first and the third syllable in tri-syllabic sequence. The results suggest that learners may be biased towards learning adjacent dependencies, and that there may be age-limitations on learning non-adjacent dependencies. They found that subjects appeared to have focused on adjacent dependency even when its predictability was significantly lower than that of non-adjacent dependency. They also found younger infants (12 month olds) could not learn non-adjacent dependency under the same condition where older infants (15 and 17 month olds) learned it.

Results from Dell et al. (2000) and Onishi et al. (2002) show that the adult phonological processing system adapts in ways that reflect learning of second order phonotactic constraints dependent on an adjacent vowel. The experiments in this chapter investigate if the same system can also learn second order phonotactic constraints dependent on non-adjacent phonemes, which could potentially be harder to learn.

2.1.2. Typological frequency and learnability

One of the key questions in phonology is why sound patterns are the way they are in natural languages; why are some sound patterns frequently attested, while others are rarely or never attested? One common approach (e.g., Ohala, 1993; Blevins, 2004) is to explain asymmetries in typology on the basis of difference in robustness of phonetic precursors to phonological sound patterns. For example, devoicing of /g/ to /k/ is more common than devoicing of /b/ or /d/, because the air-pressure above the glottis rises faster when constriction of the air-flow is made closer to the back of the oral cavity (Blevins, 2004).

Moreton (2006) argues that typological asymmetry can also be explained by asymmetry in the learnability of sound patterns; a more readily learnable sound pattern is

more likely to be innovated and passed down to the next generation of speakers. He presents supporting evidence by comparing the learnability of a sound pattern relating the height of two vowels against a sound pattern relating the vowel height to the voicing of the following consonant. Sound patterns relating height of vowels are more common than sound patterns relating vowel height to the voicing of the following consonant. Despite a similar degree of robustness in the phonetic precursor, adult subjects learned the more common pattern better than the less common pattern.

A similar conjecture is also raised in Newport and Aslin (2004). Adult subjects in their study could detect statistical regularity between non-adjacent consonants in a stream of CV syllables and use the statistical regularity for word segmentation. The subjects could also detect statistical regularity between non-adjacent vowels. However, the subjects could not detect statistical regularity between non-adjacent CV syllables. The authors note the fact that sound patterns that restrict co-occurrence of non-adjacent syllables are not attested, and argue that their results suggest that the differences in typological frequency of sound patterns may be due to the differences in learnability.

With the exception of backness disharmony, the non-adjacent phonotactic constraints discussed in this chapter have their counterparts in natural language phonology. Examples of liquid harmony are found in Bukusu (de Blois, 1975; Odden, 1994) and Pohnpeian (Rehg and Sohl, 1981). Examples of liquid disharmony are found in Albanian, Old Irish, Latin, Ossetic (Testen, 1997), and Georgian (Fallon, 1993).[8] Examples of backness harmony are found in Finnish (Karlsson, 1999), Hungarian (Siptár and Törkenczy, 2000), Tatar (Poppe, 1963), East Turki (Poppe, 1965), Kyrgyz (Herbert and Poppe, 1963),

[8] I thank Juliette Blevins for sharing her data on liquid disharmony.

Yakut (Krueger, 1962), Buriat (Poppe, 1965), and Khalkha Mongolian (Goldsmith, 1985).[9]

The counterparts of the four constraints differ in terms of their typological frequency. A review of phonology literature suggests that dependencies between non-adjacent vowels are more common than dependencies between non-adjacent consonants. Hansson (2001) acknowledges that liquid disharmony is more common than liquid harmony though both are rare sound patterns. To the best of my knowledge, backness disharmony between non-adjacent vowels is not attested.

If differences in typological frequency are due to difference in learnability, then more commonly attested phonotactic constraints should be more readily learned. Liquid disharmony should be more readily learned than liquid harmony. Backness harmony should be more readily learned than backness disharmony. Backness harmony should also be more readily learned than either liquid harmony or liquid disharmony. The experiments presented in this chapter investigate if there is difference in learnability between the four non-adjacent phonotactic constraints and if the result is consistent with the prediction from the typological frequency.

2.2. Basic idea underlying the experiments

Each experiment tests the learnability of a separate non-adjacent phonotactic constraint. Subjects experience auditory exposure to non-sense words that instantiate the experimental constraint, called "study" words. Subjects are then tested on two types of new non-sense words: phonotactically "legal" words that are consistent with the constraint, and phonotactically "illegal" words that violate the constraint. To distract subjects from

[9] The list of languages with the sound patterns is not intended to be exhaustive.

strategically identifying the experimental constraint, “filler” words which neither instantiate nor violate the experimental constraint are presented along with the study, legal, and illegal words.

Subjects are trained on the study words by performing an auditory repetition task; words are presented one at a time, and subjects are told to repeat each word they hear as quickly and accurately as possible. The experiments in this chapter are classified into two types depending on the task subjects perform for the test words. In auditory repetition experiments, subjects perform auditory repetition tasks for the test words. In grammaticality judgment experiments, subjects make a binary grammaticality judgment for each test word they hear.

A significant difference in subjects’ performance between legal and illegal words is interpreted as evidence of learning the experimental phonotactic constraint. In the auditory repetition experiments, performance is measured in terms of response latency and error rate. If subjects learn the experimental constraint, response latency for legal words will be faster than response latency for illegal words. Similarly, the error rate for legal words will be smaller than the error rate for illegal words. In the grammaticality judgment experiments, performance is measured in terms of the d' score, where the hit-rate is defined as the proportion of legal words judged as grammatical, and the false-alarm rate is defined as the proportion of illegal words judged as grammatical. If subjects learn the experimental constraint, their d'-scores will be significantly above zero.

2.3. Auditory repetition experiments

Four auditory repetition experiments each tested learnability of four non-adjacent phonotactic constraints: liquid harmony (Experiment 1), liquid disharmony (Experiment 2),

backness harmony (Experiment 3), and backness disharmony (Experiment 4).

2.3.1. Methods

2.3.1.1. Subjects

Fifteen University of Illinois students participated in each experiment and received course credit as compensation. All subjects were native speakers of English.

2.3.1.2. Materials

Five hundred twelve tri-syllabic non-sense words of the form CV.CV.CV were recorded in a sound-proof booth by a male native speaker of English and digitized at 44.1 KHz, 16 bit using the Kay Elemetrics CSL 4300B. The initial syllable of each word was fixed to either /sa/ or /ke/. The remaining two consonants were chosen from {/s/, /k/, /l/, /r/}. The remaining two vowels were chosen from {/a/, /e/, /i/, /u/}.

The inventory of 512 words was classified into the following five groups. The CA group comprised 64 words which instantiated the liquid harmony constraint (e.g., /sa.la.la/ or /ke.ra.ra/). The CD group comprised 64 words which instantiated the liquid disharmony constraint (e.g., /sa.la.ra/ or /ke.ra.la/). The VA group comprised 64 words which instantiated the backness harmony constraint (e.g., /sa.si.li/ or /ke.su.lu/). The VD group comprised 64 words which instantiated the backness disharmony constraint (e.g., /sa.si.lu/ or /ke.su.li/). The F group comprised 256 words which did not instantiate any of the four constraints (e.g., /sa.ki.la/ or /ke.ra.se/).

For each experiment session, 94 words were pseudo-randomly chosen from the above inventory: 16 study words, 18 legal words, 18 illegal words, and 42 filler words. Study and legal words were chosen from the group instantiating the experimental phonotactic constraint. Illegal words were chosen from the group instantiating the opposite

constraint. Filler words were chosen from the remaining three groups. The chosen words were distributed over five blocks as in Table 2.1.

	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 2.1. Experiment design for the auditory repetition experiments.

For example, words were selected and organized for Experiment 1 as follows. Firstly, 16 study words and 18 legal words were chosen from the CA group. The set of 16 study words recurred in all five blocks while the legal words were divided into three groups which respectively occurred in blocks 3, 4, and 5. Secondly, 18 illegal words were chosen from the CD group and divided into three groups which respectively occurred in blocks 3, 4, and 5. Finally, 40 fillers were chosen from VA, VD, or F group and divided into five groups which respectively occurred from blocks 1 through 5.

Study, legal, and illegal words in each block were counterbalanced such that half began with /sa/ and the other half began with /ke/. The filler words in each block were counterbalanced such that half began with /sa/ while the other half began with /ke/, and that half were words each with one or more phonemes occurring more than once (e.g., /sa.le.si/) while the other half were words with no phoneme occurring more than once (e.g., /sa.le.ki/). Words in each block were randomly ordered and the corresponding sound-files were played

to subjects one by one using the E-Prime software.

2.3.1.3. Procedures

A trial consisted of performing auditory repetition task for the word presented through a headphone. Subjects were asked to listen to the word and repeat what they heard as quickly and accurately as possible into the microphone placed in front. All responses were recorded to measure error-rate and latency was measured from stimulus offset to response onset.

2.3.1.4. Scoring

Responses where subjects mispronounced one or more phonemes were marked as errors. Responses that the microphone failed to detect so that subjects were forced to reiterate were marked as machine failures. A response whose latency indicates that subjects responded before the onset of the final syllable of the stimulus was marked as an early response. A response whose latency was 2.5 *SD* beyond each subject's mean latency was marked as an outlier.

Machine failures were excluded from computing the error rate. The error rate of each subject per stimulus type (study, legal, illegal, or filler) was defined as the proportion of the number of errors to the total number of trials for the stimulus type with the machine failures removed. For example, if a subject made nine errors out of 18 legal trials without any machine failures throughout the experiment, the error rate for legal words would be 0.5.

Errors, machine failures, early responses, and outliers were excluded in measuring subjects' latency. Early responses were excluded as they were assumed to indicate that the subject started to respond before fully perceiving the stimulus. The mean number of trials excluded per subject is summarized in Table 2.2 for each category. The remaining latencies were averaged by stimulus type (study, legal, illegal, or filler) and by block (blocks 1

through 5) to create 16 scores per subject: scores for study and filler words from blocks 1 through 5 and legal and illegal words from blocks 3 through 5.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Errors	5.60	8.67	5.87	9.20
Failures	0.27	0.93	1.27	0.47
Early responses	0.46	1.87	2.93	0.93
Outliers	3.80	3.40	3.53	3.26

Table 2.2. Mean number of trials per subject excluded from scoring.

2.3.2. Results

2.3.2.1. Experiment 1

Evidence of learning for liquid harmony was found in terms of both error rate and latency. The mean error rates by stimulus type are summarized in Table 2.3. Subjects made significantly fewer errors for legal words than for illegal words ($t(14) = -2.646$, $p = .019$).

	Study	Legal	Illegal	Filler
Error rate (*SD*)	0.036 (0.037)	0.030 (0.059)	0.067 (0.052)	0.025 (0.025)

Table 2.3. Mean error rate (*SD*) by stimulus type for Experiment 1.

The mean latencies by stimulus type and block are summarized in Table 2.4. A within-subjects ANOVA with block (blocks 3~5) and stimulus type as factors showed subjects responded significantly faster to legal words than to illegal words in the three blocks ($F(1,14) = 6.278$, $p = .025$). However, block and legality of the stimulus showed no

significant interaction ($F(2,28) = 1.487$, $p = .243$), suggesting no interaction between the effect of learning and the duration of familiarization. In addition, no reliable difference in latency was found between study and legal words ($F(1,14) = 0.027$, $p = .872$). Thus, the latency data suggests that subjects were able to learn the constraint and generalize it to new instances, although the effect of learning did not vary significantly with respect to the duration of familiarization.

	Study	Legal	Illegal	Filler
Block 1	283 (154)			292 (178)
Block 2	293 (176)			302 (168)
Block 3	284 (157)	271 (159)	305 (193)	298 (178)
Block 4	298 (167)	297 (190)	300 (198)	314 (190)
Block 5	284 (163)	290 (179)	321 (178)	288 (189)

Table 2.4. Mean (*SD*) latency by word type and block in milliseconds for Experiment 1.

2.3.2.2. Experiment 2

Evidence of learning for liquid disharmony was found in terms of latency but not in terms of error rate. The mean error rates by stimulus type are summarized in Table 2.5. Difference in error rate between legal and illegal words was not significant ($t(14) = 1.871$, $p = .082$).

	Study	Legal	Illegal	Filler
Error rate (*SD*)	0.071 (0.055)	0.063 (0.059)	0.030 (0.036)	0.033 (0.042)

Table 2.5. Mean error rate (*SD*) by stimulus type for Experiment 2.

The mean latencies by stimulus type and block are summarized in Table 2.6. Subjects responded significantly faster to legal words than to illegal words in the three blocks ($F(1,14) = 8.435$, $p = .012$). However, block and legality of the stimulus showed no significant interaction ($F(2,28) = 0.111$, $p = .895$). In addition, no reliable difference in latency was found between study and legal words ($F(1,14) = 1.199$, $p = .292$). Thus, the latency data suggests that subjects were able to learn the constraint and generalize it to new instances, although the effect of learning did not vary significantly among blocks.

	Study	Legal	Illegal	Filler
Block 1	428 (253)			436 (235)
Block 2	374 (252)			354 (256)
Block 3	326 (267)	322 (253)	360 (240)	353 (255)
Block 4	347 (237)	313 (257)	359 (242)	358 (229)
Block 5	325 (214)	334 (226)	348 (228)	320 (212)

Table 2.6. Mean (*SD*) latency by word type and block in milliseconds for Experiment 2.

2.3.2.3. Experiment 3

Evidence of learning for backness harmony was found neither in latency nor in error rate. The mean error rates by stimulus type are summarized in Table 2.7. Difference in error rate between legal and illegal words was not significant ($t(14) = 0.269$, $p = .792$).

	Study	Legal	Illegal	Filler
Error rate (*SD*)	0.034 (0.036)	0.037 (0.058)	0.033 (0.046)	0.047 (0.053)

Table 2.7. Mean error rate (*SD*) by stimulus type for Experiment 3.

The mean latencies by stimulus type and block are summarized in Table 2.8. Subjects did not respond significantly faster to legal words than to illegal words in the three blocks ($F(1,14) = 0.164$, $p = .691$). In addition, block and legality of the stimulus showed no significant interaction ($F(2,28) = 0.289$, $p = .751$). Thus, there was no evidence of subjects learning the constraint and generalizing it to new instances, although there was evidence of implicit memory for study words as suggested by subjects responding to studied words significantly faster than to unstudied words (legal, illegal, and filler words) in the three blocks ($F(1,14) = 21.388$, $p < .001$).

	Study	Legal	Illegal	Filler
Block 1	396 (214)			406 (214)
Block 2	352 (212)			393 (199)
Block 3	327 (204)	339 (196)	346 (226)	345 (173)
Block 4	298 (207)	304 (226)	309 (229)	320 (223)
Block 5	283 (214)	341 (214)	332 (186)	337 (211)

Table 2.8. Mean (*SD*) latency by word type and block in milliseconds for Experiment 3.

2.3.2.4. Experiment 4

Evidence of learning for backness disharmony was found neither in latency nor in error rate. The mean error rates by stimulus type are summarized in Table 2.8. Difference in error rate

between legal and illegal words was not significant ($t(14) = -0.924$, $p = .371$).

	Study	Legal	Illegal	Filler
Error rate (*SD*)	0.079 (0.079)	0.033 (0.046)	0.052 (0.049)	0.033 (0.057)

Table 2.8. Mean error rate (*SD*) by stimulus type for Experiment 4.

The mean latencies by stimulus type and block are summarized in Table 2.9. Subjects did not respond significantly faster to legal words than to illegal words in the three blocks ($F(1,14) = 0.188$, $p = .672$). In addition, block and legality of the stimulus showed no significant interaction ($F(2,28) = 0.390$, $p = .681$). Thus, there was no evidence of subjects learning the constraint and generalizing it to new instances, although there was evidence of implicit memory for study words as suggested by subjects responding to studied words significantly faster than to unstudied words (legal, illegal, and filler words) in the three blocks ($F(1,14) = 8.208$, $p = .012$).

	Study	Legal	Illegal	Filler
Block 1	431 (226)			451 (242)
Block 2	363 (217)			391 (213)
Block 3	347 (188)	358 (214)	367 (189)	374 (192)
Block 4	339 (183)	344 (193)	325 (198)	353 (197)
Block 5	309 (157)	330 (163)	319 (166)	339 (213)

Table 2.9. Mean (*SD*) latency by word type and block in milliseconds for Experiment 4.

2.3.3. Summary

Evidence of learning was found for liquid harmony and liquid disharmony but not for backness harmony and backness disharmony. Subjects responded significantly faster to legal words than to illegal words in Experiments 1 and 2. In addition, subjects made significantly fewer errors for legal words than for illegal words in Experiment 1. However, there was no significant difference in latency or error rate between legal and illegal words in Experiments 3 and 4. Evidence of learning found in Experiments 1 and 2 suggests that the adult phonological processing system can adapt to second order constraints contingent on a non-adjacent phoneme, which was suspected to be harder to learn than the second order constraints contingent on the adjacent vowel in Dell et al. (2000) and Onishi et al. (2002).

As far as the four non-adjacent phonotactic constraints are concerned, there seems to be no relation between the typological frequency of a phonotactic constraint and its learnability. If typologically more frequent constraints were more readily learnable, we would expect our subjects to learn backness harmony more readily than liquid harmony and liquid disharmony. We would also expect them to learn liquid disharmony more readily than liquid harmony. However, liquid harmony was learned as readily as liquid disharmony. The stronger evidence against the hypothesized relation is that there was no evidence of learning for backness harmony.

The results appear similar to those of Bonatti et al. (2005), which suggests that speakers may be more sensitive to statistical regularity between non-adjacent consonants than vowels in word-segmentation tasks. In their study, adult speakers of French listened to a continuous stream of tri-syllabic words of the form CV.CV.CV and later performed a word-segmentation task by choosing between a word and a tri-syllabic sequence that spans

word-boundary ("part-words"). In the consonant condition, transitional probability between consonants within a word was 1.0. On the other hand, transitional probabilities between vowels within a word, adjacent syllables, and consonants spanning a word-boundary were 0.5. The vowel condition was the same except that the constrained segments were vowels rather than consonants. They found that subjects preferred words to part-words in the consonant condition but not in the vowel condition.

However, it is also possible that subjects in the current experiments did learn the constraints on high vowels but that the effect of learning did not surface as significantly better performance for legal words than for illegal words. Evidence suggesting that the vowel constraints may have been learned comes from Moreton (2006) and Pycha et al. (2003). Moreton (2006) reports experiments where adult speakers of English learned height harmony between non-adjacent vowels as readily as voicing harmony between non-adjacent consonants. Based on Moreton's results, we would expect that backness harmony can be learned as readily as liquid harmony. Pycha et al. (2003) reports experiments where adult speakers of English learned artificial non-adjacent dependencies between the vowel in CVC stem and the vowel in VC suffix. Both harmonic and disharmonic non-adjacent vowel dependencies were equally well learned. This suggests that backness disharmony can be learned as readily as backness harmony.

However, the subjects in the studies by Moreton (2006) and Pycha et al. (2003) performed a different task for the test words from the subjects in our experiments. Subjects in Moreton (2006) made binary grammaticality judgments (Yes/No) and subjects in Pycha et al. (2003) made a choice between a grammatical item and an ungrammatical item. The difference in evidence of learning regarding the vowel constraints between the two studies and the auditory repetition experiments discussed here could be due to the difference in task. Therefore, the next section presents two experiments testing learnability of the same

constraints but where subjects perform grammaticality judgment task for the test words instead of auditory repetition task.

2.4. Grammaticality judgment experiments

The two grammaticality judgment experiments test learnability of liquid harmony (Experiment 5) and backness harmony (Experiment 6). Learnability of liquid disharmony and backness disharmony is not tested again, because the results from the auditory repetition experiments suggest that whether a constraint is harmonic or disharmonic does not influence learnability. Evidence of learning was found for both liquid harmony and disharmony. Evidence of learning was found neither for backness harmony nor disharmony.

2.4.1. Methods

2.4.1.1. Subjects

Fifteen subjects from the same population as in the auditory repetition experiments participated and received course credit as compensation for each experiment.

2.4.1.2. Materials

The set of stimuli for a subject in each experiment was exactly the same as the set used in the corresponding auditory repetition experiment. For example, the first subject in Experiment 5 was trained and tested on the same set of study, legal, illegal, and filler words used for the first subject in Experiment 1 (liquid harmony). However, the stimuli were organized in a slightly different way as summarized in Table 2.10.

Each training block comprised 16 study words which recurred in all five blocks in the corresponding auditory repetition experiment and eight filler words used in the

corresponding block in the auditory repetition experiment. For example, the filler words in the first training block for the first subject in Experiment 6 were the same filler words in Block 1 for the first subject in Experiment 3 (backness harmony). The test block comprised 18 legal words, 18 illegal words used in the corresponding auditory repetition experiment, and 16 filler words used in the last two blocks (Blocks 4 and 5) in the corresponding auditory repetition experiment. Stimuli were randomly ordered in each block.

	Practice	Train 1	Train 2	Train 3	Test	Total
Study		16	16	16		48
Legal					18	18
Illegal					18	18
Filler	2	8	8	8	16	42
Total	2	24	24	24	52	126

Table 2.10. Design for the grammaticality judgment experiments.

2.4.1.3. Procedures

At the beginning of each experiment, subjects were told that they would first practice a set of non-sense words from an imaginary language, and then would listen to another set of words and decide for each word if it sounded like a word in the imaginary language. In the first three training blocks, subjects listened to each word through headphones and were instructed to repeat each word as quickly and accurately as possible. In the test block, for each word they listened to, subjects were asked to press ‘1’ on the keyboard if the word sounded like it was from the imaginary language and press ‘0’ if it did not.

2.4.1.4. Scoring

To determine if subjects could discriminate legal words from illegal words apart from their bias, we measured *d′* for each subject using signal detection theory (e.g., Macmillan, 1993; Stanislaw and Todorov, 1999). We define hit-rate (*H*) as the proportion of trials where subjects responded '1' to legal words, and false-alarm rate (*F*) as proportion of trials where subjects responded '1' to illegal words. A subject's *d′* can then be computed by *Z(H) – Z(F),* where the function *Z(x)* is the inverse-normal transform. However, two subjects in the liquid harmony condition had a hit rate of 1.0 whose *Z(H)* results in an infinite value. As their false-alarm rates were not zero, the extreme cases were considered to be sampling variability and the log-linear rule (e.g., Hautus, 1995) was applied in computing the hit-rate and the false-alarm rate by adding 0.5 to the number of hits or false-alarms and dividing it by total number of legal words or illegal words plus one, respectively.

2.4.2. Results

Mean *d′* scores (*SD*) were 0.761 (0.793) for the subjects in the liquid harmony condition and 0.488 (0.557) for the subjects in the backness harmony condition. One sample *t*-test with the null hypothesis being "*d′* = 0.0" showed that subjects discriminated legal words from illegal words both in the liquid harmony condition ($t(14) = 3.717$, $p = .002$), and in the backness harmony condition ($t(14) = 3.399$, $p = .004$). Furthermore, independent samples *t*-test between subjects in the two conditions showed no difference in their performance in discriminating legal words from illegal words ($t(28) = 1.089$, $p = .285$).

2.4.3. Summary

Evidence of learning was found for both liquid harmony (Experiment 5) and backness harmony (Experiment 6). Moreover, the two constraints were learned equally well. This suggests that the absence of evidence of learning for backness harmony and backness disharmony in the auditory repetition experiments may be due to some factor specific to auditory repetition task but not due to subjects' failure to learn the phonotactic constraints. Recall that whether a constraint was harmonic or disharmonic did not lead to any difference in evidence of learning in the auditory repetition experiments. In addition, adult speakers of English in Pycha et al. (2003) learned both backness harmony and backness disharmony equally well. These results collectively suggest that the four non-adjacent phonotactic constraints in this chapter do not differ in their learnability despite the differences in their typological frequency.

2.5. Chapter summary

This chapter presented four auditory repetition experiments and two grammaticality judgment experiments investigating if adults can learn four non-adjacent phonotactic constraints: liquid harmony, liquid disharmony, backness harmony, and backness disharmony. The results from the six experiments collectively show that adults can learn the four non-adjacent phonotactic constraints can be learned from brief experience. In addition, contrary to what the conjectured relation between learnability of a sound pattern and its typological frequency would suggest, the results show that there is no difference in learnability between the four phonotactic constraints despite the differences in their typological frequency.

In the auditory repetition experiments, evidence of learning was found for liquid harmony and disharmony but not for backness harmony and disharmony. Subjects responded to legal words significantly faster than illegal words when tested on liquid harmony (Experiment 1) and liquid disharmony (Experiment 2). In addition, subjects made significantly fewer errors for legal words than for illegal words when tested on liquid harmony (Experiment 1). However, difference in subjects' latency or error rate between legal and illegal words was not significant when tested on backness harmony (Experiment 3) and backness disharmony (Experiment 4).

To test if the asymmetry in evidence of learning is indeed due to a difference in learnability between constraints on liquids and constraints on high vowels, learnability of liquid harmony and backness harmony was tested again in grammaticality judgment experiments. Evidence of learning was found for both constraints, and moreover the two constraints were learned equally well. Subjects showed discriminability between legal and illegal words for both liquid harmony (Experiment 5) and backness harmony (Experiment 6), and there was no difference in discriminability between the subjects in the two experiments.

The results add to the growing body of evidence on malleability of the adult phonological processing system and extend the range of the sound patterns which the adult phonological processing system can learn from brief experience. As the four constraints restrict co-occurrence of phonemes in two non-adjacent positions, they may be harder to learn than the second-order constraints contingent on the adjacent vowel in Dell et al. (2000) and Onishi et al. (2002), because identifying the dependency between non-adjacent positions necessitates the learner to scan a longer phoneme sequence and use more processing resources. The results show that the adult phonological processing system can not only learn second order phonotactic constraints contingent on the adjacent vowel but

also the constraints contingent on a non-adjacent phoneme. Moreover, while previous studies investigating learnability of non-adjacent phonological dependencies demonstrated learning using explicit judgment tasks (e.g., Newport and Aslin, 2004), the results here show that the learning of the non-adjacent phonotactic constraints also has effect on implicit tasks such as perception.

CHAPTER 3

PERCEPTUAL FACILITATION EXPERIMENTS

The experiments in Chapter 2 show that the adults can learn non-adjacent phonotactic constraints from brief experience, and as a consequence of learning, they perceive phonotactically legal words faster than phonotactically illegal words. This facilitative effect of phonotactic learning on perception of legal words is expected as shown in previous studies on phonotactic learning (e.g., Onishi et al., 2002). However, the results in Chapter 2 also show an interesting asymmetry: it is not clear why legal words are perceived faster than illegal words only when the constraints on liquids are learned but not when the constraints on high vowels are learned. In a narrow interpretation of Chomsky's distinction between competence and performance (Chomsky, 1965), while the grammatical competence of the phonological processing system has been changed by learning backness harmony, the change is not reflected in the actual performance of the phonological processing system. To explain the asymmetry in change in perceptual performance, I propose a perceptual facilitation hypothesis and present two auditory repetition experiments in support of the hypothesis.

The chapter is organized as follows. Section 3.1 introduces the perceptual facilitation hypothesis. Section 3.2 illustrates how the hypothesis explains the asymmetry found in Chapter 2. Section 3.3 presents two auditory repetition experiments in support of the hypothesis. Section 3.4 summarizes and discusses the experimental results.

3.1. Perceptual facilitation hypothesis

The perceptual facilitation hypothesis is that the knowledge of a phonotactic constraint

facilitates perception of phonotactically legal speech sounds if the constraint restricts co-occurrence of phonemes that are perceptually more confusable to each other. The name implies that the asymmetry in results of the auditory repetition experiments is better explained in terms of how perception is facilitated by the knowledge of phonotactic constraint rather than whether the phonotactic constraint is learned or not.

Non-words or sub-lexical units are perceived more accurately (e.g., Brown and Hildum, 1956) and more quickly (e.g., Auer, 1993; Vitevitch and Luce, 1998) if they are consistent with the speakers' knowledge of phonotactic constraints. This is often explained in psycholinguistic models of speech perception (e.g., McClelland and Elman, 1986; Norris, 1990; Luce et al., 2000) as increase in likelihood or activation of lexical or sub-lexical units consistent with the phonotactic constraints. The same idea is implemented in stochastic models such as Hidden Markov Models for automatic speech recognition (e.g., Rabiner and Juang, 1993; Jelinek, 1997).

The assumption underlying the approach to speech perception in the psycholinguistic models and recognition in automatic speech recognition systems is that the perception process is equivalent to the process of choice. There are multiple lexical or sub-lexical units to choose from for a given acoustic / auditory / perceptual input. If the likelihood or the activation level of a unit is higher, then it is more likely to be chosen as the result of perception / recognition. From a different perspective, if the likelihood of the correct unit for the input is higher, the unit is perceived better.

The likelihood that a phoneme will be incorrectly identified as a different phoneme yields perceptual confusion. By increasing the likelihood of the correct phoneme at a word position, potential perceptual confusion at the position is reduced. On the other hand, decreasing the likelihood of the correct phoneme increases the confusion at the position. This could be how the knowledge of a phonotactic constraint facilitates perception of legal

words to be better than illegal words. For example, given the stimulus /sa.la.la/, knowledge of liquid harmony increases the likelihood of the second /l/ and decreases the likelihood of /r/ in the same position, thereby reducing the potential confusion between /l/ and /r/ at the word position.

The argument of the perceptual facilitation hypothesis is that if there is little room for confusion at the word position, there is little room for the phonotactic knowledge to facilitate perception. The likelihood of the correct phoneme is near the ceiling of performance while the likelihood of the other candidate phonemes is near the floor. Therefore, knowledge of a phonotactic constraint facilitates perception more if the constrained phonemes are perceptually more confusable.

3.2. Phoneme confusability

In order for the perceptual facilitation hypothesis to explain the asymmetry in the auditory repetition experiments, the constrained phoneme pair /l/ and /r/ must be perceptually more confusable and their likelihood must be farther away from the ceiling and the floor than the pair /i/ and /u/. The argument that the liquids are more confusable or similar to each other than high vowels has both empirical and theoretical support.

Empirical support comes from Luce (1986), who reports a set of confusion experiments to investigate perceptual similarity between consonants and vowels in English. To this end, a male speaker of English recorded 345 CV syllables and 330 VC syllables combined out of 23 onset consonants, 15 vowels, and 22 coda consonants, covering all possible such syllables in English. The three types of phonemes are listed in Table 3.1.

onset	p, t, k, b, d, g, ʧ, ʤ, s, ʃ, z, f, θ, v, ð, h, n, m, l, r, w, j
vowel	i, ɪ, ɛ, e, æ, a, aʊ, aɪ, ʌ, ɔ, oɪ, oʊ, ʊ, u, ɚ
coda	p, t, k, b, d, g, ʧ, ʤ, s, ʃ, z, f, θ, v, ð, n, m, l, r, ŋ, ʒ

Table 3.1. List of English phonemes in confusion experiments in Luce (1986).

The syllables were presented to each of 120 adult speakers of English at three different signal-to-noise ratios (SNR): +15 dB, +5 dB, and -5 dB. The task of the subjects was to identify the consonants or vowels in each syllable. The responses from subjects for the presented phonemes at three noise levels are summarized as confusion matrices.

The confusion matrices suggest that in our experimental setting the liquids are more likely to be confused as each other than the high vowels. Table 3.2 summarizes the relevant portion of the matrices in terms of percentage of mutual confusion. For example, the original matrices specify the number of times subjects confused /l/ as /r/ and /r/ as /l/ separately, and there certainly is asymmetry in confusion. In Table 3.2, I combined the confusion frequencies in both directions and specify the percentage of confusion rather than frequency for better comparison as the number of trials for consonants was different from that for vowels in Luce (1986). As the stimuli in the auditory repetition experiments are of the form CV.CV.CV, the consonant confusability is based only on the confusion matrix for onsets.

	SNR = +15 dB	SNR = +5 dB	SNR = -5 dB
/i/ vs. /u/	0.4 %	7.0 %	34.4 %
/l/ vs. /r/	8.0 %	11.0 %	13.0 %

Table 3.2. Confusability of liquids and high vowels summarized from Luce (1986).

Table 3.2 suggests that liquids are more likely to be confused than the high vowels in our experimental setting. Recall that the stimuli were recorded at a sound-proof booth and digitized at 44.1 KHz and 16 bit rate, and that they were presented to subjects through a headphone in the booth. Therefore, the confusability reflects the confusability in our experimental setting better when the noise level is at +15 dB than when the noise level is at +5 dB or -5 dB. The liquids are confused as each other 8.0 % of the time while the high vowels are confused as each other 0.4 % of the time when noise level is at +15 dB.

Theoretical support comes from Frisch et al. (1997) whose similarity metric suggests that the liquids are phonologically more similar to each other than the high vowels. Assuming a specific feature representation scheme, similarity between two phonemes is computed as follows. All features used to describe a phoneme are listed in a set. The complete list of subsets of the set constitutes the power set for the phoneme. Similarity between two phonemes is computed by dividing the size of the intersection of two power sets by the size of the union of two power sets.

The metric is not a direct predictor of perceptual confusability per se, but Bailey and Hahn (2005) argues that such a theoretical similarity metric can be a better predictor than the metrics based on empirical data when applied to various tasks where phonological confusability and/or similarity is assumed to play an important role. Accordingly, the metric can be interpreted as a theoretical estimation of perceptual confusability.

In conjunction with the SPE feature specification (Chomsky and Halle, 1968), similarity between /l/ and /r/ is 0.5407, while similarity between /i/ and /u/ is 0.2674 according to the similarity metric.[10] Therefore, the liquids are theoretically predicted to be

[10] All similarity scores in the dissertation were computed by a Perl script written by Adam Albright (Albright, 2003) available at http://web.mit.edu/albright/www/.

more similar to each other than the high vowels. Thus, there is empirical and theoretical support to assume that the liquids are more confusable to each other than the high vowels.

3.3. Perceptual facilitation experiments

The perceptual facilitation hypothesis predicts that knowledge of a phonotactic constraint facilitates perception more if the constrained phonemes are more confusable to each other. Specifically, the difference in perceptual latency or accuracy between legal and illegal words would be greater if the constrained phonemes were more confusable.

Two auditory repetition experiments are presented below as evidence in support of the hypothesis. Experiment 7 tests the prediction that if the constrained phonemes are less confusable to each other, the difference in perceptual latency or accuracy between legal and illegal words will be smaller. Experiment 8 tests the prediction that if the constrained phonemes are more confusable to each other, the difference in perceptual latency or accuracy will be bigger.

3.3.1. Experiment 7

Experiment 7 is an auditory repetition experiment identical to Experiment 1 (liquid harmony) except that the phonotactic constraint restricts co-occurrence of /l/ and /m/ instead of /l/ and /r/. For example, /sa.la.la/ and /ke.ma.ma/ are phonotactically legal, while /sa.la.ma/ or /ke.ma.la/ are phonotactically illegal. The pair is arguably perceptually less confusable to each other than the liquids. The prediction of the perceptual facilitation hypothesis is that the difference in perceptual latency or accuracy between legal and illegal words would be smaller than the difference found in the liquid harmony experiment, and

may fail to reach significance.

3.3.1.1. Rationale

The constrained phoneme pair is chosen to be /l/ and /m/ primarily because the pair is less confusable to each other than the liquids are. The degree of confusability between /l/ and /m/ is, in fact, similar to the degree of confusability between /i/ and /u/. Moreover, the degree of confusability between /m/ and the two filler consonants that may also occupy the same position (/s/ and /k/) is similar to the degree of confusability between /r/ and the two consonants. Table 3.3 shows how perceptually confusable at SNR = +15 dB (Luce, 1986) and phonologically similar (Frisch et al., 1997) the pair /l/ and /m/ is compared to the liquid pair and the high vowel pair. Table 3.4 shows that confusability and similarity between /m/ and the two filler consonants are similar to confusability and similarity between /r/ and the two consonants.

	/l/ vs. /m/	/l/ vs. /r/	/i/ vs. /u/
confusability	1.7 %	8.0 %	0.4 %
similarity	0.1579	0.5407	0.2674

Table 3.3. Confusability and similarity between /l/ and /m/ as opposed to the liquid pair, and the high vowel pair.

	/r/ vs. /s/	/m/ vs. /s/	/r/ vs. /k/	/m/ vs. /k/
confusability	0.0 %	0.03 %	0.0 %	0.0 %
similarity	0.1638	0.1373	0.0938	0.1266

Table 3.4. Confusability and similarity against /s/ and /k/ between /r/ and /m/.

3.3.1.2. Methods

3.3.1.2.1. Subjects

As in Experiment 1, 15 adult native speakers of English were recruited. Four subjects were paid by cash, two subjects received course credit for an introductory Linguistics class, and nine subjects received course credit for an introductory course in Psychology.

3.3.1.2.2. Materials

Exactly the same set of words was used as in Experiment 1, except that all instances of /r/ were replaced by /m/.

3.3.1.2.3. Procedures

The procedure was the same as in Experiment 1.

3.3.1.2.4. Scoring

Scoring was done in the same way as in Experiment 1. Latencies for each subject were averaged by block and stimulus type with the following excluded: errors (M = 5.20 trials per subject), responses not recognized by the microphone at first attempt (M = 1.53 trials per subject), responses made before the onset of the final syllable (M = 0.27 trials per subject), and latencies 2.5 SD beyond each subject's mean latency (M = 3.40 trials per subject).

3.3.1.3. Results

The mean latencies by stimulus type and block are summarized in Table 3.5. Subjects did not respond significantly faster to legal words than to illegal words in the three blocks

($F(1,14) = 0.206$, $p = .657$). In addition, block and legality of the stimulus showed no significant interaction ($F(2,28) = 0.972$, $p = .391$). Thus, there was no evidence of subjects learning the constraint and generalizing it to new instances in terms of difference in latency between legal words and illegal words.

	Study	Legal	Illegal	Filler
Block 1	466 (144)			460 (152)
Block 2	469 (143)			468 (183)
Block 3	449 (166)	451 (154)	453 (134)	437 (117)
Block 4	448 (159)	430 (167)	449 (181)	445 (130)
Block 5	442 (153)	445 (164)	437 (152)	447 (146)

Table 3.5. Mean (*SD*) latencies by type and block in milliseconds for Experiment 7.

3.3.1.4. Summary

The constraint in Experiment 7 restricted co-occurrence of non-adjacent /l/ and /m/. The constrained phonemes were less confusable to each other than the liquids constrained in Experiment 1. The perceptual facilitation hypothesis predicts that the difference in how well words are perceived between legal and illegal words would be smaller in Experiment 7 than in Experiment 1. The result is consistent with the prediction. The difference in latency between legal and illegal words failed to reach significance, while it reached significance in Experiment 1.

3.3.2. Experiment 8

Experiment 8 is an auditory repetition experiment identical to Experiment 3 (backness harmony) except that words in the last three blocks are presented with noise in the background. The high vowel pair is perceptually more confusable to each other with noise in the background than without noise. The prediction of the perceptual facilitation hypothesis is that the difference in perceptual latency or accuracy between legal and illegal words would be larger than the difference found in Experiment 3.

3.3.2.1. Rationale

In addition to testing the validity of the perceptual facilitation hypothesis, the experiments in this chapter also tests if the asymmetry found in Chapter 2 is concerned with whether the constrained phonemes are consonants or vowels. To meet both ends, the current experiment must test a non-adjacent phonotactic constraint on a vowel pair that is more confusable than the high vowel pair, and as confusable as the liquid pair. The words must satisfy the CV.CV.CV template, and if we excluded the vowels that cannot end an English syllable (/ɪ/, /ɛ/, /æ/, /ʌ/, /ʊ/), no such vowel could be found in Luce's confusion matrix at SNR = +15 dB. However, the high vowel pair becomes more confusable as noise is added, as illustrated in Table 3.6.

	SNR = +15 dB	SNR = +5 dB	SNR = -5 dB
/i/ vs. /u/	0.4 %	7.0 %	34.4 %
cf.) /l/ vs. /r/	8.0 %	11.0 %	13.0 %

Table 3.6. Confusability between the high vowel pair at various noise levels.

Note that the confusability between the high vowels at SNR = +5 dB is similar to the confusability between the liquids at SNR = +15 dB. Therefore, in Experiment 8, confusability between the constrained vowels is increased by adding white noise to the recorded stimuli at SNR = +5 dB.

3.3.2.2. Methods

3.3.2.2.1. Subjects

Fifteen adult native speakers of English participated and received course credit as compensation.

3.3.2.2.2. Materials

Exactly the same set of words was used as in Experiment 3, except that white noise was added at SNR = +5 dB to the stimuli presented to subjects in blocks 3 through 5. Noise was not added in the first two blocks to provide subjects with the same learning opportunity as the subjects in Experiment 3.

3.3.2.2.3. Procedures

The procedure was the same as in Experiment 3.

3.3.2.2.4. Scoring

Subjects produced many errors when the stimuli were presented with noise in the background. Mean error rate in the first two blocks, where stimuli were presented without noise, was 0.074. On the other hand, mean error rate in the latter three blocks, where stimuli were presented with noise, was 0.660. 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 impossible to compare latency between legal and illegal words. Therefore, perceptual facilitation was measured in terms of difference in error rate instead of latency between legal and illegal words.

3.3.2.3. Results

The mean error rates by stimulus type and block are summarized in Table 3.7.

	Study	Legal	Illegal	Filler
Block 1	0.092 (0.081)			0.083 (0.102)
Block 2	0.079 (0.090)			0.017 (0.044)
Block 3	0.596 (0.131)	0.600 (0.225)	0.844 (0.160)	0.650 (0.143)
Block 4	0.642 (0.148)	0.733 (0.216)	0.744 (0.217)	0.642 (0.176)
Block 5	0.621 (0.143)	0.644 (0.124)	0.900 (0.105)	0.558 (0.240)

Table 3.7. Mean (*SD*) error rates by type and block in milliseconds for Experiment 7.

The mean error rate was significantly smaller for legal words than for illegal words in three blocks ($F(1,14) = 31.381$, $p < .001$).

3.3.2.4. Summary

The constraint in Experiment 8 was the same backness harmony as in Experiment 3. However, as white noise was added to stimuli presented in the last three blocks, the constrained phonemes in the test words (legal, illegal words) were more confusable to each other than the constrained phonemes in Experiment 3. The perceptual facilitation hypothesis predicts that difference in how well words are perceived between legal and

illegal words would be larger in Experiment 8 than in Experiment 3. The result was consistent with the prediction. Subjects made significantly fewer errors for legal words than for illegal words, while difference in neither latency nor error rate reached significance in Experiment 3.

3.4. Chapter summary

The chapter presented the perceptual facilitation hypothesis and two auditory repetition experiments in support of the hypothesis. The perceptual facilitation hypothesis predicts that knowledge of phonotactic constraints facilitates perception of phonotactically legal words more if the constrained phonemes are more confusable to each other. The hypothesis explains the asymmetry found in the results of the auditory repetition experiments discussed in Chapter 2. The asymmetry was that the difference in perceptual latency between legal and illegal words reached significance when the constrained phoneme pair was /l/ and /r/, but failed to reach significance when the constrained phoneme pair was /i/ and /u/. Studies on perceptual confusability and phonological similarity suggest that /l/ and /r/ is a more confusable pair than /i/ and /u/. Therefore, the hypothesis predicts that the difference in perceptual latency will be greater for liquid harmony/disharmony than for backness harmony/backness disharmony.

The prediction of the hypothesis was further tested by two auditory repetition experiments (Experiment 7 and 8) in comparison with Experiment 1 (liquid harmony) and Experiment 3 (backness harmony) in Chapter 2. The constrained phoneme pair in Experiment 7 was /l/ and /m/, which is a consonant pair less confusable than /l/ and /r/ in Experiment 1. As predicted by the hypothesis, difference in perceptual latency between legal and illegal words failed to reach significance in Experiment 7, while it reached

significance in Experiment 1. The constrained phoneme pair in Experiment 8 was /i/ and /u/, identical to Experiment 3. However, noise was added to the test words so that they became more confusable than the test words in Experiment 3. As predicted by the hypothesis, difference in perceptual accuracy between legal and illegal words reached significance in Experiment 8, difference in neither accuracy nor latency reached significance in Experiment 3.

The perceptual facilitation hypothesis provides a more detailed account of how phonotactic knowledge changes the perceptual behavior of the adult phonological processing system. The process is more subtle than "phonotactic knowledge facilitates perception of phonotactically legal sound patterns" as previously noted. The basis of the hypothesis is the claim that phonotactic knowledge reduces perceptual confusion inherent in phonotactically legal words, and possibly increases perceptual confusion in phonotactically illegal words. Therefore, if there is more room for perceptual confusion, the difference in perceptual latency or accuracy between legal and illegal words will be greater.

CHAPTER 4

CONNECTIONIST MODEL OF PHONOTACTIC LEARNING AND PERCEPTUAL FACILITATION

The results in the previous two chapters show that the adult phonological processing system can rapidly learn non-adjacent phonotactic constraints. As a result of learning, adults judge phonotactically legal novel words as more grammatical than illegal novel words, and perceive the legal words more rapidly and accurately than the illegal words. Moreover, perception of legal words is facilitated more if the constrained phonemes are more confusable, as summarized in the perceptual facilitation hypothesis. In Chapters 4 and 5, I demonstrate how this change in the adult phonological processing system can be computationally modeled.

The experiments with human subjects in Chapter 2 suggest that the same exposure to the words embodying phonotactic constraints can affect the behavior of the adult phonological processing system differently depending on which task the system performs. Specifically, the subjects were exposed to the same set of words embodying the backness harmony in both Experiment 3 and Experiment 6 in the familiarization phase. The difference, however, was that the subjects performed different tasks in the test phase: auditory repetition task in Experiment 3 and grammaticality judgment task in Experiment 6. Evidence of learning was observed in Experiment 6 but not in Experiment 3.

This difference in the effect of phonotactic learning can be explained by assuming that perception and grammaticality judgment take place in two separate stages. A speaker first identifies the phonological structure that best matches the perceptual input, and then computes the grammaticality of the identified phonological structure. The process in the first stage corresponds to perception, and the process in the latter stage corresponds to

grammaticality judgment. Phonotactic learning affects perception by changing how quickly and accurately the phonological processing system identifies the phonological structure that best matches the perceptual input; the system will be better at identifying a novel phonological structure if it has cumulated experience identifying similar phonological structures from the perceptual input. At the same time, by storing the phonological structures of utterances in a particular language, regardless of whether they are the results of auditory perception or visual perception, the system learns the properties that hold common among the stored phonological structures and explicitly refer to the properties to decide if a new utterance is a member of the same language or not.

Accordingly, I present two separate models of the effect of phonotactic learning on the behavior of the adult phonological processing system: a connectionist model of the effect of phonotactic learning on perception in Chapter 4 and a Bayesian belief network model of the effect of phonotactic learning on grammaticality judgment in Chapter 5. However, as will be shown in Chapter 5, the two models can be seamlessly integrated, illustrating how speakers first identify the phonological structure that best matches the perceptual input and then directly compute the grammaticality of the identified phonological structure.

The connectionist model in this chapter demonstrates how phonotactic learning affects the perceptual behavior of the adult phonological processing system. In particular, it is a connectionist formulation of the perceptual facilitation hypothesis. The model is a single-layered [11] recurrent perceptron that learns to map perceptual input to its

[11] The term “single-layered” is sometimes used ambiguously in the literature. Some use it to describe artificial neural networks with a single hidden layer. Others use it to describe networks without any hidden layers, the only “single” layer being the output layer besides

corresponding phonological representation. In conjunction with several assumptions on input/output representation and interpretation of the model's behavior, the model correctly duplicates the interaction between confusability and phonotactic knowledge observed in the auditory repetition experiments. Furthermore, it provides the input for the Bayesian belief network described in the next chapter, which duplicates the lack of such interaction in the grammaticality judgment experiments.

The chapter is organized as follows. Section 4.1 summarizes the basic idea and assumptions underlying the model. Section 4.2 briefly introduces the basics of artificial neural networks crucial to understanding the current model. Section 4.3 describes how the basic idea is implemented in the model as a single layered recurrent perceptron. Section 4.4 gives a simple illustration of the model's ability to simulate the interaction between perceptual confusability and phonotactic knowledge. Section 4.5 presents the model's simulation of the auditory repetition experiments. Section 4.6 provides a summary of the chapter.

4.1. Basic idea and assumptions

The model implements the perceptual facilitation hypothesis such that knowledge of a phonotactic constraint facilitates perception more if the constrained phonemes are more confusable to each other. In the auditory repetition experiments, the degree of perceptual facilitation is measured by the difference in subjects' mean latency and/or accuracy between legal and illegal words. The model estimates how well, either in terms of accuracy or latency, a subject will perceive a word after he/she has learned the phonotactic constraint.

---

the input layer. I use the term to refer to artificial neural networks without hidden layers.

We let the model make the estimation for all words presented to subjects in the experiment, and compute the mean estimates for legal words and for illegal words. We approximate the degree of perceptual facilitation from phonotactic learning by the difference between the two means. The model expresses the hypothesis by predicting that the difference will be greater if the constrained phonemes are more confusable.

To develop a model that makes the estimate for a given word, we make specific assumptions regarding the following: (1) how perceptual input is represented, (2) what the model does when the input is presented, and (3) how to interpret the model's behavior to derive the estimate. The first assumption will be elaborated in section 4.1.1. The latter two assumptions will be elaborated in section 4.1.2.

4.1.1. Representation of the perceptual input

The perceptual information about a word unfolds to the listener over time. If we divided the temporal duration of a word into frames, at each time-frame the listener would process perceptual information regarding the portion of the input spanning the time-frame. We assume that the portion of the input at each time-frame corresponds to a constituent phoneme of the word.

There is a possible set of phonemes that matches the input at the given time-frame. That is, the input is ambiguous, although some phonemes are more likely than others. The goal of speech perception is to resolve the ambiguity and identify the best candidate phoneme from multiple phonemes with different likelihoods. The likelihood of a phoneme is approximated by how often listeners identify or confuse the input as that particular phoneme relative to other phonemes in the language.

Therefore, we represent the input to the model at each time-frame as a vector of

relative confusability. Each component of the vector represents a phoneme in the language. The value of each component reflects the relative confusability of the phoneme represented by the component with respect to the actual phoneme presented as the input. The relative confusability is computed in the same way as how phoneme similarity is computed in Luce et al. (2000). Suppose the actual phoneme is represented by the i^{th} vector component, and S_{ij} denotes how often it is confused as the phoneme represented by the j^{th} vector component. The value of the k^{th} component, act_k, is computed as in (1).

$$act_k = \frac{S_{ik}}{\sum_j S_{ij}} \quad (1)$$

For example, suppose that a language has three phonemes *{A, B, C}*, and that out of ten times *A* was presented, listeners of the language reported that they heard the three phonemes seven times, twice, and once, respectively. If the input phoneme presented at the current time-step were *A*, the vector representation of the input would be <0.7, 0.2, 0.1>. As the input at each time-frame spans a phoneme, a word will be presented to the model as a sequence of such vectors, where each vector represents the perceptual confusion that must be reduced to correctly identify each component phoneme.

4.1.2. Task of the model and interpretation of the model output

The model changes the likelihoods in the input vectors in two ways. Firstly, the model weights the likelihood of each phoneme at each time-frame. Secondly, the model stores the likelihoods of the phonemes in the previous time-frames and uses that information to change the likelihoods of the phonemes in the current time-frame.

The first change in likelihood reflects how the model evaluates the likelihood of each phoneme specified in the input vector to identify the correct phoneme. Positively weighting the likelihood of phoneme *X* means that its presence in the input vector promotes identifying the correct phoneme. Negatively weighting the likelihood of the same phoneme means that its presence in the input vector hinders identifying the correct phoneme. The second change in likelihood reflects how the model applies the knowledge of sequential phonotactic constraints, such as the non-adjacent phonotactic constraints in our experiments, to facilitate perception.

The change in likelihood is made for each input vector sequentially presented to the model. Since there is a separate vector for each component phoneme of a word, the final likelihood of phoneme *X* in the n^{th} vector represents how likely the model is to choose *X* as the n^{th} phoneme of the word. This is the model's estimate of how likely a subject is to perceive or choose *X* as the n^{th} phoneme of the stimulus word.

We assume that the likelihood of choosing a phoneme at a given time-frame is inversely related to how fast a subject responds to the phoneme. Similar assumptions on the relation between choice probability and latency have been made in connectionist models of cognitive processes. Studies such as Cleeremans and McClelland (1991), Luce et al. (2000), and Gureckis and Love (2005) directly assume that the behavioral latency is inversely related to the choice probability.

Other studies (e.g., Roelofs, 1997; Mirman et al., 2005) assume a less obvious relation. Mirman et al. (2005) assumes that a threshold value of choice probability must be reached in order for the chosen response to take place. The choice probability increases gradually over cycles and the simulated latency is computed by counting the number of cycles after which the choice probability reaches the threshold. Because the number of cycles to threshold decreases as the probability increases, higher choice probability at a

given time-frame implies a shorter latency.

Following Luce, R. D. (1986), Roelofs (1997) assumes that the choice probability at the n^{th} time-frame is equivalent to the hazard rate expressing the probability of making the choice at the current time-frame while the choice was not made in the previous time-frames. That is,

$$h_n(choice) = P(choice, t = n \mid \neg choice, t = i, i < n) \quad (2)$$

Assuming the duration of each time-frame is Δt, The probability of the choice being made by the n^{th} time-frame, *P(choice, t = n)*, and the expected latency, *E(choice)*, are computed as follows.

$$P(choice, t = n) = h_n(choice) \times \prod_{t=1}^{n-1} 1 - h_t(choice) \quad (3)$$

$$E(choice) = \sum_{n=1}^{\infty} P(choice, t = n) \cdot n \cdot \Delta t \quad (4)$$

We follow the simple assumption that behavioral latency is inversely related to the choice probability. In addition, we make the intuitive assumption that behavioral accuracy is positively related to the choice probability. Therefore, if the model's estimate of the choice probability of the n^{th} phoneme of the word is higher, we interpret it to mean that the subject perceives the phoneme more rapidly and more accurately. To approximate how quickly and accurately a subject perceived the whole word, we multiply the choice probabilities computed for each constituent phoneme of the word.

4.2. A very brief introduction to artificial neural networks

We implement the ideas and assumptions introduced in the previous section with a single layered recurrent perceptron. The basics of artificial neural networks crucial to understanding the perceptron are introduced below. For a more comprehensive reference, see Haykin (1999), for example.

4.2.1. Perceptron

The perceptron is the simplest form of feed-forward neural networks, which is a linear classifier that maps its input vector to a binary output value by applying a linear threshold function to the weighted sum of input. That is, binary classification is equivalent to the sign of *f(x)*, where *f(x)* is computed by

$$f(x) = w \bullet x - \theta = \sum_{j=1}^{|x|} w_j x_j - \theta \qquad (5)$$

where x denotes the input vector, w denotes the vector of weights of each component of x, and θ denotes the threshold.

The idea that this classifier is a type of artificial neural networks becomes clear if we picture it as a directed graph of nodes as in Figure 4.1. In Figure 4.1, the nodes labeled $input_j$ ($1 \leq j \leq |x|$) denote the corresponding input features and their values are externally specified by x_j. The edges emanating from these nodes are labeled with the corresponding weights w_j. The node labeled *bias* always takes the value of one and its edge is weighted by $-\theta$, which represents the threshold. The value of the node labeled *output* is the result of the

classification. The value of each node is multiplied by the weight of the edge emanating from the current node and is passed onto the destination node, in this case the output node. The value of the output node is the sum of these incoming values. The term *artificial neural network* is understandable from this graph representation if we assume the nodes represent *artificial neurons* and the edges represent connections between the neurons, establishing a *network* of artificial neurons.

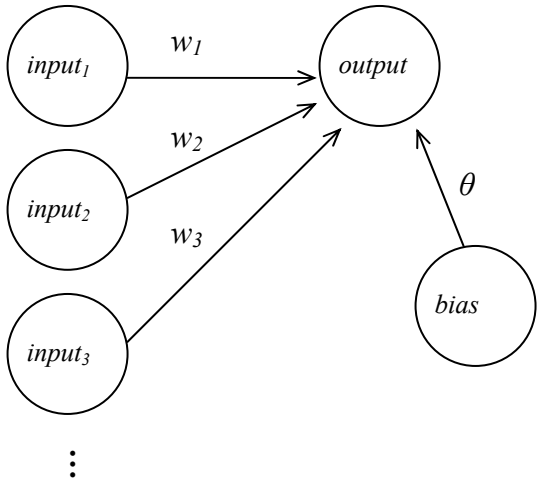

Figure 4.1. Representation of the linear classifier as a directed graph.

The model in this dissertation consists of perceptrons of a slightly different type. Instead of applying the sign function to the weighted sum, we apply the sigmoid function to the weighted sum, where the sigmoid function, *σ(x)* is defined as follows.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{6}$$

The sigmoid function is used in the current model instead of the sign function mainly for three reasons. Firstly, it is a continuous function of the weighted sum to a value within the range between 0 and 1. Secondly, it introduces non-linearity in the model. These two properties, illustrated in Figure 4.2, are important for the current model because we approximate the performance measures in the auditory repetition experiments based on the result of perceptron classification. Latency and accuracy are both continuous measures of performance and the ceiling effect found in the auditory repetition experiments suggests that there may be non-linearity in task performance.

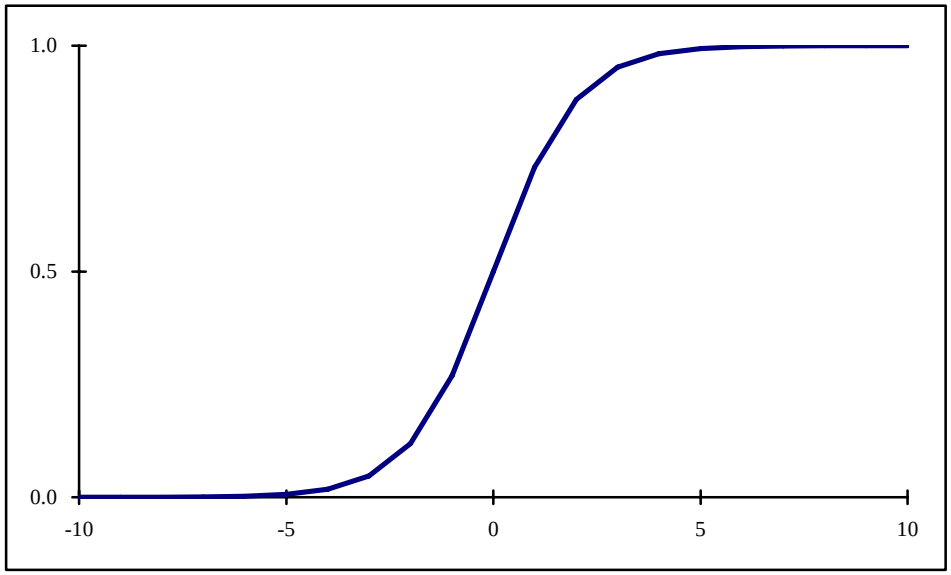

Figure 4.2. The sigmoid curve.

Finally, its derivative expressed in (7) is easy to compute in terms of its output.

$$\frac{d}{dx}\sigma(x) = \sigma(x)(1-\sigma(x)) \qquad (7)$$

This property is useful because both the sigmoid function of weighted sum and its derivative must be computed for training the current model using the delta rule described in the next section.

4.2.2. Training the perceptron

Since the input values are externally given, the perceptron relies on its connection weights for classification. Therefore, learning to correctly classify examples is equivalent to identifying the set of weights that yields optimal classification performance. Many such learning algorithms have been proposed for perceptron and its variants in addition to the initial algorithm proposed by Rosenblatt (1958). The algorithm used for the current model is the delta rule which is a supervised learning algorithm using gradient descent to identify the set of weights that minimizes the amount of classification error on the training data.

Gradient descent is an algorithm that finds a local minimum of the function by taking steps proportional to the negative of the gradient at the current position. Simply put, we go down the slope of the curve until we reach the bottom. The delta rule reduces classification error of the perceptron by computing the partial derivative of the error function with respect to each weight and updating the weight by some proportion of the negative of that derivative. Suppose we have a perceptron that classifies input vectors of dimension n and is activated by the sigmoid function $\sigma(x)$ defined in (6). We define in (8) the classification error function with respect to $y(x)$ which is the desired classification of the input vector x.

$$E = \frac{1}{2}\left(y(x) - \sigma\left(\sum_{i=1}^{n} w_i x_i \right) \right)^2 \qquad (8)$$

The gradient of the error with respect to the weight of the connection from the j^{th} component of the input is computed as in (9).

$$\begin{aligned}
\frac{\partial E}{\partial w_j} &= \frac{\partial(\frac{1}{2}(y(x)-\sigma(\sum_{i=1}^{n} w_i x_i))^2)}{\partial w_j} \\
&= -(y(x)-\sigma(\sum_{i=1}^{n} w_i x_i))\cdot\sigma'(\sum_{i=1}^{n} w_i x_i)\cdot\frac{\partial(\sum_{i=1}^{n} w_i x_i)}{\partial w_j} \\
&= -(y(x)-\sigma(\sum_{i=1}^{n} w_i x_i))\cdot\sigma(\sum_{i=1}^{n} w_i x_i)\cdot(1-\sigma(\sum_{i=1}^{n} w_i x_i))\cdot x_j
\end{aligned} \tag{9}$$

The delta rule changes the j^{th} weight by some proportion α of the negative of the gradient as in (10). The proportion α is generally considered the learning rate.

$$w_j = w_j - \alpha\cdot\frac{\partial E}{\partial w_j} \tag{10}$$

4.2.3. Multi-layered perceptrons

A perceptron described above classifies a single output feature from a vector of input features. A layer of *n* such perceptrons would mean that *n* output features will be classified simultaneously from the input vector. Many artificial neural networks have multiple layers of perceptrons. That is, a network may have one layer of multiple neurons, each of which corresponds to the output neuron of a separate perceptron. The neurons in this layer may in turn function as input neurons of another set of perceptrons whose output neurons form another layer. For example, consider the network with two layers of perceptrons in Figure

4.3.

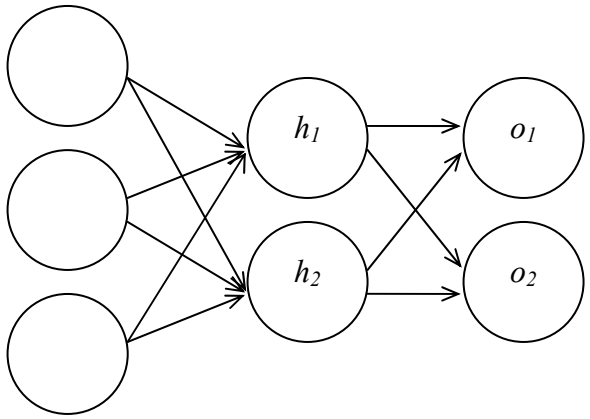

Figure 4.3. An example of two layered perceptron.

The neurons labeled h_1 and h_2 in the first layer, and o_1 and o_2 in the second layer are all perceptrons on their own, exemplifying a network with multiple layers with multiple perceptrons in each layer. The values, or often called *activations* in the connectionist literature, of h_1 and h_2 in the first layer are used as input to o_1 and o_2 in the second layer.

Perceptrons of two or more layers are used primarily to classify linearly inseparable data set. Suppose we plot each member of the data set, or each *sample,* in a vector space and label it either ‘yes’ or ‘no’. The sample space is *linearly separable* if there is a hyperplane such that on the samples on one side of the hyperplane is labeled ‘yes’ and the samples on the other side is labeled ‘no’. For example, a two dimensional sample space is linearly separable if there is a straight line that correctly divides the samples according to their labels. The dot-product of the weight vector and the feature vector can be interpreted as the hyperplane in the feature space. Since the perceptron relies on this dot-product for classification, it follows that perceptrons cannot classify linearly inseparable data sets.

The most commonly cited example of linearly inseparable two-dimensional sample space is that of the XOR (eXclusive OR) function, where a sample is labeled yes if and only if the value of one binary feature is different from the value of the other. The samples of the XOR function are plotted in Figure 4.4. The XOR function is linearly inseparable because there is no straight line that can correctly divide the samples.

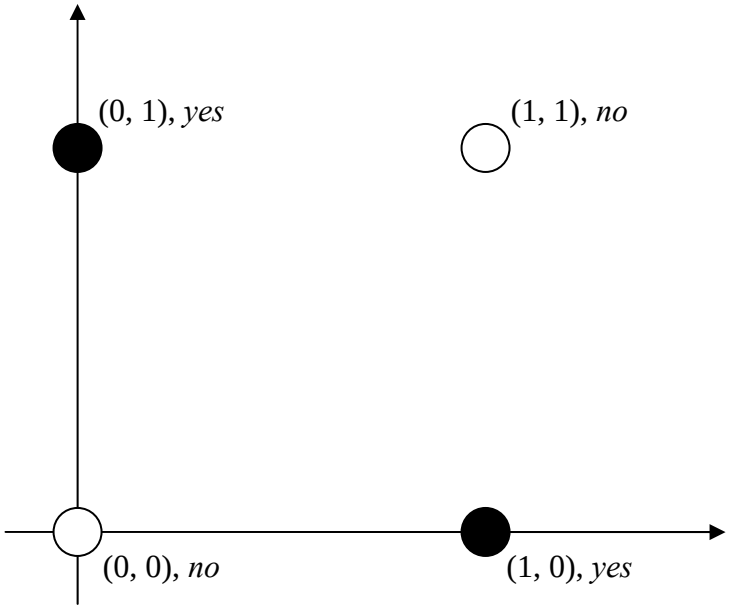

Figure 4.4. Hypothesis space of the XOR function

Adding a layer of perceptrons between input and output has the effect of enlarging the hypothesis space. Therefore, using a multi-layered perceptron has become a popular method to avoid working in small hypothesis space lest the samples be linearly inseparable. For example, by adding the two perceptrons in a hidden layer, the two-layered perceptron in Figure 4.5 with a threshold function for each perceptron correctly computes the XOR function.

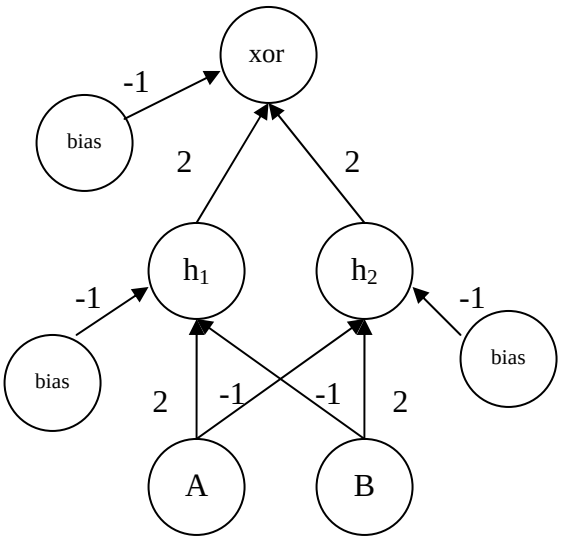

Figure 4.5. A two-layered perceptron that computes XOR function. The input nodes A and B are directly activated as either one or zero. The nodes h_1, h_2, and xor have a threshold of one; their final activation is 1 if the sum input activation is above one, and zero if the sum input activation is below one.

Despite its advantage in classifying linearly inseparable samples, adding hidden layers of perceptrons complicates the structure of the network and it becomes harder to comprehend the behavior of the network. In addition, identifying the optimal number of hidden perceptrons is time-consuming since it is identified by trial-and-error.

The current model consists of a single layer of multiple perceptrons. It does not have hidden layers for two reasons: the behavior of the model is easier to understand and control without the hidden layers, and adding a hidden layer seems excessive since the model performs well without one as will be shown below.

4.2.4. Recurrent neural networks

The (multi-layered) perceptrons illustrated above are directed acyclic graphs; no neuron in the network has path along the connections to itself. Artificial neural networks that correspond to directed acyclic graphs are called *feed-forward* neural networks. On the other hand, neural networks that have directed cycles are called *recurrent* neural networks.

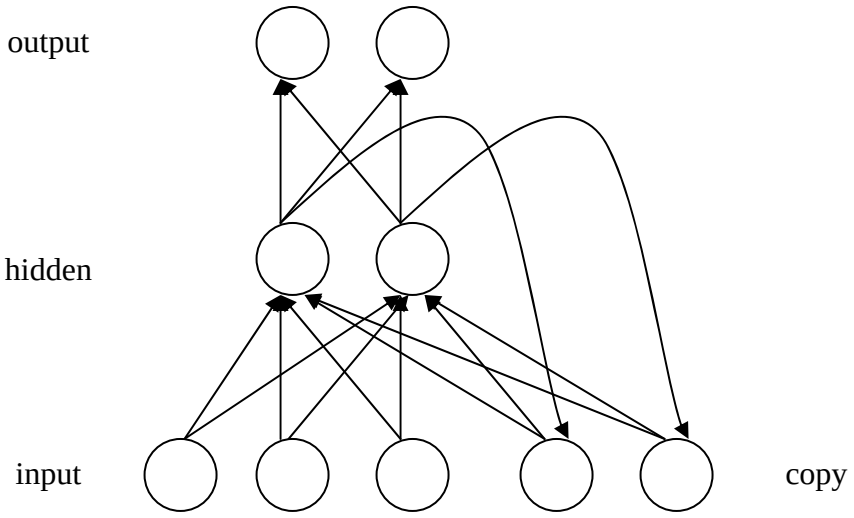

Figure 4.6. A simple recurrent network.

Recurrent networks are frequently used in time-series prediction tasks, where data consisting of input-output pairs are temporally ordered. For each input presented per time-step, the network predicts the corresponding output. A feed-forward network should suffice if the input presented at the current time-step alone can predict the output. However, when history of input, output, or some abstraction of the input is necessary for predicting the output, the relevant information from the previous time-step(s) must be stored and fed back to the network as input at the current time-step. Figure 4.6 illustrates a simple recurrent

network (Elman, 1990) where activation pattern in the hidden layer at time *t-1* is stored in the copy layer and fed back into the network as input at time *t*.

The current model approximates auditory repetition task as a time-series prediction task. It processes a sequence of temporally ordered vectors encoding perceptual confusion to identify the component phoneme at the corresponding word position from left to right based on the history of processing output. As a consequence, we simulate the auditory repetition experiments with a recurrent network rather than a feed-forward network.

4.3. Implementation of the model

The basic ideas and assumptions elaborated in section 4.1 are implemented as a single-layered recurrent perceptron. The details of the implementation are described below.

4.3.1. Input layer

The input layer consists of a set of nodes where each node represents a phoneme in the language. The nodes are activated by the value of the corresponding component of the input vector at the given time-frame. For example, suppose there are three phonemes *{A, B, C}* in the language and there is a confusion matrix as in Table 4.1.

	A	*B*	*C*	Total
A	70	20	10	100
B	10	80	10	100
C	15	10	75	100
Total	95	110	95	300

Table 4.1. An example confusion matrix

The confusion matrix, for example, reads speakers of the language confused *A* as *B* 20 times out of 100 times *A* was presented. The matrix allows us to encode the sequence *A*-*C* as the sequence of two vectors <0.7, 0.2, 0.1> and <0.15, 0.1, 0.75>. The input layer in this example would consist of three nodes representing *A*, *B*, and *C*, respectively. The nodes will be activated by <0.7, 0.2, 0.1> at the first time-frame, and <0.15, 0.1, 0.75> at the second time-frame.

4.3.2. Output layer

Recall that the model changes the likelihoods in the input vector at each time-frame, and that the result of change is stored to change the contents of the vector at future time-frames. For this purpose, the output layer is a sequence of *blocks* of nodes, as in Figure 4.7.

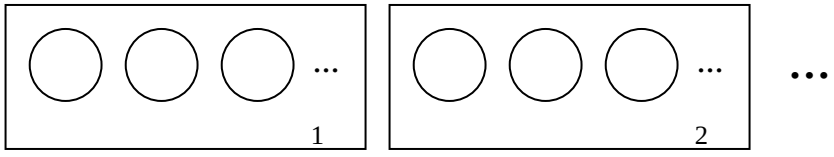

Figure 4.7. Blocks of nodes in the output layer.

Each block exists to represent the change in the contents of the input vector presented at a particular time-frame. Accordingly, there are as many nodes in each block of the output layer as there are in the input layer. Since the input presented at the n^{th} time-frame encodes the confusion information regarding the n^{th} phoneme of the word, the n^{th} block of the output layer also represents the n^{th} word position.

The result of change in likelihood of a phoneme is reflected as the activation of the

corresponding output node. For example, if the model changed the n^{th} input vector <0.7, 0.2, 0.1> to <0.9, 0.03, 0.07>, the nodes in the n^{th} block in the output layer would be activated by <0.9, 0.03, 0.07>.

4.3.3. Connections

The nodes in the output layer are, in fact, perceptrons. The input nodes are fully connected to the output nodes. The nodes in each block of the output layer are fully connected to the nodes in the blocks that follow the current block in sequence. This is illustrated in Figure 4.8.

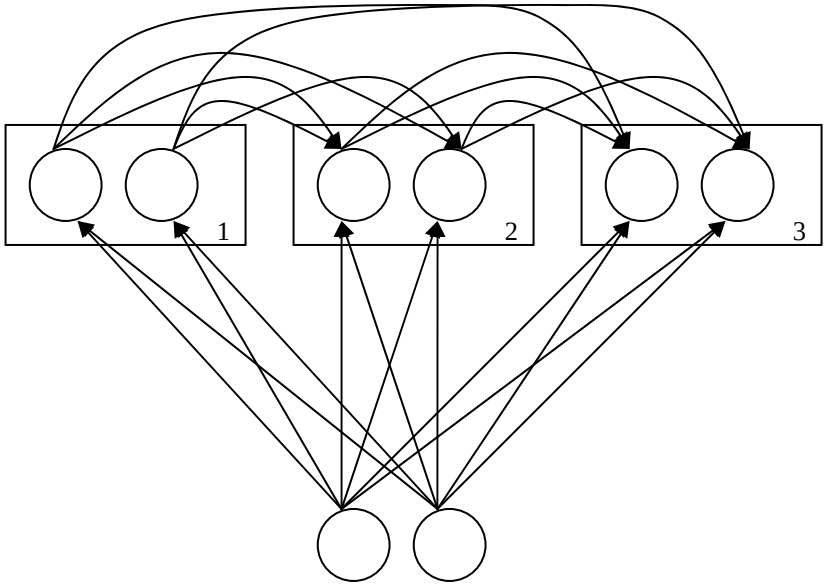

Figure 4.8. An example single layered recurrent perceptron.

The model changes the likelihoods in the input vector by spreading the activation from the input nodes and the output nodes in the previous blocks to the output nodes in the block

whose label matches the index of the current time-frame. For example, at time-frame 3, the activation from the input nodes and the output nodes in blocks 1 and 2 will spread to the output nodes in block 3. The output nodes in the block matching the current time-frame are activated according to the sigmoid function of weighted sum of incoming activations. The activation pattern is stored to activate the nodes in the following blocks in the later time-frame until the last time-frame pertaining to a word.

The connections between the input nodes and the output nodes encode how the model evaluates the likelihood or the confusion information of each phoneme specified in the input vector to identify the correct phoneme. A positive connection between input node *X* and output node *Y* implies that the likelihood of phoneme *X* given the perceptual input encourages the model to choose *Y* as the phoneme matching the input. A negative connection between the two implies that the likelihood of phoneme *X* discourages the model to choose *Y* as the phoneme matching the input.

The connections between the output nodes in different blocks encode how the model applies the knowledge of sequential phonotactic constraints. A positive connection between output node *X* in block *m* and output node *Y* in block *n* implies that the likelihood of *X* as the m^{th} phoneme of the word increases the likelihood of *Y* as the n^{th} phoneme of the word. A negative connection between the two implies that the likelihood of *X* as the m^{th} phoneme of the word decreases the likelihood of *Y* as the n^{th} phoneme of the word.

4.3.4. Luce ratio

In section 4.1.2, we assumed that the probability of choosing phoneme *X* at the n^{th} time-frame is inversely related to the latency at which the subject perceives *X* as the n^{th} phoneme in the word. We also assumed the choice probability is positively related to subjects'

perceptual accuracy. Furthermore, we assumed that the subject's latency to the stimulus word is inversely related to the product of the choice probabilities over individual constituent phonemes.

We apply the Luce choice rule (Luce, 1959) to compute the choice probability from the activation pattern at the end of the last time-frame pertaining to a word. For a given set of N nodes, each representing one of N available choices, the probability of making the choice represented by the i^{th} node is computed by (11), where act_i denotes the amount of activation for the i^{th} node.

$$\frac{act_i}{\sum_{j=1}^{N} act_j} \tag{11}$$

In our case, the choice is made between the nodes within a block of the output layer. Therefore, the choice probability for the n^{th} phoneme X is computed by the ratio, henceforth the Luce ratio, of activation of output node X to the sum of activation over all output nodes in the n^{th} block. The product of the Luce ratio over all constituent phonemes approximates the choice probability of the word, henceforth the word score. Greater word score implies faster perceptual latency and higher perceptual accuracy.

4.4. A simple illustration

The single layered recurrent perceptron described above can simulate the effect of perceptual confusion on the degree of perceptual facilitation. The same perceptron, or a set of perceptrons with identical structure and a parallel set of weights, predicts that phonotactic knowledge would facilitate perception more if the constrained phonemes are

more confusable to each other. That is, the predicted difference in word score between legal and illegal words would be greater if the phonotactic constraint restricts co-occurrence of a more confusable phoneme pair.

We illustrate this with a simple hypothetical example. Suppose a language where there are two phonemes *A* and *B* in the language, and a word in the language is two phonemes long. The set of all possible words in the language would be {*AA*, *AB*, *BA*, *BB*}. Let us further assume that a speaker of the language participates in an experiment and learns a phonotactic constraint that favors repetition (*AA*, *BB*) and disfavors non-repetition (*AB*, *BA*).

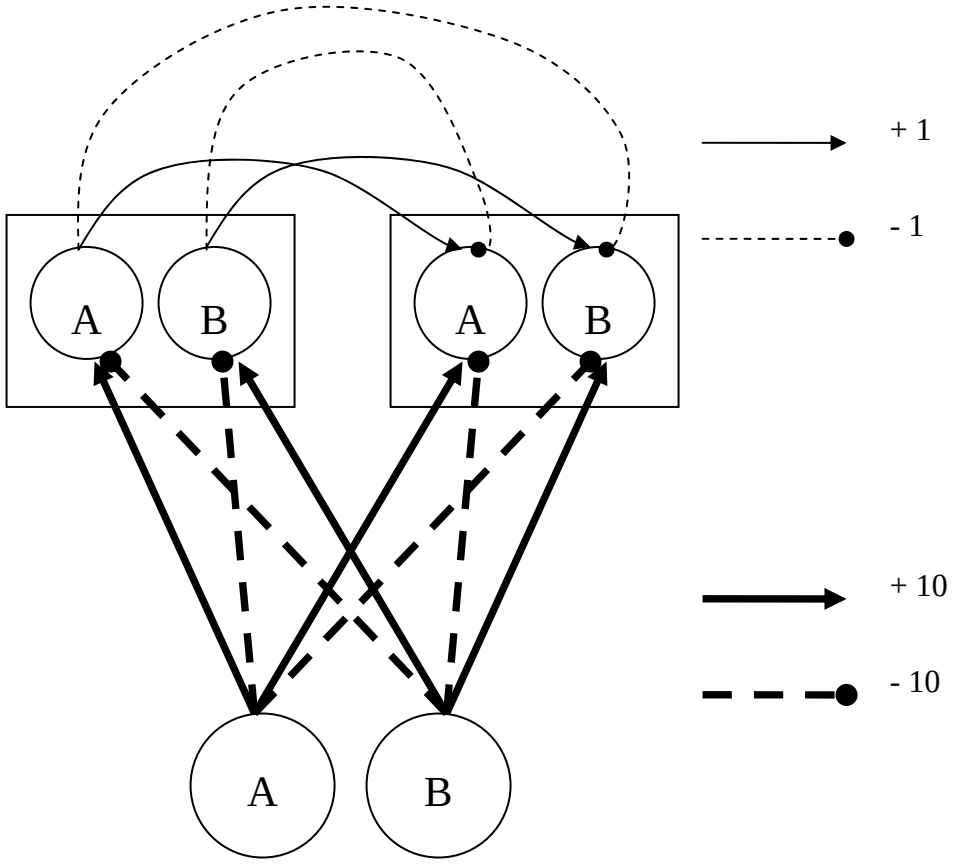

Figure 4.9. A recurrent perceptron for simple illustration.

The above set of assumptions can be interpreted to mean that the speaker's knowledge consists of two parts. The speaker knows how to evaluate the confusion information to identify the correct phoneme from his/her experience with the language. That is, given a signal that may be confused as either *A* or *B* with a certain probability, the speaker knows how to weight the probabilities. The speaker also knows the phonotactic constraint as a consequence of learning from the experiment. Based on this interpretation, the speaker's knowledge can be implemented as the recurrent perceptron in Figure 4.9.

In Figure 4.9, the nodes represent the phonemes in the language that they are labeled after. The two nodes at the bottom constitute the input layer, while the four nodes (perceptrons) on top constitute the output layer. The two blocks in the output layer represent the two word positions, with the block to the left representing the first word position. The connections from the input nodes to the output nodes encode the speaker's knowledge on how to weight the confusion information to identify the correct phoneme. A positive connection from an input node implies that the phoneme represented by the node should be considered likely given the confusion information, while a negative connection implies that the phoneme should be considered unlikely. The connections from the output nodes in the first block to the nodes in the second block encode the phonotactic knowledge. As the phonotactic constraint favors repetition, the nodes with the same phoneme label are positively connected. On the other hand, as the constraint disfavors non-repetition, the nodes with different labels are negatively connected. The absolute value of the connection weights is much larger for connections between input and output nodes than for connections between output nodes of different blocks. This reflects the fact that the speakers' knowledge on disambiguating the confusable perceptual input is based on their life-long experience with their native language while their phonotactic knowledge is

acquired from a very brief laboratory experience.

Two types of words, legal and illegal, are presented to the above perceptron, and the perceptron computes the word score for each word presented. The degree to which the phonotactic knowledge facilitates perception of phonotactically legal words is approximated by the difference in mean word score between legal and illegal words. We are interested in how the degree of perceptual facilitation changes as a function of the confusability between the constrained phonemes *A* and *B*. Therefore, we vary the confusability between *A* and *B* and plot the difference in mean word score between legal and illegal words.

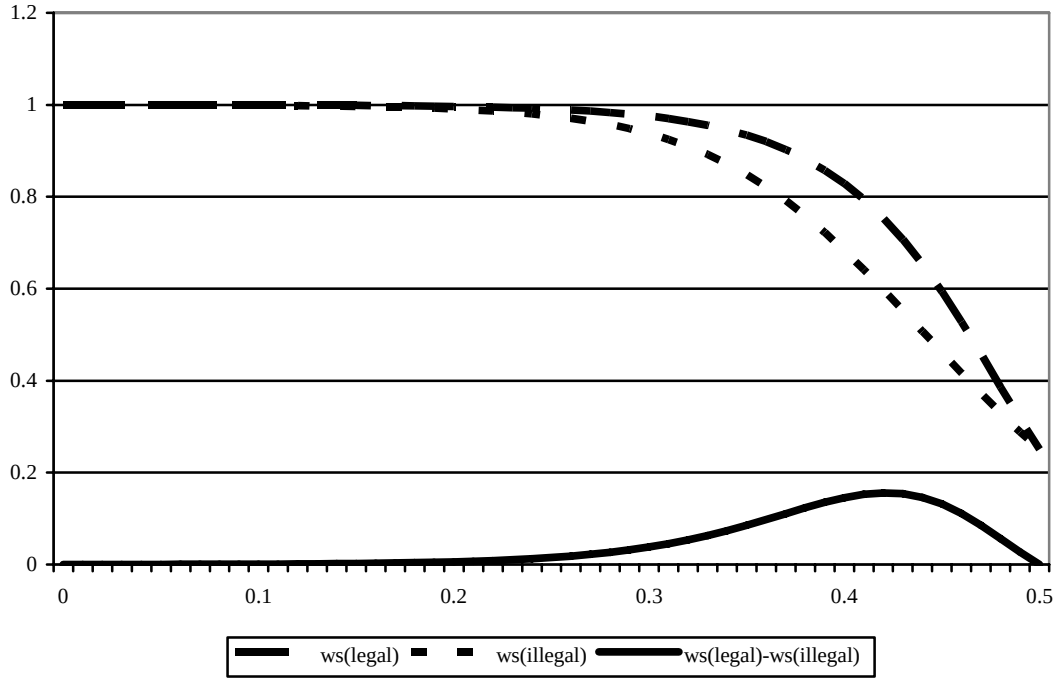

Figure 4.10. Predicted degree of perceptual facilitation as a function of confusability between constrained phonemes. The horizontal axis shows the confusability between the constrained phonemes. The vertical axis shows the predicted degree of perceptual facilitation.

The result is summarized in Figure 4.10. Confusability of zero means that a phoneme is never mistakenly perceived as the other. Confusability of 0.5 means that a phoneme is mistakenly perceived as the other half of the time. Since we are considering only two phonemes, this is when the confusability is at its peak. *ws(legal)* and ws*(illegal)* denote mean word scores for legal and illegal words, respectively. *ws(legal)* - *ws(illegal)* denotes the difference between the two means. The change in *ws(legal)* - *ws(illegal)* indicates that the perception of phonotactically legal words is increasingly facilitated as the two constrained phonemes become more confusable up to a certain point. The decrease after the point possibly reflects the fact that the confusion is too severe for the phonotactic knowledge to play any role.

To summarize, the recurrent perceptron predicts in this simple illustration that the knowledge of a phonotactic constraint would facilitate perception more if the constrained phonemes are more confusable. This suggests that the recurrent perceptron has the potential to simulate the results found in the auditory repetition experiments. In the following section, I illustrate that the perceptron can indeed simulate the results from the auditory repetition experiments.

4.5. Simulation

We extend the approach in the simple illustration to simulate four auditory repetition experiments: Experiment 1 (/l/-/r/ harmony[12]), Experiment 7 (/l/-/m/ harmony), Experiment

[12] In the remaining portion of the chapter, I denote liquid harmony and backness harmony as /l/-/r/ harmony and /i/-/u/ harmony to emphasize the identity of the constrained

3 (/i/-/u/ harmony), and Experiment 8 (noisy version of /i/-/u/ harmony). The basic idea underlying the simulation is identical to the one underlying the simple illustration study. Difference in subjects' latency between legal and illegal words is approximated by difference in mean word score between legal and illegal words. The study shows that the difference in mean word score is greater when the perceptron simulates experiments where the constrained phonemes are perceptually more confusable to each other. Specifically, the difference is greater in /l/-/r/ harmony simulation and noisy version of /i/-/u/ harmony simulation than in /l/-/m/ harmony simulation and /i/-/u/ harmony simulation, respectively.

4.5.1. Methods

We simulate each experiment with two types of recurrent perceptrons: one whose connection weights are predefined and the other whose weights are trained. Recall that the connection weights of a perceptron can be interpreted as encoding a speaker's knowledge on how to resolve confusion in the perceptual input to identify the correct phonological representation of the input. Using the perceptron with a predefined set of weights embeds the assumption that the speaker's life-time experience with English and his/her brief exposure to study words amidst the filler words resulted in a particular form of knowledge that causes perception of legal words to be better than illegal words. By training a perceptron, on the other hand, we do not make assumptions about whether the relevant knowledge is acquired from the experience. Rather, the emphasis is on whether the knowledge can be acquired from the experience.

Details on how the weights are predefined are described in section 4.5.1.1, and

---

phonemes.

details on how the perceptrons are trained are described in section 4.5.1.2. All other aspects of the simulation are the same for both types of recurrent perceptrons: the structure of the perceptron, how the perceptual input is presented, and how the perceptron calculates the word score for each stimulus to approximate subjects' latency or accuracy.

Each recurrent perceptron has the following structure. The input layer consists of 40 nodes, each of 39 representing a phoneme in English and one indicating silence. The output layer consists of six blocks of perceptrons, with 40 perceptrons in each block. The n^{th} block in the output layer represents the n^{th} phoneme in the word, and each of the 40 perceptrons in the block represents presence or absence of a particular phoneme in the corresponding word position. We are limiting the length of the word to six phonemes as the stimuli in the auditory repetition experiments were six phonemes long. Each output node is fully connected to the nodes in the input layer and the output nodes in the previous blocks.

At each time-frame, input nodes are activated according to the Luce confusion matrix. Assuming the n^{th} phoneme in the word is *Y*, the node *X* is activated by relatively how often *Y* is confused as *X* in the confusion matrix. For example, if *Y* is confused as *X* 70 times out of 100 times *Y* was presented to Luce's subjects, input node *X* is activated by 0.7 at the n^{th} time-frame. Recall, however, that there are nine confusion matrices in Luce (1986): one for phonemes in each of three word positions (onset, nucleus, and coda) presented to subjects at each of three noise levels (SNR = +15 dB, +5 dB, -5 dB). By default, we refer to the confusion matrices at SNR = +15 dB to activate input nodes. To simulate the noisy version of /i/-/u/ harmony experiment, we refer to the matrices at SNR = +5 dB for test words. This is because the subjects were exposed to stimuli without background noise in the first two blocks, while they were exposed to stimuli with background noise at SNR=+5dB in the final three blocks. When the n^{th} phoneme in the word is a consonant, we refer to the confusion matrix for onset phonemes to activate input

nodes at the n^{th} time-frame. When the n^{th} phoneme in the word is a vowel, we refer to the confusion matrix for nucleus phonemes. Input nodes representing the phonemes not listed in the relevant matrix is not activated at all.

Subjects' latency to a stimulus word is approximated by the word score calculated as in the simple illustration. Luce-ratios of the constituent phonemes are computed at the individual word position and multiplied over all of the six positions. We compare the mean word score for the legal words against the mean word score for the illegal words. Difference in the two means approximates the degree of facilitation in perception of legal words from learning the experimental phonotactic constraint.

4.5.1.1. Predefining connection weights

For a predefined recurrent perceptron, the weights are specified similar to how they were specified in the simple illustration. Input nodes and output nodes are connected positively (+10) if their labels are the same, and negatively (-10) if their labels are different. Knowledge of the phonotactic constraint is encoded in the connections between the output nodes in different blocks.

The perceptrons used for different auditory repetition experiments differ with respect to how the knowledge of the experimental phonotactic constraint is encoded, although the basic idea of the method is the same. We first identify the four nodes representing the four position-specific constrained phonemes. For example, to encode liquid harmony, we identify the nodes representing /l/ and /r/ in the third block, and the nodes representing /l/ and /r/ in the fifth block. This is because the constrained liquids occupy the third and the fifth word positions in the stimuli (e.g., /sa.**l**a.**r**a/). We then connect the two nodes in the former block to the two nodes in the latter block, and weight the connections according to the phonotactic constraint. If the constraint favors co-occurrence

of the phonemes represented by the two connected nodes, the connection is positively weighted. Otherwise, the connection is negatively weighted. In the example of liquid harmony, the node representing /l/ (/r/) in the third block and the node representing /l/ (/r/) in the fifth block are positively connected. On the other hand, the node representing /l/ (/r/) in the third block and the node representing /r/ (/l/) in the fifth block are negatively connected.

We start out by weighting the connections encoding the phonotactic knowledge either +1 or -1, as in the simple illustration. As will be shown in section 4.5.2.1, this yields a legality effect that is in the right direction, but not large enough to reach statistical significance. This is because the number of phonemes in the language and the length of a word are far larger than those in the simple illustration. Therefore, we also run the simulation with perceptrons whose weights encoding the phonotactic knowledge is either +5 or -5.

4.5.1.2. Training the recurrent perceptron

The recurrent perceptron can be trained using the delta rule described in section 4.2.2. We try two different ways to train the perceptron. In the first approach, we start with a perceptron whose weights are randomly initialized, and train it on the study words and the filler words in the first two blocks of the auditory repetition experiment that is being simulated. In the second approach, we first train the perceptron on a list of English words, and then train it on the set of study words and filler words in the first two blocks of the auditory repetition experiment it simulates. The first approach makes no assumption about the role of subjects' knowledge of English phonotactics in learning the experimental phonotactic constraints. On the other hand, the second approach assumes that their prior knowledge of English phonotactics can be encoded as the weights of the perceptron trained

on a list of English words, and that learning the experimental constraint is equivalent to fine-tuning their knowledge of English phonotactics to experimental stimuli. We refer to the first type of perceptrons as naïve perceptrons and the second type of perceptrons as English perceptrons.

The English perceptron was first trained on 3788 words selected from the CMU pronunciation dictionary, release 0.6 (Weide, 1998). The words covered all CV.CV.CV words in the dictionary excluding the ones containing the coda consonants (/ŋ/ and /ʒ/). The learning rate was set to 0.3 and training was complete after 1000 iterations over the word list. The perceptron achieved 98.47 % accuracy in predicting the correct phoneme sequence for a given word. The perceptron predicted the phoneme sequence by concatenating the phonemes represented by the maximally activated node in each block of the output layer. Prediction accuracy was defined as the percentage of words for which the perceptron predicted correct phoneme sequences.

Both the English perceptron and the naïve perceptron were trained on the study words and the filler words in the first two blocks of the auditory repetition experiment that was simulated. The filler words in the remaining three blocks were not used for training since the blocks contained legal and illegal words. In simulating the noisy /i/-/u/ harmony, the input nodes were activated according to the Luce's confusion matrix at SNR = +15 dB for the training set, while they were activated according to the matrix at SNR = +5 dB for the test set. Learning rate was fixed to 0.3, while the number of iterations over the training set varied between 10, 50, and 100 for the naïve perceptron, and 10, 50, 100, 300, 500, and 1000 for the English perceptron.

4.5.2. Results

4.5.2.1. Predefined perceptrons

The differences in mean word score between legal and illegal words for each experiment when the magnitude of weights of the connections encoding the experimental phonotactic constraint (henceforth, phonotactic weights) are +1 or -1 are summarized in Figure 4.11. The mean legal and illegal word scores are summarized in Table 4.2.

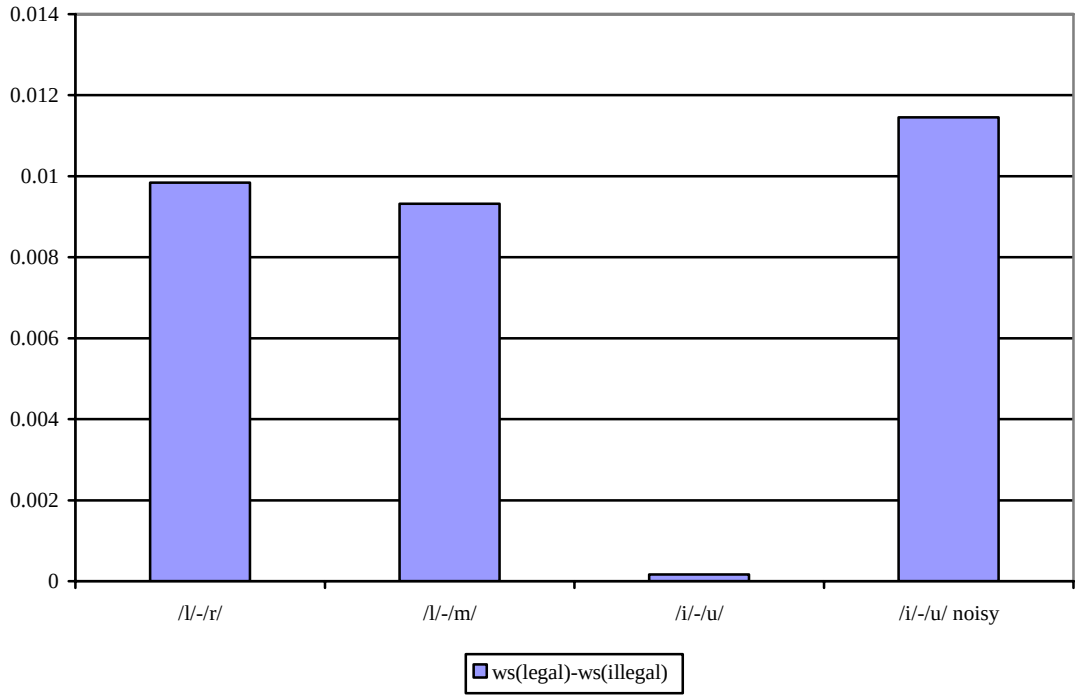

Figure 4.11. Difference in mean word score between legal and illegal words with phonotactic weights fixed to +1 or -1. The vertical axis shows the difference in predicted mean score between legal and illegal words.

The prediction of the perceptron is consistent with the results from the corresponding experiments. Difference in latency or accuracy between legal and illegal words is predicted to be greater in /l/-/r/ harmony and noisy /i/-/u/ harmony than in /l/-/m/ harmony and /i/-/u/

harmony, respectively. The only concern is that the difference in word score between legal and illegal words is too small. In fact, the difference is not significant in either of the consonant harmony simulations ($t(14) = 0.533$, $p = .602$ for /l/-/r/ harmony, $t(14) = 0.504$, $p = .621$ for /l/-/m/ harmony). The small difference is partly due to the fact that the magnitude of the phonotactic weights is too small compared with the magnitude of the weights for evaluating the confusion information.

	ws(legal)	*ws(illegal)*	*ws(legal)-ws(illegal)*
/l/-/r/ harmony	0.4987	0.4888	0.0098
/l/-/m/ harmony	0.4992	0.4898	0.0093
/i/-/u/ harmony	0.6859	0.6857	0.0002
Noisy /i/-/u/ harmony	0.8853	0.8738	0.0115

Table 4.2. Mean word scores for legal and illegal words with phonotactic weights fixed to +1 or -1.

One thing noticeable about the word scores is that they are much higher when simulating noisy /i/-/u/ harmony than the other experiments. They are also higher when simulating the plain /i/-/u/ harmony experiment than the two consonant harmony experiments. The reason is that in the confusion study done by Luce (1986), at SNR = +15 dB, subjects mistakenly perceived /a/ as /ɔ/ 40.9 % of the time while they correctly perceived it as /a/ 38.4 % of the time. Since the input nodes are activated according to the confusion matrix, this large confusion causes the estimated word score to drop significantly if the word contains the vowel /a/. There are less legal and illegal words containing the vowel /a/ in /i/-/u/ harmony experiments than in the two consonant harmony experiments. This is because the last two vowels for legal and illegal words are always either /i/ or /u/ in the /i/-/u/ harmony experiments. On the other hand, /a/ can occupy either of the last two vowel positions for

legal and illegal words in the consonant harmony experiments. In addition, /a/ was confused as /ɔ/ 24.2 % of the time, while they correctly perceived it as /a/ 37.8 % when the phonemes were presented at SNR = +5 DB. The probabilities are much larger in simulating noisy /i/-/u/ harmony experiment, because the input nodes are activated according to the confusion matrix at SNR = +5 dB. This influence of confusion on word score is another cause of the small difference between legal and illegal words, especially in simulating the consonant harmony experiments.

One way to overcome these factors is to increase the magnitude of phonotactic weights. Therefore, we also look at the results from the perceptrons whose phonotactic weights are +5 or -5. The differences in mean word score between legal and illegal words for each experiment using perceptrons whose phonotactic weights are +5 or -5 are summarized in Figure 4.12. The mean legal and illegal word scores are summarized in Table 4.3.

	ws(legal)	*ws(illegal)*	*ws(legal)-ws(illegal)*
/l/-/r/ harmony	0.4988	0.4100	0.0888
/l/-/m/ harmony	0.4992	0.4844	0.0147
/i/-/u/ harmony	0.6859	0.6738	0.0121
Noisy /i/-/u/ harmony	0.8870	0.4662	0.4208

Table 4.3. Mean perceptron probabilities for legal and illegal words with phonotactic weights fixed to +5 or -5.

For the consonant harmony experiments, the difference in mean word score between legal and illegal words was significant in /l/-/r/ harmony simulation ($t(14) = 4.494$, $p = .001$), However, the difference was not significant in /l/-/m/ harmony simulation ($t(14) = 0.749$, $p = .466$). Between the two experiments, the difference was significantly larger in /l/-/r/

harmony simulation than in /l/-/m/ harmony experiment ($t(14) = 12.089$, $p < .001$).

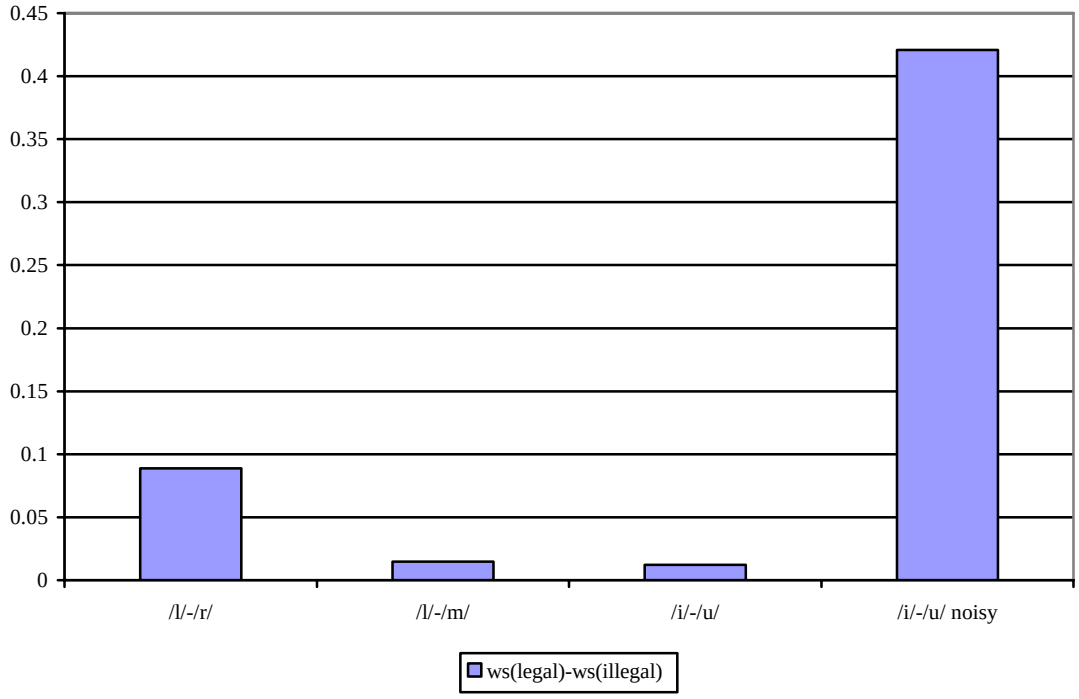

Figure 4.12. Difference in mean word score between legal and illegal words with phonotactic weights fixed to +5 or -5. The vertical axis shows the difference in predicted mean score between legal and illegal words.

For the vowel harmony experiments, the difference in mean word score between legal and illegal words was significant in both the experiment without noise ($t(14) = 25.780$, $p < .001$) and the experiment with noise ($t(14) = 20.970$, $p < .001$). Between the two experiments, the difference was significantly larger in noisy version of /i/-/u/ harmony simulation than in plain /i/-/u/ harmony experiment ($t(14) = 20.830$, $p < .001$).

4.5.2.2. Naïve perceptrons

The differences in mean word score between legal and illegal words are summarized in Figure 4.13 and Table 4.4. The mean word score for legal words was significantly higher than the mean word score for illegal words in all four simulations regardless of how long the perceptron was trained.

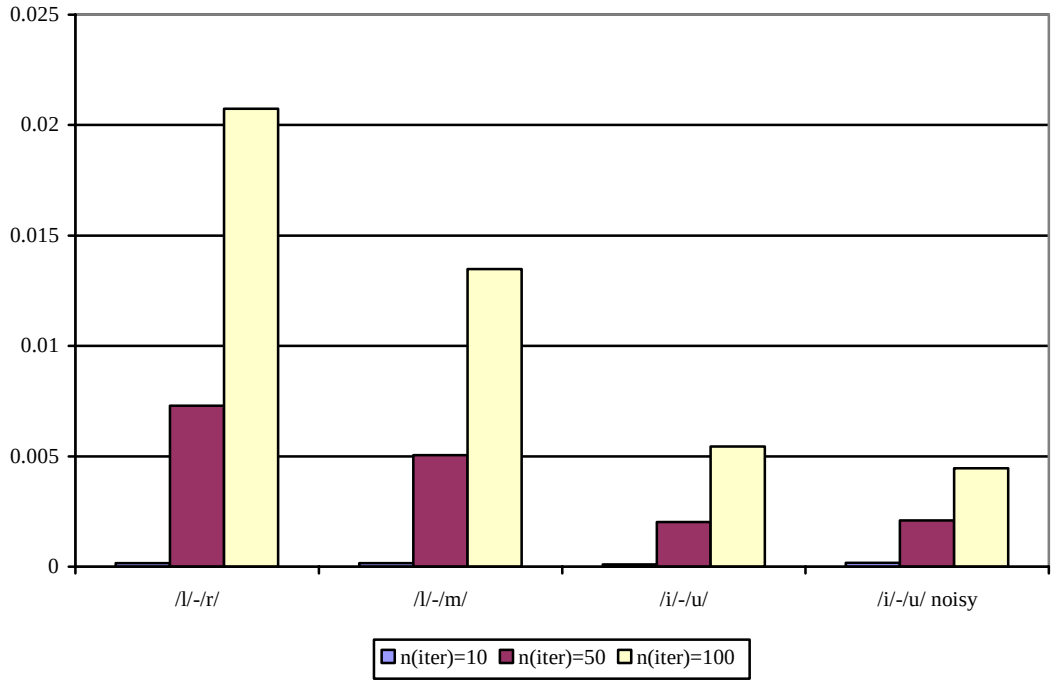

Figure 4.13. Difference in mean word score between legal and illegal words predicted by naïve perceptrons at different number of iterations. The vertical axis shows the difference in predicted mean score between legal and illegal words.

The difference in the predicted degree of perceptual facilitation between conditions fluctuates. For consonant harmony simulations, the difference between /l/-/r/ harmony and /l/-/m/ harmony tends to increase as the perceptron is trained more, reaching significance

when iterated 100 times over the training set (*n*(iter) = 10: *t*(14) = 0.065, *p* = .949; *n*(iter) = 50: *t*(14) = 1.953, *p* = .071; *n*(iter) = 100: *t*(14) = 2.229, *p* = .043). On the other hand, for vowel harmony simulations, the difference between the noisy /i/-/u/ harmony and the plain /i/-/u/ harmony tends to decrease as perceptron is trained more (*n*(iter) = 10: *t*(14) = 2.468, *p* = .027; *n*(iter) = 50: *t*(14) = 0.090, *p* = .929; *n*(iter) = 100: *t*(14) = -0.467, *p* = .648).

Exp	*n*(iter)	*ws(legal)*	*ws(illegal)*	Difference	Statistical significance
/l/-/r/	10	0.0007	0.0005	0.0002	$t(14) = 7.582, p < .001$
	50	0.0195	0.0122	0.0073	$t(14) = 9.839, p < .001$
	100	0.0547	0.0340	0.0207	$t(14) = 9.459, p < .001$
/l/-/m/	10	0.0007	0.0006	0.0002	$t(14) = 6.970, p < .001$
	50	0.0202	0.0152	0.0051	$t(14) = 6.595, p < .001$
	100	0.0555	0.0420	0.0135	$t(14) = 6.620, p < .001$
/i/-/u/	10	0.0007	0.0006	0.0001	$t(14) = 5.024, p < .001$
	50	0.0197	0.0177	0.0020	$t(14) = 2.854, p = .013$
	100	0.0546	0.0491	0.0055	$t(14) = 2.950, p = .011$
/i/-/u/ noisy	10	0.0008	0.0007	0.0002	$t(14) = 6.337, p < .001$
	50	0.0115	0.0093	0.0021	$t(14) = 5.094, p < .001$
	100	0.0244	0.0199	0.0045	$t(14) = 4.701, p < .001$

Table 4.4. Mean word scores for legal and illegal words predicted by naïve perceptrons. Difference is equal to *ws(legal) – ws(illegal)*.

The perceptron predicts perception will be facilitated significantly more in the noisy condition when iterated ten times. However, the direction of the prediction is reversed when the perceptron is trained more, which seems to be the result of differences in the magnitude of predicted word scores between the two vowel harmony conditions. When iterated over 100 times, the predicted mean word scores in /i/-/u/ harmony are 0.0546 for legal words

and 0.0491 for illegal words. On the other hand, the predicted scores in the noisy condition are 0.0244 for legal words and 0.0199 for illegal words. On the average, when iterated over 100 times, the predicted word scores in the noisy condition (M = 0.2214) are approximately 42.7 % smaller than the predicted word scores in the plain /i/-/u/ harmony condition (M = 0.0518).[13] Because the overall magnitude of word scores is smaller in the noisy condition, the absolute difference between the legal word scores and the illegal word scores could also be smaller. In terms of the ratio of the mean legal word scores to the mean illegal word scores, legal words are predicted to be perceived 1.22 times better than illegal words in the noisy condition, while legal words are predicted to be perceived 1.11 times better than illegal words in the plain /i/-/u/ harmony condition.

4.5.2.3. English perceptrons

The differences in mean word score between legal and illegal words are summarized in Figure 4.14 and Table 4.5. The English perceptrons had to be trained much longer than the naïve perceptron before the legality effect could be observed. The difference in mean word score between legal and illegal words reached significance after 1000 iterations for /l/-/r/ harmony and after 500 iterations for the noisy version of /i/-/u/ harmony. In the consonant harmony simulations, the model initially predicted illegal words to be more probable than legal words until 300 iterations for /l/-/r/ harmony, and 1000 iterations for /l/-/m/ harmony. This is due to the frequency of co-occurrence patterns in the CV.CV.CV words selected from the CMU pronunciation dictionary.

---

[13] This is because the perceptual input is far noisier when we refer to the confusion matrices at SNR = +5 dB; not only is there more confusion between /i/ and /u/, but there is more confusion between any two phonemes in general.

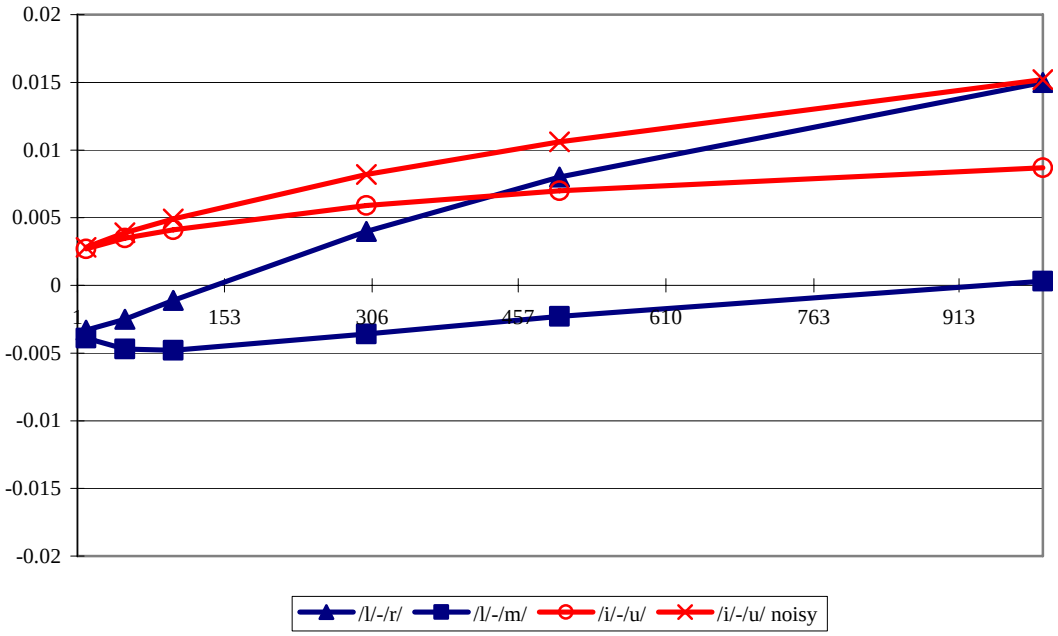

Figure 4.14. Difference in mean word score between legal and illegal words predicted by English perceptrons at different number of iterations. The horizontal axis shows the number of iterations. The vertical axis shows the predicted mean difference in word score between legal and illegal words.

As summarized in Table 4.6, there were more English words that violated the consonant harmonies in the training set than those that instantiated the consonant harmonies. Eighty three words violated /l/-/r/ harmony, while 46 words instantiated the constraint. Seventy four words violated /l/-/m/ harmony, while 23 words instantiated the constraint. On the contrary, 63 words violated /i/-/u/ harmony, while 124 words instantiated the constraint. As a result, the English perceptrons were biased against the consonant harmonies before they were trained on the words in the experiments, and they required more training iterations to offset the bias. On the contrary, they were biased towards the vowel harmony, and therefore

they predicted legal words to be more probable than illegal words immediately after ten iterations.

The difference in the predicted degree of perceptual facilitation was not significant between conditions; neither in consonant harmony simulations (*n*(iter) = 10: *t*(14) = -25.350, *p* < .001; *n*(iter) = 50: *t*(14) = -14.385, *p* < .001; *n*(iter) = 100: *t*(14) = 0.429, *p* = .674; *n*(iter) = 300: *t*(14) = 0.778, *p* = .450; *n*(iter) = 500: *t*(14) = 1.043, *p* = .315; *n*(iter) = 1000: *t*(14) = 1.575, *p* = .138) nor in vowel harmony simulations (*n*(iter) = 10: *t*(14) = -0.002, *p* = .998; *n*(iter) = 50: *t*(14) = 0.086, *p* = .993; *n*(iter) = 100: *t*(14) = 0.160, *p* = .876; *n*(iter) = 300: *t*(14) = 0.350, *p* = .731; *n*(iter) = 500: *t*(14) = 0.490, *p* = .632; *n*(iter) = 1000: *t*(14) = 0.747, *p* = .468). Except for the consonant harmony simulations when iterated ten times and fifty times over the training set,[14] the differences appear to be in the right direction; the predicted degree of perceptual facilitation is greater when the constrained phonemes are more confusable to each other.

[14] Notice that the direction is in the opposite of what it should be in two ways: (1) illegal novel words are predicted to be perceived better than legal novel words in both /l/-/r/ harmony and /l/-/m/ harmony (see Table 4.5), and (2) the predicted degree of facilitation of perception of illegal words is predicted to be greater in /l/-/m/ harmony than in /l/-/r/ harmony. The reason for this is two fold. First, the English word list contains more words that violate the two consonant harmony patterns, and ten to fifty training iterations are not enough to offset the bias towards illegal words. Second, the bias against consonant harmony is greater in /l/-/m/ co-occurrence pattern (74 illegal words vs. 23 legal words) than in /l/-/r/ co-occurrence pattern (83 illegal words vs. 46 legal words).

Exp	n(iter)	ws(legal)	ws(illegal)	Difference	Statistical significance
/l/-/r/	10	0.0869	0.0902	-0.0033	$t(14) = -0.811$, $p = .431$
	50	0.1321	0.1346	-0.0025	$t(14) = -0.519$, $p = .612$
	100	0.1697	0.1708	-0.0011	$t(14) = -0.207$, $p = .839$
	300	0.2683	0.2643	0.0040	$t(14) = 0.653$, $p = .524$
	500	0.3305	0.3225	0.0080	$t(14) = 1.258$, $p = .229$
	1000	0.4254	0.4104	0.0150	$t(14) = 2.290$, $p = .038$
/l/-/m/	10	0.0802	0.0841	-0.0039	$t(14) = -0.959$, $p = .354$
	50	0.1241	0.1288	-0.0047	$t(14) = -0.948$, $p = .359$
	100	0.1627	0.1675	-0.0048	$t(14) = -0.848$, $p = .411$
	300	0.2634	0.2670	-0.0036	$t(14) = -0.548$, $p = .592$
	500	0.3264	0.3287	-0.0023	$t(14) = -0.329$, $p = .747$
	1000	0.4224	0.4221	0.0003	$t(14) = 0.052$, $p = .960$
/i/-/u/	10	0.1074	0.1047	0.0027	$t(14) = 0.715$, $p = .486$
	50	0.1499	0.1464	0.0035	$t(14) = 0.813$, $p = .430$
	100	0.1878	0.1837	0.0041	$t(14) = 0.896$, $p = .385$
	300	0.2863	0.2804	0.0059	$t(14) = 1.190$, $p = .254$
	500	0.3480	0.3410	0.0070	$t(14) = 1.400$, $p = .183$
	1000	0.4423	0.4336	0.0087	$t(14) = 1.712$, $p = .109$
/i/-/u/ noisy	10	0.0308	0.0280	0.0028	$t(14) = 1.839$, $p = .087$
	50	0.0460	0.0421	0.0039	$t(14) = 1.818$, $p = .091$
	100	0.0609	0.0560	0.0049	$t(14) = 1.849$, $p = .086$
	300	0.1042	0.0960	0.0082	$t(14) = 2.085$, $p = .056$
	500	0.1345	0.1239	0.0106	$t(14) = 2.290$, $p = .038$
	1000	0.1859	0.1707	0.0152	$t(14) = 2.635$, $p = .020$

Table 4.5. Mean word scores for legal and illegal words predicted by English perceptrons. Difference is equal to *ws(legal) - ws(illegal)*.

Co-occurrence pattern	Number of words with the co-occurrence pattern
C_2=/l/, C_3=/l/	22
C_2=/r/, C_3=/r/	24
C_2=/l/, C_3=/r/	23
C_2=/r/, C_3=/l/	60
C_2=/l/, C_3=/m/	24
C_2=/m/, C_3=/l/	50
C_2=/m/, C_3=/m/	1
V_2=/i/, V_3=/i/	118
V_2=/u/, V_3=/u/	6
V_2=/u/, V_3=/i/	56
V_2=/i/, V_3=/u/	7

Table 4.6. Frequency of co-occurrence patterns in the 3788 $C_1V_1.C_2V_2.C_3V_3$ words extracted from the CMU pronunciation dictionary.

4.5.3. Summary

Three different types of perceptrons were used to simulate the auditory repetition experiments. The predefined perceptrons embodied the assumption that subjects in the four experiments have learned the experimental constraints equally well. The emphasis was primarily on whether the model predicts that phonotactic knowledge will facilitate perception more if the constrained phonemes are more confusable. On the other hand, the emphasis for the naïve perceptrons and the English perceptrons was also on whether the model can learn the experimental constraints, in addition to whether it can simulate the perceptual facilitation hypothesis. Moreover, the English perceptrons embodied the

assumption that learning an artificial phonotactic constraint can be interpreted as fine-tuning the speakers' prior knowledge of phonotactics of their language.

Results of the simulation studies using the predefined perceptrons were consistent with the results from the auditory repetition experiments. The model correctly predicts that perception will be facilitated more in /l/-/r/ harmony experiment and the noisy /i/-/u/ harmony experiment than in /l/-/m/ harmony experiment and the plain /i/-/u/ harmony experiment, respectively.

Simulation studies using the naïve perceptrons showed that the model can simulate the learning process using the delta rule. After being trained briefly on study words and filler words, the model correctly predicts that legal words will be perceived more quickly and/or more accurately than illegal words. Although the statistics are not significant, the model also predicts that perception will be facilitated more in the experiment where the constrained phonemes are more confusable when the duration of training was brief.

Results of the simulation studies using the English perceptrons were in the right direction. When trained on study words and filler words over hundreds of iterations, the model correctly predicts that legal words will be perceived more quickly and/or more accurately than illegal words for /l/-/r/ harmony and the noisy version of /i/-/u/ harmony. The weights learned from the 3788 selected CV.CV.CV words seemed to have overfit the data and therefore required far more iterations to offset the overfitting. Although the statistics are not significant, the model also predicts that perception will be facilitated more in the experiment where the constrained phonemes are more confusable.

4.6. Chapter summary

Chapter 4 presented a single layered perceptron to model how the adult phonological

processing system learns phonotactic constraints and how the acquired phonotactic knowledge affects the system's perceptual behavior. The perceptron consists of an input layer and an output layer. A node in the input layer represents a phoneme in the language, and its activation value at each time-frame reflects the likelihood of the phoneme it represents given the perceptual input. Nodes in the output layer are organized into blocks. Each block represents a word-position and a node in the block represents a phoneme in the language that may occupy the position. The nodes between the two layers and the nodes between any two blocks are fully connected. The latter type of connections model the knowledge of sequential phonotactic constraints such as the non-adjacent phonotactic constraints studied in the dissertation. By spreading activation along the connections, the initial likelihoods of phonemes are modified as they reach each word position. How well the input word will be perceived is estimated from the resulting activation pattern at the output layer.

Four auditory repetition experiments in support of the perceptual facilitation hypothesis were simulated. Difference in subjects' perceptual latency or accuracy between legal and illegal words was approximated by difference in the perceptron's mean word score between legal and illegal words. Simulation results with three types of perceptrons were presented. Phonotactic knowledge was manually specified in the predefined perceptrons. Naïve perceptrons started with randomly initialized connection weights and were trained on study words and filler words. English perceptrons were first trained on a list of CV.CV.CV words in the CMU pronunciation dictionary and then trained on study words and filler words. Simulation results were consistent with the perceptual facilitation hypothesis for all three types of perceptrons; they predicted that perception will be facilitated more in /l/-/r/ harmony experiment and the noisy /i/-/u/ harmony experiment than in /l/-/m/ harmony experiment and the plain /i/-/u/ harmony experiment, respectively. In

addition, the naïve perceptrons and the English perceptrons simulated the learning process; after being trained on the nonsense words to which the human subjects were exposed, the perceptrons predicted the phonotactically legal novel words will be perceived better than the phonotactically illegal novel words.

CHAPTER 5

BAYESIAN BELIEF NETWORK MODEL OF PHONOTACTIC LEARNING AND GRAMMATICALITY JUDGMENT

The connectionist model in the previous chapter demonstrated how phonotactic learning and its effect on perception can be simulated. In this chapter, I demonstrate how phonotactic learning and its effect on grammaticality judgment can be simulated using a Bayesian belief network coupled with the connectionist model in the previous chapter. The connectionist model returns the set of phoneme sequences that best match the perceptual input corresponding to the nonsense words to which the human subjects were exposed in the familiarization phase. The Bayesian belief network learns the probability distribution that governs the set of phoneme sequences identified by the connectionist model and uses the distribution to compute the phonotactic probability of novel phoneme sequences.

The Bayesian belief network learns the probability distribution by discovering conditional dependencies between position-specific phonemes using the K2 algorithm and estimate the conditional probabilities between the dependent phonemes using maximum likelihood estimate so that the observed probability of the words in the training set is maximized. Knowledge of phonotactic constraints, co-occurrence restrictions in particular, is represented as a set of dependency relations between phonemes at different word positions. The trained belief network can compute how likely a new word is to be observed in the language characterized by the phonotactic constraints. When coupled with the single layered recurrent perceptron discussed in Chapter 4, the belief network correctly predicts the results of the grammaticality judgment experiments in Chapter 2.

The chapter is organized as follows. Section 5.1 introduces a probabilistic interpretation of grammaticality judgment. Section 5.2 introduces the basic ideas of

Bayesian belief networks. Section 5.3 describes how a Bayesian belief network can be trained using the K2 algorithm and maximum likelihood estimate. Section 5.4 presents simulation studies of grammaticality judgment experiments using the Bayesian belief network coupled with the single layered recurrent perceptron. Section 5.5 summarizes the chapter.

5.1. Probabilistic interpretation of grammaticality judgment

Based on their knowledge of phonotactic constraints, speakers of a language can decide if a phonological form is grammatical or not. Traditionally, a form is judged ungrammatical if it violates one or more constraints. However, recent studies suggest that speakers' judgments of grammaticality are better characterized as probabilistic behavior. Speakers judge a form to be more grammatical if its components appear in the language more frequently (e.g., Vitevitch et al., 1997). In addition, Coleman and Pierrehumbert (1997) argues that speakers' judgments depend more on the overall acceptability of the word rather than how phonotactically illegal or less frequent a particular component is. Therefore, an ideal model of phonological competence and one that simulates speakers' grammaticality judgment behavior should be able to compute the probability of the whole word based on its knowledge of phonotactics.

Suppose a speaker of language L perceives the incoming speech signal as a word W. We assume that the speaker judges grammaticality of W based on the probability that W is a member of L. That is, the speaker's grammaticality judgment is based on (1).

$$P(W \in L \mid W = \arg\max_{w} P(w \mid signal)) \quad (1)$$

By Bayes theorem, (1) is equivalent to (2).

$$\frac{P(W = \arg\max_{w} P(w \mid signal) \mid W \in L) \cdot P(W \in L)}{P(W = \arg\max_{w} P(w \mid signal))} \quad (2)$$

We make two more assumptions. Firstly, $P(W = \arg\max_{w} P(w \mid signal)) = 1$ for all words presented to the speaker. This is equivalent to assuming that the speaker recognizes a word always by identifying the most likely candidate given the signal. Secondly, $P(w \in L)$ is identical for all words presented to the speaker. This is because the term captures the speaker's inherent bias in deciding if a word is a member of the language. We are assuming that the speaker remains equally biased the whole time. Considering these two assumptions, computing (1) is simplified to computing (3).

$$P(W = \arg\max_{w} P(w \mid signal) \mid W \in L) \quad (3)$$

That is, we approximate a speaker's grammaticality judgment of a word by the probability of observing the word in the language.

We compute this probability as follows. We represent words in the same way as they are represented in the output layer of the single layered recurrent perceptron. That is, a node represents a phoneme in the language and the n^{th} block of nodes represents the n^{th} word position. This allows us to use the output of the connectionist model as the input for the grammaticality judgment. We make the nodes Boolean by setting the value of the maximally activated node within a block to one while suppressing the value of all other

nodes in the block to zero. This Boolean activation pattern represents the most likely phoneme sequence given the perceptual input. We train a Bayesian belief network over the observed Boolean activation patterns and let the resulting network compute the probability for future activation patterns. In sum, to approximate a grammaticality judgment, we compute the joint probability of conjunction of Boolean node values using a Bayesian belief network that is trained on a list of words in the language.

5.2. Bayesian belief networks

A Bayesian belief network (e.g., Mitchell, 1997; Russell and Norvig, 2003) is a directed graph that describes the joint probability distribution over a set of random variables using the notion of conditional independence. A node in the graph represents a random variable. Two nodes can be connected by a directed edge. If there is a directed path from node X to node Y, Y is called a descendant of X. This notation serves to indicate conditional independence relation between the variables. An edge from a *parent* node to a *child* node indicates that the child variable is conditionally independent of all non-descendant variables given the parent variable. A conditional probability table is associated with a node X that specifies $P(X \mid parents(X))$ for all values of X and *parents(X)*. To compute joint probability of the instance $<x_1, x_2, x_3, \ldots, x_n>$ of the set of variables $<X_1, X_2, X_3, \ldots, X_n>$, we compute the following.

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i)) \qquad (4)$$

For example, consider the network over four Boolean variables $<X, Y, W, Z>$ in Figure 5.1.

The four nodes represent the variables they are labeled after. The table to the right of each node is the conditional probability table. The edges indicate that *X* and *Z* are parents of *Y*, and that *Y* is conditionally independent of the non-descendant *W* given *X* and *Z*. The joint probability of the instance <*X=T, Y=F, W=T, Z=F*>, for example, is computed according to equation (4) as follows.

$$\begin{aligned} & P(X=T, Y=F, W=T, Z=F) \\ & = P(X=T) \cdot P(Y=F \mid X=T, Z=F) \cdot P(W=T) \cdot P(Z=F) \\ & = 0.85 \times 0.4 \times 0.25 \times 0.6 \\ & = 0.051 \end{aligned} \tag{5}$$

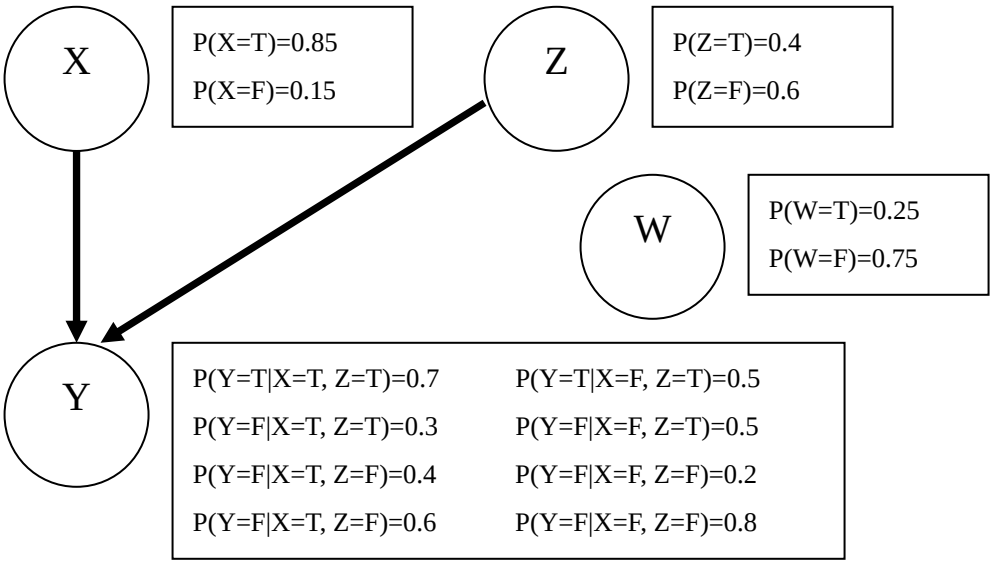

Figure 5.1. An example Bayesian belief network.

An intuitive interpretation of the relation between parents and children is that the relation is a *causal* relation. Variables other than the parent variables have no effect in determining the

value of the child variable. Therefore, the edge between two nodes can be interpreted as implying a causal relation between the two variables represented by the nodes.

In our case, we want to have a Bayesian belief network that computes the probability of a phonological form and captures the phonotactic constraints in terms of causal relation. For example, a network that encodes liquid harmony would be one that assigns higher probability to /sa.la.la/ than /sa.la.ra/ and one that has an edge from the node representing the first liquid to the node representing the second liquid in the word. We can manually build such a network, but we want a system that learns the relevant parameters (conditional probability tables, and presence or absence of conditional dependency between nodes) from scratch.

5.3. Learning a Bayesian belief network

Inducing a Bayesian belief network from data consists of connecting the nodes whose variables they represent are in causal relations and estimating the conditional probability table for each node. That is, learning consists of finding the structure of the network and estimating the network conditional probabilities. We follow the approach in Cooper and Herskovits (1992).

5.3.1. Finding the network structure

The most obvious way to learn the network structure is to enumerate all possible network structures and identify the one that best fits the data. That is, given the dataset *D*, learning is equivalent to solving the following equation where B_s is a possible network structure and B_{max} is the one that best fits the data.

$$B_{\max} = \arg\max_{B_s} P(B_s \mid D) \qquad (6)$$

One major problem with this approach is that the number of possible structures increases exponentially as a function of the number of nodes, and the time-complexity of solving equation (6) is exponential.

The K2 algorithm (Cooper and Herskovits, 1992) is a heuristic method that greedily searches for a structure that maximizes $P(B_s,D)$[15] in polynomial time. The basic idea of the K2 algorithm is the following. We start with a network where none of the nodes has parents. We incrementally add a parent to a node if adding the parent leads to a network with a higher $P(B_s,D)$ than the previous network, until the probability stops increasing. The precise pseudocode of the K2 algorithm in Cooper and Herskovits (1992) that incrementally adds parents and prints out the parents of each node is restated as follows.

---

[15] By definition of conditional probability,

$$P(B_s \mid D) = \frac{P(B_s, D)}{P(D)}$$

Since change in B_s does not lead to change in $P(D)$,

$$\arg\max_{B_s} P(B_s \mid D) = \arg\max_{B_s} \frac{P(B_s, D)}{P(D)}$$

Therefore, the structure that maximizes $P(B_s, D)$ also maximizes $P(B_s|D)$.

1. **procedure** K2;
2. {Input: A set of n nodes, an ordering on the nodes, an upper bound u on the number of parents a node may have, and a database D containing m cases.}
3. {Output: For each node, a printout of the parents of the node.}
4. **for** i := 1 **to** n **do**
5. $\pi_i := \emptyset$;
6. $P_{old} := g(i, \pi_i)$;
7. OKToProceed := **true**
8. **while** OKToProceed and $|\pi_i| < u$ **do**
9. let z be the node in Pred(x_i) - π_i that maximizes $g(i, \pi_i \cup \{z\})$;
10. $P_{new} := g(i, \pi_i \cup \{z\})$;
11. **if** $P_{new} > P_{old}$ **then**
12. $P_{old} := P_{new}$;
13. $\pi_i := \pi_i \cup \{z\}$;
14. **else** OKToProceed := **false**;
15. **end** {while};
16. **write**('Node:', x_i, 'Parents of this node:', π_i);
17. **end** {for};
18. **end** {K2};

The function $g(i, \pi_i)$ in line 6 of the pseudocode is intuitively interpreted as the probability of the database D given that the parents of x_i are π_i. The function is defined as follows:

$$g(i,\pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}! \qquad (7)$$

where:

π_i : set of parents of node x_i

$q_i = |\phi_i|$

ϕ_i : List of all possible instantiations of the parents of x_i in database D. That is, the Cartesian product of all possible values of the parent variables.

$r_i = |V_i|$

V_i : List of all possible values of the variable x_i

α_{ijk} : Number of instances in D where x_i is instantiated with its k^{th} value, and the parents x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

$N_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$: The number of instances in the database where the parents of x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

5.3.2. Estimating the conditional probabilities

Once the structure of the network is identified using the K2 algorithm, we must estimate the conditional probabilities associated with each node to fully define a Bayesian belief network. We follow Cooper and Herskovits (1992) in estimating the probabilities assuming the following:

1. The variables are discrete.
2. Given a belief network, each instance in database occurs independently.
3. No variable has missing values.
4. Given a belief network structure, the likelihood of a particular conditional probability assignment follows a uniform distribution.

Given the four assumptions, we estimate the probability that the variable x_i has value v_{ik} while its parent(s) π_i are instantiated with the j^{th} instantiation (w_{ij}) in the list of all possible parent instantiations as follows.

$$P(x_i = v_{ik} \mid \pi_i = w_{ij}) = \frac{\alpha_{ijk} + 1}{N_{ij} + r_i} \quad (8)$$

where:

ϕ_i : List of all possible instantiations of the parents of x_i in database D. That is, the Cartesian product of all possible values of the parent variables.

$r_i = |V_i|$

V_i : List of all possible values of the variable x_i

α_{ijk} : Number of instances in D where x_i is instantiated with its k^{th} value, and the parents x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

$N_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$: The number of instances in the database where the parents of x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

5.4. Simulation

We simulate the two grammaticality judgment experiments (Experiments 5 and 6) in Chapter 2 using a Bayesian belief network over the output nodes of the perceptron. The underlying assumption is that subjects first identify the most likely phoneme sequence given the auditory stimulus and then compute the probability that the sequence is a member of the artificial language exemplified by the study words to decide if it is grammatical or not. That is, we assume the process involves computing (9) and (10)

$$W^{\max} = \arg\max_{W^i} P(W_1^i, W_2^i, \ldots, W_n^i \mid input) \quad (9)$$

$$P(W \in L \mid W = W^{\max}) \tag{10}$$

where W^i denotes the i^{th} word in the set of all possible words and W^i_j denotes j^{th} phoneme of the word.

Following the set of assumptions in section 5.1, (10) is equivalent to (11).

$$P(W = W^{\max} \mid W \in L) \tag{11}$$

The basic idea is that (9) is computed by the perceptron and (11) is computed by a Bayesian belief network over the binary activation pattern in the output layer of the perceptron. The network is first trained on the list of most likely phoneme sequences identified by the perceptron given the study words. The trained network then computes probabilities for predictions from the perceptron given legal and illegal words. Subjects' discriminability between legal and illegal words is approximated by the difference in mean probabilities between legal and illegal words. Details of the simulation are described below.

5.4.1. Methods

We use a separate perceptron-Bayesian belief network pair for each grammaticality judgment experiment we simulate. Recall that we simulated auditory repetition experiments with three types of perceptrons: predefined perceptrons, naive perceptrons, and English perceptrons. Theoretically, as long as two different perceptrons identify the same phoneme sequence as the most likely sequence, coupling the belief network with either perceptron will lead to the same result. However, the three types of perceptrons do not always make

the same predictions, so we simulate the grammaticality judgment experiments separately for each of them.

For each trial in the experiment, the perceptron first makes its prediction for the trial stimulus. The stimulus word is represented in the same way as in the simulation studies for the auditory repetition experiments. The prediction of the perceptron is then converted to a vector of binary activation pattern. After all phonemes of the input word are presented and the output nodes are activated accordingly, the most highly activated node is identified for each block in the output layer. The activation value of the identified nodes is converted to one, while the values of all other nodes are converted to zero. The converted vector is used as input to the Bayesian belief network. The predicted vectors for the words presented to the subjects in the training blocks are used to train the belief network. The trained belief network computes the probability of each vector for the words in the test block to approximate subjects' grammaticality judgment.

In fact, the process of converting the activation pattern to a vector of binary activation pattern is equivalent to identifying the most likely phoneme for each word position. The *softmax* function (e.g., Bridle, 1989) is one way to interpret the output of a neural network as a posterior probability for a categorical variable. For a network of n output nodes, the posterior probability of the variable represented by the i^{th} output node is defined as in (12), where $input_j$ denotes the weighted sum of activations flowing into the j^{th} output node.

$$\frac{e^{input_i}}{\sum_{j=1}^{n} e^{input_j}} \qquad (12)$$

Recall that the output nodes are activated according to the sigmoid function of input

activation. Since the sigmoid function monotonically increases, the amount of input to an output node would be greater than the amount of input to any other nodes in the block if its activation is the highest in the block. Since the value of the softmax function is greater if the amount of input is greater, the posterior probability of a variable would be the greatest of all variables if the corresponding node is the most highly activated node of all nodes.

Each block of output nodes represents a specific word position, and each node within a block captures presence or absence of a particular phoneme in that word position. The value of the softmax function of an output node representing a phoneme x in the n^{th} block can be interpreted as the posterior probability that the n^{th} phoneme in the word is x given the perceptual input. Accordingly, the phoneme represented by the most highly activated node in the block is the most likely phoneme for the corresponding word position. Thus, setting the value of the most highly activated node in each block to one and setting the value of all other nodes to zero captures the process of identifying the most likely phoneme for each word position.

For each experiment, all study words and filler words in the training blocks are presented to the perceptron one by one. The prediction of the perceptron for each word is converted to a vector of binary activation pattern, where each component is a Boolean variable representing presence or absence of a phoneme in a particular word position. The resulting batch of such vectors is used to train the Bayesian belief network used for the given experiment. The structure of the network is induced by the K2 algorithm described in section 5.3.1. The conditional probabilities are estimated using the method described in section 5.3.2.

Once the Bayesian belief network is trained, all words in the test block are presented to the perceptron one by one. The prediction of the perceptron for each test word is first converted to a vector of binary activation pattern as in the training phase. The

Bayesian belief network then computes the probability of the converted vector. To approximate the subject's discriminability between legal and illegal words, we compute the difference in mean probabilities between legal and illegal words.

5.4.2. Results

5.4.2.1. Coupled with the predefined perceptrons

Figure 5.2 and Table 5.1 summarize the mean probabilities computed by the trained Bayesian belief network coupled with the predefined perceptrons for legal and illegal words in both liquid harmony and backness harmony.

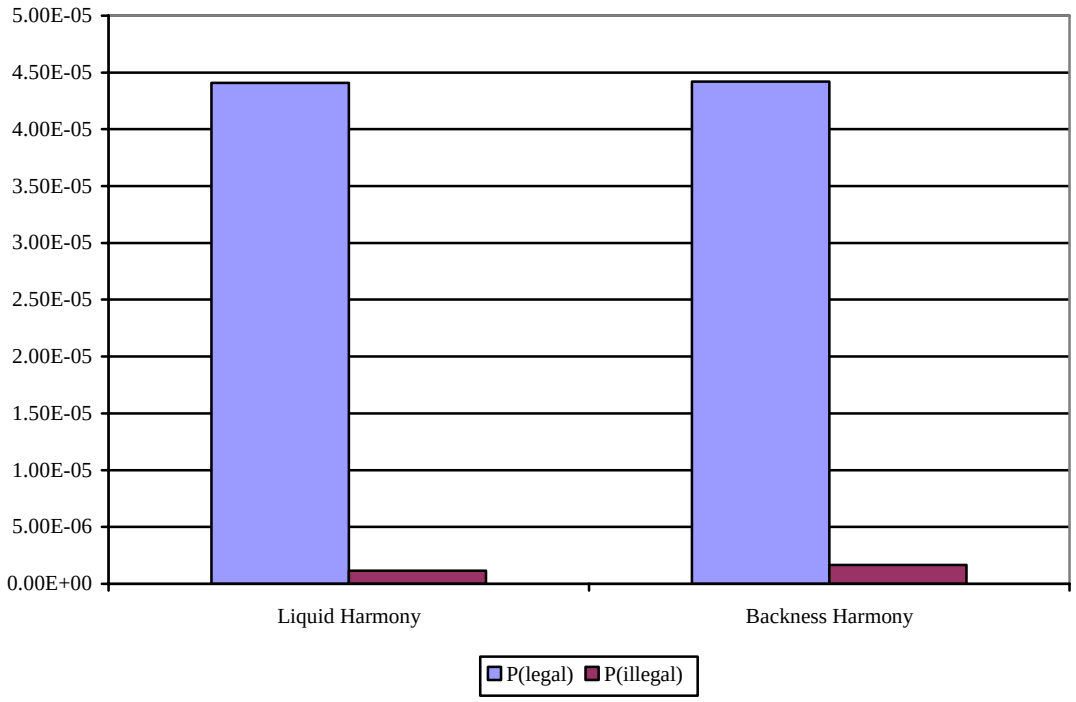

Figure 5.2. Mean Bayesian belief network probabilities for legal and illegal words. The vertical axis shows the predicted mean probability.

	Liquid Harmony	Backness Harmony
Legal	4.41E-05	4.42E-05
Illegal	1.15E-06	1.65E-06

Table 5.1. Mean Bayesian belief network probabilities for legal and illegal words.

The trained Bayesian belief network assigned higher probabilities for legal words than for illegal words in both liquid harmony ($t(14) = 12.956$, $p < .001$) and backness harmony ($t(14) = 16.827$, $p < .001$). Moreover, the difference in probability between legal and illegal words was not different across the two conditions ($t(28) = 0.071$, $p = .944$). This contrasts with the simulation studies for the auditory repetition experiments where the difference between legal and illegal was significantly different between the two conditions ($t(28) = 3.88$, $p = .001$). The perceptron word scores which approximate subjects' latencies in the corresponding simulation study are repeated in Figure 5.3 and Table 5.2 for comparison.

	Liquid Harmony	Backness Harmony
Legal	0.498787	0.685874
Illegal	0.410023	0.673771

Table 5.2. Mean perceptron word scores for legal and illegal words.

This contrast suggests that subjects' grammaticality judgment is less affected by how perceptually confusable the constrained phonemes are to each other than their auditory repetition task.

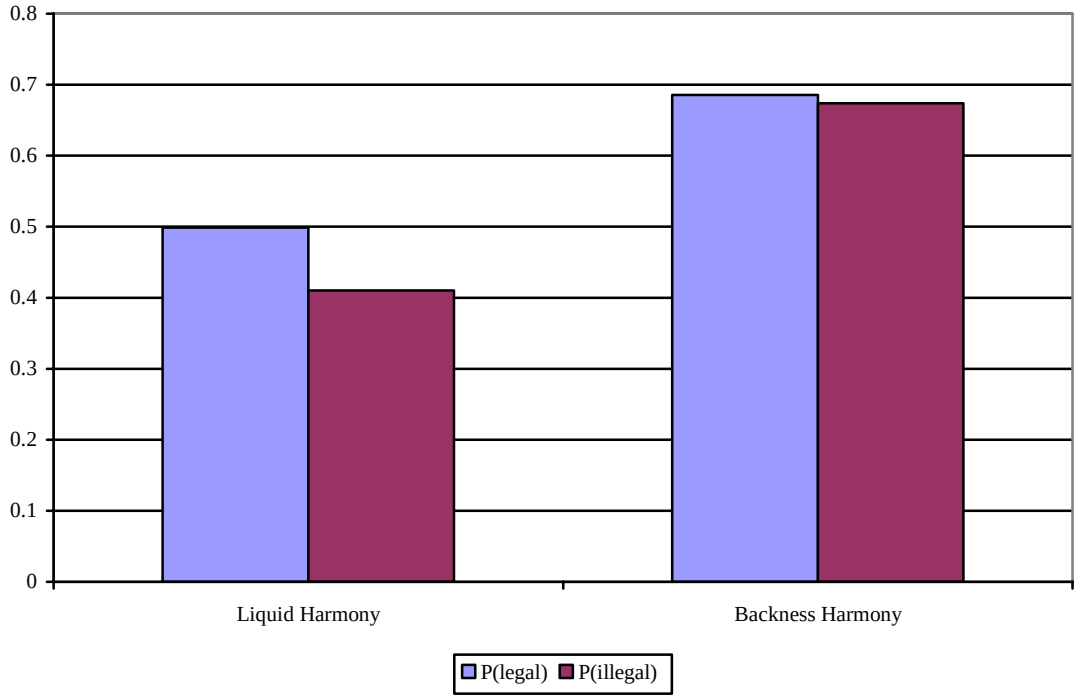

Figure 5.3. Mean perceptron word scores for legal and illegal words. The vertical axis shows the predicted word scores.

5.4.2.2. Coupled with the naïve perceptrons

Figure 5.5, Figure 5.6 and Table 5.3 summarize the mean probabilities computed by the trained Bayesian belief network coupled with the naive perceptrons for legal and illegal words in both liquid harmony and backness harmony experiments.

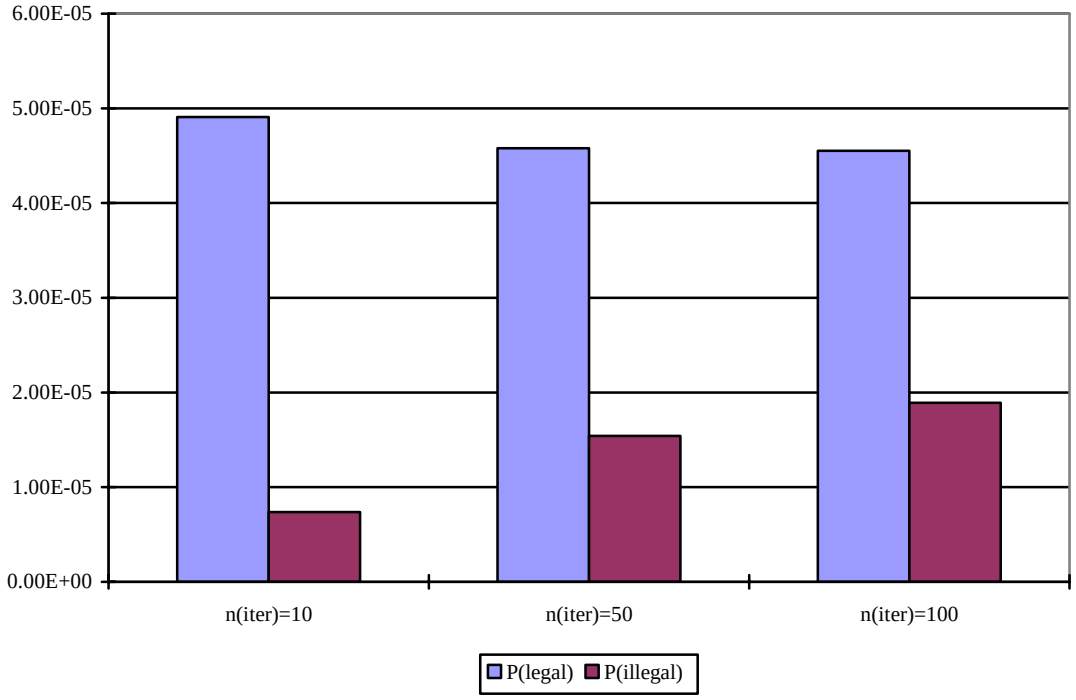

Figure 5.4. Mean probabilities for legal and illegal words in liquid harmony simulation using belief network coupled with the naive perceptrons. The vertical axis shows the predicted mean probability.

Exp	*n*(iter)	*P(legal)*	*P(illegal)*	Difference	Statistical significance
liquid harmony	10	4.91E-05	7.36E-06	4.18E-05	$t(14) = 13.391$, $p < .001$
	50	4.58E-05	1.54E-05	3.04E-05	$t(14) = 7.600$, $p < .001$
	100	4.55E-05	1.89E-05	2.65E-05	$t(14) = 6.041$, $p < .001$
Backness harmony	10	4.96E-05	2.34E-06	4.73E-05	$t(14) = 10.087$, $p < .001$
	50	4.69E-05	1.70E-06	4.52E-05	$t(14) = 11.547$, $p < .001$
	100	4.70E-05	1.69E-06	4.53E-05	$t(14) = 13.054$, $p < .001$

Table 5.3. Summary of probabilities computed by the Bayesian belief network coupled with naïve perceptrons. Difference is equal to *P(legal)* - *P(illegal)*.

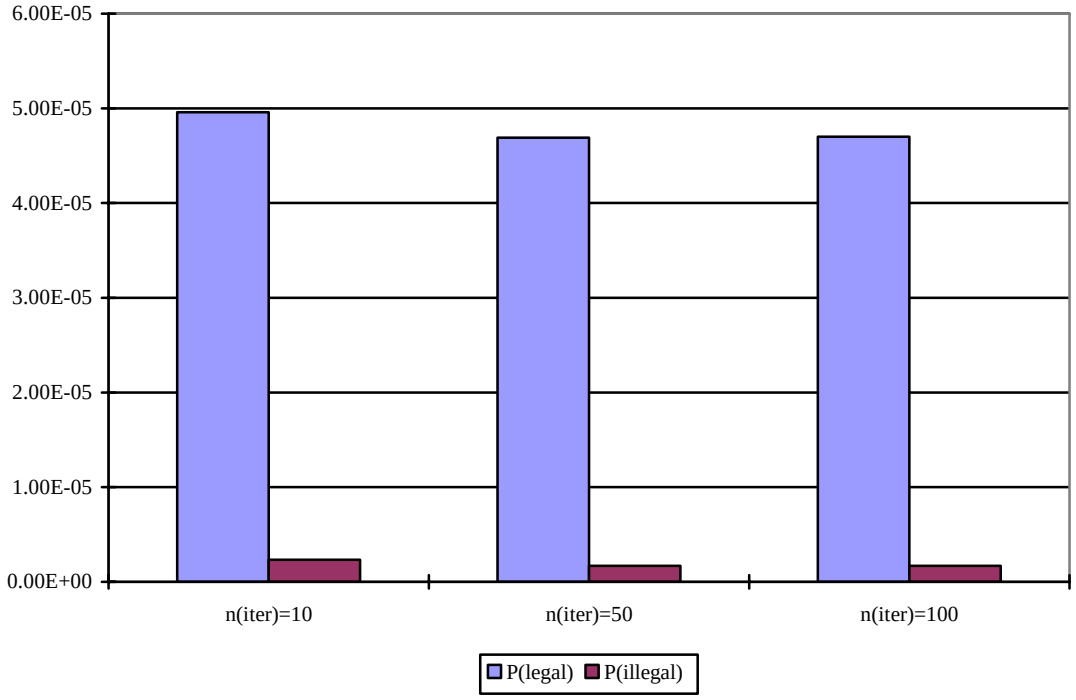

Figure 5.5. Mean probabilities for legal and illegal words in backness harmony simulation using belief network coupled with the naive perceptrons. The vertical axis shows the predicted mean probability.

The Bayesian belief network assigned higher probabilities to legal words than to illegal words in both conditions regardless of how long the perceptron was trained. The size of difference in probability between legal and illegal words was not different across the two conditions when the paired perceptron was trained over ten iterations ($t(28) = -0.974$, $p = .338$). This contrasts with the simulation studies for the auditory repetition experiments where the size of difference between legal and illegal was significantly different between the two conditions (n(iter) = 10: $t(28) = 2.076$, $p = .047$; n(iter) = 50: $t(28) = 5.127$, $p < .001$; n(iter) = 100: $t(28) = 5.333$, $p < .001$).

However, the size of difference in probability assigned by the belief network

between legal and illegal words was significantly different across the two conditions when the paired perceptron was trained longer (*n*(iter) = 50: *t*(28) = -2.646, *p* = .013; *n*(iter) = 100: *t*(28) = -3.357, *p* = .002). This is because the perceptron becomes more sensitive to the phonotactic constraint as iteration increases, and hyper-corrects some illegal words to the corresponding legal words. The belief network assigns higher probability to such hyper-corrected words, which leads to increase in mean probability for illegal words as shown in Figure 5.4. The tendency to hypercorrect is more salient in liquid harmony simulation because the weights encoding the phonotactic knowledge plays a bigger role as the confusion between the constrained phonemes is greater than in backness harmony simulation. Therefore, the size of difference in probability between legal and illegal words reduces in liquid harmony, whereas the size of difference remains relatively unchanged in backness harmony.

Exp	*n*(iter)	*P(legal)*	*P(illegal)*	Difference	Statistical significance
liquid harmony	10	4.79E-05	1.70E-06	4.62E-05	$t(14) = 13.762$, $p < .001$
	50	4.42E-05	1.21E-06	4.30E-05	$t(14) = 12.434$, $p < .001$
	100	4.44E-05	1.20E-06	4.32E-05	$t(14) = 12.318$, $p < .001$
Backness harmony	10	4.63E-05	2.24E-06	4.41E-05	$t(14) = 11.128$, $p < .001$
	50	4.62E-05	1.70E-06	4.45E-05	$t(14) = 12.678$, $p < .001$
	100	4.57E-05	1.68E-06	4.40E-05	$t(14) = 14.080$, $p < .001$

Table 5.4. Summary of probabilities with the errors removed. Difference is equal to *P(legal) – P(illegal)*.

Table 5.4 summarizes the probabilities computed by the same belief network with the errors removed. With the errors removed, the difference is no longer significant across the two conditions (*n*(iter) = 10: *t*(28) = 0.408, *p* = .686; *n*(iter) = 50: *t*(28) = -0.292, *p* = .772;

n(iter) = 100: *t*(28) = -0.164, *p* = .871), while the mean probability for legal words remains significantly larger than the mean probability for illegal words in both conditions.

5.4.2.3. Coupled with the English perceptrons

Figure 5.6, Figure 5.7 and Table 5.5 summarize the mean probabilities computed by the trained Bayesian belief network coupled with the English perceptrons for legal and illegal words in both liquid harmony and backness harmony. The Bayesian belief network assigned higher probabilities for legal words than for illegal words in both liquid harmony and backness harmony. Moreover, the size of difference in probability between legal and illegal words was not different across the two conditions (*n*(iter) = 10: *t*(28) = 0.003, *p* = .997; *n*(iter) = 50: *t*(28) = -0.036, *p* = .971: *n*(iter) = 100: *t*(28) = 0.065, *p* = .949).

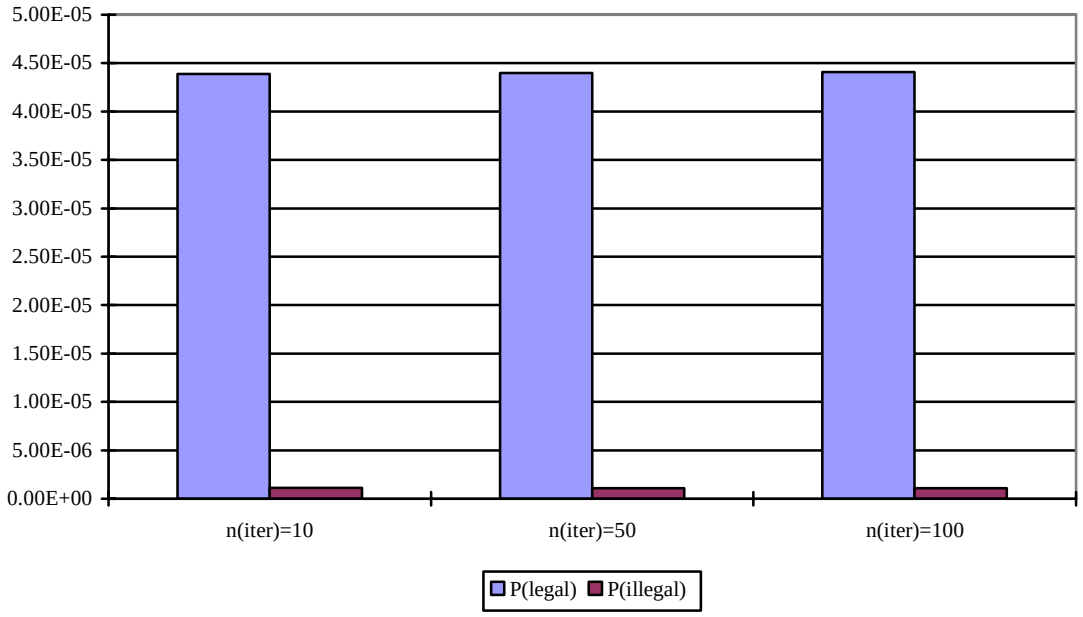

Figure 5.6. Mean probabilities for legal and illegal words in liquid harmony simulation using belief network coupled with the English perceptrons. The vertical axis shows the predicted mean probability.

Exp	*n*(iter)	*P(legal)*	*P(illegal)*	Difference	Statistical significance
liquid harmony	10	4.39E-05	1.11E-06	4.28E-05	$t(14) = 12.982$, $p < .001$
	50	4.40E-05	1.10E-06	4.29E-05	$t(14) = 12.037$, $p < .001$
	100	4.41E-05	1.10E-06	4.30E-05	$t(14) = 12.117$, $p < .001$
Backness harmony	10	4.47E-05	1.91E-06	4.28E-05	$t(14) = 14.104$, $p < .001$
	50	4.47E-05	1.63E-06	4.31E-05	$t(14) = 14.861$, $p < .001$
	100	4.43E-05	1.63E-06	4.27E-05	$t(14) = 14.798$, $p < .001$

Table 5.5. Summary of probabilities computed by the Bayesian belief network coupled with the English perceptrons. Difference is equal to *P(legal) – P(illegal)*.

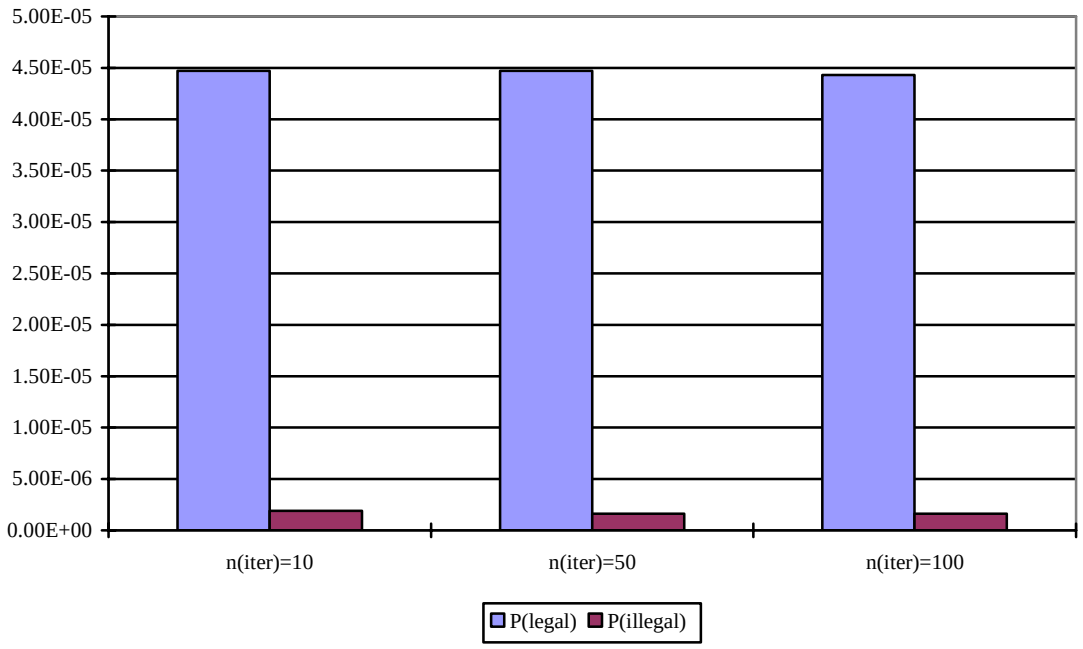

Figure 5.7. Mean probabilities for legal and illegal words in backness harmony simulation using belief network coupled with the English perceptrons. The vertical axis shows the predicted mean probability.

5.4.3. Summary

The Bayesian belief network coupled with a single layered recurrent perceptron simulates the grammaticality judgment experiments correctly. It learns to compute word probabilities from the predictions made by the perceptron given the words in the training block. After learning is complete, the belief network correctly assigns higher probabilities to legal words than to illegal words in simulating both liquid harmony experiment and backness harmony experiment. Furthermore, the predicted size of legality effect on grammaticality judgment is not significantly different between the two experiments. This is consistent with the results from the grammaticality judgment experiments in Chapter 2 and contrasts with the results from the simulation of auditory repetition experiments where the size of legality effect does differ between experiments.

The success of simulation is in part due to the fact that the probability of each word was computed over the most likely phoneme sequence identified by the perceptron. The original activation pattern reflecting the confusability information is converted to a binary activation pattern before it is passed onto the belief network. This conversion reflects the assumption that as long as a phoneme sequence best matching the perceptual input is identified, the listener judges grammaticality of the word based on its formal property (how consistent it is with the phonotactics of the language), but not based on with what certainty the spoken word was recognized.

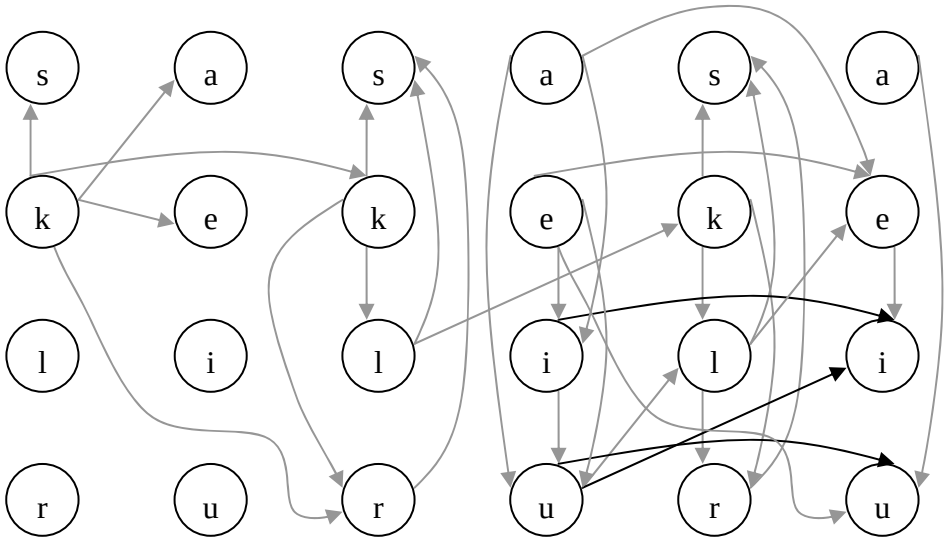

Figure 5.8. Graphical illustration of segmental dependency learned by the K2 algorithm for backness harmony.

One useful feature of the Bayesian belief network is that an edge between the nodes implies that there is a dependency between the phonemes that the nodes represent. For example, consider Figure 5.8 which summarizes the structure of the network identified by the K2 algorithm for a backness harmony experiment. For example, note that the edges between the two non-adjacent /i/'s and /u/'s indicate that there is dependency between the non-adjacent vowels. This suggests that the probabilities $P(W_5=/i/|W_3=/i/)$, $P(W_5=/i/|W_3=/u/)$, and $P(W_5=/u/|W_3=/u/)$ were factors in computing the overall word probability $P(W=W_1,W_2,W_3,W_4,W_5,W_6)$. Had the dependency not been learned, the word probability would have been computed agnostic to the co-occurrence between the non-adjacent vowels. In addition, note that the belief network learned dependencies between other nodes as well. This means computing the word probability does not solely rely on the dependency

between the non-adjacent vowels. If we assumed the word probability are directly reflected in speakers' grammaticality judgment behavior, this could explain the tendency noted by Coleman and Pierrehumbert (1997) that the judgment is not solely dependent on the unacceptability of a specific part of the word.

5.5. Chapter summary

This chapter presented a Bayesian belief network that learns phonotactic constraints as the probability distribution on phoneme co-occurrence. The set of phoneme sequences that are the result of perception constitutes the training set for the belief network. The probability distribution is learned by discovering conditional dependencies between position-specific phonemes using the K2 algorithm and estimating the conditional probabilities between the dependent phonemes using maximum likelihood estimate so that the observed probability of the words in the training set is maximized. Based on the learned probability distribution, the Bayesian belief network computes the phonotactic probability of novel phoneme sequences necessary to judge their grammaticality.

When coupled with the connectionist model of perceptual facilitation, the Bayesian belief network duplicates the results of the two grammaticality judgment experiments in Chapter 2. Subjects in both liquid harmony (Experiment 5) and backness harmony (Experiment 6) condition learned to discriminate between legal and illegal words equally well. Similarly, the belief network consistently assigned higher probabilities to legal words than to illegal words in both conditions equally well. By appending the Bayesian belief network to the connectionist model, I implemented the idea that perception and grammaticality judgment take place in two separate stages: a speaker first identifies the phonological structure that best matches the perceptual input and then computes the

grammaticality of the identified phonological structure. It was shown in Chapter 4 that the connectionist model can alone simulate the effect of phonotactic learning on perception. The results in this chapter show that when the phonotactic probability of the output of the connectionist model is computed by the appended Bayesian belief network, the effect of phonotactic learning on grammaticality judgment can be simulated.

In addition to its ability to successfully simulate the grammaticality judgment experiments, the belief network is an attractive candidate for the model of phonological competence because it is straightforward to identify the segmental dependencies that affect speakers' grammaticality judgment. By printing out the parent-child relations identified by the K2 algorithm, it is possible to track the conditionally dependent segments that factored in computing the probability of a word as a whole.

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

6.1. Summary

The dissertation addressed two questions regarding phonotactic learning in the adult phonological processing system. How malleable is the adult phonological processing system? How does cumulating processing experience cause the system to change its phonological processing behavior? In particular, this dissertation investigated whether the adult phonological processing system can learn co-occurrence restrictions on non-adjacent phonemes and how its behavior changes as a result of learning by presenting eight experiments and two computational models.

The phonotactic constraints in this dissertation restricted co-occurrence of two non-adjacent phonemes that are one phoneme apart from each other. In terms of order of complexity, they are second order phonotactic constraints contingent on a non-adjacent phoneme. Due to their non-adjacency, they were suspected to be harder to learn than the second order constraints contingent on an adjacent vowel examined in previous studies on malleability of the adult phonological processing system (Dell et al., 2000; Onishi et al., 2002). The six experiments discussed in Chapter 2 show that adults can learn non-adjacent phonotactic constraints from brief perception and production experience. The results add to the growing body of evidence for malleability of the adult phonological processing system, and extend the range of phonotactic constraints that can be learned.

The four non-adjacent phonotactic constraints discussed in Chapter 2 differ with respect to the typological frequency of their natural language counterparts. Following the conjecture relating typological frequency of sound patterns with their learnability (Newport

and Aslin, 2004; Moreton, 2006), one would expect the four constraints to differ in their learnability such that constraints of higher typological frequency would be more readily learned than constraints of lower typological frequency. However, the six experiments in Chapter 2 show that the four constraints are equally well learned despite the difference in their typological frequency.

The auditory repetition experiments in Chapters 2 and 3 suggest the necessity for a more detailed account of how phonotactic learning facilitates perception of phonotactically legal words. I proposed the perceptual facilitation hypothesis such that phonotactic knowledge facilitates perception of legal words more if the constrained phonemes are more confusable to each other. The basic idea is that phonotactic knowledge facilitates perception of phonotactically legal words by reducing the inherent perceptual confusion. Assuming non-linearity in perception performance, there is more room for the phonotactic knowledge to reduce confusion inherent in legal words and increase confusion inherent in illegal words, if the phonotactic constraint targets a perceptually more confusable phoneme pair.

The single layered recurrent perceptron presented in Chapter 4 is a connectionist model of how learning a phonotactic constraint facilitates perception of phonotactically legal words. Phonotactic knowledge is encoded among the connection weights. The activation pattern reflecting perceptual confusion inherent in the input word spreads along the connections and is modified according to the phonotactic knowledge. Initial activation pattern is modified in such a way that the constituent phonemes of phonotactically legal words end up with higher activation while the constituent phonemes of illegal words end up with lower activation. The higher activation assigned to the constituent phonemes of legal words implies that legal words will be perceived more accurately and quickly than illegal words. Simulation results with perceptrons initialized in various ways are generally consistent with the prediction of the perceptual facilitation hypothesis and the results of the

auditory repetition experiments in the dissertation.

The Bayesian belief network in Chapter 5 is presented as an explicit stochastic model of how phonotactic constraints are learned from the words instantiating the constraint and constitute the grammatical knowledge of the speaker. The phonotactic knowledge is represented by conditional dependency relation between two or more phonemes and (conditional) probability table associated with each phoneme. The belief network can learn phonotactic constraints using the K2 algorithm and maximum likelihood estimate, and compute probability of a word accordingly. Coupled with the single layered recurrent perceptron, the trained network correctly simulates the two grammaticality judgment experiments in Chapter 2. It consistently assigns higher probability to phonotactically legal words than to illegal words equally well for both liquid harmony and backness harmony.

6.2. Future directions

There are many research questions to be investigated stemming from the experiments and computational models in the dissertation. I briefly discuss the following four in particular: non-adjacency of phonotactic constraints, locus of facilitation in auditory repetition task, level of phonological representation, and application of the Bayesian belief network in automatic language identification problem.

6.2.1. Non-adjacency of phonotactic constraints

Non-adjacent phonotactic constraints in the dissertation were assumed to be harder to learn than the constraints on adjacent phonemes because it required learner to process at least

three phonemes at once to detect the dependency between the non-adjacent phonemes, where as processing two phonemes at once would have sufficed to detect the dependency between adjacent phonemes. However, in the framework of autosegmental phonology (Goldsmith, 1976), all of the constraints can be interpreted as constraints on adjacent slots if we assumed a separate tier for consonants and vowels as assumed by McCarthy (1979, 1981). A possible interpretation, as also pointed out by Newport and Aslin (2004), is that the subjects learned the constraints because the constrained units are adjacent within the tier. The question that follows is whether adults can also learn phonotactic constraints over non-adjacent consonant slots or vowel slots.

6.2.2. Locus of facilitation in auditory repetition task

The assumption underlying the perceptual facilitation hypothesis and the single layered recurrent perceptron was that the subjects' latency measured in auditory repetition experiments only reflects their performance in perception. The assumption was primarily based on Onishi et al. (2002) where subjects' latency to auditory repetition task for phonotactically legal words became faster after they listened to study words and rated clarity of articulation. As there was no production task involved while subjects were trained on study words, the reduction in their latency in the test phase was arguably due to the effect of learning on perception rather than production. However, it is not clear at this point whether learning a phonotactic constraint simultaneously affects both perception and production, or whether there is a transfer of phonotactic knowledge between perception and production. In addition, the subjects in our experiments were trained both by listening and repeating study words. Therefore, it is not clear whether reduction in auditory repetition latency examined in this dissertation is due to effect of learning on perception, production,

or both.

6.2.3. Level of phonological representation

In the computational models, a (non-)word was represented as a sequence of component phonemes. The perceptron approximated subjects' perceptual latency or accuracy to an input non-word by the activation pattern assigned to the component phoneme sequence. The Bayesian belief network approximated subjects' judgment of grammaticality of the non-word by computing the joint probability of observing the component phoneme sequence in a language characterized by the phonotactics. The emphasis in the dissertation was on the component phoneme sequence because the constraints whose learnability was tested in the experiments were restrictions on which two non-adjacent phonemes can or cannot co-occur. However, experiments also show that speakers are also sensitive to dependency between sub-segmental features (Goldrick, 2004) and supra-segmental information such as whether the phonotactically constrained syllable is word-initial or not (Vitevitch et al., 1997). Therefore, ideally, non-words must be represented at multiple phonological levels, and a model should successfully integrate the information from all levels of representation.

6.2.4. Application in automatic language identification

Automatic language identification is the problem of identifying the language being spoken from a sample of utterances (Muthusamy et al., 1994). Phonetic and phonological features such as spectral characteristics (e.g., Cimarusti and Ives, 1982; Li, 1994), phonotactics (e.g., Hazen and Zue, 1993; Zissman and Singer, 1994), prosody (e.g., Itahasi et al., 1994;

Thyme-Gobbel and Hutchins, 1996) have been used, often in combination, to solve the problem. A common approach to language identification problem is to build a system that returns a score for each language integrated over the individual scores with respect to each feature for the given utterance. The system returns the language with the highest score as the answer to the problem.

Zissman and Berkling (2001) argue that systems that incorporate higher level linguistic information such as phonotactics perform more accurately than systems that do not. The Bayesian belief network presented in Chapter 5 can be trained to compute the phonotactic probability of a phoneme sequence. Moreover, compared with the standard *n*-gram model of phonotactics, the belief network can represent the probability distribution more efficiently. The *n*-gram model must store the joint probabilities for all possible *n*-grams. However, once the conditional independence between variables is identified, the joint probabilities can be represented with a smaller number of conditional probabilities (Russell and Norvig, 2003). Therefore, the Bayesian belief network has the potential to be used as a component of the language identification system that computes phonotactic probability.

REFERENCES

Albright, A. (2003). SimilarityCalculator.pl, *Segmental Similarity Calculator, using the "shared natural classes" method of Frisch, Broe, and Pierrehumbert (1997)*.

Albright, A. (2006). Gradient phonotactic effects: Lexical? grammatical? both? neither? Paper presented at the 81st Annual Meeting of the Linguistic Society of America, Albuquerque, New Mexico, January 7, 2006.

Arakawa, S. (1977). *Gairaigo ziten*, 2nd edition. Tokyo: Kadokawa.

Auer, E. T. (1993). *Dynamic processing in spoken word recognition: The influence of paradigmatic and syntagmatic states*. Ph.D. dissertation, University at Buffalo, Buffalo, NY.

Bailey, T. M., and Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*, 568-591.

Bailey, T. M., and Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language, 52*, 339-262.

Blevins, J. (2004). *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*, *6*, 451-459.

Bridle, J. (1989). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman-Soulie & J. Hérault (Eds.) *Neurocomputing: Algorithms, Architectures, and Applications*. New York: Springer-Verlag.

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

Chambers, K. E., Onish, K. H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*, B69-B77.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

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

Cimarusti, D., and Ives, R. B. (1982). Development of an automatic identification system of spoken languages: Phase I. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, 1661-1663.

Cleeremans, A., and McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, *3*, 235-253.

Coleman, J. S., and Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology, Third meeting of the ACL special interest group in computational phonology* (pp. 49-56). Somerset, NJ: Association for Computational Linguistics.

Cooper, G., and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, *9*, 309-347.

de Blois, K. F. (1975). Bukusu generative phonology and aspects of Bantu structure. (Annałes no. 85, Sciences humaines, Series in 8^{o}.) Tervuren: Musée royal de l'Afrique centrale.

Dell, G. S., Reed, K. D., Adams, D. R., and Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1355-1367.

Dell, G. S., and Warker, J. A. (2004). The tongue slips into (recently learned) patterns. In H. Quene and V. van Heuvan (Eds.), *On speech and language: Studies for Sieb G. Nooteboom* (pp. 45-56). Utrecht, the Netherlands: Netherlands Graduate School of Linguistics.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Evans, J. L., Viele, K., Kass, R. E., and Tang, F. (2002). Grammatical morphology and perception of synthetic and natural speech in children with specific language impairments. *Journal of Speech, Language, and Hearing Research*, *45*, 494-504.

Fallon, P. D. (1993). Liquid dissimilation in Georgian. In Andreas Kathol and michael Bernstein (Eds.), *Proceedings of the 10th Eastern States Conference on Linguistics*, 105-116.

Friederici, A. D., and Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception and Psychophysics*, *54*, 287-295.

Frisch, S. A., Broe, M., and Pierrehumbert, J. (1997). Similarity and phonotactics in Arabic. *Rutgers Optimality Archive* [Online], ROA-223-1097. Available at http://www.web-slingerz.com/cgi-bin/oa_list.cgi.

Frisch, S. A., Large, N. R., and Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*, 481-496.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 27-52.

Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*, 583-604.

Goldsmith, J. A. (1976). *Autosegmental Phonology*. Ph.D. dissertation, MIT, Cambridge, Massachusetts.

Goldsmith, J. A. (1985). Vowel harmony in Khalkha Mongolian, Yaka, Finnish, and Hungarian. In C. Ewen and J. Anderson (Eds.), *Phonology Yearbook 2* (pp. 253-275). New York: Cambridge University Press.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431-436.

Gomez, R. L., and Maye, J. (2005). The developmental trajectory of non-adjacent dependency learning. *Infancy 7*, 183-206.

Greenberg, J. (1950). The patterning of root morphemes in Semitic. *Word*, *5*, 162-181.

Grossberg, S., Boardman, I., and Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 483-503.

Gureckis, T. M., and Love, B. C. (2005). A critical look at the mechanisms underlying implicit sequence learning. *Proceedings of the 27th Annual Conference of Cognitive Science Society*.

Hansson, G. O. (2001). *Theoretical and typological issues in consonant harmony*. Ph.D. dissertation, University of California-Berkeley, Berkeley, California.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers*, *27*, 46-51.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. New Jersey: Prentice Hall.

Hazen, T. J., and Zue, V. W. (1993). Automatic language identification using a segment-based approach. *Proceedings of Eurospeech*, *2*, 1303-1306.

Herbert, R., and Poppe, N. (1963). *Kirghiz Manual*. Bloomington, Indiana: Indiana University Publications.

Itahashi, S., Zhou, J., Tanaka, K. (1994). Spoken language discrimination using speech fundamental frequency. *Proceedings of International Conference on Spoken Language Processing*, *4*, 1899-1902.

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

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402-420.

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630-645.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, *20*, 219-273.

Karlsson, F. (1999). *Finnish: An essential grammar*. New York: Routledge.

Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Oxford: Blackwell.

Kessler, B., and Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, *37*, 295-311.

Krueger, J. (1962). *Yakut Manual*. Bloomington, Indiana: Mouton & Co.

Leigh, P., and Charles-Luce, J. (2002). Developmental effects of phonotactic probability and lexicality on duration in speech production. In C.T. McLennan, P. A., Luce, G. Mauner, and J. Charles-Luce (Eds.), *University at Buffalo Working Papers on Language and Perception*, *1*, 603-653.

Li, K.-P. (1994). Automatic language identification using syllabic spectral features. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, *1*, 297-300.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Ph.D. dissertation, Indiana University, Bloomington, Indiana.

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

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

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren and C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 21-57). Hillsdale, NJ: Erlbaum.

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

Mattys, S. L., Jusczyk, P. W., Luce, P. A., and Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465-495.

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*, 91-121.

McCarthy, J. (1979). *Formal Problems in Semitic Phonology and Morphology*. Ph.D. dissertation, MIT, Cambridge, Massachusetts.

McCarthy, J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12*, 373-418.

McCarthy, J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry, 172*, 207-263.

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

McLennan, C. T., Luce, P. A., La Vigne, R. and Charles-Luce, J. (2005). Changes in adult speech perception and implicit learning of non-adjacent phonotactic dependencies. In Valerie Hazan and Paul Iverson (Eds.), *Proceedings of ISCA Workshop on Plasticity in Speech Perception*, London, U.K., June 15-17, 2005.

Mirman, D., McClelland, J. L., and Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language, 52*, 424-443.

Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.

Mody, M., Studdert-Kennedy, M., and Brady, S. (1997). Speech perception deficit in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64*, 199-231.

Moreton, E. (2006). Phonotactic learning and phonological typology. Paper presented at the 37th meeting of the Northeast Linguistics Society, Urbana, Illinois, October 13-15, 2006.

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

Morrison, G. S. (2004). Are functional constraints active synchronically? Poster presented at the Ninth Conference on Laboratory Phonology, Urbana, IL, June 25, 2004.

Morrison, G. S., and Kirchner, R. (2007). Phonetic naturalness and phonological learnability. In Tracy O'Brien (Ed.), *University of Alberta Working Papers in Linguistics*, *3*.

Muthusamy, Y. K., Barnard, E., and Cole, R. A. (1994). Reviewing automatic language identification. *IEEE Signal Processing Magazine*, *4*, 33-41.

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

Norris, D. (1990). A dynamic-net model of human speech recognition. In G.T.M. Altmann (Ed.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, Cambridge, MA: MIT Press, 87-105.

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

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (1997). The possible-word constraint in continuous speech. *Cognitive Psychology*, *34*, 191-243.

Odden, D. (1994). Adjacency parameters in phonology. *Language*, *70*, 289-330.

Ohala, J. J. (1993). The phonetics of sound change. In Charles Jones (Ed.), *Historical linguistics: problems and perspectives* (pp. 237-278). London: Longman.

Okada, K., (2005). *Phonological Processing in Speech Perception and Production: FMRI Investigation*. Ph.D. dissertation, University of California-Irvine, Irvine, California.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13-B23.

Peperkamp, S., Skoruppa, K., and Dupoux, E. (2005). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia, and C. Zaller (Eds.), *Proceedings of the 30th annual Boston University Conference on Language Development* (pp. 464-475). Somerville, MA: Cascadilla Press.

Pitt, M. A., and McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347-370.

Poppe, N. (1963). *Tatar Manual: Descriptive Grammar and Texts with a Tatar-English Glossary*. Bloomington, Indiana: Mouton & Co.

Poppe, N. (1965). *Introduction to Altaic Linguistics*. Wiesbaden: Otto Harrassowitz.

Prince, A., and Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. MS, Rutgers University and University of Colorado.

Pycha, A., Nowak, P. Shin, E. and Shosted, S. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In Gina Garding and Mimu Tsujimura (Eds.), *Proceedings of the XXIIth West Coast Conference on Formal Linguistics* (pp. 423-435). Somerville, MA: Cascadilla Press.

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

Rehg, K. L., and Sohl, D. G. (1981). *Ponapean Reference Grammar*. Honolulu: The University Press of Hawaii.

Roca, I., and Johnson, W. (1999). *A Course in Phonology*. Oxford: Blackwell.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249-284.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 6, 386-408.

Russell, S., and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2nd edition). New Jersey: Prentice Hall.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month old infants. *Science*, *274*, 1926-1928.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606-621.

Saffran, J. R., and Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, *39*, *3*, 484-494.

Siptár, P., and Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford: Oxford University Press.

Snowling, M. J. (2004). The science of dyslexia: A review of contemporary approaches. In Martin Turner and John Paul Rack (Eds.), *The Study of Dyslexia*. New York: Kluwer Academic Publishers.

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavioral Research Methods, Instruments, and Computers*, *31*, 137-149.

Stemberger, J. P. (1982). *The lexicon in a model of language production*. Ph.D. dissertation, University of California-San Diego, San Diego, California.

Storkel, H. L., Jonna, A., and Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language and Hearing Research*, *49*, *6*, 1175-1192.

Testen, D. (1997). *Ossetic phonology*. In Alan Kaye (Ed.), *Phonologies of Asia and Africa*, *Volume 2* (pp. 707-731). Winona Lake, Ind.: Eisenbrauns.

Thyme-Gobbel, A. E., and Hutchins, S. E. (1996). On using prosodic cues in automatic language identification. *Proceedings of International Conference on Spoken Language Processing*, *3*, 1768-1772.

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

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

Vitevitch, M. S., and Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, *9*, 325-329.

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374-408.

Vitevitch, M. S., and Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, 36, 481-487.

Wagner, R. K., and Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, *2*, 192-212.

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

Weide, R. L. (1998). The CMU pronunciation dictionary, release 0.6, available at http://speech.cs.cmu.edu/cgi-bin/cmudict.

Whalen, C. A., and Dell, G. S. (2006). Speaking outside the box: Learning of non-native phonotactic constraints is revealed in speech errors. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver, Canada.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In Gina Garding and Mimu Tsujimura (Eds.), *Proceedings of the XXIIth West Coast Conference on Formal Linguistics* (pp. 533-546). Somerville, MA: Cascadilla Press.

Zissman, M. A., and Singer, E. (1994). Automatic language identification of telephone speech messages using phoneme recognition and n-gram modeling. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, *1*, 305-308.

Zissman, M. A., and Berkling, K. M. (2001). Automatic language identification. *Speech Communication*, *35*, 115-124.

AUTHOR'S BIOGRAPHY

Hahn Koo was born in Honolulu, Hawaii, on December 30, 1977. He grew up in Seoul, Republic of Korea and graduated from Seoul National University in 2002 with a degree in English language and literature. While pursuing his graduate study in Linguistics, he worked as a research intern at the Center for Human Interaction Research, Motorola Labs in 2004 and 2005. Following the completion of his Ph.D., Koo will begin to work for Nuance Communications as a text-to-speech language specialist.