Harmony versus Distance in Phonetic Enhancement*

Adam Wayment[1], Luigi Burzio, Donald Mathis, Robert Frank

Department of Cognitive Science, Johns Hopkins University

1. Introduction

Pioneering work in phonology claimed that phoneme inventories involve representations over minimally contrastive, or distinctive, features (Halle & Jakobson, 1956). It is now understood (e.g. Clements, 2003) that most languages only make use of a small fraction of the number of possible contrasts. Instead of combining freely, some distinctive features tend to cluster with others. The prototypical example is the clustering of backing and rounding in vowels: [+back, +round] vowels tend to contrast only with [−back, −round] vowels, with no further combinations of these features present in the inventory .

The principle of Phonetic Enhancement (PE; Stevens, Keyser, and Kawasaki 1986) has been proposed as an explanation of this co-variance (redundancy) of distinctive features, finding the basis in acoustic similarity: a feature that is not used contrastively can be recruited to enhance the effect of some other feature with similar acoustics. Since rounding and backing both lengthen the front cavity, these features have similar acoustics by way of a lower F2. Therefore, rounding can be used to enhance a contrast in backing.

Within this account, it has generally been assumed that PE occurs to satisfy a desideratum on the communicative process: attaining greater perceptual distance between phonemes, as in the theory of *Adaptive Dispersion*, (Lindblom 1986, Flemming 1995, 2004, de Boer 2001). The theory of Adaptive Dispersion incorporating PE, has been variously implemented in Optimality Theoretic phonology. Specific implementations differ in the resources they exploit in the determination of perceptual distance. In the approach of Flemming (1995, 2004) and Boersma (1998), markedness hierarchies are

* We would like to thank audiences at Johns Hopkins University and NELS-37 at Champaign-Urbana for their insightful feedback on this research.

[1] This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

Emily J. Elfner and Martin Walkow (eds.): NELS 37, 253-266.
GLSA Amherst.

implicated, while in Steriade's (1995, 2001) P(erceptual)-Map approach it is faithfulness hierarchies that are used. In the theory of Wilson (2000, 2001), Targeted Constraints both penalize short distance in the way of Markedness and select the repair in the way of Faithfulness. Despite the differences, these accounts share the goal of characterizing not only static inventories, but simultaneously also the patterns of segmental neutralizations. Taking certain specific phonological contexts like unstressed positions and syllable codas to provide weaker perceptual cues and thus to effectively shrink the perceptual space, the neutralizations associated with these positions (vowel inventory reductions, coda devoicing) are seen to derive from the same demands at work with the inventory at large: ensuring sufficient perceptual distance among possible segments.

These important results notwithstanding, the theory of Adaptive Dispersion suffers from a notable flaw, as first pointed out in Ohala (1980). In itself, distance would be enhanced even by a clustering of features that are acoustically orthogonal, and not just by those that are acoustically similar. For instance a 'diagonal' inventory consisting of {i, □} will satisfy greater distance than a 'vertical' one consisting of {□, a}. Yet, unlike the latter, the former is not attested. This has required supplementing Dispersion or Phonetic Enhancement with counterbalancing principles that require maximal utilization of distinctive features, e.g. Lindblom's (1986) notion that perceptual contrast is favored only until it becomes 'sufficient', or Clements' (2003) principle of 'Feature Economy' by which features must be utilized to produce the maximal number of contrasts.

In this work we present an approach that does not require such a delicate balancing act, and avoids the stipulation of counterbalancing principles entirely. We argue that both distance and Phonetic Enhancement follow from a certain type of harmony maximization, and that, unlike a direct quest for distance, this maximization does not predict clustering of acoustically orthogonal features. Our approach is based on Burzio's (2002a, b, 2005) entailment framework, which combines properties of both Optimality Theory (Prince and Smolensky 2004) with certain ideas from neural net computation. This approach is able to capture distance-based interactions among mental representations independently of the specific role of perception and has been advanced by Burzio as an account of phenomena beyond that realm, in Phonology proper as well as Morphology. If the present claims are correct, this approach will then prove superior to previous attempts, both narrowly, as an account of PE, and broadly, as a unified account of previously unrelated phenomena.

2. Methods for Entailment Networks

In this section, we review the main tenets of the entailment framework and show how a set of entailments can be directly implemented in a Hopfield network (Hopfield 1982). This formalization will allow us both to test the ability of the entailment framework to predict clustering of components, and to conduct a specific simulation of phonetic enhancement.

2.1 Attraction under the Representational Entailments Hypothesis

The hypothesis in (1) on the nature of mental representations has been shown capable of explaining diverse phenomena in phonology and morphology, by implementing attraction-over-distance (Burzio 2002a,b, 2005; Burzio and Tantalou in press;).

(1) **Representational Entailments Hypothesis** (REH): Mental representations of linguistic expressions contain sets of entailments. E.g. a representation consisting of A, B corresponds to the entailments: A→B, B→A (if A then B; if B then A).

This hypothesis can be formalized for the general case where a representation has an arbitrary number of components as in (2), which ensures that each component entails every other component.

(2) **Def**: A representation R=(C, E) consists of
i. A set of components C={A_1, A_2, ... A_n};
ii. A set of (logical) entailments, E, such that for all $1 \leq i, j \leq n$, the entailment $A_i \rightarrow A_j$ is an element of E.

For example, a representation, **U**, of the back, round, high vowel [u] has feature-value components {[+back], [+round], [+hi]} as well as the following entailments:

(3)

[+back]→[+back]	[+round]→[+back]	[+hi]→[+back]
[+back]→[+round]	[+round]→[+round]	[+hi]→[+round]
[+back]→[+hi]	[+round]→[+hi]	[+hi]→[+hi]

An immediate consequence of the REH is that the entailments for distinct representations may agree or disagree. The following notion of entailment violation formalizes such "disagreement":

(4) **Def**: An entailment of **R**, $A_i \rightarrow A_j$ is violated by **R′** if A_i is a component of **R′**, but A_j is not a component of **R′**. An entailment is satisfied if it is not violated.

For example, the entailment [+hi]→[+back] is violated by a representation with components [+hi] and [−back].

With this notion of entailment violation in place, it is possible to view the entailment framework in OT-theoretic terms, where each entailment is a violable constraint. The grammatical system can then be seen as an optimization process that seeks to *minimize entailment violation.* By adopting an optimization axiom, the entailment framework can express the pressure for similar representations to neutralize or, as we shall say, engage in *attraction.* Consider representations U of [u] = {[+back], [+round], [+hi]} with the set of entailments listed in (3) and Y of [y] = {[−back],

[+round], [+hi]}.[2] **Y** differs from **U** in one feature, backness, and this gives rise to two entailment violations: [+hi]→[+back] and [+round]→[+back]. Such entailment violations, we claim, is the source of an attractive pressure between these representations, leading to neutralization.

To account for the results of Adaptive Dispersion theory, our framework should predict that the attractive pressure should be sensitive to multidimensional distance, falling off as representations become more distinct (less similar). Continuing the example just considered, if we add representation **I** of [i] = {[−back], [−round], [+hi]}, which is more distant from **U**, having two less common components, our measure of entailment violation should yield the result that **U** exerts a stronger force on **Y** than **U** exerts on **I**. A simple count of entailment violations does not however yield the desired result: just as was the case for **Y**, **I** also violates two of the entailments of **U**: [+hi]→ [+back] and [+hi]→[+round]. Observe though that if the [−back] component of **Y** were altered to [+back] two fewer entailments of **U** would be violated, as the representations would be rendered identical, whereas if the [−back] component of **I** were altered to [+back], the resulting representation, namely {[+back], [−round], [+hi]} continues to violate two **U**-entailments: [+hi]→[+round] and [+back]→[+round]). Therefore, we propose to define the strength of attraction as in (5) below, which correctly captures the role of distance: attraction becomes stronger as representations become more similar.

(5) The strength of the attraction (neutralization) force between two representations is proportional to the maximal change in entailment violation that occurs by altering any single differing component.

2.2 Entailment Networks

While a novelty from the point of view of generative linguistics, the conception of representation in (2) above resonates with concepts long familiar in other areas of cognitive science. In particular, (2) is a virtual restatement of 'Hebbian learning' (Hebb 1949), popularly paraphrased as 'neurons that fire together, wire together.' The 'firing together' of neurons is analogous to the co-occurrence of two components such as **A** and **B** within the same representation, while the 'wiring together', which refers to the transmission of activation via synaptic connections, is analogous to the entailments (if neuron **A** is active, neuron **B** will also be active, given a connection between them). The Representational Entailments Hypothesis is thus essentially that the brain applies Hebbian learning to linguistic experience, just as neuroscientists believe it does in general.[3] The entailment framework requires that each component entails every other component, so for the domain of artificial neural networks, the Representational Entailment Hypothesis corresponds to the type of computation carried out by a Hopfield net (Hopfield 1982), in which each unit is connected to all other units.

---

[2] We refrain from listing the entailments of **Y**, as they are uniquely recoverable from the components.

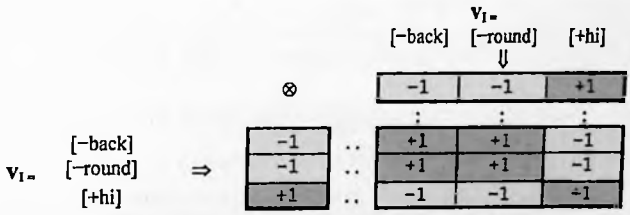
[3] We are not, of course, claiming that the components of linguistic representations are represented by the firing or connectivity of individual neurons, but rather that the computational principles involved in the linguistic system are the same as those seen at this lower level.

In this and the subsequent sections, we review results demonstrated in (Wayment, Burzio, Mathis & Frank, in preparation) which show that the relationship between networks and entailments is stronger than analogy. This section shows how Hebbian learning over a localist, binary encoding instantiates the entailments predicted by the REH in the connections between units of a Hopfield network. We will label as *Entailment Networks* those Hopfield networks that satisfy the conditions required (see (8) and (15) below) to preserve entailment structure through Hebbian learning.

Previous entailment accounts (Burzio 2002a,b) used binary phonological and semantic features as the primitive kind of component. However, the primitive elements of connectionist networks are the activation values of individual units. In order to bridge between these two conceptions, we must therefore show how to define entailments over numerical vectors, and then link the construction of such entailments to Hebbian learning.

Suppose each component of a representation **R** is encoded as a single unit which has exactly two possible activation states: +1 or −1, corresponding, respectively, to the + or − feature values of the component. A representation **R** with *n* components is thus encoded as a specific *n*-dimensional vector, $\mathbf{v_R}$, where all elements of $\mathbf{v_R}$ are +1 or −1. The set of entailments for **R** can be recorded in an *entailment matrix*, **M**, constructed by taking the tensor product of $\mathbf{v_R}$ with itself (i.e., each element $\mathbf{M}_{ij} = \mathbf{v}_{R(i)} \cdot \mathbf{v}_{R(j)}$, where $\mathbf{v}_{R(i)}$ is the i^{th} component of $\mathbf{v_R}$). For example, the representation I of [i] = {[−back], [−round], [+hi]} is encoded as $\mathbf{v_I}$ = [−1 −1 +1]; the first dimension of $\mathbf{v_I}$ encodes backness, the second rounding, and so on. The entailment matrix for I is given by $\mathbf{v_I} \otimes \mathbf{v_I}$ below in (6).

(6) $\mathbf{M} = \mathbf{v_I} \otimes \mathbf{v_I} =$



By interpreting one index of **M** as the antecedent of an entailment and the other index as the consequent, the entailment matrix is directly relatable to the REH. Essentially, **M** encodes all entailments of **I**: +1 entries in the matrix correspond to entailments that assert identical values for feature components; $\mathbf{M}_{ij}$ = +1, asserts [αf_i]→[αf_j]. −1 entries in the matrix correspond to entailments that assert different values for feature components; $\mathbf{M}_{ij}$ = −1, asserts [αf_i]→[-αf_j]. Therefore, the entry in the entailment matrix from rounding to backing $\mathbf{M}_{2,1}$ = +1 rightly asserts [αround]→[αback], preserving the entailment [−round]→[−back]. Likewise, all other entailments prescribed by (2) are preserved. Thus, the entailments among a localized, binary encoding of the components of a representation can be recorded in an entailment matrix constructed via a tensor product.

Linking entailment matrices to Hebbian learning in a Hopfield network becomes a trivial matter because they both involve tensor products. With units that have +1 or −1 activation values, the Hebbian learning rule (Hebb, 1949; see also Smolensky & Legendre 2006, Ch 9) for changing the connections between units is:

(7) $\Delta \mathbf{W} = \eta \cdot \mathbf{i} \otimes \mathbf{t}$

where η is the learning rate, $\mathbf{i}$ is the input vector and $\mathbf{t}$ is the target vector, the desired output. In some Hopfield networks, input and target vectors are presumed to be identical, since all units are connected, so the Hebbian learning rule for Hopfield networks is $\Delta \mathbf{W} = \eta \cdot \mathbf{i} \otimes \mathbf{i}$. Trivially, constructing an entailment matrix is thus a single instance of Hebbian learning in such a network with η=1. Because entailment matrices preserve the entailment structure of a representation, Hebbian learning in a Hopfield network also preserves the entailment structure of a representation in the connections between the units. Thus, the following linking hypothesis is justified:

(8) **Linking entailments with Hebbian learning**: Let the components of a representation **R** be encoded as a localist, binary activation vector, $\mathbf{v_R}$. Then, the set of entailments for **R** predicted by the REH can be instantiated in a Hopfield network with the Hebbian learning rule $\Delta \mathbf{W} = \eta \cdot \mathbf{v_R} \otimes \mathbf{v_R}$.

2.3 Attraction as Harmony in Entailment Networks

Not only do Entailment Networks encode the entailments of representations, but they also exhibit attraction properties similar to entailment violation under the well-known network measure of *Harmony* (Smolensky 1986, Smolensky & Legendre 2006). Just as entailment violation measures how much one representation agrees with another representation's entailments, Harmony measures the degree to which a pattern of activation "agrees" with the weights of a network. The standard quadratic Harmony function is given below:

(9) **Def**: The harmony of pattern **A** given weight matrix **W**:

$$\mathcal{H}_{\mathbf{W}}(\mathbf{A}) = \mathbf{A} \cdot \mathbf{W} \cdot \mathbf{A}^{\mathrm{T}} = \sum_{i,j} a_i w_{ij} a_j .$$

For a given **W**, different patterns of activation may have different harmony values. By smoothly varying the change in patterns over the space of possible activation values, one can draw a *harmony landscape* (see **Figure 1** below).

Previously, we adopted an optimization axiom into the REH framework, similar to the one embodied in OT, that postulates that the system of mental representations for linguistic expressions is organized so as to minimize entailment violation. It is not obvious in what way Entailment Networks might minimize entailment violation. However, a variety of networks are known to be sensitive to Harmony: given the right update equations, during the processing of an input pattern, the pattern of activation will drift from lower harmony to higher harmony activation states (see Smolensky &

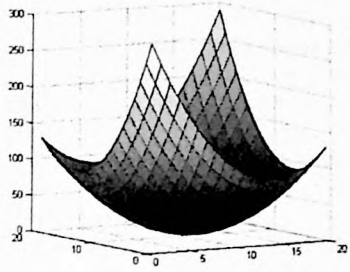


Figure 1. *Harmony landscape.* The peaks in this landscape are the attractors of this network.

Legendre (2006), Ch 9 for a thorough discussion of architectures and activation functions that result in a harmony maximizing network). Because the weight matrix generated by (8) is symmetric, Entailment Networks have symmetric connections, therefore they are harmony maximizing on repeated incremental updates. The harmony maxima in a landscape like those in **Figure 1** therefore constitute the attractors of the Entailment network, the states into which such a network will settle as activation is allowed to propagate, since at these points, small deviations in the patterns of activation only lead to lower harmony. Therefore, by linking entailment violation with harmony, as stated in (10) below, we can be confident that Entailment Networks are indeed sensitive to the attraction properties predicted by the REH.

(10) **Linking entailment violation with harmony:** The harmony of a pattern **R′** given the weight matrix $\mathbf{W_R}$ defined by (8) can be completely described in terms of entailment violation as follows:

$$\begin{aligned} \mathcal{H}_{\mathbf{W_R}}(\mathbf{R}') &= (\#\text{ of satisfied entailments}) - (\#\text{ of entailment violations}) \\ &= (\text{total }\#\text{ of entailments}) - 2\times(\#\text{ of entailment violations}) \end{aligned}$$

While space prevents us from providing an explicit proof of (10), we will just note that each term of the summation in (9) corresponds to determining whether or not an entailment is satisfied. A pattern's harmony is increased by 1 for every entailment satisfied and decreased by 1 for every entailment violated. (10) ensures that an Entailment Network which maximizes harmony, also optimizes with respect to minimizing entailment violation. Further, we can now see why the force of attraction as defined in (5) is determined by changes in entailment violations resulting from small modifications to a representation: the computation of Hopfield networks proceeds by following the derivative of the Harmony function. As a result, if a small change in activation results in a large increase in harmony, there is a stronger tendency for a pattern to move toward the attractor.

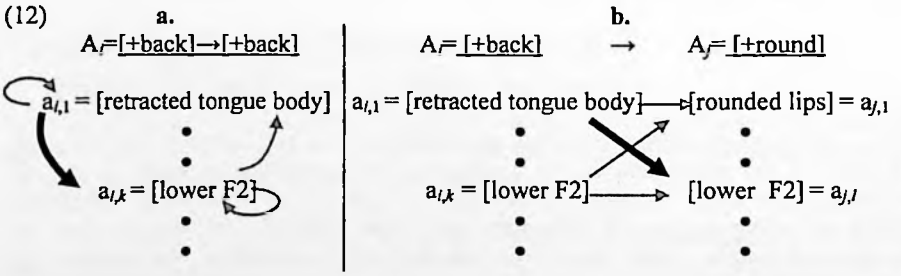
2.4 Adding Sub-components for Similarity and a Binding Corollary

To this point, distinctive features have been treated as the sole constituents of the representations of phonemes. We have illustrated attraction effects in terms of phonemes that are more or less similar in regard to their distinctive features. However, phonetic enhancement makes clustering predictions based on the similarity of the acoustic correlates of the distinctive features, not the featural similarity of phonemes. Therefore, in order to model PE, we must augment the entailment framework in order to distinguish more and less similar distinctive features. Our solution is to assume a recursive step, under which each component is also a representation (as defined in (2)) with its own set of (*sub*-)components. In the case of the distinctive features that make up a phoneme, these sub-components represent the acoustic (and possibly articulatory) correlates of the distinctive features. As components now have sub-components, the entailment between components must be defined as the collection of entailments among the sub-components:

(11) **Extended REH**: A representation $\mathbf{R}=(C,E)$ consists of
- *i.* A set of components, $C=\{A_1, A_2, \ldots A_n\}$;
- *ii.* A set of entailments, E, where for all $1\leq i, j\leq n$, the entailment $A_i \rightarrow A_j$ is an element of E.
- *iii.* Each component, A_i, is a set of m_i sub-components, $A_i=\{a_{i,1}, a_{i,2} \ldots a_{i,m_i}\}$.
- *iv.* Each entailment of E, $A_i \rightarrow A_j$, is a set of sub-entailments, $\{a_{i,k} \rightarrow a_{j,l} \mid \forall k,l$ such that $1\leq k\leq m_i, 1\leq l\leq m_j\}$.

To explore the effects of similarity under (11), consider expanding two entailments of $\mathbf{U}=\{$[+back],[+round], [+hi]$\}$: [+back]→[+back] (in 12a) and [+back]→[+round] (in 12b), where [+back] and [+round] have an identical sub-component, [lower F2]:

(12)



(12a) illustrates that each component gives rise to self-entailments, which under (11*iv*) resemble the sort seen in (3), where 'every (sub-)component entails every other.' Additionally (11*ii,iv*) provide cross-component entailments, as in (12b). Crucially [+back] is similar to [+round] because they have a sub-component [lower F2] in common; the more identical sub-components, the greater the similarity. In the case of similarity, some cross-component entailments will be identical to some self-entailments.

In (12) $a_{i,1} \rightarrow a_{i,k} = a_{i,1} \rightarrow a_{j,l}$, since both $a_{i,k}$ and $a_{j,l}$ are the sub-component [lower F2]. The dark arrows in (12) highlight this equality.

In this approach, entailment strength is additive: two identical entailments are interpreted as a single entailment which is twice as strong. Therefore, the more identical sub-components shared by two components, the stronger the entailment between them. Hence, the entailment framework predicts that the degree to which components will tend to form a cluster in a system of representations depends only on their similarity. Thus, as we will note below, the entailment framework has the potential to derive constituency at different levels of representation; that is, why features bind together to form phonemes and why phonemes bind together to form morphemes.

(13) **The binding corollary of the REH** (see Burzio 2005, 77-81):
For an entailment, $A_i \rightarrow A_j$, the greater the similarity of A_i and A_j, the stronger the entailment between them. Such stronger entailment predicts A_i and A_j tend to bind together, acting as a single unit, i.e. a *constituent.*

We claim that phonetic enhancement is an instance of the binding corollary, specifically:

(14) **PE-Binding hypothesis**: Distinctive features with similar acoustic correlates constitute similar representations, so they entail one another more strongly and tend to bind together and act like a single unit.

If verified, the PE-Binding hypothesis would explain why phoneme inventories tend to self-organize into systems that have a single-contrast across mutually enhanced features. Confirming the PE-hypothesis would also validate the Binding Corollary, which would then make a general clustering prediction subsuming PE. Therefore, the remainder of this paper seeks to validate this claim through a neural network simulation which directly implements the REH.

2.5 REH-Training for Binding in Entailment Networks

As a system of attractors, the REH and the binding corollary should be amenable to a Hopfield network simulation. However, the Hebbian learning training procedure given above (8) was defined on the assumption that the relevant components of representations could be fully described in terms of binary feature values. The previous section extended the REH for feature values that could be more or less similar. This necessitates a change to the network model, which will allow entailment networks to represent the entailments among sub-components.

To make the discussion concrete, we first present our scheme for encoding distinctive features as vectors which correspond to distributed patterns of activation in a network. In the simulations below, we will contrast similar features [back] and [round] with orthogonal features [back] and [hi]. Confining ourselves to acoustic similarity, the sub-components of a distinctive feature serve as an encoding of their acoustic correlates.

Thus, the feature values are encoded in *feature value vectors*, which are distributed patterns of activation corresponding to their acoustic correlates. The acoustic correlates of the features [back], [round], and [high] for an idealized male speaker (Stevens 2000) are as follows. Rounded vowels and unrounded vowels differ by about 300 HZ on F2. Front vowels and back vowels differ by about 1000 HZ on F2. For the sake of experimental clarity, we assume there is no height-F2 interaction in this idealized vowel space. Thus, we only examine high and mid vowels, which differ by about 200 HZ on F1.

Table 1. *Idealized vowel formants.* Approximate spectral peaks of the first and second formant for an idealized male speaker, ignoring height-F2 interactions.

Vowel	[hi]	[back]	[round]	F1	F2
i	+	–	–	300	2150
y	+	–	+	300	1850
ɶ	+	+	–	300	1150
u	+	+	+	300	850
e	–	–	–	500	2150
ø	–	–	+	500	1850
ɸ	–	+	–	500	1150
o	–	+	+	500	850

Table 2. *Thermometer Encoding.* The 7-dimensional feature value vectors used in Experiments 1 and 2.

features	ΔF1	ΔF2	F1 Therm.		F2 Therm.				
[-back]	0	500	0	0	1	1	1	1	1
[+back]	0	-500	0	0	-1	-1	-1	-1	-1
[−round]	0	150	0	0	1	1	1	-1	-1
[+round]	0	-150	0	0	-1	-1	-1	1	1
[−hi]	100	0	1	1	0	0	0	0	0
[+hi]	-100	0	-1	-1	0	0	0	0	0

A 'thermometer' encoding scheme (**Table 2**) was used to encode the (detailed) acoustics of the features [hi], [back], and [round] from **Table 1**. Each formant is associated with a number of subcomponents. The activation of a single therm-unit corresponds to a raising (1) or lowering (−1) of the respective formant by 100 Hz from a baseline formant frequency of (F1=40, F2=1500). The total amount of activation distributed over the therm-units is thus related to the amount of deviation from that midpoint

If the components of a representation are vectors $A_i=\{a_{i,1}, a_{i,2} \ldots a_{i,m_i}\}$ then (by 8) each entailment $A_i \rightarrow A_j$ of the Extended-REH(11) is given by the matrix formed by taking the tensor product between them: $A_i \otimes A_j$. Thus, $\mathbf{M}_{[+back]\rightarrow[+round]}$=[+back]⊗[+round] = [0 0 1 1 1 1 1]⊗[0 0 -1 -1 -1 1 1]. By postulating that entailment strength is additive, each entailment is viewed as a step in a special training procedure which we call *REH-Training*. REH-training designates that a network is trained on the sub-entailments for each possible pair of components in a representation. For instance, the steps of the training program for a representation {[+back],[+round]} are

(15) REH-Training for **R**={[+back], [+round]}:

1. [+back]→ [+back]: $\Delta\mathbf{W}=\eta$[+back]⊗ [+back]
2. [+back]→[+round]: $\Delta\mathbf{W}=\eta$[+back]⊗ [+round]
3. [+round]→[+back]: $\Delta\mathbf{W}=\eta$[+round]⊗[+back]
4. [+round]→ [+round]: $\Delta\mathbf{W}=\eta$[+round]⊗ [+round]

The final weight matrix that results from this training (with a uniform learning rate, η = 1) for a representation with constituents $\{A_1, A_2, \ldots, A_n\}$ is derived in (Wayment, Burzio, Mathis & Frank, in preparation), but the results are reported here:

(15) $$\mathbf{W}_{final} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_i \otimes A_j$$

By construction of this final matrix (15), we have shown how to instantiate a set of entailments in the connections of a network for representations rich enough to test the PE-Binding hypothesis.

The results reported in this section demonstrate that Entailment Networks are correctly viewed as a *direct* implementation of the Representation Entailments Hypothesis. Thus, the simulation results reported in this paper bear directly on testing the predictions of the REH on the system of linguistic representations. By virtue of its applicability to mental representations in general rather than just to perception, the REH accounts for the role of multidimensional distance beyond perception. At the same time, it will also cover the phenomena reviewed above that invoke 'perceptual' distance.

3. Phonetic Enhancement Experiments

In order to detect instances of phonetic enhancement, and to develop a specific experimental linking hypothesis for the binding corollary, we introduce a standard measure of similarity in connectionist representations:

(16) **Def**: The similarity of two patterns of activation **A** and **B** is related to their 'dot' or inner product $\mathbf{A}\cdot\mathbf{B} = \|\mathbf{A}\|\cdot\|\mathbf{B}\| \cos \theta_{\mathbf{A},\mathbf{B}}$.

Thus, the similarity of the encodings of the distinctive features from **Table 2** is

(17)

[+back] ·[+round] = 1	[+back] ·[+hi] = 0	[+round]·[+hi] = 0
[+back] ·[−round] = -1	[+back] ·[−hi] = 0	[+round]·[−hi] = 0
[−back] ·[+round] = -1	[−back] ·[+hi] = 0	[−round]·[+hi] = 0
[−back] ·[−round] = 1	[−back] ·[−hi] = 0	[−round]·[−hi] = 0

A precise measure of similarity is necessary for our experimental linking hypothesis:

(18) **Linking hypothesis for the Binding corollary**: Under uniform training conditions using the REH-training procedure (15), feature combinations which are more similar (as defined in 16) will form stronger attractors (be more harmonic cf. (10)) than feature combinations which are less similar.

'Uniform training' is an important control because a Hopfield network which receives equal exposure to all possible feature combinations will therefore not privilege a combination based on frequency or some other artifact of training. Any differences must

arise from how the system treats different combinations, not because of some bias in the input training data (cf. Richness of the Base, Prince & Smolensky 2004).

3.2 Experiment 1: Mutual Enhancement of Backing and Rounding

Based on the similarities reported in (17), the hypothesis in (18) predicts that when a Hopfield network is trained with each of the four possible combinations of [back] and [round] using the REH-training procedure for each combination, the patterns of activation corresponding to the combinations [+back, +round] and [−back, −round] should be stronger attractors than either of [+back, −round] or [−back, +round]. **Figure 2** shows the resulting harmony landscape after performing REH-training on each possible combination of backing and rounding, given the encoding in **Table 2**. Harmony was measured over all possible combinations of backing and rounding. Clearly, the phonetically enhanced pairs are more harmonic than the non-enhanced alternatives.

Although [αback, −αround] combinations are local maxima, the network prefers the [αback, αround] combinations in at least the following ways: they have higher harmony, the slope of the harmony surface is steeper, and they have a larger basin of attraction. We claim that this dichotomy in harmony behavior represents a preference in the network, a binding of [αback] and [αround].

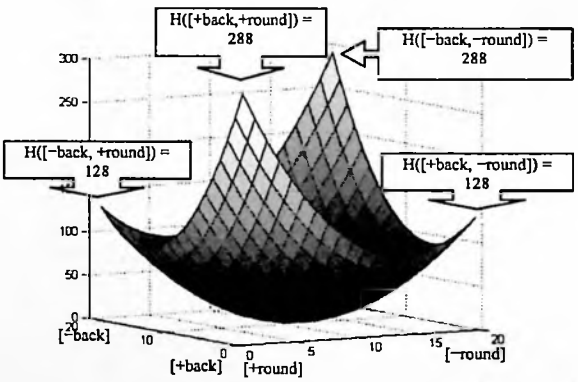


Figure 2. *Rounding and Backing harmony landscape*

3.3 Experiment 2: Avoiding the Diagonal of Height and Backing

When the steps just described are taken with the features [back] and [hi], the hypothesis in (18) tells us that no combination should be more harmonic than any other since all combinations have equal internal similarity: zero. Our simulations confirm this result. **Figure 3** contains the resulting harmony landscape after performing REH-training on [back] and [hi]. Failure to show a preference for binding is explained as follows under the REH: since F1 and F2 components of the thermometer encoding scheme were chosen to be orthogonal, there is no shared internal structure between [back] and [hi], therefore

there is no binding: the patterns of activation corresponding to the combinations of feature value vectors are thus equally preferred.

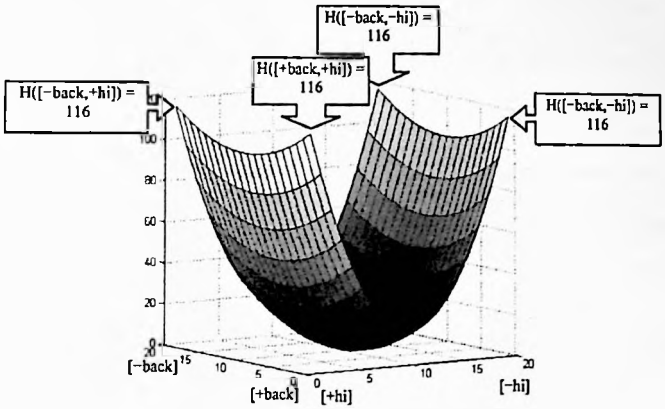


Figure 3. *Height and Backing harmony landscape.*

4. Conclusion

Experiment 1 has shown that the REH (Burzio 2002a, b, 2005) predicts that when a similarity relation holds between two components of a representation, these will tend to cluster across representations, as if bound together by stronger entailments. When applied to PE, where acoustically similar distinctive features cluster, our account is superior to the traditional one. If left unchecked, the latter would predict clustering even in the case of acoustically orthogonal features, since that would enhance overall acoustic distance. As **Experiment 2** shows, our model exhibits a preference for one diagonal over another only when the dimensions are not orthogonal. When they are orthogonal, uniform training—a conceptual analog to 'Richness of the Base' in OT, will provide no basis for choosing diagonals. Thus, privileging a diagonal could only come from some super-ordinate plan that would disregard uniform training. As diagonal inventories are unattested, evidently the brain has no such plan, obeying principles that are more strictly computational (maximal harmony) than 'functional' (maximal perceptual distance).

Furthermore, our account is also more general, predicting comparable similarity-based effects outside of acoustics/perception. As noted, the REH was in fact first introduced to deal with otherwise puzzling patterns of allomorphy, such as 'Non-Derived Environment Blocking' and 'Lexical Conservatism' (Burzio 2002a), where overall similarity plays a role in neutralizing allomorphs. Morphological syncretism reveals comparable effects (Burzio 2005; Burzio and Tantalou in press). Because the REH maintains that representations at *all levels* of linguistic expression contain sets of entailments, the binding corollary is expected to apply at all levels as well, so among the predictions of our approach is of course that binding of components under similarity should occur beyond the case of distinctive features. A preliminary review yields

promising results in this area as well. Consider that languages are routinely characterized as having inventories not only of phonemes, but also of diphthongs as well as of C clusters. Heterogeneous CV or VC diphones are not usually reported, consistent with the expected role of similarity, Cs and Vs being maximally dissimilar. These preliminary indications suggest that the entailment framework may provide a general means to understanding compositionality in phonetics, phonology, and morpho-phonology.

References

Boersma, P. 1998. Functional Phonology. Holland Academic Graphics, The Hague.

Burzio, L. 2002a. Surface-to-Surface Morphology: when your Representations turn into Constraints. In Many Morphologies, ed. P. Boucher, 142-177. Cascadilla Press.

Burzio, L. 2002b. Missing Players: Phonology and the Past-tense Debate. Lingua 112:157-199.

Burzio, L. 2005. Sources of Paradigm Uniformity. In Paradigms in Phonological Theory, eds L. Downing, T. Hall, and R. Raffelsiefen. 65-106. Oxford: Oxford University Press.

Burzio, L. and N. Tantalou. In press. Modern Greek Accent and Faithfulness Constraints in OT. Lingua.

Clements, G.N. 2003. Feature economy in sound systems. Phonology 20.3:287-333.

de Boer, B. 2001. The Origins of Vowel Systems. Oxford UP.

Flemming, E. 1995. Auditory Representations in Phonology Ph.D. Dissertation, UCLA.

Flemming, E. 2004. Contrast and Perceptual Distinctiveness. In Phonetically Based Phonology eds B. Hayes, R. Kirchner and D. Steriade, 232-276. Cambridge University Press.

Halle, M. and R. Jakobson. 1956. Fundamentals of Language. Mouton, The Hague.

Hopfield, J.J. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 79, 2554-2558.

Lindblom, B. 1986. Phonetic Universals in Vowel Systems. In Experimental Phonology, eds. J. J. Ohala and J. J. Jaeger, New York: Academic.

Ohala, J. 1980. Moderator's introduction to the symposium on phonetic universals in phonological systems and their explanation. Proceedings of the Ninth International Congress of PhoneticSciences 3:181-185. 1975 Institute of Phonetics. University of Copenhagen.

Smolensky, P. 1986. Information processing in dynamical systems: Foundations of harmony theory. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations, eds. D. E. Rumelhart, J. L. McClelland & the PDP Research Group, 192-281. Cambridge. MIT Press/Bradford Books.

Smolensky, P. and G. Legendre. 2006. The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar Vol. 1: Cognitive Architecture: vol. 2: Linguistic and Philosophical Implications. MIT Press.

Steriade, D. 1999. Lexical Conservatism in French Adjectival Liaison. In Formal Perspectives in Romance Linguistics, eds. B. Bullock, M. Authier and L. Reed , 243-270. John Benjamins: Amsterdam.

Steriade, D. 2001. The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Ms. UCLA.

Stevens, K. 2000. Acoustic Phonetics. MIT Press.

Stevens, K., S. Keyser, and H. Kawasaki. 1986. Towards a phonetic and phonological theory of redundant features. In Invariance and Variability in Speech Processes, eds. J. Perkell and D. Klatt. Lawrence Erlbaum, Hillsdale.

Wayment, A., L. Burzio, D. Mathis and R. Frank. In preparation, Ms draft 2006. Phonetic Enhancement as Harmony Maximization.

Wilson, C. 2000. Targeted Constraints: An Approach to Contextual Neutralization in Optimality Theory. Ph.D. Dissertation, Johns Hopkins University.

Wilson, C. 2001. Consonant Cluster Neutralization and Targeted Constraints. Phonology 18.1:147- 197.

Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218

wayment@cogsci.jhu.edu