Adam Wayment, Luigi Burzio, Donald Mathis, Robert Frank — Johns Hopkins U.

Neural net meets phonetic enhancement — LSA Meeting. January 8, 2006

Burzio (2002a, b), (2005): **Attraction**, crucial property of representations. Attract one-another in ways that depend on distance/ similarity. Attraction is pervasively present:

Perception/ phonetics:	Categorical perception; ‘perceptual magnet’ effect
Phonology:	At work in ‘Dispersion’ of segmental inventories; segmental neutralizations; Non-Derived Environment Blocking; Lexical Conservatism: OT’s IO- FAITH, OO-FAITH = Attraction; segmental assimilations = Attraction
Morphology:	Syncretism = Attraction

Representational Entailments Hypothesis (REH).
E.g.. An entailment matrix for a representation R = A, B, C

Same architecture as a Hopfield net (entailments = connectivity)

	A [	B [	C [
A Ψ	1	1	1
B Ψ	1	1	1
C Ψ	1	1	1

REH yields attraction:

				R1's entailments violated by -C
R1:	A	B	C	
R2:	A	B	-C	A Ψ C; B Ψ C
R2':	A	-B	-C	A Ψ C; ~~B Ψ C~~

REH raises Q:

R1:	A	B	C	Subconstituent: A, B
R2:	A	B	-C	Mutual entailment A]B stronger

Are all entailments created equal?
Issue of predicting sub-constituents

R1:	A	B	C	Subconstituent: A, C
R3:	A	-B	C	Mutual entailment A]C stronger

Traditional ontology:	**Clustering subconstituents:**
There are: phonemes	Features
???no n-phones??	Phonemes
There are: morphemes	Phonemes & Meaning Types of morphemes: Concat.; Non-concat; Reduplicative.

No attempt in the tradition to identify the clustering principles, with one exception:

Phonetic Enhancement: Features cluster when their acoustic properties are similar or concurring, and thus mutually enhancing. (Stevens, Keyser and Kawasaki 1986; Flemming 1995)

Burzio (2005, 77-81) argues REH contains a **general solution** to clustering problem that would subsume Phonetic Enhancement: **Similarity** results in both: i) **Attraction** of representations, and ii) **Binding of subconstituents** within representations.

Strength of entailment A] B depends on similarity A/ B. If so, the binding effect should be amenable to **simulation in a Hopfield net**.

Phonetic Enhancement in a Hopfield Network. General Methods

Representations

- All representations correspond to distributed patterns of unit activation.
- A distinctive feature value (e.g. [+round]) is a specific activation vector over the units of the entire network (representing acoustic/ articulatory properties).
- Phonemes are combinations of feature value vectors.
    - Feature value vectors are combined using vector addition.

(1) If phoneme **P** consists of distinctive feature values $\mathbf{f}_1$ and $\mathbf{f}_2$, then $\mathbf{P} = \mathbf{f}_1 + \mathbf{f}_2$.

Applying the REH to phonemes. A phoneme **P** consisting of $\mathbf{f}_1$ and $\mathbf{f}_2$ has the following set of entailments:

(2)
$$\begin{array}{ll} \mathbf{f}_1 \rightarrow \mathbf{f}_1 & \mathbf{f}_2 \rightarrow \mathbf{f}_1 \\ \mathbf{f}_1 \rightarrow \mathbf{f}_2 & \mathbf{f}_2 \rightarrow \mathbf{f}_2 \end{array}$$

Training Entailments. Let **A** and **B** be patterns of activation. The weight matrix, **W**, corresponding to the entailment **A** |**B** is given by the **outer product** (aka **tensor product**) of **A** and **B**:

(3)
$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_1 \cdot b_1 & a_1 \cdot b_2 & \dots & a_1 \cdot b_n \\ a_2 \cdot b_1 & a_2 \cdot b_2 & \dots & a_2 \cdot b_n \\ \dots & \dots & \dots & \dots \\ a_n \cdot b_1 & a_n \cdot b_2 & \dots & a_n \cdot b_n \end{bmatrix}$$

Training any patterns **A** and **B** in a fully-connected Hopfield network corresponds to $\Delta\mathbf{W}=\eta\cdot\mathbf{A}\otimes\mathbf{B}$, where η is the learning rate.

REH training. Training the combination of $\mathbf{f}_1$ and $\mathbf{f}_2$ in (2) in a Hopfield network corresponds to the following training steps:

(4)
$$\begin{array}{ll} \mathbf{f}_1 \rightarrow \mathbf{f}_1\text{: } \Delta\mathbf{W}=\eta\cdot \mathbf{f}_1 \otimes \mathbf{f}_1 & \mathbf{f}_2 \rightarrow \mathbf{f}_1\text{: } \Delta\mathbf{W}=\eta\cdot \mathbf{f}_2 \otimes \mathbf{f}_1 \\ \mathbf{f}_1 \rightarrow \mathbf{f}_2\text{: } \Delta\mathbf{W}=\eta\cdot \mathbf{f}_1 \otimes \mathbf{f}_2 & \mathbf{f}_2 \rightarrow \mathbf{f}_2\text{: } \Delta\mathbf{W}=\eta\cdot \mathbf{f}_2 \otimes \mathbf{f}_2 \end{array}$$

When starting with initial weights all zero, the final weight matrix for training the combination of $\mathbf{f}_1$ and $\mathbf{f}_2$ is:

(5)
$$\mathbf{W}_{\mathbf{f1,f2}} = 0.25\cdot(\mathbf{f}_1 \otimes \mathbf{f}_1 + \mathbf{f}_1 \otimes \mathbf{f}_2 + \mathbf{f}_2 \otimes \mathbf{f}_1 + \mathbf{f}_2 \otimes \mathbf{f}_2)$$

Harmony and Attraction. **Harmony** (Smolensky 1986; Smolensky & Legendre 2005) is the degree to which activation patterns (representations) “agree” with the weights of a network.

- The Harmony of pattern **A** given weight matrix **W**:

(6)
$$\mathcal{H} = \mathbf{A}^{\mathrm{T}} \cdot \mathbf{W} \cdot \mathbf{A} = \sum_{i,j} a_i w_{ij} a_j$$

- Hopfield networks are **harmonizing** networks: given the right update rule, less harmonic patterns tend toward more harmonic patterns.
- **Attraction in a Hopfield network**: harmony hill-climbing; moving from lower harmony to higher harmony states.

- By measuring the harmony of different patterns of activation we can generate a **harmony landscape**, which identifies the attractors of the network.

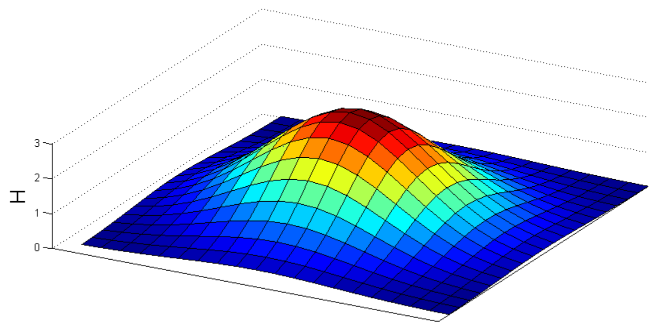

Figure 1. *Harmony landscape*. The harmony peak is the attractor in this network.

Recap:

Notions	Implementation in Hopfield Networks
Feature Value	Specific pattern of activation (vector)
Phoneme	Addition of feature value vectors
Entailments	Connections (Weights)
REH	Training Program (5).
Attraction	Harmony maximization

Phonetic Enhancement Hypothesis for Entailment-Networks: After REH training in a fully-connected Hopfield Network, pairs of feature values that are more similar will form higher harmony states than pairs of feature values that are less similar.

- Similarity of vectors **A** and **B** is definable as their **inner**(dot) **product:**

(7) $$\mathbf{A}\cdot\mathbf{B} = \sum_i a_i b_i = |\mathbf{A}|\cdot|\mathbf{B}|\cdot\cos\theta_{\mathbf{AB}}$$

Case Study: Backing and Rounding in Vowels

The features [+back] and [+round] are understood to cluster in many languages because they contribute similarly to lowering of F2 and are thus mutually enhancing (Stevens, Keyser, & Kawasaki 1986, Flemming 1995, Stevens 2000):

Enhancement:

$$\begin{array}{cc} [+\text{back}, & +\text{round}] \\ \downarrow & \downarrow \\ \text{lo F2} & \text{lo F2} \end{array} \Rightarrow \text{lowest F2} \qquad \begin{array}{cc} [-\text{back}, & -\text{round}] \\ \downarrow & \downarrow \\ \text{hi F2} & \text{hi F2} \end{array} \Rightarrow \text{highest F2}$$

No Enhancement:

$$\begin{array}{cc} [-\text{back}, & +\text{round}] \\ \downarrow & \downarrow \\ \text{hi F2} & \text{lo F2} \end{array} \Rightarrow \text{mid F2} \qquad \begin{array}{cc} [+\text{back}, & -\text{round}] \\ \downarrow & \downarrow \\ \text{lo F2} & \text{hi F2} \end{array} \Rightarrow \text{mid F2}$$

Acoustic facts for an idealized male speaker (Stevens 2000):

- Rounded vowels and unrounded vowels differ by about 300 HZ on F2.
- Front vowels and back vowels differ by about 1000 HZ on F2.
- Hi and mid vowels 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
	+	+	!	300	1150
u	+	+	+	300	850
e	!	!	!	500	2150
□	!	!	+	500	1850
o	!	+	!	500	1150
□	!	+	+	500	850

"Thermometer" encoding of analog formant values of non-low vowels:

- Measures formant deviation from the relevant midpoint: F1 = 400, F2=1500.
- The total amount of activation over the units corresponds to the magnitude of deviation; one group of units corresponds to positive change, another negative.

Features	ΔF1	ΔF2	F1 THERM				F2 THERM									
[−back]	0	500	0	0	0	0	0	0	0	0	0	1	1	1	1	1
[+back]	0	-500	0	0	0	0	1	1	1	1	1	0	0	0	0	0
[−round]	0	150	0	0	0	0	0	0	0	0	0	1	1	1	0	0
[+round]	0	-150	0	0	0	0	0	0	1	1	1	0	0	0	0	0
[−high]	100	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
[+high]	-100	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0

Table 2. *Thermometer Encoding*. The feature value vectors used in our simulations.

- **Similarity**: [−back]·[−round] = 3 and [+back]·[+round] = 3, with θ = 39.23º, whereas [−back]·[+round] = 0 and [+back]·[−round] = 0, with θ = 90º.

- **Phonetic Enhancement Hypothesis** (from above): when doing REH training on the F2 thermometer representations, we predict stronger attractors at [+back, +round] and [−back, −round].

<u>**Results** for phonetic enhancement experiment</u>. Although all patterns receive equal REH training, attractors indeed develop at the locations of phonetic enhancement.

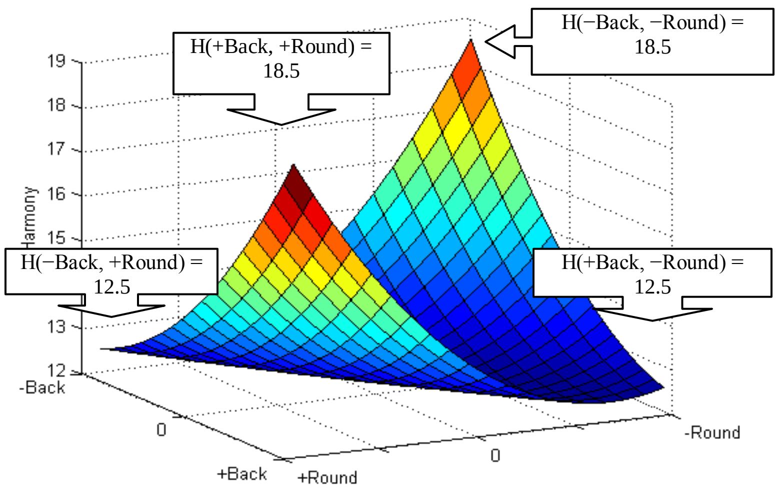

Figure 3a. *Rounding and Backing harmony landscape.*

Q.: Does phonetic enhancement result in greater distance/ dispersion (Flemming 1995) as well as greater harmony?

A.: Yes. We can project this harmony landscape back onto the relevant acoustic space. Along F2, the highest harmony states [+back, +round] and [−back, −round] are the most extreme, so indeed REH training yields **dispersion**.

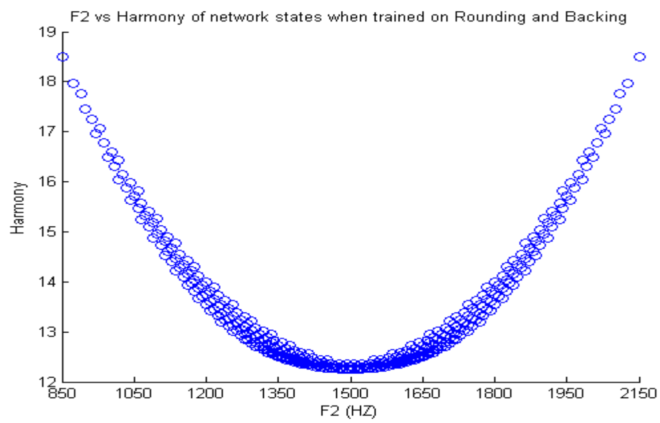

Figure 3b. *Projection of Harmony landscape **3a** onto F2 space.*

<u>**Binding**/ clustering and **similarity**</u>. Binding together of features only occurs when there is shared internal structure. If features are **orthogonal**, then there is no binding.

- When training on all combinations of [± high] and [± back], each is equally harmonic.

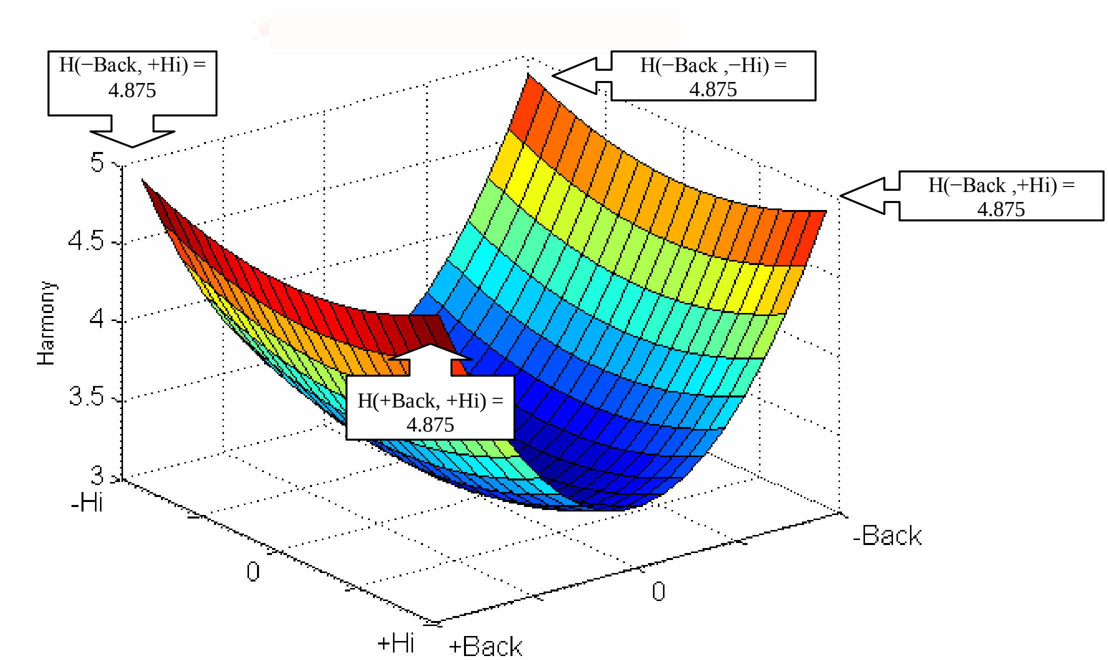

Figure 4a. *Height and Backing harmony landscape.*

- Acoustic dispersion is still at work, since extreme F1, F2 values are more harmonic than interior points.

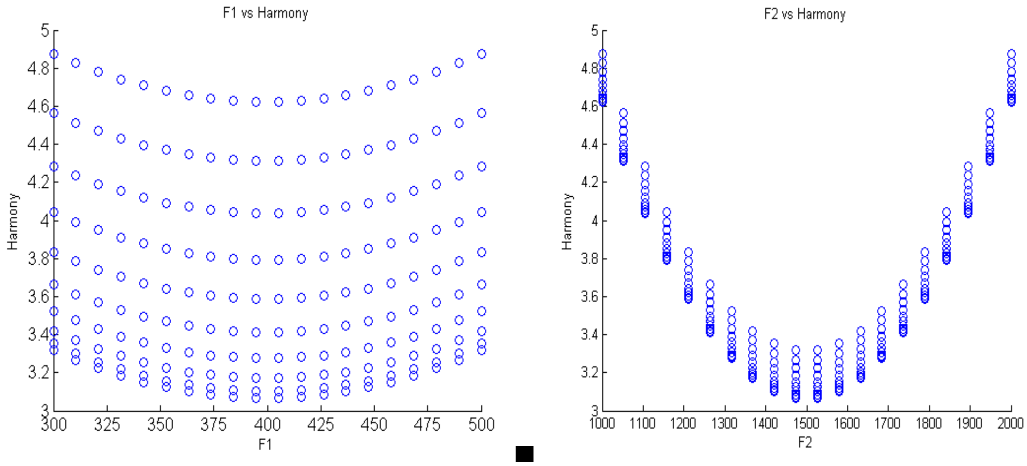

Figure 4b,c. *Projection of harmony landscape **4a** onto F1(**b**) and F2(**c**).*

<u>**Why** does REH training work?</u> The harmony of patterns **A**,**B** in a Hopfield Network whose weight matrix results from REH training of **A**,**B** is given in (8).

(8)

$$
\begin{aligned}
H(\mathbf{AB}) &= \mathbf{A}^T\cdot\mathbf{W}_{\mathbf{A\&B}}\cdot\mathbf{B} \\
&= \mathbf{A}^T\cdot(\,\mathbf{A}\otimes\mathbf{A}+\mathbf{B}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{B}+\mathbf{B}\otimes\mathbf{A}\,)\cdot\mathbf{B} \\
&= \mathbf{A}^T\cdot[\,\mathbf{A}\otimes(\mathbf{A}+\mathbf{B})+\mathbf{B}\otimes(\mathbf{A}+\mathbf{B})\,]\cdot\mathbf{B} \\
&= \mathbf{A}^T\cdot[\,(\mathbf{A}+\mathbf{B})\otimes(\mathbf{A}+\mathbf{B})\,]\cdot\mathbf{B} \\
&= [\,\mathbf{A}^T\cdot(\mathbf{A}+\mathbf{B})\,]\otimes[\,(\mathbf{A}+\mathbf{B})^T\cdot\mathbf{B}\,] \\
&= (\mathbf{A}^T\cdot\mathbf{A}+\mathbf{A}^T\cdot\mathbf{B})\cdot(\mathbf{A}^T\cdot\mathbf{B}+\mathbf{B}^T\cdot\mathbf{B}) \\
&= (|\mathbf{A}|^2+|\mathbf{A}|\cdot|\mathbf{B}|\cdot\cos\theta_{\mathbf{AB}})\cdot(|\mathbf{A}|\cdot|\mathbf{B}|\cdot\cos\theta_{\mathbf{AB}}+|\mathbf{B}|^2)
\end{aligned}
$$

Binding under similarity: Ignoring length factors, harmony is related to the cos $\theta_{\mathbf{AB}}$ which is maximal when $\theta_{\mathbf{AB}} = 0^o$, so the more similar the vectors, the higher the harmony.

Dispersion: Because the thermometer encoding measures deviation from neutral position, it rewards more extreme locations by making the vectors longer, leading to higher harmony.

……………

The thus demonstrated Binding Corollary should now be applicable across the board.
In progress: 'Generalized Similarity': combines sequential proximity and structural similarity.

Morphological typology results from binding under Generalized Similarity. Given syllable theory, segments are generally prevented from being both adjacent and from the same major class, but they bind together into morphemes under either form of similarity.

subconst.	**morpheme**	**seq. adjacency**	**structural similarity**
segment	Concatenative: *dis-en-tangle-ment*	Y	* (Syllable Theory)
segment	Root and pattern: *k t b* *u i*	* (Syllable Theory)	Y
> σ	Reduplicative: *bu-bulud*	Y	Y

When chunks get syllable-size or larger, though, both sequential adjacency and structural similarity can combine, predicting reduplication.

Conclusion: The binding together of a representation's components under similarity follows as a corollary of the Representational Entailments Hypothesis. As was shown by the Hopfield net model, when applied to features and their acoustic substance, this corollary yields the effect known as Phonetic Enhancement. When applied across the board, it sheds light on the ontology of units more generally, and on the typology of morphological systems in particular.

Authors' e-mail

Adam Wayment:	awayment@jhu.edu
Luigi Burzio:	burzio@jhu.edu
Donald Mathis:	mathis@jhu.edu
Robert Frank:	rfrank@jhu.edu

References

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

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

Burzio, L. (2005) 'Sources of Paradigm Uniformity', in Laura J. Downing, T. A. Hall, Renate Raffelsiefen, eds. *Paradigms in Phonological Theory*. Oxford: Oxford University Press: 65-106.

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

Smolensky, P., and G. Legendre, (2005). *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar Vol. 1: Cognitive Architecture; vol. 2: Linguistic and Philosophical Implications.* MIT Press.

Smolensky, P. (1986) 'Information processing in dynamical systems: Foundations of harmony theory', in D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books. 194-281.

Stevens, K. (2000) *Acoustic Phonetics.* MIT Press.

Stevens, K. N, S. J. Keyser, and H. Kawasaki (1986). 'Towards a phonetic and phonological theory of redundant features', in Joseph S. Perkell and Dennis KH. Klatt (eds) *Invariance and Variability in Speech Processes*. Lawrence Erlbaum, Hillsdale.