Gradient similarity in Lezgian laryngeal harmony: representation & computation

Huteng Dai

Department of Linguistics and RuCCs, Rutgers University

March 1, 2020

Take-home message

Given a well-defined **representational structure** and gradient **similarity metric**, gradient representation is **not** orthogonal to universal feature system and discrete symbolic computation.

The slides and code can be found on `http://hutengdai.com`

Similarity we live by

- Similarity defines the natural classes that interact in **phonology**.
- “perceptual distinctness”, “perceptibility”, “contrast” ……

#	Related research programs	Selected works
1.	Output-driven Phonology	Tesar (2014)
2.	Base-Reduplicant Correspondence	McCarthy & Prince (1995)
3.	Paradigm Uniformity	Benua (1997)
4.	**Agreement by Correspondence**	Rose & Walker (2004)
5.	Dispersion Theory	Flemming (2013)
6.	P-map	Steriade (2001)
7.	Similarity avoidance principle	Frisch et al. (2004)
8.	Contrastive Hierarchy	Dresher (2009)
9.	Learning bias	Wilson (2006)
10.	Exemplar phonology	Bybee (2003)

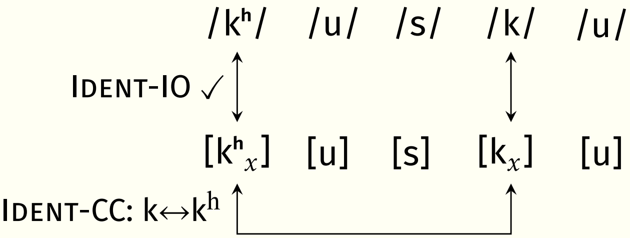
Lesson from ABC

	/kʰ/	/u/	/s/	/k/	/u/
IDENT-IO ✓	↕			↕	
	[kʰ]	[u]	[s]	[k]	[u]

Bolivian Aymara (Rose & Walker, 2004)

IDENT-IO[SG]

Lesson from ABC



Bolivian Aymara (Rose & Walker, 2004)

CORR(Tʰ↔T) > IDENT-IO[SG]

Lesson from ABC

/k^h/ /u/ /s/ /k/ /u/

IDENT-IO: k↔k^h

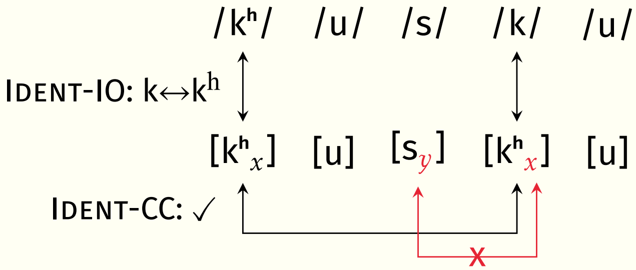
[$k^h{}_x$] [u] [s] [$k^h{}_x$] [u]

IDENT-CC: ✓

Bolivian Aymara (Rose & Walker, 2004)

IDENT-CC[SG], CORR(T^h↔T) > IDENT-IO[SG]

Lesson from ABC



Bolivian Aymara (Rose & Walker, 2004)

IDENT-CC[SG], CORR(Tʰ↔T) > IDENT-IO[SG]

Similarity and ABC

Similarity is encoded in the **correspondence (Corr) hierarchy**:

$$\textsc{Corr}[T\leftrightarrow T] \gg \textsc{Corr}[T\leftrightarrow D] \gg \textsc{Corr}[K\leftrightarrow T] \gg \textsc{Corr}[K\leftrightarrow D] \gg \ldots$$

T'=Ejective, T=Voiceless, T^h=Aspirated, D=Voiced, D'=Implosive, T vs. K: the difference on Place.

(Rose & Walker, 2004)

The probabilistic nature of similarity

- CORR hierarchy is grounded on **categorical** featural similarity metrics:

$$\textit{similarity}(x, y) = \frac{\text{the number of shared features between } x \text{ and } y}{\text{the total number of shared and nonshared features}}$$

as in natural classes-based metrics (Frisch et al., 2004)

- A Bayesian perspective:
  - Similarity is the **belief** that two segments x and y are **(non-)identical**;
  - This belief is updated by the observed shared features.

(Tenenbaum & Griffiths, 2001; Jaynes, 2003)

Structural assumption in feature system

- The distance from [+] to [-] is 1 step for any feature.
  - Any pairs of phonemes with the same amount of shared features have exactly the same similarity;
  - If T↔T' is sufficiently similar to be in agreement, then T↔T^h, T↔D, and T↔K must be in agreement as well.

Lezgian laryngeal harmony

- *T'↔T is a categorical constraint in Lezgian ($N = 0$), and always triggers laryngeal harmony, while Tʰ↔T and T↔D are sufficiently **dissimilar** to escape the impetus to agree.
  - *Underrepresented co-occurrences* ($O/E < 1$)

    T↔T', T'↔T, T'↔D, T'↔Tʰ, D↔T', D↔Tʰ, Tʰ↔D, Tʰ↔T', ...
  - *Overrepresented co-occurrences* ($O/E \geq 1$)

T'↔T'	[q'at͡s'un]	'get dirty'	T↔T	[qaqa]	'ready'
Tʰ↔Tʰ	[tʃʰipʰ]	'fool'	D↔D	[midad]	'grieve'

(Ozburn & Kochetov, 2018)

Lezgian laryngeal harmony

- *T'↔T is a categorical constraint in Lezgian ($N = 0$), and always triggers laryngeal harmony, while $\mathrm{T^h}$↔T and T↔D are sufficiently **dissimilar** to escape the impetus to agree.
  - *Underrepresented co-occurrences* ($O/E < 1$)

    T↔T', T'↔T, T'↔D, T'↔$\mathrm{T^h}$, D↔T', D↔$\mathrm{T^h}$, $\mathrm{T^h}$↔D, $\mathrm{T^h}$↔T', ...
  - *Overrepresented co-occurrences* ($O/E \geq 1$)

T'↔T'	[q'at͡s'un]	'get dirty'	T↔T	[qaqa]	'ready'
$\mathrm{T^h}$↔$\mathrm{T^h}$	[tʃ^{h}ip^h]	'fool'	D↔D	[midad]	'grieve'
$\mathrm{T^h}$↔T	[k^{h}ut͡sun]	'to flush'	T↔D	[et͡sigun]	'put'

(Ozburn & Kochetov, 2018)

Challenge to categorical similarity metrics

The calculated similarity neither aligns with the **co-occurrence** constraints, nor fits the distribution of **speech errors**.

(Rose & King, 2007)

Inventory	minimally dissimilar pairs	Languages
T’, T, T^h, D	***T↔T’**, ✓T↔D, ✓T^h↔T ...	Lezgian, Ndebele
T’, T, D	***T↔T’**, ✓T↔D	Amharic, Chaha, Chontal
T’, T, T^h	***T↔T’**, *T↔T^h	Peruvian & Bolivian Aymara
T’, T, D’	***T↔T’**, ✓T’↔D’	Tzotzil, Tzutujil, Yucatec
T’, T, D, D’	***T↔T’**, ***D’↔D**, ✓T↔D ...	Hausa
T, D, D’	***D’↔D**, ✓T↔D	Bumo Izon, Kalabari Ijo

(Ozburn & Kochetov, 2018; Hansson, 2010, adapted)

Analysis: the special status of [cg]

- Cross-linguistically, different features play different roles in similarity.
- Only the difference on [cg] **always** triggers harmony
- Hypothesis: the distance from [+cg] to [-cg] is systematically shorter than in other Laryngeal features.

(Gallagher & Coon, 2009; Kochetov & Ozburn, 2014)

Acoustic cues

Cross-linguistically, the difference of VOT and preceding vowel duration on [CG] is less distinctive than [VOICE] and [SG].

(Beguš, 2017; Gallagher, 2010a)

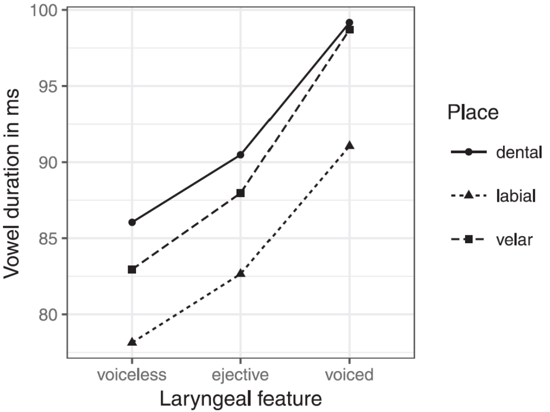


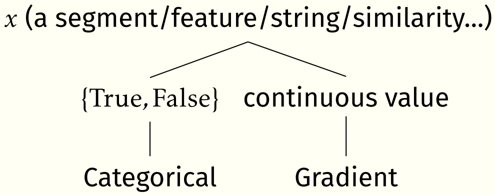
FIG. 2. Estimates of the effects of Laryngeal Features and Place of articulation on preceding vowel duration in ms (from a linear mixed effects model).

Georgian (Beguš, 2017)

Representation

What's representation?

- Representation is the **abstraction** of phonetic and/or phonological knowledge;



Gradient representation: Pros and Cons

- Pros: weighted/gradient (sub-)featural representation can easily handle language-specific granularity;

  (Ladefoged, 1969, 1972, 1973; Keating, 1985; Smolensky & Goldrick, 2016)

- Cons:
  - High degree of freedom (cf. tone numbers);
  - The empirical/laboratory evidence is not always available;
  - Trade-off between granularity and **generality**;

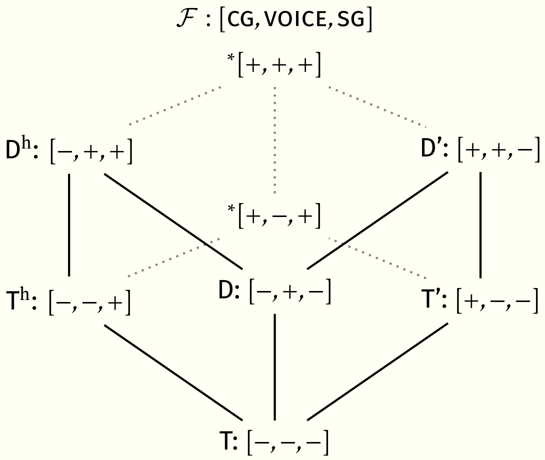
    e. g. universal feature system; typology; similarity metrics; modular representation, etc.

    see criticism in Mackenzie (2009)

Proposal: weighted featural lattice

- Lattice: the universal feature system:
  - Partial order (relative distance);
  - Bounds;

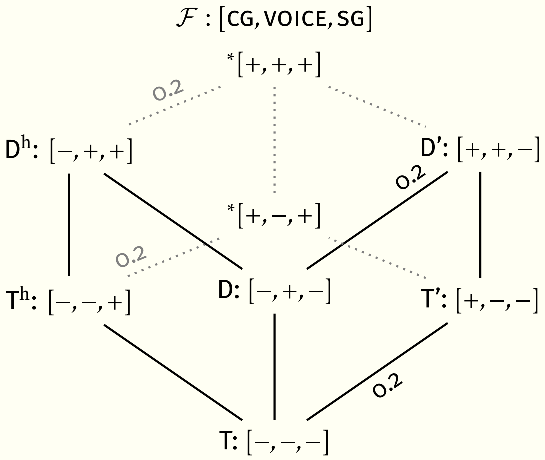
(Tesar, 2014; Magri, 2018)



Proposal: weighted featural lattice

- Weight: the phonetic distance between [+] and [-]
- $w_{[\text{CG}]} < w_{[\text{VOICE}]}, w_{[\text{SG}]}$
- Restriction: $0 < w_f < 1$

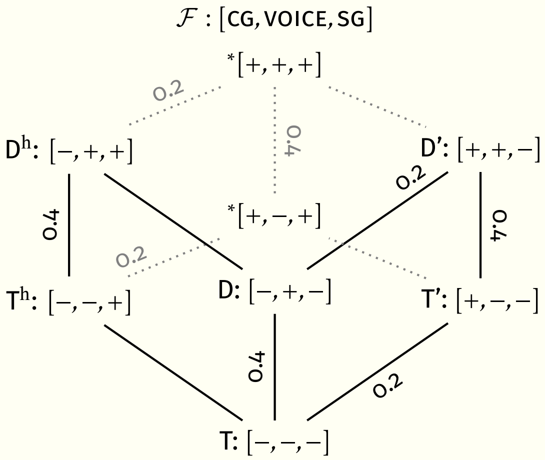
  "How likely two **features** are non-identical, given the observed phonetic cues?"



Proposal: weighted featural lattice

- Weight: the phonetic distance between [+] and [-]
- $w_{[\text{CG}]} < w_{[\text{VOICE}]}, w_{[\text{SG}]}$
- Restriction: $0 < w_f < 1$

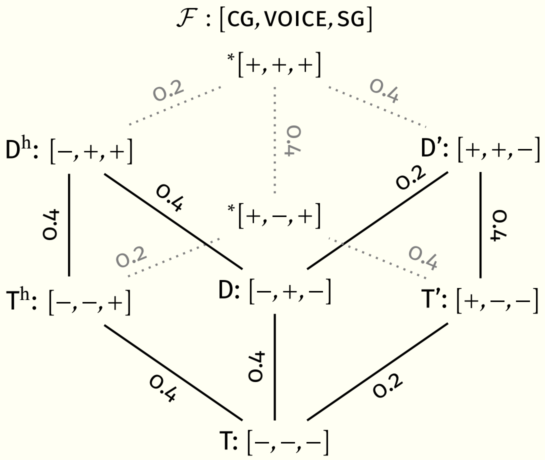
  "How likely two **features** are non-identical, given the observed phonetic cues?"



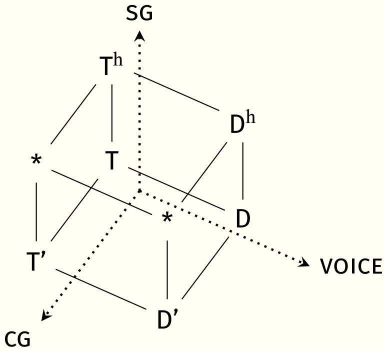
Proposal: weighted featural lattice

- Weight: the phonetic distance between [+] and [-]
- $w_{[\text{CG}]} < w_{[\text{VOICE}]}, w_{[\text{SG}]}$
- Restriction: $0 < w_f < 1$

  "How likely two **features** are non-identical, given the observed phonetic cues?"

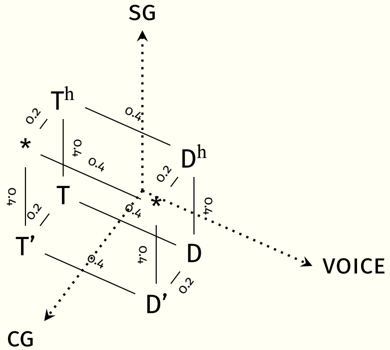


Representational space



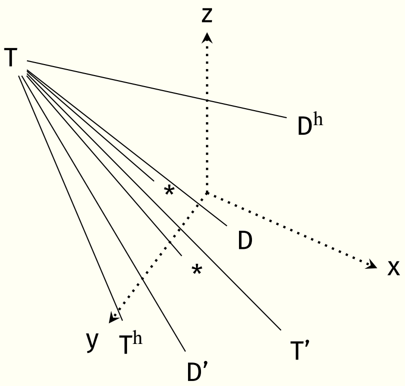
phonological structure $\Longleftrightarrow$ discrete lattice

Representational space



phonetic information $\Longleftrightarrow$ weighted lattice: scaled by [0.2, 0.4, 0.4]

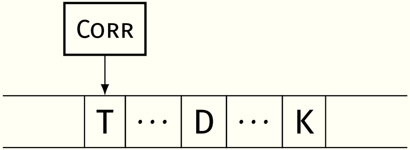
Unconstrained representational space



e. g. *Hilbert space* (Smolensky et al., 2014)

Alternative: segmental similarity scale

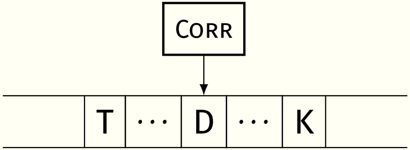
- One-dimensional totally-ordered (weighted) similarity scale;
- The relative similarity is encoded by **adjacency**.



(Rose & Walker, 2004, P.505)

Alternative: segmental similarity scale

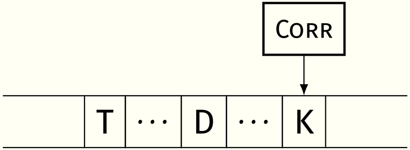
- One-dimensional totally-ordered (weighted) similarity scale;
- The relative similarity is encoded by **adjacency**.



(Rose & Walker, 2004, P.505)

Alternative: segmental similarity scale

- One-dimensional totally-ordered (weighted) similarity scale;
- The relative similarity is encoded by **adjacency**.

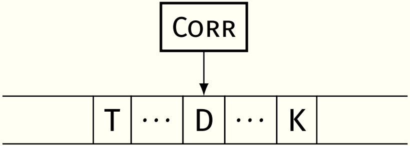


(Rose & Walker, 2004, P.505)

Alternative: segmental similarity scale

CORR[T↔D] ≫ CORR[K↔T] ≫ CORR[K↔D]

✓ *similarity*(T,D) > *similarity*(T,K)

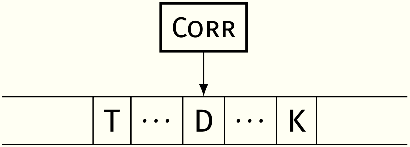


Alternative: segmental similarity scale

CORR[T↔D] ≫ CORR[K↔T] ≫ CORR[K↔D]

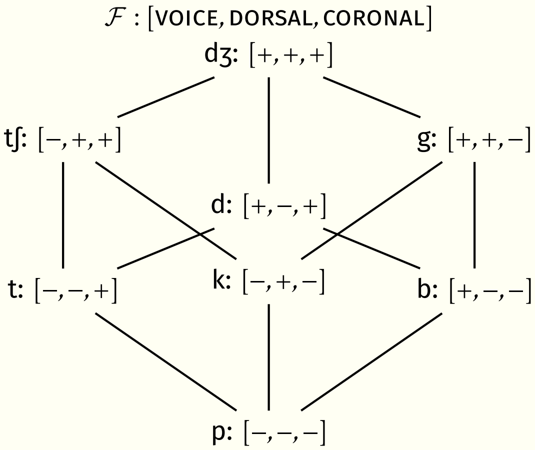
✓*similarity*(T,D) > *similarity*(T,K)

*similarity(K,D) > similarity(K,T)



Alternative: segmental similarity scale

A lattice represents a higher dimensional space which captures the insights lost in an one-dimensional similarity scale.



Computation

Phonetic distance

- The phonetic distance between two segments x and y is computed over a weighted featural lattice $\mathbf{w}$:

$$\textit{distance}_{\mathbf{w}}(x, y) = \sum_{f \in \mathcal{F}} w_f \cdot \delta_f(x, y), \qquad \text{(summed weights of nonshared features)}$$

$$\delta_f(x, y) = \begin{cases} 0, & \text{if } x \text{ and } y \text{ share the feature } f \\ 1, & \text{else} \end{cases}$$

(Wilson & Obdeyn, 2009)

Similarity as Bayesian probability

- Phonological similarity is the belief that x and y are (non-)identical, which is updated by the observed **phonetic distance** between two segments.

$$\begin{aligned} \textit{similarity}_{\mathbf{w}}(x,y) &= 1 - \textit{dissimilarity}_{\mathbf{w}}(x,y) \\ &= 1 - \frac{\textit{distance}_{\mathbf{w}}(x,y)}{\sum_{f \in \mathcal{F}} w_f} \end{aligned}$$

  - This function converts the *phonetic distance* to a probability in $[0,1]$.
  - $\sum_{f \in \mathcal{F}} w_f$ is the **maximal distance** between two segments.

Similarity of LARYNGEAL pairs

C1↓C2→	T’	T	D’	D	T^h	D^h
T’	1	0.8	0.6	0.4	0.4	0
T	0.8	1	0.4	0.6	0.6	0.2
D’	0.6	0.4	1	0.8	0	0.4
D	0.4	0.6	0.8	1	0.2	0.6
T^h	0.4	0.6	0	0.2	1	0.6
D^h	0	0.2	0.4	0.6	0.6	1

Tractable: the set of thresholds is always **finite** whatever the alphabet is.

Agreement by similarity

- IDENT-CC(LARYNGEAL, *SIMILARITY* $\geq k$):
  if *similarity* $\geq k$ in a 2-long **subsequence**,
  penalize any difference in LARYNGEAL.
- IDENT-IO[LARYNGEAL]:
  penalize any input-output difference in LARYNGEAL.
- The critical threshold is determined by the lowest similarity that triggers harmony.

ID-CC(LARYN., *S.* ≥ 1)

↓

ID-CC(LARYN., *S.* ≥ 0.8)

↓

ID-IO[LARYNGEAL]

↓

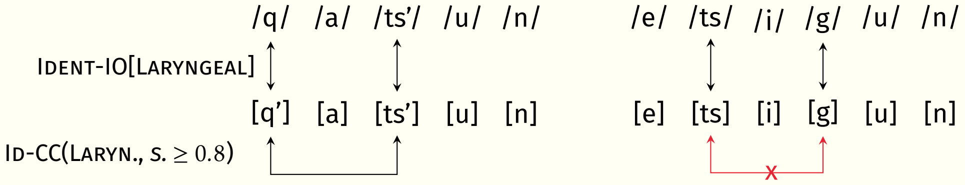
ID-CC(LARYN., *S.* ≥ 0.6)

↓

ID-CC(LARYN., *S.* ≥ 0.4)

Formal language-theoretic computation

- The 2-long **subsequences** in *[qats'un] include {q…a, **q…ts'**, q…u, q…n, a…ts', a…u, a…n, ts'…u, ts'…n, u…n} (Heinz, 2010)



(Q & A: "Agreement by similarity vs. by projection")

Constraint-based analysis

/qats'un/	ID-CC(LARYN., $s. \geq 0.8$)	ID-IO[LARYN.]	ID-CC(LARYN., $s. \geq 0.6$)
a. q…ts' [s. = 0.8]	*!		*
b. ☞ q'…ts' [s. = 1]		*	

/etsigun/	ID-CC(LARYN., $s. \geq 0.8$)	ID-IO[LARYN.]	ID-CC(LARYN., $s. \geq 0.6$)
a. dz…g [s. = 1]		*!	
b. ☞ ts…g [s. = 0.6]			*

Classical OT (Prince & Smolensky, 2004)

Typology

The typology of laryngeal harmony is predicted by varying critical thresholds.

Inventory	Thresholds	Pairs	Languages
T', T	0.8	*T↔T'	Gitksan, Chol
T', T, D'	0.8	*T↔T', ✓T'↔D', ✓T↔D'	Tzotzil, Tzutujil, Yucatec
T', T, D, D'	0.8	*T↔T', *D'↔D, ✓T↔D, ...	Hausa
T', T, T^h, D	0.8	*T↔T', ✓T↔D, ✓T^h↔T ...	Ndebele, Lezgian
T, D, D'	0.8	*D'↔D, ✓T↔D, ✓T↔D'	Bumo Izon, Kalabari Ijo
T', T, D	0.8	*T↔T', ✓T↔D, ✓T'↔D	Amharic
	0.4	*T↔T', *T↔D, *T'↔D	Chaha
T', T, T^h	0.4	*T↔T', *T↔T^h, *T'↔T^h	Peruvian & Bolivian Aymara

(Ozburn & Kochetov, 2018; Hansson, 2010, adapted)

Theoretical implications

Balance between structure and substance

- Substance-free approach: similarity relevant for motivating phonological processes is based on underlying abstract phonological representations;

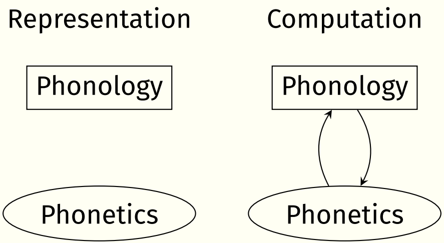
  e.g. Contrastive hierarchy (Mackenzie, 2009, 2011), GSR (Smolensky & Goldrick, 2016)

- Phonetically grounded approach: supplement the universal feature system with **language-specific** (sub)features, such as [long VOT], to account for perceptual similarity.

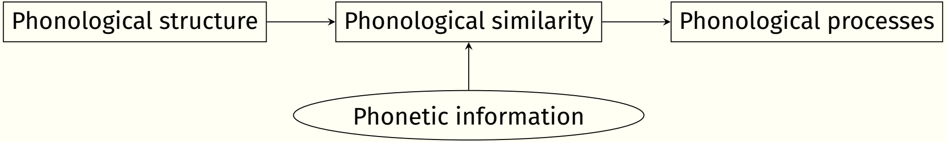
  (Gallagher, 2010a,b, 2012; Lionnet, 2017)

Modular representation

- Lattice:
  - the phonological, abstract, symbolic, universal structure of feature system;
- Weight:
  - the phonetic, fine-grained, gradient, language-specific **information**;
  - not in UR, and **only** available to SR in the **computation** of input-output and surface correspondence;



The interplay of phonology and phonetics



The similarity is computed *w.r.t.* both **phonological** structure and **phonetic** information, and this information is further used in phonological computation.

Future works

- The weighting relation is **testable** in laboratory settings;

  e.g. confusion matrix; neural featural encoding

- Learning from distribution; (Wilson & Obdeyn, 2009; Mayer, 2020)
- Phonological similarity in **Sign Language**. (Keane et al., 2017)

Take-home message

Given a well-defined **representational structure** and **similarity metric**, gradient representation is **not** orthogonal to universal feature system and discrete symbolic computation.

Find the slides and code on `http://hutengdai.com`, and feel free to contact me for questions and collaborations!

Acknowledgement

I thank Adam McCollum, Adam Jardine, Bruce Tesar, Brian Pinsky, Jason Shaw, Keith Johnson, Robin Karlin and audiences at LSA 2020, BLSW 2020, and Rutgers Phonology and Phonetics Research Group (PhonX), for their comments and insights. My special thanks are extended to Alan Yu for providing the valuable recordings of Lezgian.

RUTGERS

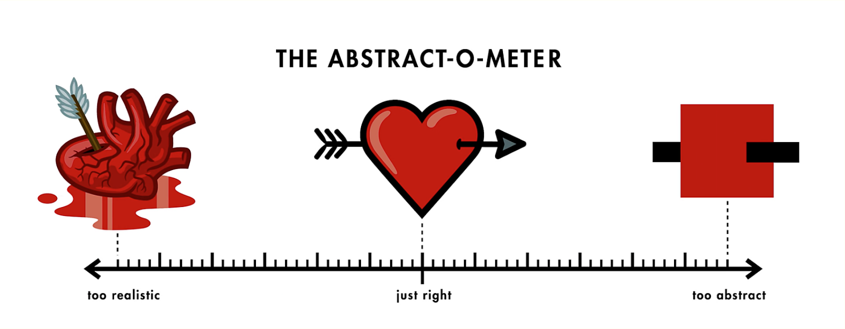
Q&A

Gradient representation as abstraction

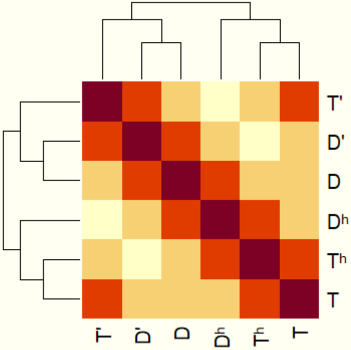
- The weights encode the phonetic information from closely-related dimensions:
  - Distributional (Hall, 2009; Mayer & Daland, 2019; Mayer, 2020)
  - Substantial (Vitevitch & Luce, 1999; Mielke, 2012; Redmon, in prep.)

Abstractness of gradient representation

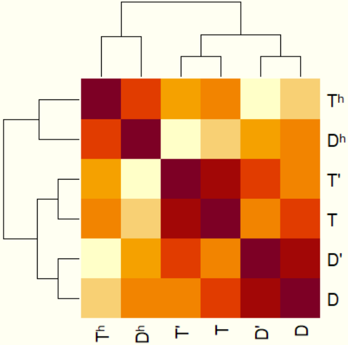
- Phonetic measurement ≠ weight on featural lattice:
  - Phonetic invariance doesn't exist; (Pierrehumbert, 2016; Zellou & Tamminga, 2014)
  - Real-numbered representation is still an **abstraction**!



Categorical vs. weighted similarity metrics

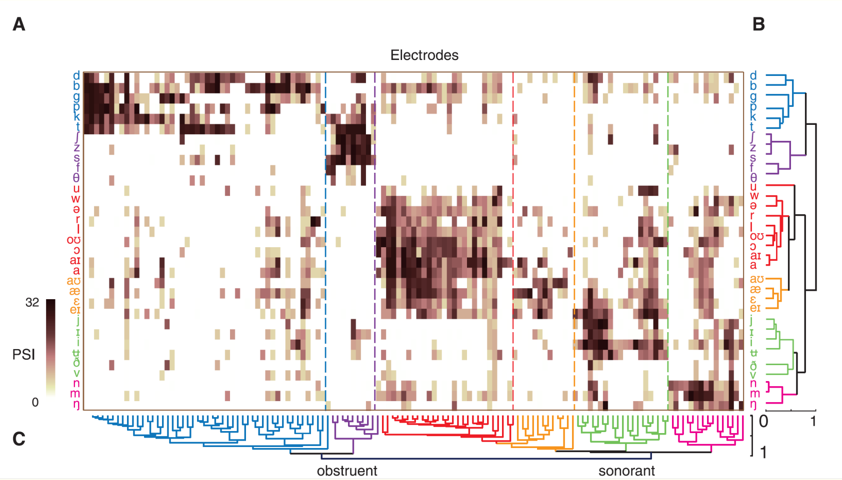


Categorical similarity metrics ☹



Gradient similarity metrics ☺

Neural featural encoding



Mesgarani et al. (2014)

GSR in UR: the road not taken

- Despite the ambition of correlating GSR and phonetics, Gradient Symbolic Representation is a **substance-free** approach *per se* – the weights are represented in **UR** and are not grounded by phonetics.
- Reasons:
  - **Tractable** computing power $\leftrightarrow$ **learnable** grammar;
  - **Restrictive** representational space;
  - **Testable** weighting relation;
  - Well-defined similarity metrics – necessary for any IDENT constraint.

Reference

Beguš, G. (2017). Effects of ejective stops on preceding vowel duration. *The Journal of the Acoustical Society of America*, *142*(4), 2168–2184.

Benua, L. (1997). *Transderivational identity: Phonological relations between words* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge University Press.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology* (Vol. 121). Cambridge University Press.

Flemming, E. S. (2013). *Auditory representations in phonology*. Routledge.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*(1), 179–228.

Gallagher, G. (2010a). *The perceptual basis of long-distance laryngeal restrictions* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Gallagher, G. (2010b). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology*, *27*(3), 435–480.

Gallagher, G. (2012). Perceptual similarity in non-local laryngeal restrictions. *Lingua*, *122*(2), 112–124.

Gallagher, G., & Coon, J. (2009). Distinguishing total and partial identity: Evidence from chol. *Natural Language & Linguistic Theory*, *27*(3), 545–582.

Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony* (Unpublished doctoral dissertation). The Ohio State University.

Hansson, G. Ó. (2010). *Consonant harmony: Long-distance interactions in phonology* (Vol. 145). Univ of California Press.

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry*, *41*(4), 623–661.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge university press.

Keane, J., Sehyr, Z. S., Emmorey, K., & Brentari, D. (2017). A theory-driven model of handshape similarity. *Phonology*, *34*(2), 221–241.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (p. 115-132). Orlando, FL: Academic Press.

Kochetov, A., & Ozburn, A. (2014). Categorical and gradient laryngeal harmony in lezgian. In *Uc berkeley phonology lab annual report* (p. 460).

Ladefoged, P. (1969, September). The measurement of phonetic similarity. In *International Conference on Computational Linguistics COLING 1969: Preprint no. 57.* Sånga Säby, Sweden. Retrieved from `https://www.aclweb.org/anthology/C69-5701`

Ladefoged, P. (1972). Phonetic prerequisites for a distinctive feature theory. *Papers in linguistics and phonetics to the memory of Pierre Delattre*, 273–285.

Ladefoged, P. (1973). The features of the larynx. *Journal of phonetics*, *1*(1), 73–83.

Lionnet, F. (2017). A theory of subfeatural representations: the case of rounding harmony in laal. *Phonology*, *34*(3), 523–564.

Mackenzie, S. (2009). *Contrast and similarity in consonant harmony processes* (Unpublished doctoral dissertation). University of Toronto.

Mackenzie, S. (2011). Contrast and the evaluation of similarity: Evidence from consonant harmony. *Lingua*, *121*(8), 1401–1423.

Magri, G. (2018). Output-drivenness and partial phonological features. *Linguistic Inquiry*, *49*(3), 577–598.

Mayer, C. (2020). An algorithm for learning phonological classes from distributional similarity. *Phonology*.

Mayer, C., & Daland, R. (2019). A method for projecting features from observed sets of phonological classes. *Linguistic Inquiry*, 1–85.

McCarthy, J. J., & Prince, A. (1995). Faithfulness and reduplicative identity. *Linguistics Department Faculty Publication Series*, 10.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010.

Mielke, J. (2012). A phonetically based metric of sound similarity. *Lingua*, *122*(2), 145–163.

Ozburn, A., & Kochetov, A. (2018). Ejective harmony in lezgian. *Phonology*, *35*(3), 407–440.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*.

Prince, A., & Smolensky, P. (2004). Optimality theory: Constraint interaction in generative grammar malden. *MA: Blackwell*.

Redmon, C. (in prep.) *Lexical acoustics: Linking phonetic systems to the higher-order units they encode* (Unpublished doctoral dissertation). University of Kansas.

Rose, S., & King, L. (2007). Speech error elicitation and co-occurrence restrictions in two ethiopian semitic languages. *Language and Speech*, *50*(4), 451–504.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 475–531.

Smolensky, P., & Goldrick, M. (2016). Gradient symbolic representations in grammar: The case of french liaison. *Ms. Available as ROA*, *1286*.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive science*, *38*(6), 1102–1138.

Steriade, D. (2001). The phonology of perceptibility effects: the p-map and its consequences for constraint organization. *Ms., UCLA*.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*(4), 629–640.

Tesar, B. (2014). *Output-driven phonology: Theory and learning* (No. 139). Cambridge University Press.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of memory and language*, *40*(3), 374–408.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science*, *30*(5), 945–982.

Wilson, C., & Obdeyn, M. (2009). *Simplifying subsidiary theory: statistical evidence from arabic, muna, shona, and wargamay.* (ms. Johns Hopkins University)

Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in philadelphia english. *Journal of Phonetics*, *47*, 18–35.