Modelling (degrees of) tone sandhi in a feed-forward phonological analysis

Aleksei Nazarov

Utrecht University

WECOL 2026

Overview

- Incomplete neutralization: a lexical contrast is collapsed in the sound system but retained in phonetic detail (paradox!)
- Degrees of incompleteness: based on factors like cognitive load and intention, more or less phonetic differences can be retained
- Proposal: sound systems mix different levels of abstractness, but strength of interaction factors is influenced by external circumstances
- Modelling: Harmonic Grammar model with “scaling factors” for weights allows to marry grammar and incompleteness

Incomplete neutralization

- Classic case: Dutch, German and other languages have final devoicing, but still fine-grained phonetic cue differences between final underlyingly voiced and voiceless sounds (e.g., Warner et al. 2004)

		UR	SR	phonetic realization
Dutch	*'wheel'*	/rad/	|rat|	[rat$_{\mathbf{d}}$]
	'rat'	/rat/	|rat|	[rat$_{\mathbf{t}}$]

- Surface representation has lost voicing contrast
- How does information about contrast reach phonetic realization?

Further empirical case studies

- Japanese long vs. lengthened vowels (Braver & Kawahara 2016, Braver 2019)
  - Length contrast in vowels: ki ‘tree’ vs. kii ‘key’
  - Monosyllabic NPs are lengthened: ki-mo ‘tree also’ vs. kiː ‘tree’
  - However, lengthened short vowels are still systematically shorter than long vowels in same prosodic position: duration(*kiː*) < duration(*kii*)
- Mandarin underlying vs. derived Tone 3 (Du & Durvasula 2022; Zeng et al. 2025)
  - Huai’an: Contrast between T1<>T3: pa^{1} ‘play with’ vs. pa^{3} ‘smash (to eat)’
  - Sandhi prohibits *T1-T1, creates new T3: /u^{2} pa^{1} xa^{1}/ → [u^{2} **pa^{3}** xa^{1}]
  - Derived T3 = Underlying T3 in phonology, but not in phonetic realization

Degrees of neutralization

- Zeng et al. (2025): participants in production experiment are presented with distractor task related to tone – show stronger neutralization of Mandarin T1 vs. T3 contrast in sandhi (tonal contact), compared to control

- Nelson & Heinz (2025): show that pragmatic context has an effect on neutralization – more intent to contrast two words leads to less neutralization, more distinction between members of neutralized contrast

Current contribution

- Goal: account for Mandarin case study with standard architecture (i.e., lexicon → phonology → phonetics), in contrast to Nelson & Heinz (2025), Zeng et al. (2025)
- Modelling: Harmonic Grammar modelling using explicit phonetic constraints (Flemming 2001, Braver 2019) and mixing of levels of abstraction (cf. Van Oostendorp 2008)
- Context-specific variation: weights within Harmonic Grammar model are modulated by scaling factors (Coetzee & Kawahara 2013)

Data: Huai’an Mandarin

- Case study that is prosodic and non-paradigm-based
- Based on between-word tonal adjustments
- Tone inventory (Du & Durvasula):
  - Tone 1: high falling
  - Tone 2: high rising
  - Tone 3: low rising
  - Tone 4: high level
- Degrees of neutralization not studied by Du & Durvasula, but found for standard Mandarin (Zeng et al. 2025)

Figure from Du & Durvasula

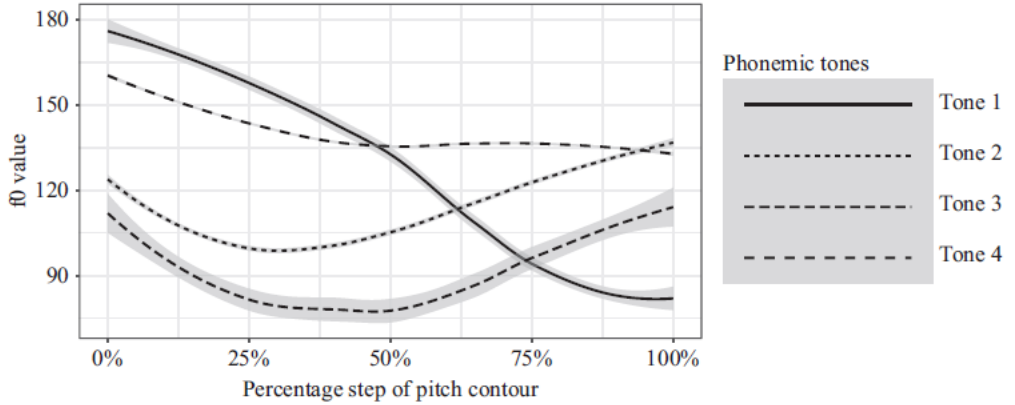


Figure 1. *Tonal contour of phonemic tones in Huai’an.*

Phonological side: Neutralizing tone sandhi

- *T1-T1: first syllable changes to T3 (optional process)

/u^{2} pa^{1} xa^{1}/ → [u^{2} pa^{3} xa^{1}] ‘Mr. Wu^{2} plays with shrimp’
MH HL HL MH LM HL

- *T3-T3: first syllable changes to T2 (obligatory)

/u^{3} pa^{3} xa^{1}/ → [u^{2} pa^{3} xa^{1}] ‘Mr. Wu^{3} smashes shrimp (to eat)’
LM LM HL MH LM HL

- T1 sandhi feeds T3 sandhi: derived T3 counts as true T3 (neutralization)

/u^{3} pa^{1} xa^{1}/ → (u^{3} pa^{3} xa^{1} →) [u^{2} pa^{3} xa^{1}] ‘Mr. Wu^{3} plays with shrimp’
LM HL HL LM LM HL MH LM HL

Examples from Du & Durvasula

Phonetic side: T1/T3 contrast remains expressed

- T3 derived from /T1/ is realized differently from underlying T3 (closer to the high tone of T1)

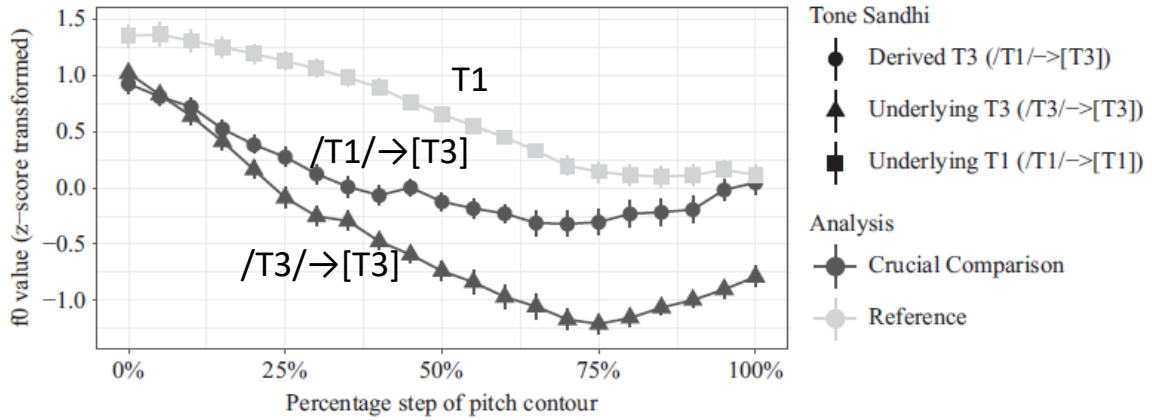


Figure 3. *Comparison of second-syllable contours in Experiment 1 (Tone 1; error bars indicate standard error).*

Figure taken from Du & Durvasula

Analytic framework

- Language-specific abstraction – at least some phonological representational elements are emergent from distributional learning (e.g., Mielke 2004, Dresher 2009, Mayer 2020)

e.g. language-specific (un)natural classes can get their own “bespoke” feature labels; but also “deeper” features – like [T] for “behaves like a voiceless stop” in Canadian Raising: wr[əɪr$_{[+T]}$]er – r[aɪr$_{[-T]}$]er

- Constraints are at least partially language specific (cf. Hayes and Wilson 2008) and can mix different levels of abstraction

e.g., in syntax [I hope this N finds you well], [the Adj-er the Adj-er] (Goldberg 2025); in phonology: *[aɪ]C$_{[+T]}$

  - Useful for interplay between systematic syntax versus idioms and fixed expressions – but also for interplay between lexical, phonotactic, and phonetic factors

Levels of abstraction in this case

- We need access to a surface, near-phonetic representation of tone
  - In this case, 6 levels from lowest to highest, corresponding to graph in Du & Durvasula (2022) – 0 through 5; but likely to be continuous
- We need a way to refer to tone categories T1, T2 etc., since tone sandhi prohibits certain sequences of tone categories
  - *T1-T1 is changed to T3-T1, *T3-T3 is changed to T2-T3
- From distributional learning and incomplete neutralization data, speakers may have abstracted deeper, contrastive categories of tones:
  - [α], normally corresponds to [T1], but to [T2] or [T3] in sandhi
  - [γ] corresponds to [T3], but to [T2] in sandhi

Harmonic Grammar

- Framework used mainly in phonological analysis (e.g. Potts et al. 2010)
- Closely related to Optimality Theory: each input has multiple output **candidates**, they get **scores** on constraints (=phonological predicting **factors**)
- Evaluation method: winning candidate = argmax(weights · scores)

/input/	w = 20	5	1	1	Harmony
	F1	F2	F3	F4	
Output 1	-1			-1	-21
☞ **Output 2**		-1	-13		**-18**
Output 3	-1				-20

Details: data

- 3 actual data points based on Du and Durvasula (F0 realization quantized to 3 time points and 5 height points):

/(u^{2}) pa$^{\gamma:3}$ xa$^{\alpha:1}$ / → [(u^{2}) páàà$^{\gamma:3}$ xa$^{\alpha:1}$] Mr. Wu^{2} plays with shrimp

/(u^{2}) pa$^{\alpha:1}$ xa$^{\alpha:1}$ / → [(u^{2}) páàā$^{\alpha:3}$ xa$^{\alpha:1}$] Mr. Wu^{2} smashes shrimp (to eat)

/u$^{\gamma:3}$ pa$^{\alpha:1}$ xa$^{\alpha:1}$ / → [u$^{\gamma:2}$ páàā$^{\alpha:3}$ xa$^{\alpha:1}$] Mr. Wu^{3} smashes shrimp (to eat)

- 4 hypothetical inputs ("test set") to ensure these are mapped onto attested forms:

/pa$^{\alpha:3}$ xa$^{\alpha:1}$/ /pa$^{\alpha:3}$/ (underlying T3 with [α] should behave like T1)
/pa$^{\gamma:3}$ xa$^{\alpha:1}$/ /pa$^{\alpha:1}$/ (underlying T1 with [γ] should behave like T3)

- Surface values considered: {u^{2},u^{3}}, {pa^{1},pa^{3}} *x* all quantized F0 values between T3 (**v́ v̏ v̀**) and T1 (**v̋ v̄ v̄**)

Details: constraints

- Tonal constraints:

*T1 *[α,T3] *T1-T1 *T3-T3	weight = 20
Ident(T)	weight = 1

- F0 realization constraints link T1, T3 and α to desired quantized F0 value (0-5) at start, middle, or end of the syllable
- Violations: distance between true and desired value is squared, then negative summed as per Braver (2019) (e.g. distance 2 for 1st syllable and 3 for 2nd syllable would yield -4-9=-13)

T1=v̋**(5)**vv	T1=v**v̄(3)**v	T1=vv**v̄(3)**	T3=**v́(4)**vv	weight = 5
α=v̋**(5)**vv	α=v**v̄(3)**v	α=vv**v̄(3)**		weight = 3
T3=v**v̌(0)**v	T3=vv**v̀(1)**			weight = 1

Results without gradualness

Underlying T3: canonical melody

	/u$^{\beta:2}$ pa$^{\gamma:3}$ xa$^{\alpha:1}$/	w = 20	20	20	5	5	3	1	1	Harmony
		*T1-T1 *T3-T3	*T1	*[α, T3]	**T3=v́v v**	T1=v́vv T1=vv̄v T1=vvv̄	α=v́vv α=vv̄v α=vvv̄	**T3=vv̀v** **T3=vvv̀**	Id(T)	
T3 -> T1	**u^{2} pā$_{5}$ā$_{3}$ā$_{3}$^{1} xa^{1}**	-1	-2						-1	-61
T3 w/ T1 melody	**u^{2} pā$_{5}$ā$_{3}$ā$_{3}$^{3} xa^{1}**		-1		-1			-13		-38
blended melody	**u^{2} pá$_{4}$à$_{2}$ā$_{3}$^{3} xa^{1}**		-1					-8		-28
T3 melody	☞ **u^{2} pá$_{4}$ā̀$_{0}$à$_{1}$^{3} xa^{1}**		-1							**-20**

Canonical T3 melody wins (max harmony) because no tone change (Ident(T)), no identical tones in a row, and no F0 realization violations.

Derived T3: incomplete neutralization

	/u$^{\beta:2}$ pa$^{\alpha:1}$ xa$^{\alpha:1}$/	w = 20	20	20	5	5	3	1	1	Harmony
		*T1-T1 *T3-T3	*T1	*[α,T3]	T3=v́vv	T1=v̋vv T1=vv̄v T1=vvv̄	**α=v̋vv** **α=vv̄v** **α=vvv̄**	T3=vv̏v T3=vvv̀	ID(T)	
Original T1	u^2 pa̋$_5$ā$_3$ā$_{3j}{}^1$ xa$_j{}^1$	**-1**	**-2**							-60
T3 w/ T1 melody	u^2 pa̋$_5$ā$_3$ā$_{3j}{}^3$ xa$_j{}^1$		-1	-1	**-1**			**-13**	-1	-59
blended melody	☞ u^2 pá$_4$à$_2$ā$_{3j}{}^3$ xa$_j{}^1$		-1	-1			**-2**	**-8**	-1	**-55**
T3 melody	u^2 pá$_4$ȁ$_0$à$_{1j}{}^3$ xa$_j{}^1$		-1	-1			**-14**		-1	-83

Blended melody wins (max harmony) because leaving original T1 leads to T1-T1 sequence, and canonical T1 and T3 melodies for abstract T3 lead to greater penalties on F0 realization constraints. Winner compromises between T3 and α.

Results with gradualness

Scaling factors

- Following work like Coetzee & Kawahara (2013), I assume that context-induced variation comes from perturbation of the weights in the model by addition of a scaling factor

  new weight = old weight + scaling factor

- Here, I assume that a scaling factor in the range [-2,2] is added to the weights of the factors to do with the highest abstraction level [α,β,γ]
- Higher scaling factor correlates with greater semantic intent (Nelson & Heinz 2025), lower cognitive load (Zeng et al. 2025)

Derived T3: incomplete neutralization, s.f. -1

	/$u^{\beta:2}$ $pa^{\alpha:1}$ $xa^{\alpha:1}$/	w = 20	20-1=19	20	5	5	2-1=1	1	1	Harmony
		*T1-T1 *T3-T3	*T1	*[α,T3]	T3=v́vv	T1=v̋vv T1=vv̄v T1=vvv̄	**α=v̋vv** **α=vv̄v** **α=vvv̄**	T3=vv̀v T3=vvv̀	ID(T)	
Original T1	u^2 $pa\acute{\!}_5\bar{a}_3\bar{a}_{3j}{}^3$ $xa_j{}^1$		-1	-1	**-1**			**-13**	-1	-56
T1-biased blend	u^2 $pa\acute{\!}_4\grave{a}_2\bar{a}_{3j}{}^3$ $xa_j{}^1$		-1	-1			**-2**	**-8**	-1	-50
T3-biased blend	☞ u^2 $pa\acute{\!}_4\grave{a}_2\bar{a}_{2j}{}^3$ $xa_j{}^1$		-1	-1			**-6**	**-2**	-1	**-48**
T3 melody	u^2 $pa\acute{\!}_4\grave{a}_0\grave{a}_{1j}{}^3$ $xa_j{}^1$		-1	-1			**-14**		-1	-54

Because of lower weight of [α]-based phonetic realization factors, there is still blending but the melody is closer to T3

Derived T3: incomplete neutralization, s.f. -2

/u$^{\beta:2}$ pa$^{\alpha:1}$ xa$^{\alpha:1}$/	w = 20	20-2=18	20	5	5	2-2=0	1	1	Harmony
	*T1-T1 *T3-T3	*T1	*[α,T3]	T3=v́vv	T1=v̋vv T1=vv̄v T1=vvv̄	**α=v̋vv** **α=vv̄v** **α=vvv̄**	T3=vv̀v T3=vvv̀	ID(T)	
Original T1 u^2 pá$_5$ā$_3$ā$_{3j}$^{3} xa$_j$^{1}		-1	-1	**-1**			**-13**	-1	-55
T1-biased blend u^2 pá$_4$à$_2$ā$_{3j}$^{3} xa$_j$^{1}		-1	-1			**-2**	**-8**	-1	-47
T3-biased blend u^2 pá$_4$à$_2$ā$_{2j}$^{3} xa$_j$^{1}		-1	-1			**-6**	**-2**	-1	-41
T3 melody ☞ u^2 pá$_4$ȁ$_0$à$_{1j}$^{3} xa$_j$^{1}		-1	-1			**-14**		-1	**-39**

No blending because [α]-based phonetic realization factors have weight of 0

Discussion

- Proof of concept: we can explain incomplete neutralization with purely phonological factors
- Necessary components: multiple levels of abstraction, interface between these, scaling factors
- Future: is this model actually better than having a complex architecture where phonology directly refers to semantics or to language processing? (model comparison work)
- Future: find levels of abstraction unsupervised/semi-supervised; explore learnability challenges (weights now found manually)

Thank you!

References

- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives* (LOT International Series 11) [Doctoral dissertation, University of Amsterdam]. The Hague: Holland Academic Graphics.
- Braver, A. (2019). Incomplete neutralization as paradigm uniformity with weighted phonetic constraints. *Phonology 36*(1), 1–36.
- Braver, Aaron & Shigeto Kawahara (2016). Incomplete neutralization in Japanese monomoraic lengthening. In Adam Albright & Michelle A. Fullwood (eds.) Proceedings of the 2014 Meeting on Phonology.
- Coetzee, A., & Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language and Linguistic Theory 31*, 47–89.
- Dresher, B.E. (2009). *The Constrative Hierarchy in Phonology*. Cambridge: Cambridge University Press.
- Du, N., & Durvasula, K. (2022). Phonetically incomplete neutralisation can be phonologically complete: evidence from Huai'an Mandarin. *Phonology 39*(4), 559–595.
- Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology 18*(1), 7–44.
- Goldberg, A. (1995). Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.
- Goldberg, A. (2025). Usage-based constructionist approaches and Large Language Models.
- Hayes, B. & C. Wilson. (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. Linguistic Inquiry 39 (3): 379-440.
- Mayer, C. (2020). An algorithm for learning phonological classes from distributional similarity. Phonology 37(1), 91-131.

References

- Nelson, S & Heinz, J. 2022. “Incomplete Neutralization and the Blueprint Model of Production”. In Peter Jurgec, Liisa Duncan, Emily Elfner, Yoonjung Kang, Alexei Kochetov, Brittney K. O’Neill, Avery Ozburn, Keren Rice, Nathan Sanders, Jessamyn Schertz, Nate Shaftoe, and Lisa Sullivan (eds.), Proceedings of the 2021 Annual Meeting on Phonology. Washington, DC: Linguistic Society of America.
- Nazarov, Aleksei. 2019. Formalizing the connection between opaque and exceptionful generalizations. *Toronto Working Papers in Linguistics* 41:1.
- Nelson, S. & Heinz, J. 2025. “The Blueprint Model of Production”. *Phonology 42*, e12.
- Van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua 118*(9), 127–142.
- Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17.2:237–74.
- Pater, Joe. 2010. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 123–54. London: Equinox.
- Potts, C., Pater, J., Jesney, K., Bhatt, R., & Becker, M. (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology 27*(1), 77–117.
- Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other subphonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics 32*(2), 251–267.
- Zeng, Yuyu, Chang Wang, Jie Zhang. 2025. Cascading activation in spoken word production drives incomplete neutralization: An internet-based study of Mandarin 3rd tone sandhi. *Journal of Phonetics* 115: 101428.

Appendix

Generalizability

- /γ,T3/ vs. /α,T1/ undergo **both** phonological neutralization (derived T3 triggers T3 sandhi) **and** phonetic differentiation (derived T3 is realized as a compromise between T1/T3 melodies)
- Behaviour of ROTB inputs shows that no fully neutralizing alternations are expected even in novel words:
  - /γ,T1/ behaves like /γ,T3/: /pa$^{γ:1}$/ → [pa$^{γ:3}$], /pa$^{γ:1}$ xa$^{α:1}$ / → [páa̰à$^{γ:3}$ xa$^{α:1}$]
  - /α,T3/ behaves like /α,T1/: /pa$^{α:3}$/ → [pa$^{α:1}$], /pa$^{α:3}$ xa$^{α:1}$/ → [páàā$^{α:3}$ xa$^{α:1}$]
- Crucially: candidates that change underlying [α] and [γ] not considered
  - Not digestible by phonological system? Lexical entry constraints? Work in progress