Gradient Phonotactics against Naturalness

Gašper Beguš

Harvard University
begus@fas.harvard.edu

Berkeley Linguistics Society
43rd Annual Meeting
UC Berkeley, Feb 4, 2017

Gradient phonotactics

- Two aspects of OT widely discussed: how to handle **naturalness** and **gradience** (Frisch et al. 2004, Antilla 2008, Coetzee and Pater 2008)
- Little has been said about intersection of the two: **unnatural gradient phonotactics**

Gradient phonotactics

- Two aspects of OT widely discussed: how to handle **naturalness** and **gradience** (Frisch et al. 2004, Antilla 2008, Coetzee and Pater 2008)
- Little has been said about intersection of the two: **unnatural gradient phonotactics**
- **Question:** Can gradience be unnatural?

Gradient phonotactics

- Two aspects of OT widely discussed: how to handle **naturalness** and **gradience** (Frisch et al. 2004, Antilla 2008, Coetzee and Pater 2008)
- Little has been said about intersection of the two: **unnatural gradient phonotactics**
- **Question:** Can gradience be unnatural?
- Trends in the lexicon that operate in an unnatural direction

Naturalness

- Unnatural phonotactics primarily in light of *surfeit of the stimulus*
- What is "unnatural"?

- **Natural**: phonetically motivated
- **Unmotivated**: lack phonetic motivations
- **Unnatural**: operating in the opposite direction from universal phonetic tendency

- UPT: phonetic processes motivated by articulatory (or perceptual) mechanisms that passively and universally operate in speech production and are typologically common.

Introduction

- Many open questions:
  - Are natural alternations possible?
  - Restricted CON: some constraints not learnable?
- How unnatural processes arise?
- Is sound change natural (Blust 2005)?

Outline

Outline

Outline

Outline

Outline

Tarma

- Tarma Quechua, a dialect of Quechua spoken in Tarma, Junín, Peru (Adelaar 1977, Puente Baldoceda 1977, Nazarov 2008)
- **Distribution of [voice] in [DOR] and [LAB] stops**
- [+voice]: intervocalically, post-consonantally, but not post-nasally

Data

- From Adelaar 1977 and Nazarov 2008

b, g / C____; C $\neq$ N

b, g / V____V

p, k / elsewhere

- Examples:
- Lexical variation

#__	[pirwa]
__R, T	[rikra]
N__	[wampu]
V__V	[kuba]
R, T__	[takba]

Data

- From Adelaar 1977 and Nazarov 2008

b, g / C____; C $\neq$ N

b, g / V____V

p, k / elsewhere

- Examples:
- Lexical variation

#__	[pirwa]
__R, T	[rikra]
N__	[wampu]
V__V	[kuba]
R, T__	[takba]

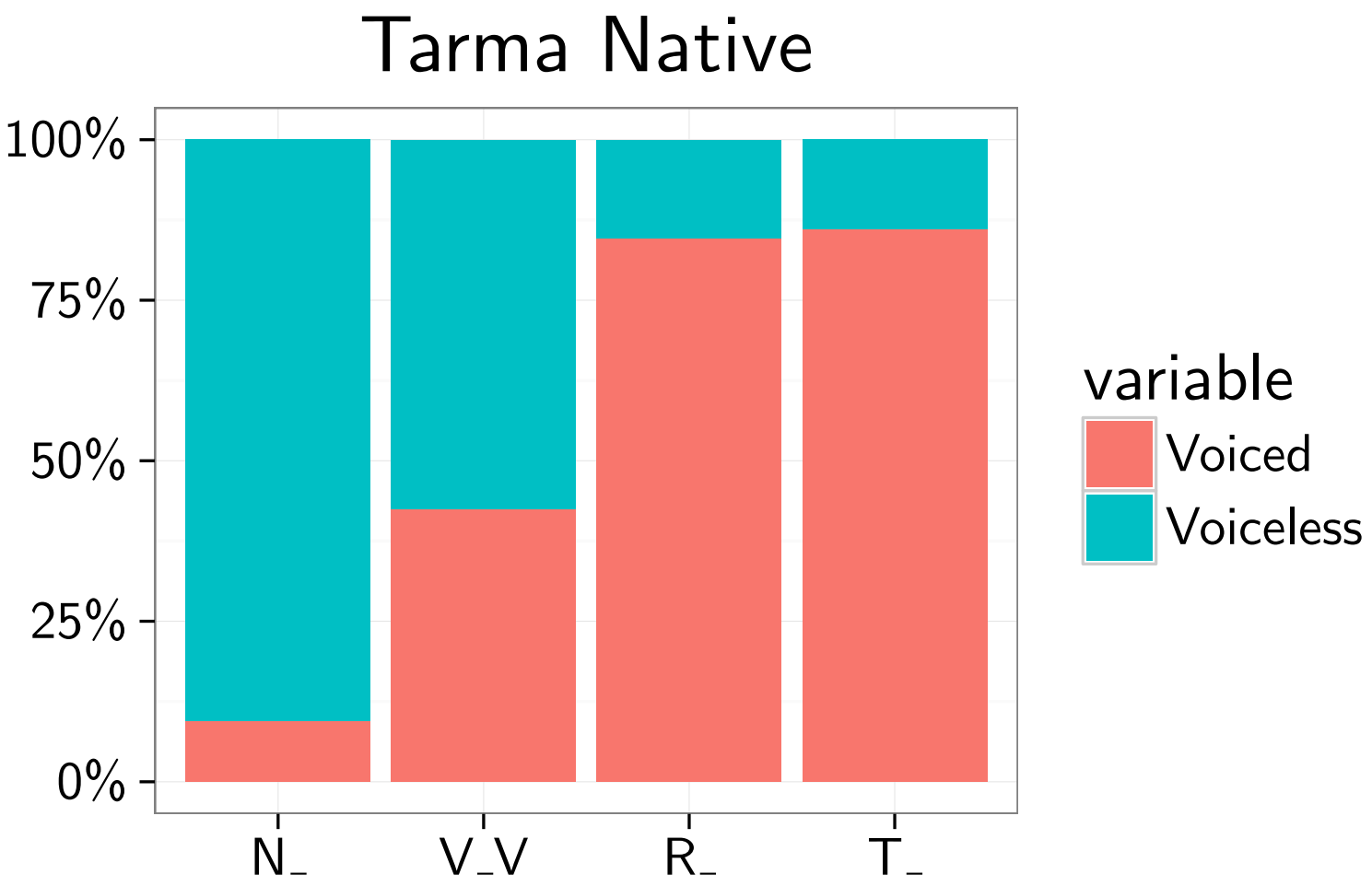
Data

- Distribution of voicing
- Native vocabulary analyzed by Nazarov (2008): [+voice] almost never N__, half of the time in V__V and almost always R,T__

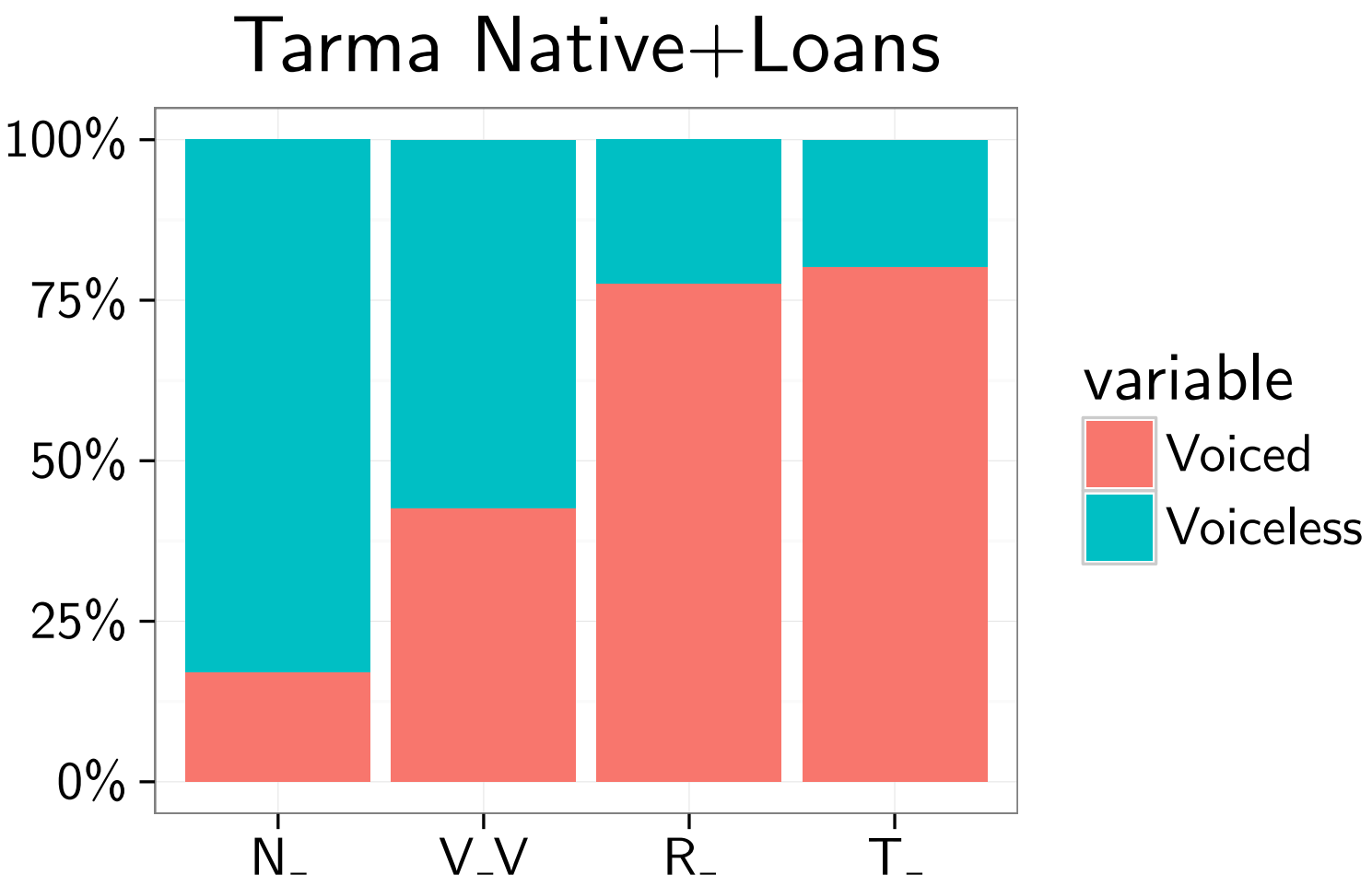
Data

- Recount (cf. Nazarov 2008) to include non-native vocabulary and statistical tests
- All tokens with [DOR] or [LAB] in TQ vocabulary (Adelaar 1977)
- 1119 tokens: 910 in native vocabulary, 289 in loans from Spanish

Data



Data


Tarma Native+Loans
100%
75%
50%
25%
0%
N_
V_V
R_
T_
variable
Voiced
Voiceless

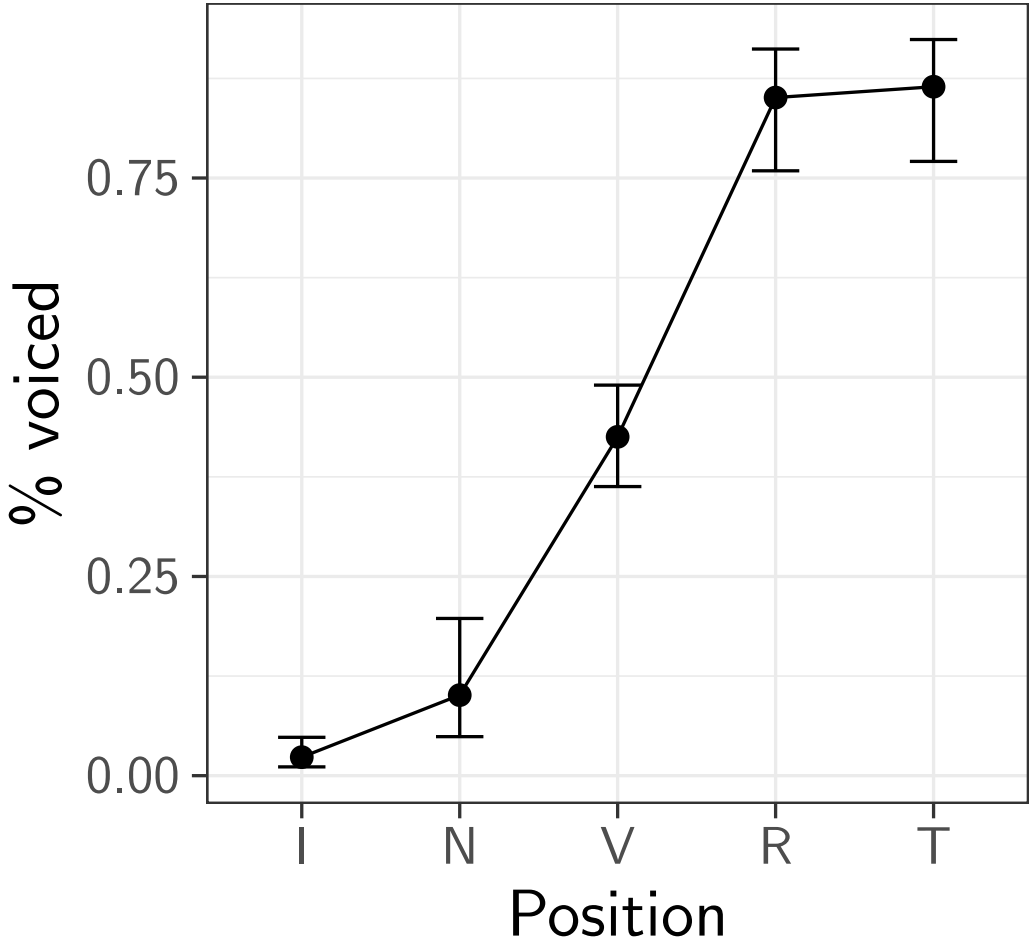
Data

- Data was fit to a logistic regression model in R (R Core Team) with `glm()`
- DV: Voice (binomial)
- IV: Context (treatment-coded with intervocalic as the reference level) and Place (treatment-coded with velar as reference level) of articulation with no interactions (native vocabulary)

Data

	Est.	SE	z value	Pr(>|z|)
(Intercept)	-0.045	0.172	-0.260	0.7952
V__V vs. R__	2.044	0.332	6.164	0.0000
V__V vs. T__	2.155	0.353	6.101	0.0000
V__V vs. N__	-1.884	0.421	-4.478	0.0000
V__V vs. #__	-3.437	0.407	-8.437	0.0000
labial vs. velar	-0.502	0.214	-2.344	0.0191

Data



Data

- *Unnatural distribution of voice feature*
  N__ < V__V
  V__V < T__
  N__ < T__
- Same distribution if we include loanwords (without predictor loanword in the model)
- Post-nasally voiced stops universally preferred

  (Hayes and Stivers 2000)
- After voiceless stops voiced stops universally dispreferred (voice disagree)
- Intervocalically, voiced stops universally preferred

Data

- *Unnatural distribution of voice feature*
  N__ < V__V
  V__V < T__
  N__ < T__
- Same distribution if we include loanwords (without predictor loanword in the model)
- Post-nasally voiced stops universally preferred

  (Hayes and Stivers 2000)
- After voiceless stops voiced stops universally dispreferred (voice disagree)
- Intervocalically, voiced stops universally preferred
- Trend in the lexicon in the **opposite direction**: favors dispreferred segments

Data

- **Stop-stop clusters**: Only two items with no voicing: [patpatya], [pikpiʃ]
- Reduplication
- bakbaNya, kutbi, lutbi, matga, mutgi, patga, pikpi_, pukba, putga, takba, tikba, tikbi, utga, utgu, witgu, xukbuN

Data

- Post-consonantally voicing after (Nazarov 2008):
  t, ʧ, ʈʂ, k, s, ʃ, x, l, lʲ, r, j, w

Pre-Tarma	Tarma	Pre-Tarma	Tarma
aʧka	aʧga	maʈʂka	maʈʂga
aʈʂpa	aʈʂba	arku	argu
arku	argu	kuʧka	kuʧga
awkis	awgis	luʃpi	luʃbi
aypa	ayba	pilpa	pilba
ʧaspu	ʧasbu	luxpi	luxbi
ʧilpi	ʧilbi	mutki	mutgi
takpa	takba	tikpa	tikba
lutpi	lutbi	tikpi	tikbi

Data

- Another locus of unnaturalness: **kb**, **tb**, **tg**
- Intervocalic stop-stop cluster: VC_1C_2V when $C_2 = [b]$ or $[g]$

TT	**TD**	**DT**	**DD**
17	69	0	0

Native only:

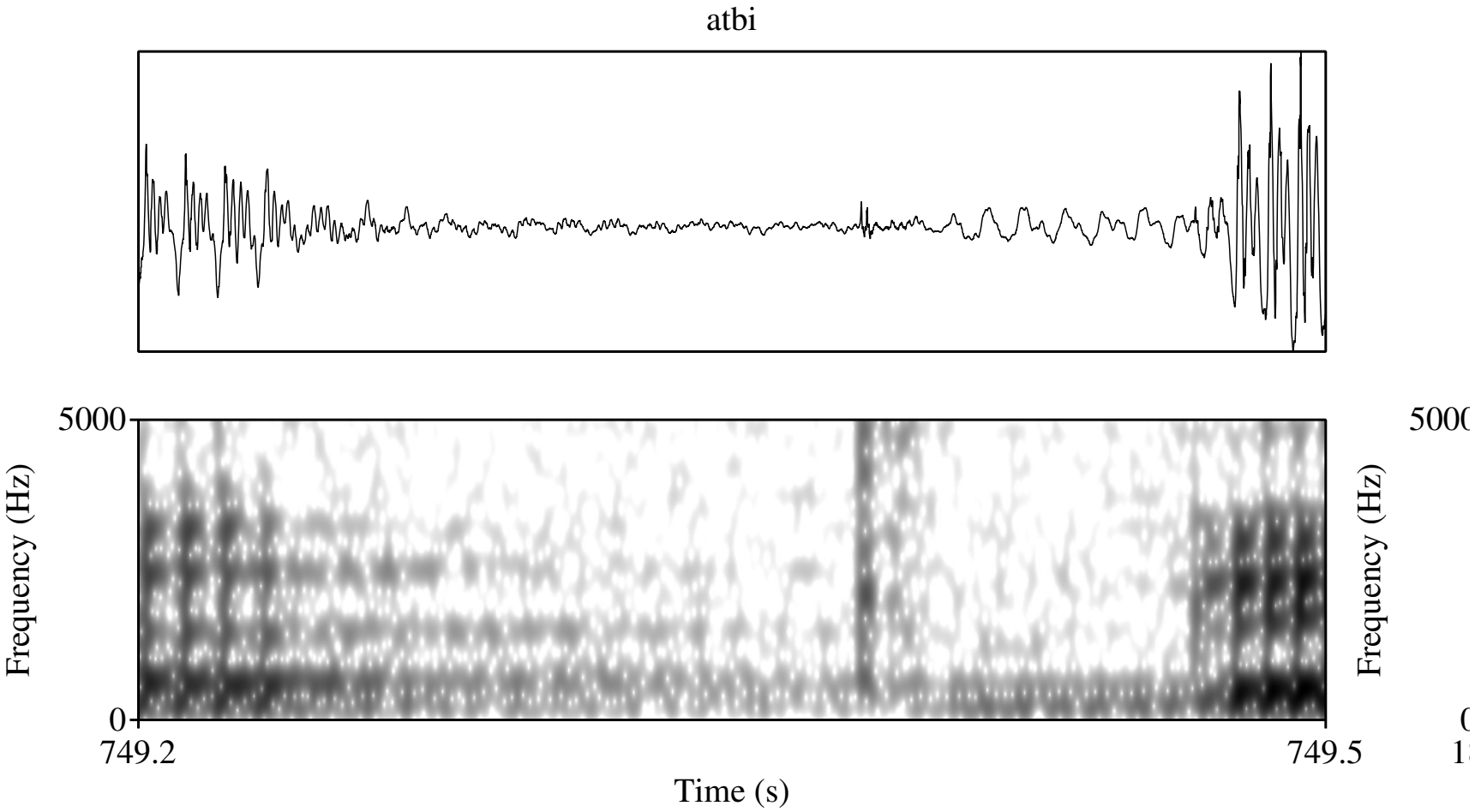
TT	**TD**	**DT**	**DD**
11	68	0	0

- Clusters that agree in voice preferred
- Pre-voicing preferred
- Tarma Quechua gradience in the **opposite direction**
- Less voice in second member clusters with first T ($\beta = 1.8$, z $= 5.6$, $p < 0.0001$)

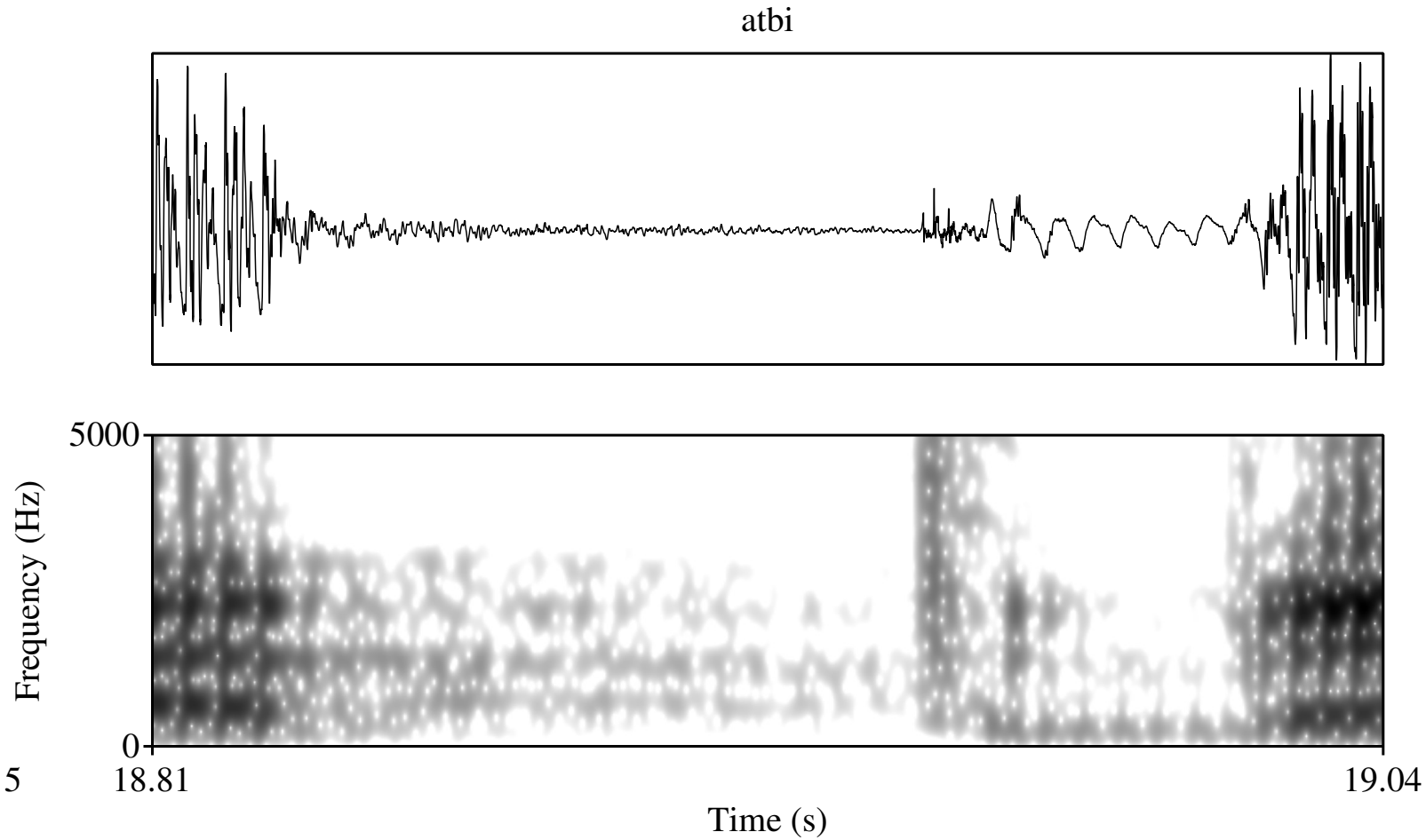
Data

- Phonetically real
- Traces of productivity

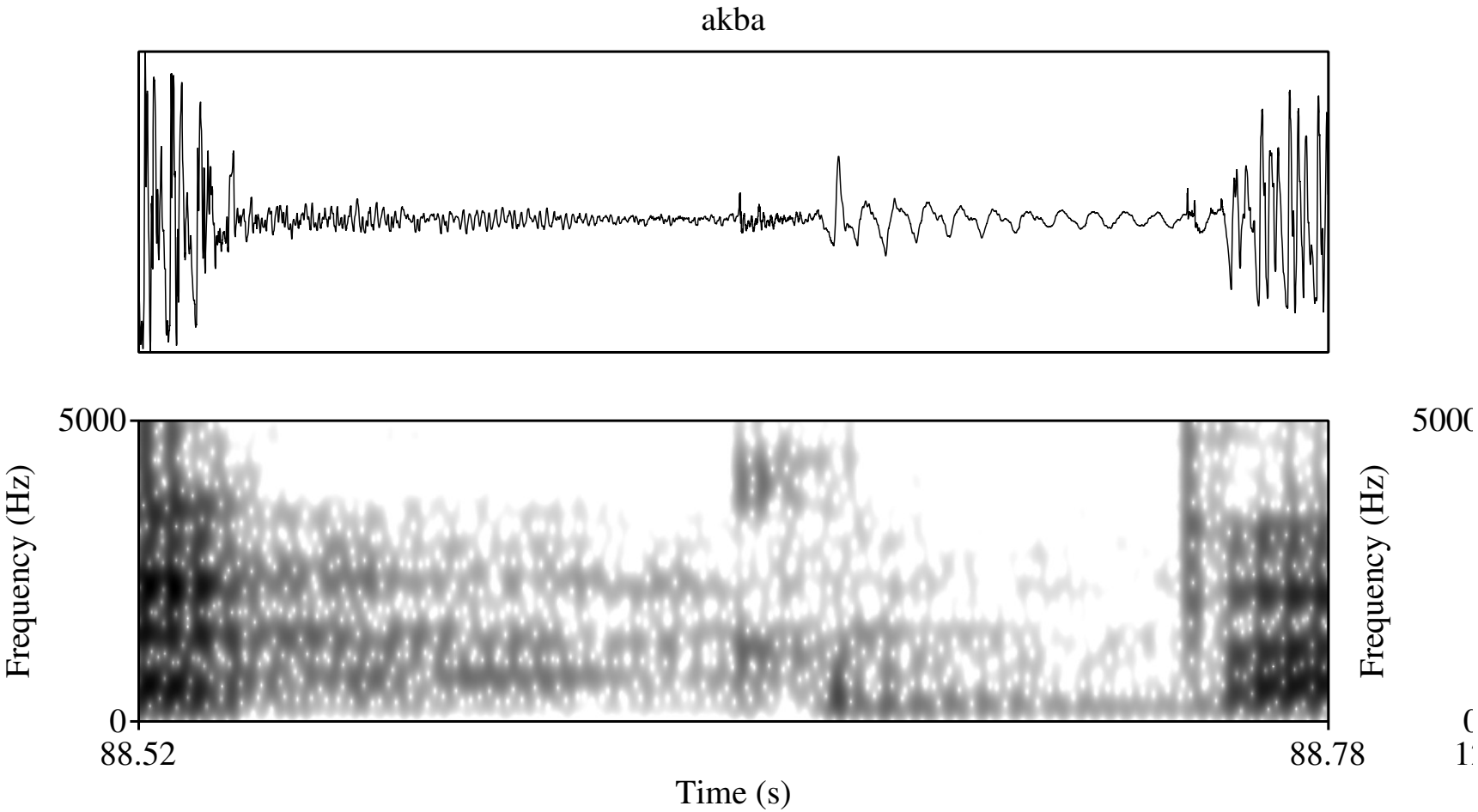
[atbi]



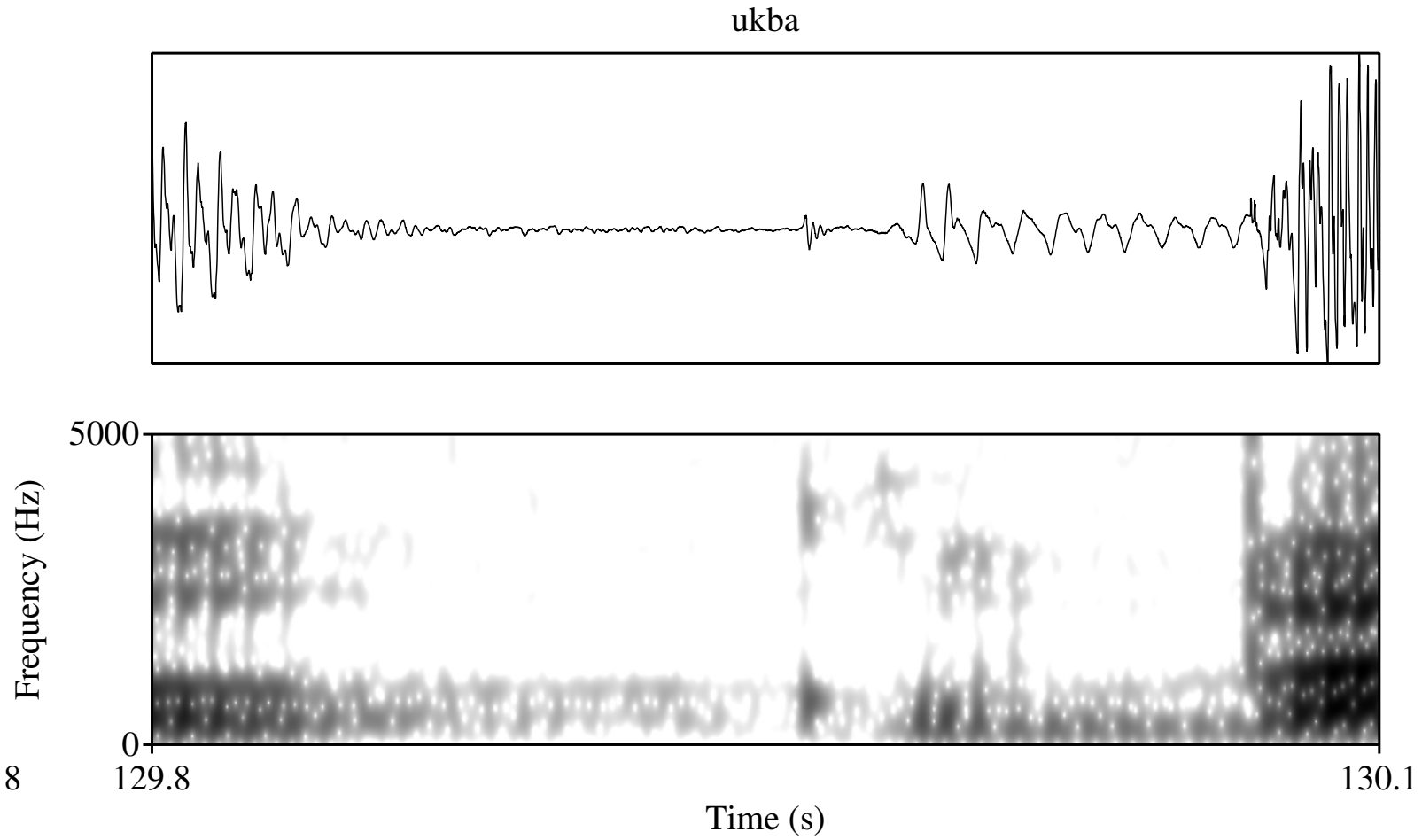
[atbi]



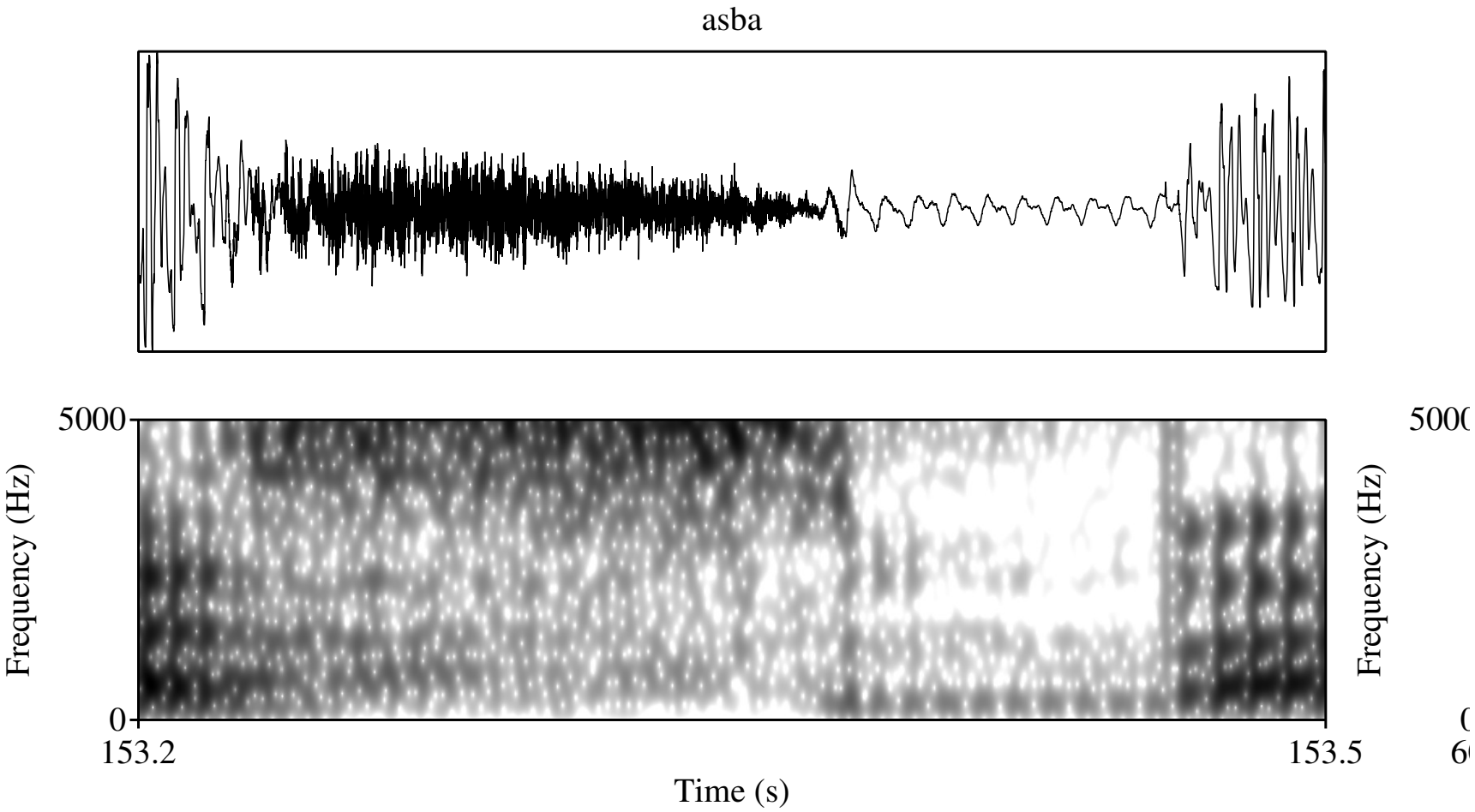
[ukba]



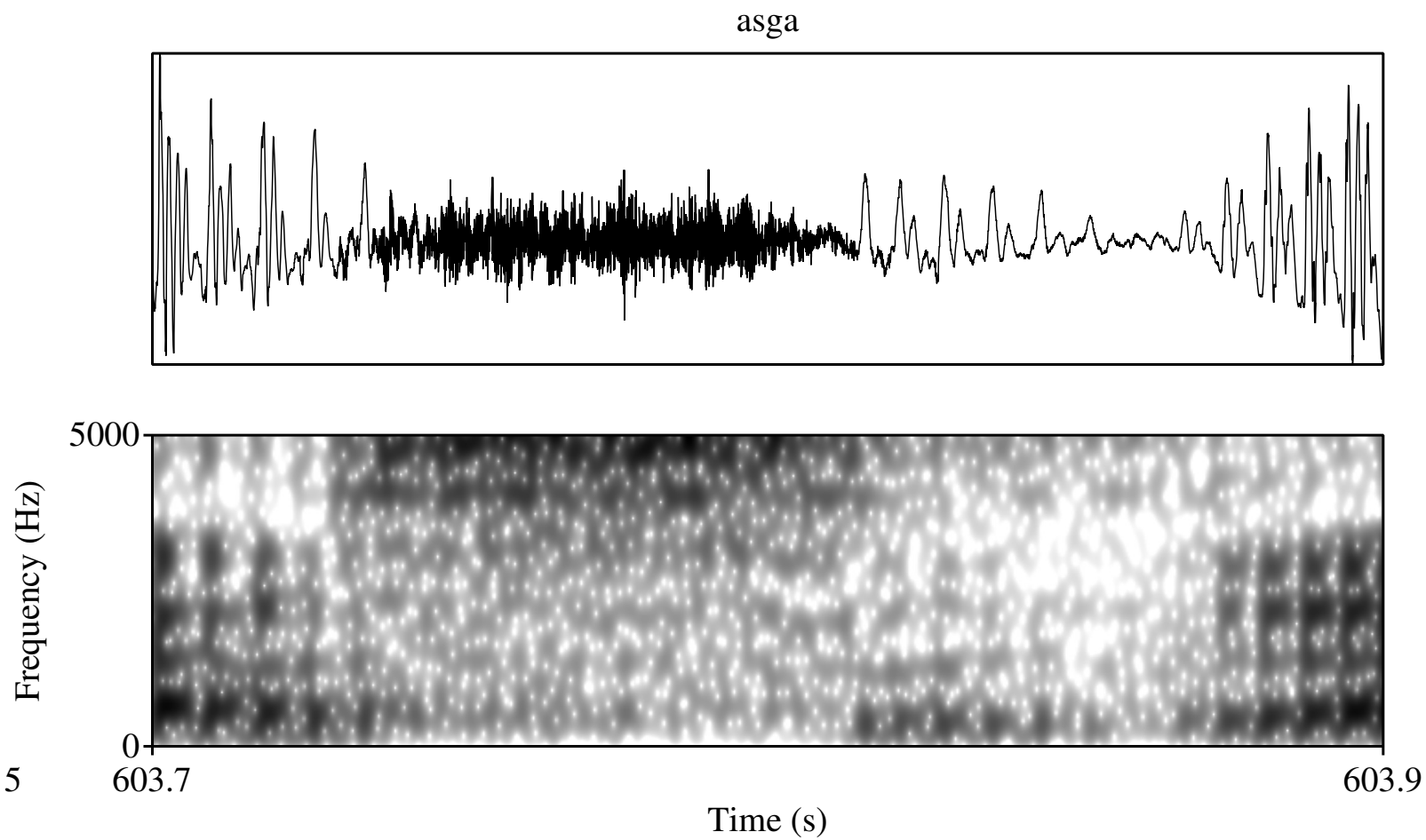
[ukba]



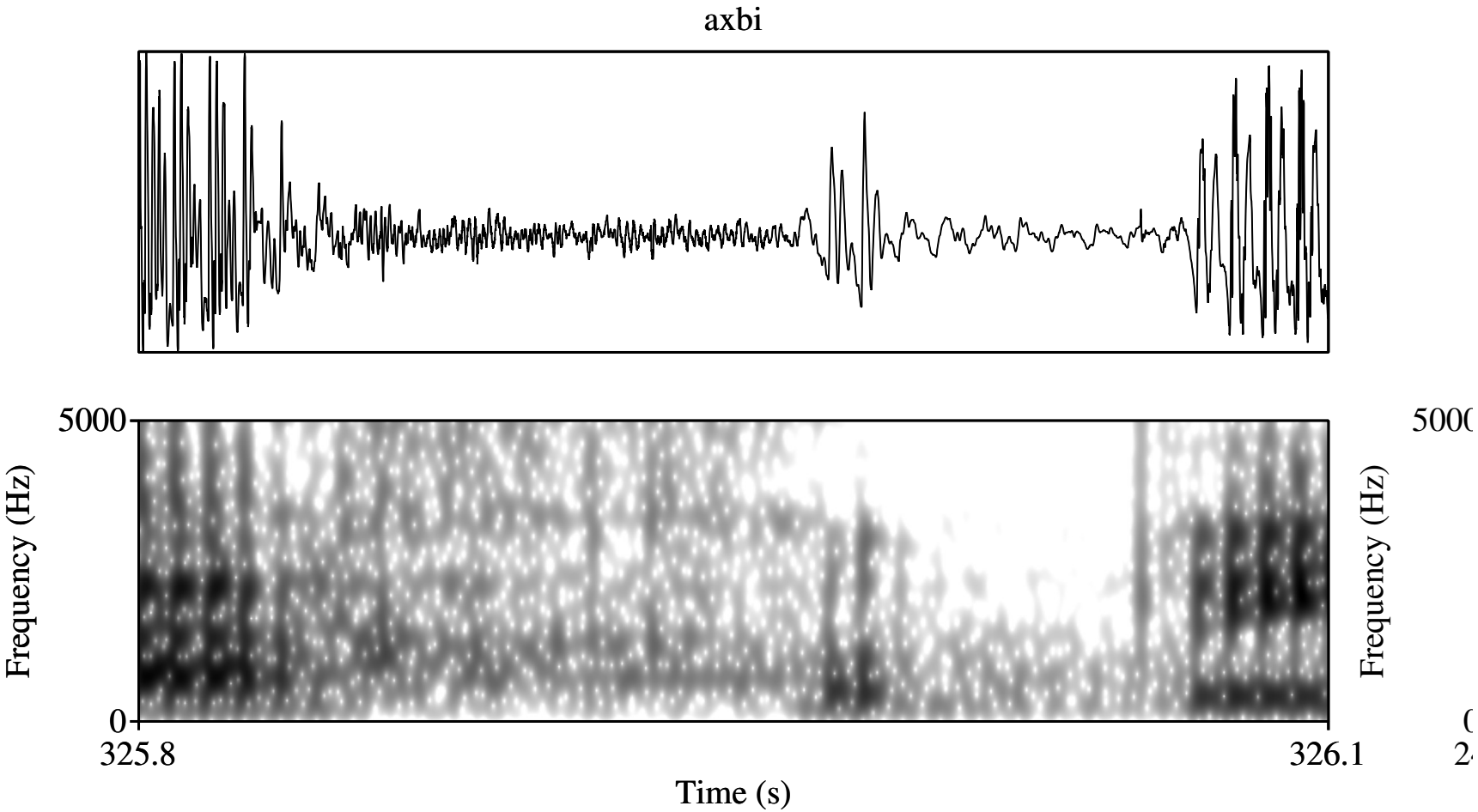
[asba]



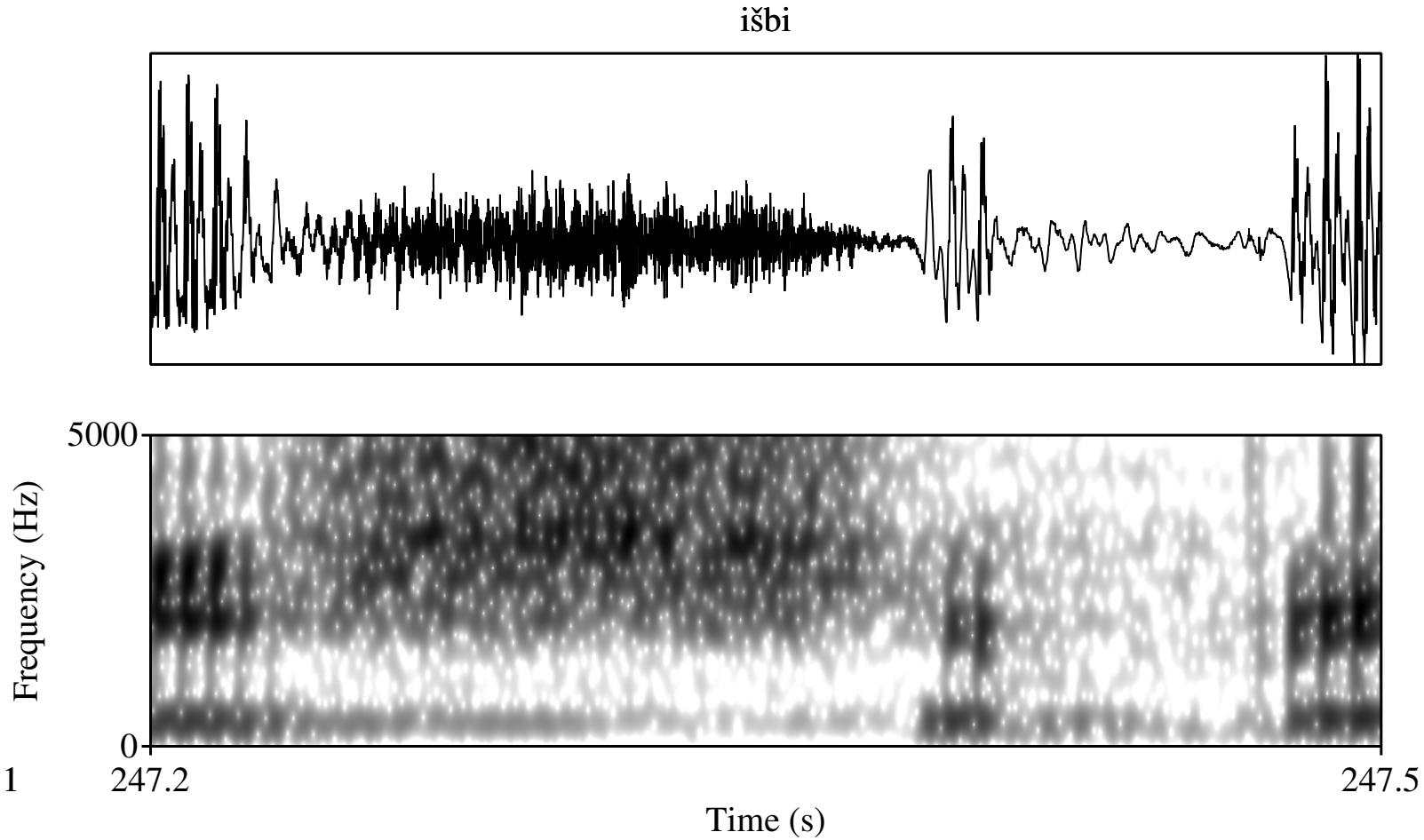
[asba]



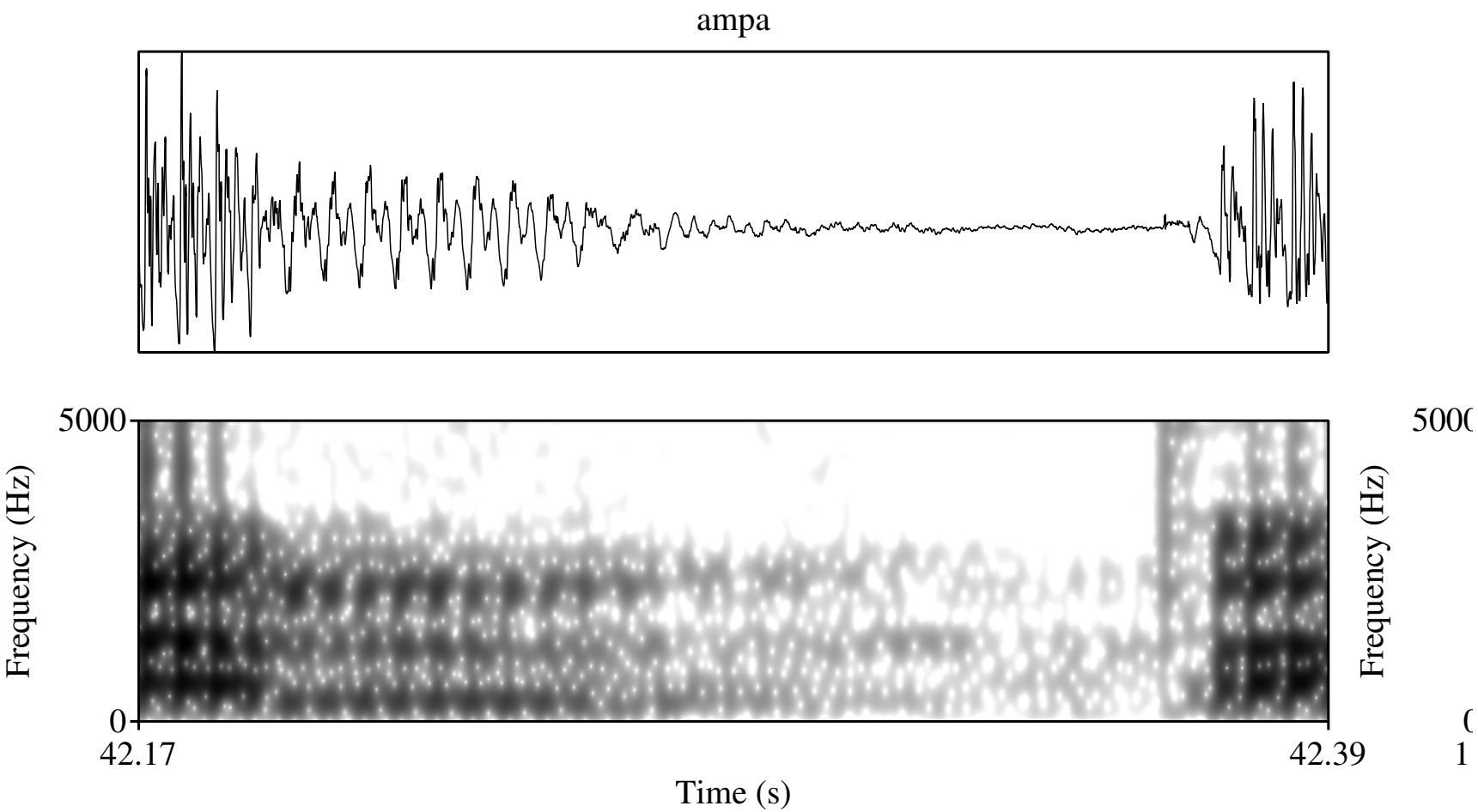
[axbi]



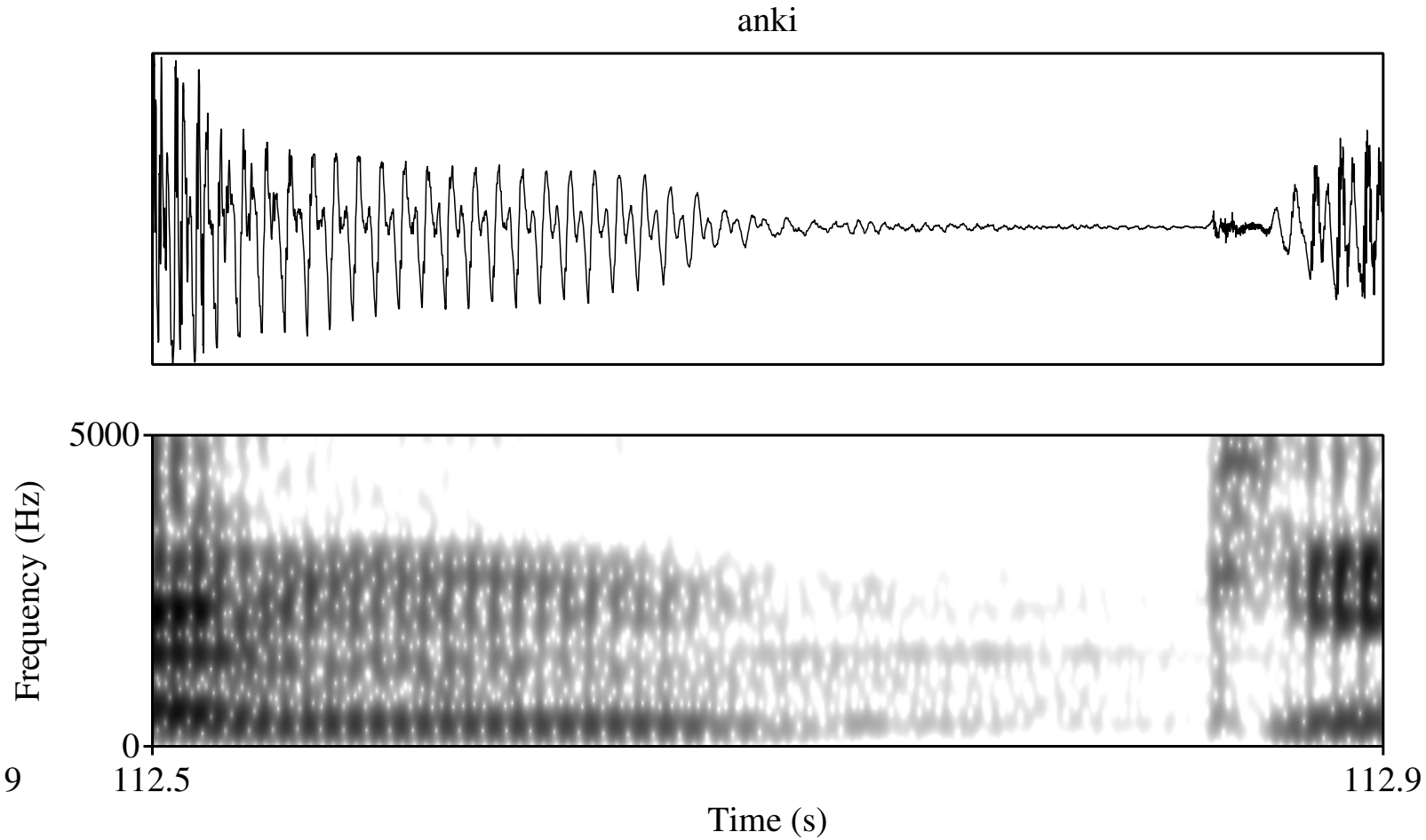
[isbi]



[ampa]



[aŋka]



Tarma

- Traces of productivity
- Loanwords

Spanish	Tarma	gloss
cuculi	kuguliː	white-winged dove
cotpe	kutbi	an animal from mount.
sauco	sawgu	magic tree
vaca	waːga	cow
taruga	taruka	deer'
dios se lo pague	yusulpaːki	thank you

Tarma

- Traces of productivity
- Loanwords

Spanish	Tarma	gloss
cuculi	kuguliː	white-winged dove
cotpe	kutbi	an animal from mount.
sauco	sawgu	magic tree
vaca	waːga	cow
taruga	taruka	deer'
dios se lo pague	yusulpaːki	thank you

- Note that devoicing was never a sound change in TQ

Tarma

- Morphophonological alternation (Creider 1968:12-13)

-ba	-pa	genitive
-bax	-pax	purposive
-bita	-pita	procedentive
-bis	-pis	'even, too'

(1) a. wawxi-gi-ba wayi-n
'the house of your brother'

b. wayi-n-pa pasa-ʃun
'we're going to walk by way of his house'

c. tamya-ya-n nuqa-ntʃik-baq
'it is raining now for us'

Tarma

- Morphophonological alternation (Creider 1968:12-13)

-ba	-pa	genitive
-bax	-pax	purposive
-bita	-pita	procedentive
-bis	-pis	'even, too'

(1) a. wawxi-gi-ba wayi-n
'the house of your brother'

b. wayi-n-pa pasa-ʃun
'we're going to walk by way of his house'

c. tamya-ya-n nuqa-ntʃik-baq
'it is raining now for us'

Tarma

- Morphophonological alternation (Creider 1968:12-13)

-ba	-pa	genitive
-bax	-pax	purposive
-bita	-pita	procedentive
-bis	-pis	'even, too'

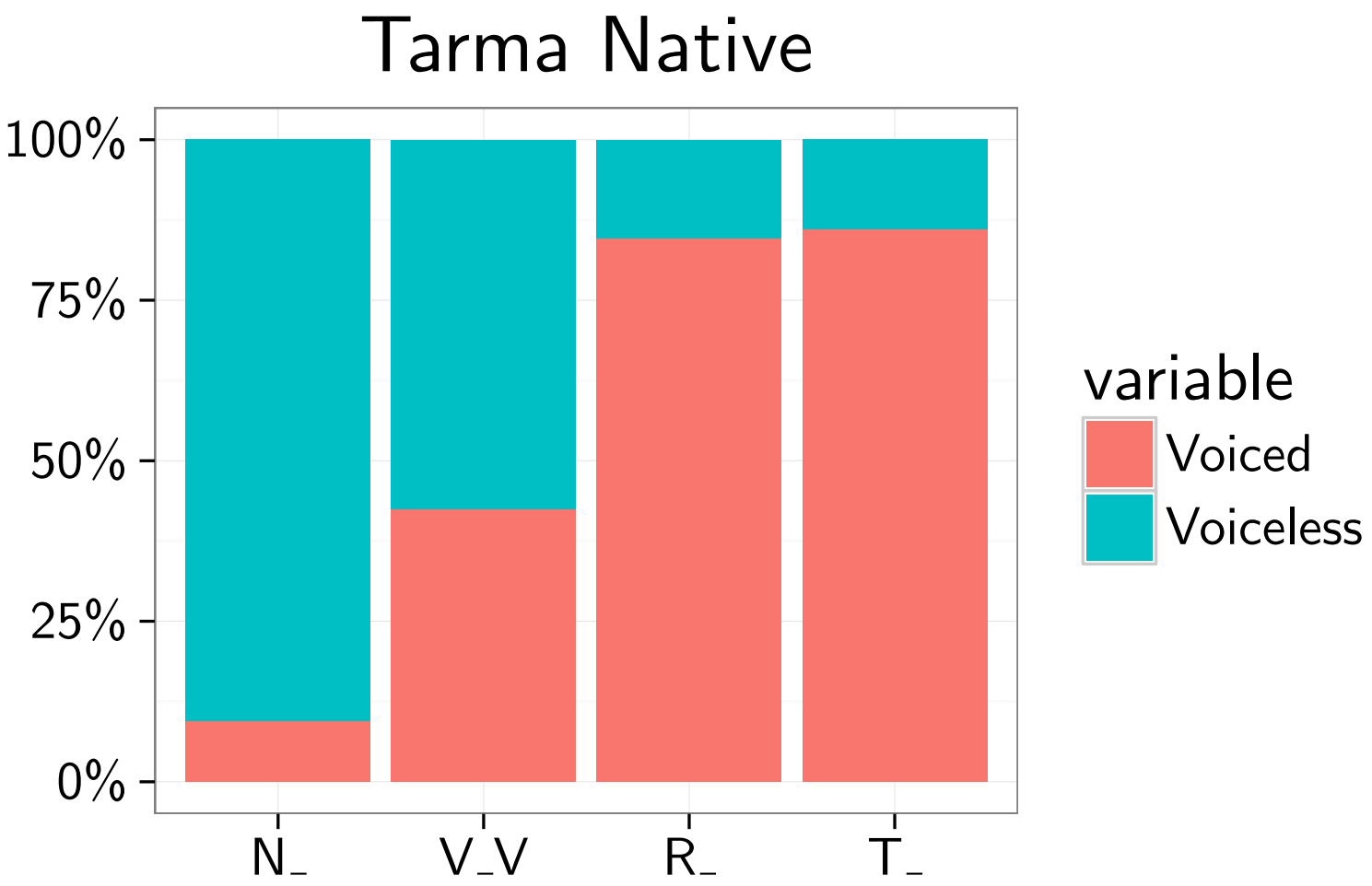
(1) a. wawxi-gi-ba wayi-n
'the house of your brother'

b. wayi-n-pa pasa-ʃun
'we're going to walk by way of his house'

c. tamya-ya-n nuqa-ntʃik-baq
'it is raining now for us'

- Limited to some suffixes: -guna, -bura, -gama, and -gasga show no alternation

Interim summary



Interim summary

- Significant unnatural trend in the lexicon
- Unnatural gradient phonotactic restriction
- Traces of productivity in morphophonology

Berawan

- Restriction against [+voice] in V__V
- Significant trend in the lexicon: full contrast in [voice] initially, intervocalically the natural element dispreferred
- Voicing intervocalically (Westbury and Keating 1986)

Intervocalic Devoicing

- Intervocalic (or postvocalic) **voicing** is common

$$T > D \ / \ V\_\_(V)$$

- Articulation

Keating (1986)

Kaplan (2010), Steriade (2010)

- Very common sound change
- The most frequent type of lenition (alongside spirantization, approximatization and others)

Kaplan (2010)

- 26 of 153 languages have intervocalic lenition

Gurevich (2004), Kaplan (2010)

Berawan

- Analysis of the lexicon in Burkhardt (2014)
- Counts of occurrence: initial, intervocalic

Berawan

- Labial

	voiceless		voiced	
dialect	#__	V__V	#__	V__V
Batu Belah	52	10	40	2
Long Teru	46	13	38	2
Long Jegan	49	10	40	2
Long Terawan	41	11	48	1

Berawan

- Alveolar

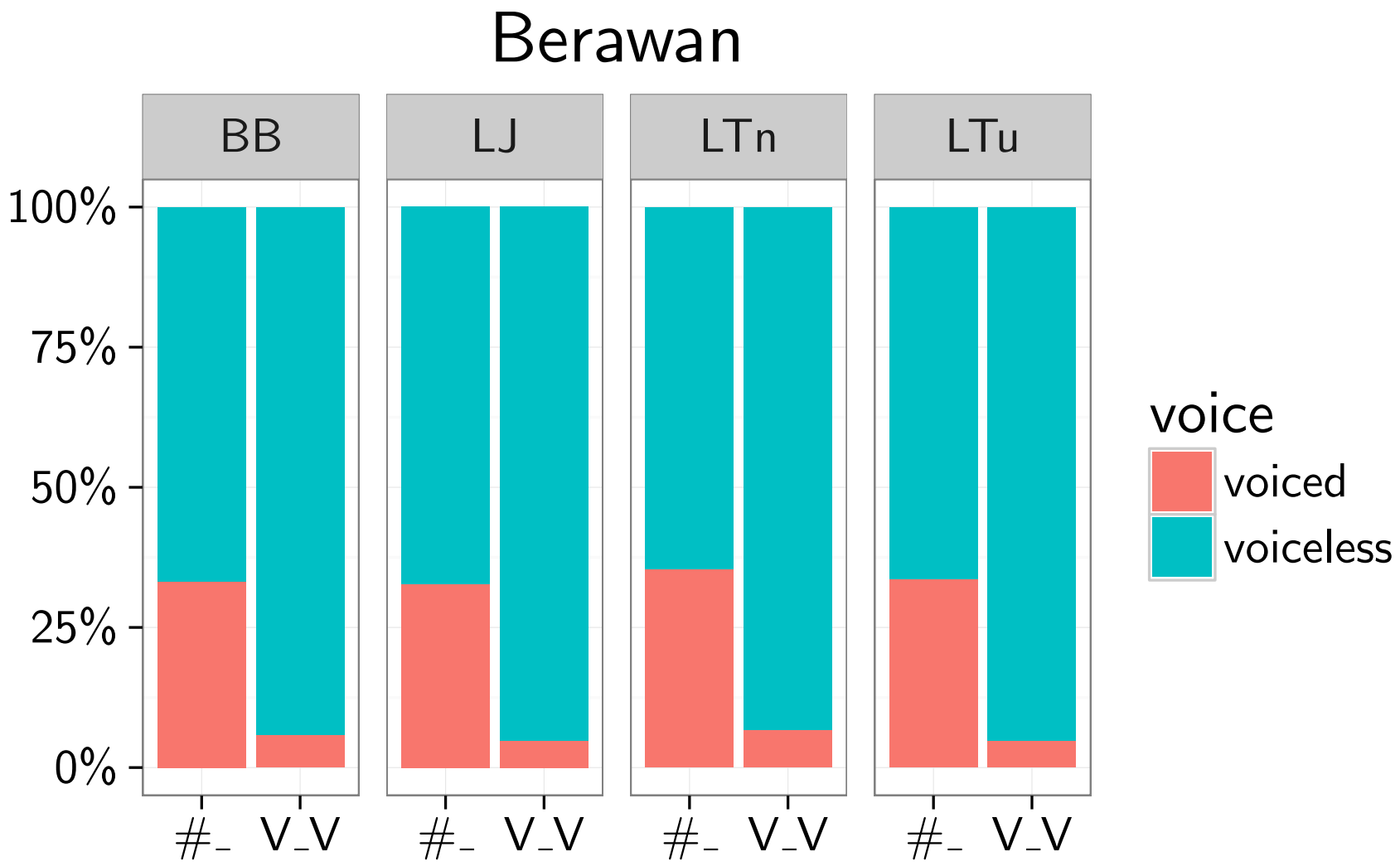
	voiceless		voiced	
dialect	#__	V__V	#__	V__V
Batu Belah	56	32	22	4
Long Teru	54	31	22	2
Long Jegan	55	32	22	3
Long Terawan	60	25	21	3

Berawan

- Velar

	voiceless		voiced	
dialect	#__	V__V	#__	V__V
Batu Belah	43	54	13	0
Long Teru	40	55	11	1
Long Jegan	44	58	10	0
Long Terawan	50	19	14	0

Data



Data

- Data was fit to a logistic regression model in R (R Core Team) with `glm()` and `brglm()` from `brglm` package
- DV: Voice (binomial)
- IV: Context (treatment-coded with initial as the reference level) and Place (treatment-coded with velar as reference level) of articulation with no interactions (native vocabulary)

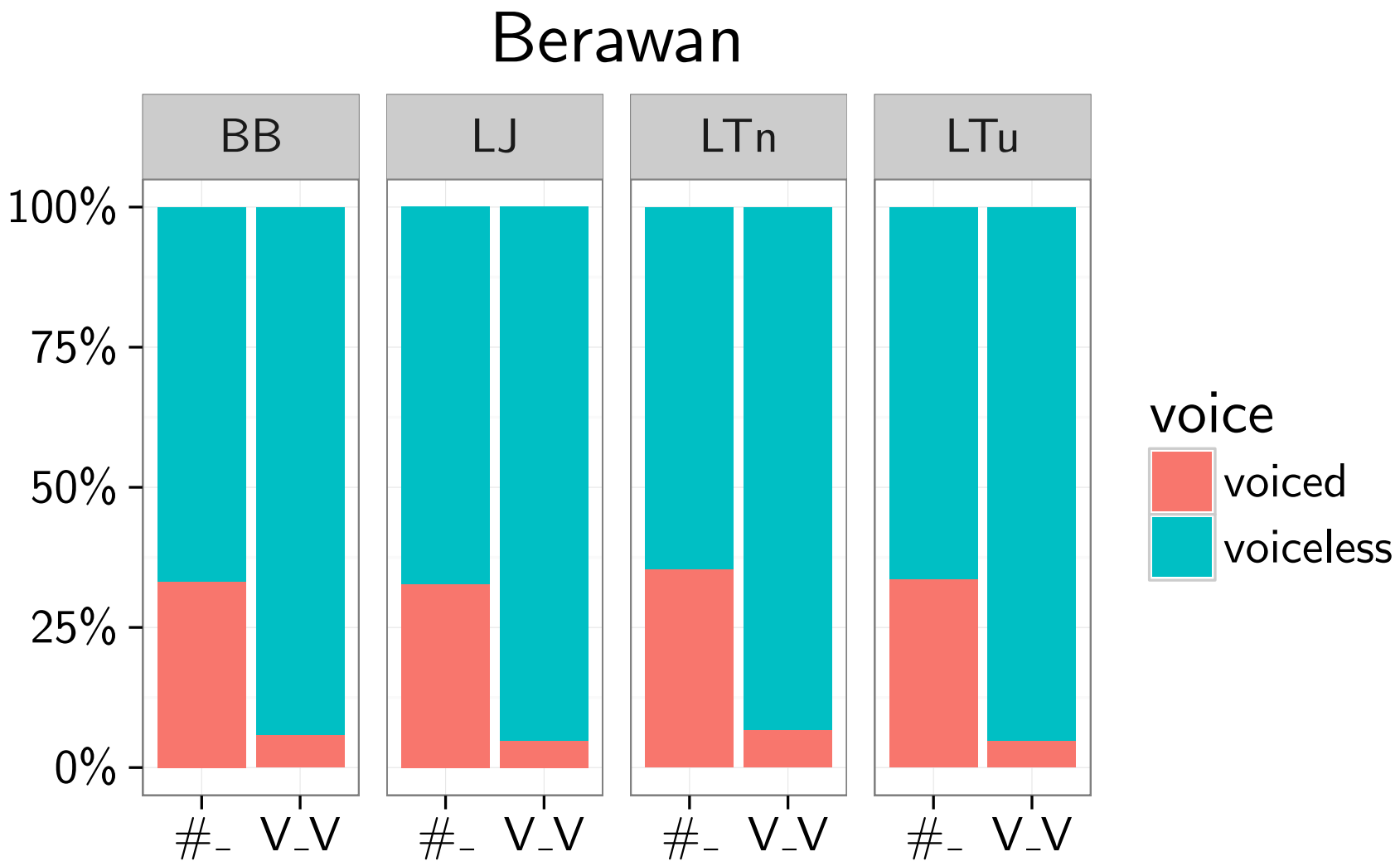
Data

	Est.	z score	Pr(>|z|)
BB	-1.773	-3.906	0.0001
LTu	-2.028	-4.139	0.0000
LJ	-1.898	-3.860	0.0001
LTn	-1.893	-3.479	0.0005

Data

- Productivity: loanwords
- Brunei Malay [pəsigupan] > Pre-Berawan *səgupan > BB [səkupen]
- Brunei Malay sigup > Pre-Berawan *sigup > BB [sigup]

Data



Data

- Significant unnatural trend in the lexicon
- Unnatural gradient phonotactic restriction against voiced stops intervocalically

Data

- How did Berawan and Tarma Quechua phonotactics arise?

Outline

Models

- Blurring process
- A > B / X = UPT
- How do we get B > A / X
- Can sound change be unnatural (Blust 2005): Berawan data results from a single sound change
- BLURRING CYCLE
  B > C / −X
  B > A
  C > B

Models

- Blurring process
- A > B / X = UPT
- How do we get B > A / X
- Can sound change be unnatural (Blust 2005): Berawan data results from a single sound change
-

BLURRING CYCLE	BLURRING CHAIN
B > C / −X	B > C / X
B > A	C > D
C > B	D > A

Blurring Process

- Post-nasal devoicing (Dickens 1984), Hyman (2001)
- In all twelve cases in which PND is reported as a sound change or synchronic alternation, it arises though a combination of 3 sound changes
-

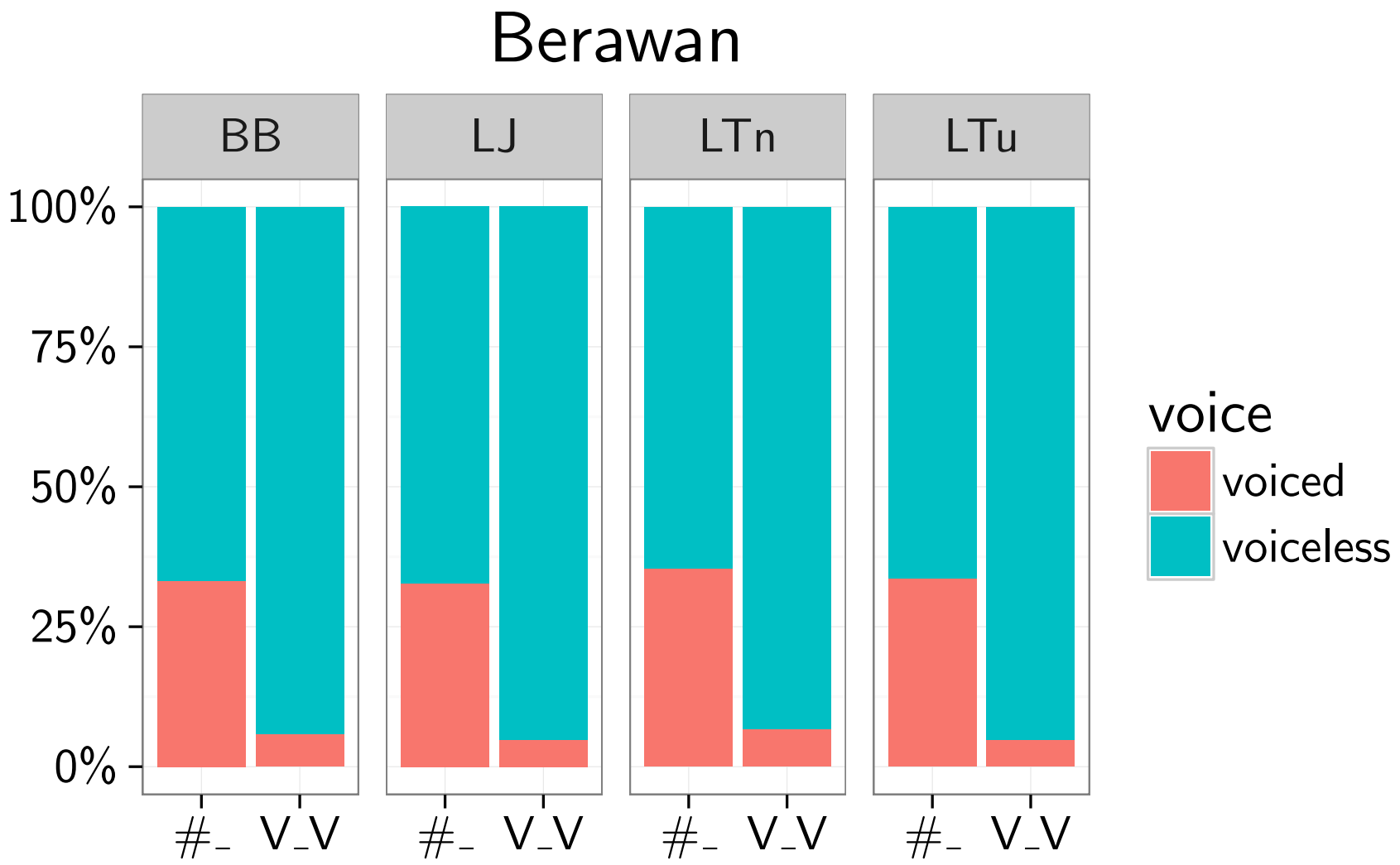
Blurring Cycle		Example		
		Avestan	band	dasa
D > Z / [−nas]	d > ð/[−nas]__	Sogdian	βand	ðasa
D > T	d > t	Yaghnobi	vant	*ðasa
Z > D	ð > d	Yaghnobi	vant	das

Blurring Process

- Blurring Process
- A historical model for explaining unnatural phenomena
- Addition to current strategies (hypercorrection)
  (a) A set of segments enters complementary distribution
  (b) A sound change occurs that operates on the changed/unchanged subset of those segments
  (c) Another sound change occurs that blurs the original complementary distribution

Outline

Berawan



Berawan dialects

- **g (from ʀ) > k / V__V**

*bigiu	>	bikiw
*gigiəq	>	giki‿
*magi	>	maki
*igiəŋ	>	ikiŋ
*ugat	>	ikit

Berawan dialects

- **g (from ʀ) > k / V__V**

*bigiu	>	bikiw
*gigiəq	>	giki ʔ
*magi	>	maki
*igiəŋ	>	ikiŋ
*ugat	>	ikit

Berawan dialects

- **g (from ʀ) > k / V__V**

*bigiu	>	bikiw
*gigiəq	>	giki ʔ
*magi	>	maki
*igiəŋ	>	ikiŋ
*ugat	>	ikit

- **b > k / V__V**

*abiəŋ	>	akiŋ
*bibi	>	biki
*bəlibiəw	>	bəlikiəw
*bibuy	>	bikuy
*dibiən	>	dikin

Berawan dialects

- **g (from ʀ) > k / V__V**

*bigiu	>	bikiw
*gigiəq	>	giki ʔ
*magi	>	maki
*igiəŋ	>	ikiŋ
*ugat	>	ikit

- **b > k / V__V**

*abiəŋ	>	akiŋ
*bibi	>	biki
*bəlibiəw	>	bəlikiəw
*bibuy	>	bikuy
*dibiən	>	dikin

Long Terawan

- Word-initially, stops remain voiced

*gəm	>	gəm
*gigun	>	gikuŋ
*gimot	>	gimok
*gitaq	>	gitaʔ
*bitok	>	bitok
*buliən	>	bulin
*busak	>	busek
*buttan	>	buten

Long Terawan

- Word-initially, stops remain voiced

*gəm	>	gəm
*gigun	>	gikuŋ
*gimot	>	gimok
*gitaq	>	gitaʔ
*bitok	>	bitok
*buliən	>	bulin
*busak	>	busek
*buttan	>	buten

Crucial problem

- ***b > k / V__V**

	Scenario 1	Scenario 2
1.	b > p	b > g
2.	p > k	g > k

- Problem: if devoicing occurs before change of place, why no change of the original *p?

Crucial problem

- ***b > k / V__V**

	Scenario 1	Scenario 2
1.	b > p	b > g
2.	p > k	g > k

- Problem: if devoicing occurs before change of place, why no change of the original *p?

*hapuy > apuy, not **akuy

Crucial problem

- ***b > k / V__V**

	Scenario 1	Scenario 2
1.	b > p	b > g
2.	p > k	g > k

- Problem: if devoicing occurs before change of place, why no change of the original *p?

  *hapuy > apuy, not **akuy

- If change of place first, why only word-internally?

Berawan dialects (Blust 2005)

- Berawan stops

D > T / V___V

	#__	**V__V**
*b	b	k
*d	d	r
*g	g	k

Berawan dialects (Blust 2005)

- Berawan stops

D > T / V___V

	#__	**V__V**
*b	b	k
*d	d	r
*g	g	k

Berawan dialects (Blust 2005)

- Berawan stops

D > T / V___V

	#__	**V__V**
*b	b	k
*d	d	r
*g	g	k

Berawan dialects (Blust 2005)

- Blurring chain

BLURRING CHAIN	Berawan
B > C / X	D > Z / V__V
C > D	Z > S
D > A	S > T

Berawan dialects

- Berawan stops: stage 1

Pre-Berawan	Berawan	
	#_	V_V
b	b	β
d	d	ð
g	g	ɣ

Berawan dialects

- Berawan stops: stage 1a

Pre-Berawan	Berawan	
	#_	V_V
b	b	β
d	d	r
g	g	ɣ

Berawan dialects

- Berawan stops: stage 2

Pre-Berawan	Berawan	
	#_	V_V
b	b	ϕ
d	d	r
g	g	x

Berawan dialects

- Berawan stops: stage 3

Pre-Berawan	Berawan	
	#_	V_V
b	b	x
d	d	r
g	g	x

Berawan dialects

- Berawan stops: stage 4

Pre-Berawan	Berawan	
	#_	V_V
b	b	k
d	d	r
g	g	k

Phonetics

- *babuj > *bibuj > *biϕuj > *bixuj > [bikuj]
- Explains dissymmetry: alveolar — labial, velar
- Solves the chronology problem

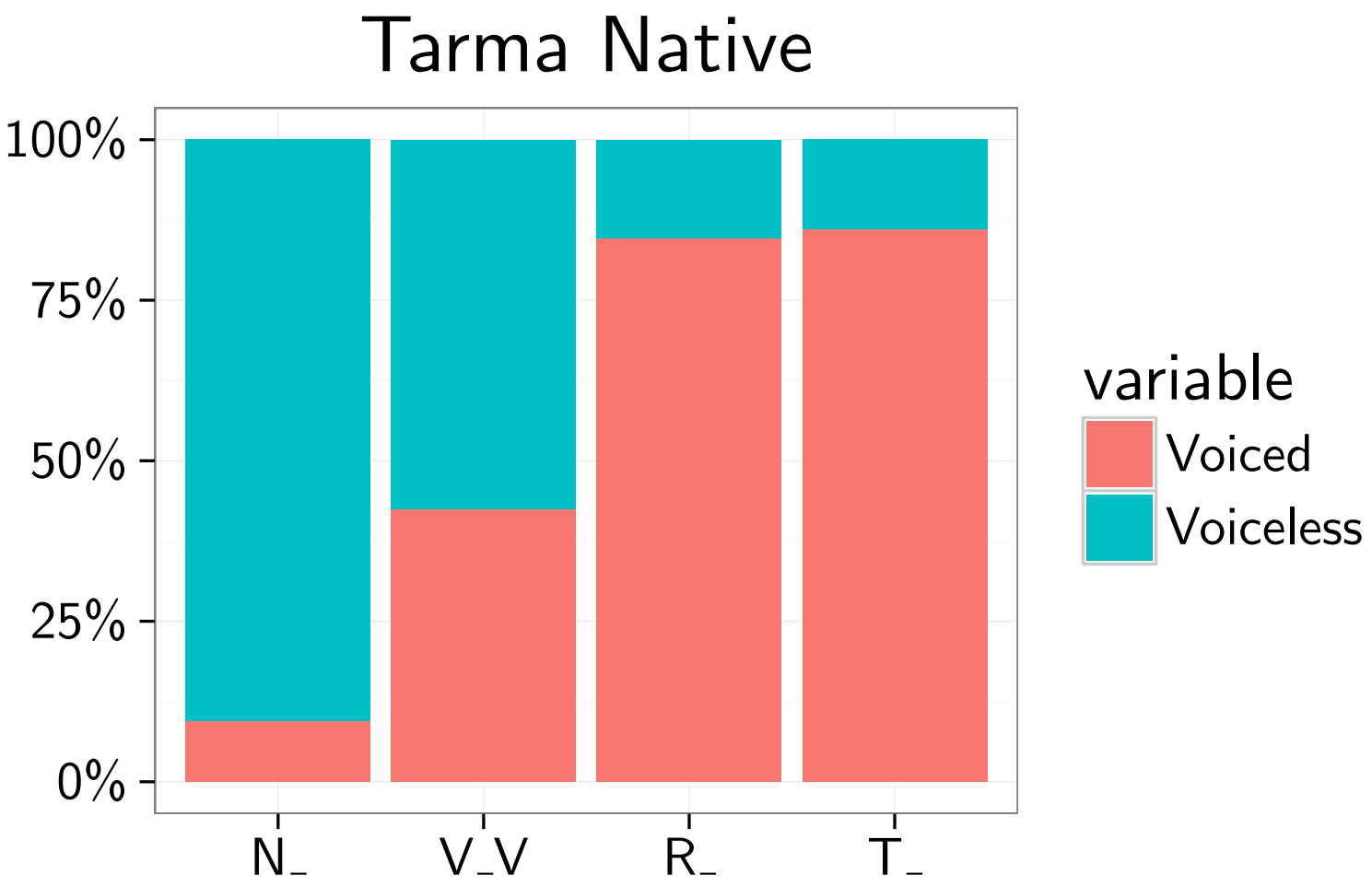
	Scenario 1	**Scenario 2**
1.	b > p	b > g
2.	p > k	g > k

- Better motivates change of place: [ϕ] > [x] easier to motivate than [b] > [g]
- Natural sound changes

Explanation

BLURRING CHAIN	Tarma Quechua
B > C / X	T > S / [−nas,−#]__[−cons]
C > D	S > Z
D > A	Z > D

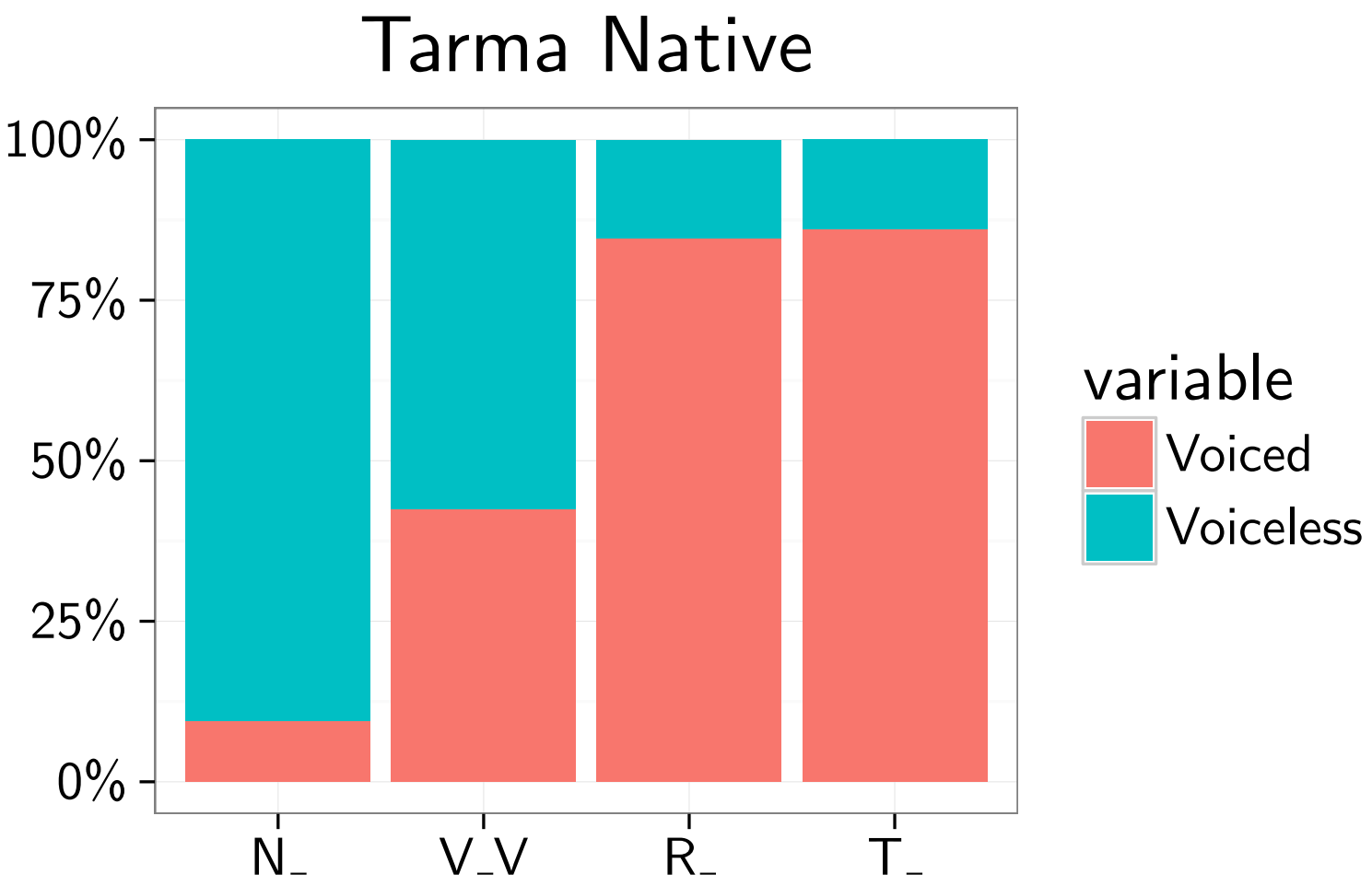
Explanation



Explanation

- T > S / [−nas,−#]__[−cons]
  - *aptay > [hax^{w}tay]
  - *upyay > [uxyay]
  - *paki > *p^{h}aki > Imbabura Quechua [faki]
  - Voiced stops still surface as fricatives
- S > Z / V __
  - Voicing of fricatives prevocalically is a common change
  - *cops amagats* [bz]
  - <uxδa-> [uxða-] < *uxθa- < *ukt^{h}a-
- Z > D
  - Not categorical: voiced stops still surface as fricatives
  - Variation in voiceless series: [ɸlawta], [plauta] (Puente Baldoceda 1977)

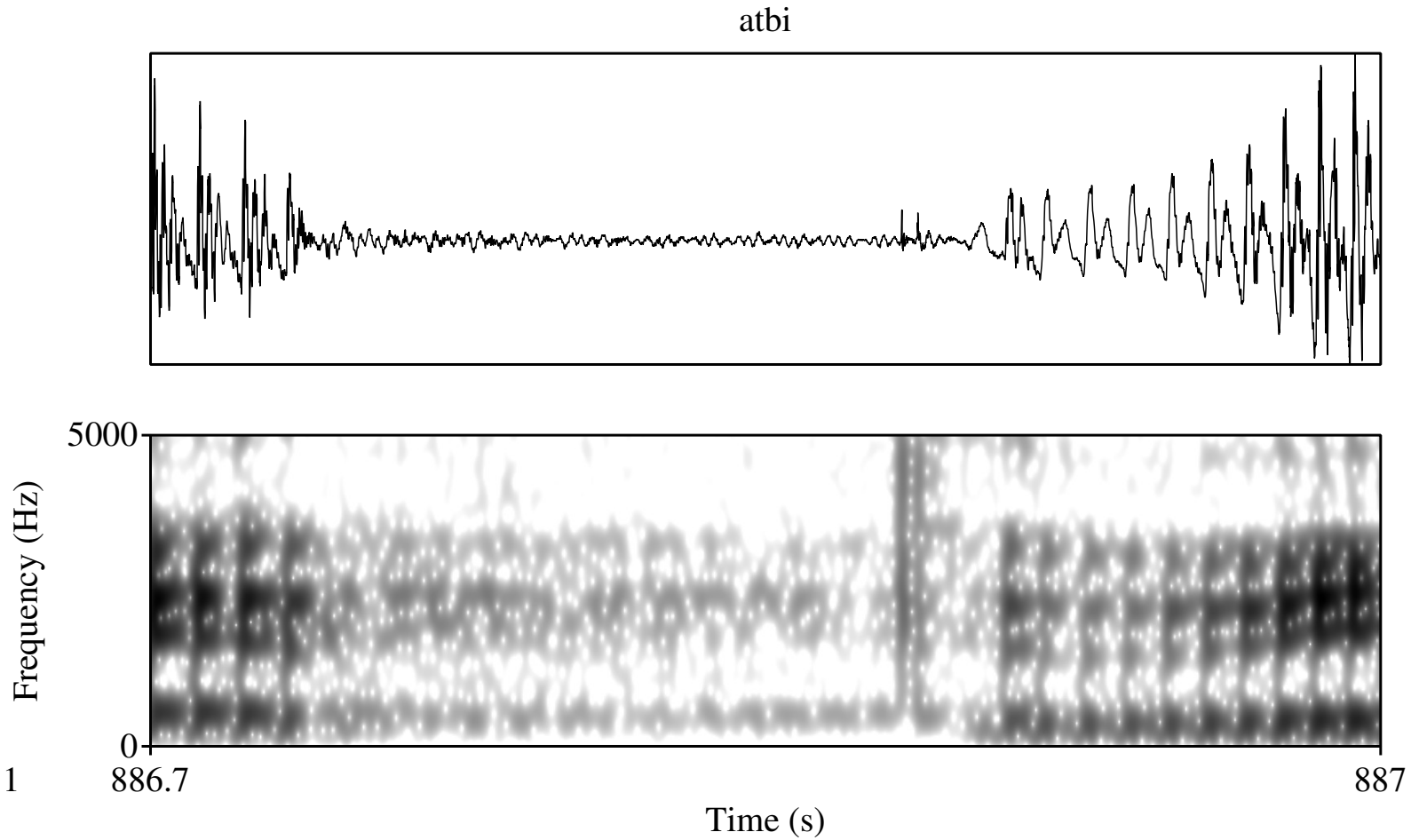
Explanation



Phonetics

- Stops surface as fricatives (apparent free variation)

[atbi]



Phonetics

- This strongly suggests that there was a stage in the development with **voiced fricatives**
- Occlusion to stops not operated categorically

Implications of the new historical explanation

- One of rare reported cases of unnatural sound change explained as a combination of sound changes
- Allows us to maintain the long-held position that sound change is always natural
- Alternative historical device for explaining unnatural sound changes (cf. hypercorrection, Ohala 1993)
- Typology within Channel Bias

Outline

Future work

- Where to go from here?
- *Harmony* (HG) $\rightarrow$ probability, but given richness of the base, we cannot derive a system in which the unnatural element is more frequent

Future work

- Where to go from here?
- *Harmony* (HG) $\rightarrow$ probability, but given richness of the base, we cannot derive a system in which the unnatural element is more frequent
- Grounds for experimental data

Conclusion

- Gradient phonotactic restrictions *can* be unnatural
- Significant unnatural trends in the lexicon with traces of productivity
- A new historical device to explain unnatural phenomena (blurring process)

References

Adelaar, Willem F. H. 1977. *Tarma Quechua: Grammar, texts, dictionary.* Lisse: The Peter de Ridder Press.

Adelaar, Willem F. H., and Muyske, Pieter C. 2004. *The Languages of the Andes*. Cambridge: CUP.

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26(1): 9-41.

Anttila, Arto. 2008. Gradient phonotactics and the Complexity Hypothesis. *Natural Language & Linguistic Theory*, 26(4), 695-729.

Beguš, Gašper. 2015a. Post-nasal devoicing as a sound change. Presentation at the 89th Annual LSA Meeting, Portland, OR.

———. 2015b. Intervocalic Devoicing in Kiput and Berawan Dialects. Presentation at the 22nd AFLA, McGill University.

Blevins, Juliette. 2004. *Evolutionary Phonology.* Cambridge: CUP.

———. 2008. Natural and unnatural sound patterns : a pocket field guide. In Klaas Willems and Ludovic De Cuypere (eds.) *Naturalness and iconicity in language*. 121–148. Amsterdam: Benjamins.

References

Blust, Robert. 1974. A Murik vocabulary, with a note on the linguistic position of Murik. *The Sarawak Museum Journal* 22.43:153–189.

———. 2005. Must sound change be linguistically motivated? *Diachronica* 22 (2): 219–269.

———. 2013. *The Austronesian Languages*. Canberra: Asia-Pacific Linguistics.

Burkhardt, Jürgen M. The reconstruction of the phonology of Proto-Berawan. PhD Thesis, Johann-Wolfgang-Goethe-Universität zu Frankfurt am Main.

Coetzee, Andries W., Susan Lin, and Rigardt Pretorius. 2007. Post-Nasal Devoicing in Tswana. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 861–4.

Dickens, Patrick J. 1984. The History of So-Called Strengthening in Tswana. *Journal of African Languages and Linguistics* 6:97–125.

Frisch, Stefan, Janet Pierrehumbert and Michael Broe. 2004. Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22. 179-228.

References

Hayes, Bruce, and Tanya Stivers. 2000. *Postnasal Voicing*. Unpublished ms.

Hyman, Larry M. 2001. The Limits of Phonetic Determinism in Phonology: *NC Revisited. In *The Role of Speech Perception in Phonology*, edited by Elizabeth Hume and Keith Johnson, 141–186. San Diego, CA: Academic Press.

Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic inquiry* 44(1): 45-75.

Hyman, Larry M. 2001. The Limits of Phonetic Determinism in Phonology: *NC Revisited. In *The Role of Speech Perception in Phonology*, ed. by Elizabeth Hume and Keith Johnson, 141–186. San Diego, CA: Academic Press.

Kümmel, Martin. 2007. *Konsonantenwandel*. Wiesbaden: Reichert.

Kaplan, Abby. 2010. *Phonology Shaped by Phonetics: The Case of Intervocalic Lenition.* PhD diss., University of California, Santa Cruz.

Locke, John. 1983. *Phonological Acquisition and Change*. New York: Academic Press.

Maria Gouskova, Elisabeth Zsiga, and One Tlale Boyer. 2011. Grounded constraints and the consonants of Setswana. *Lingua* 121 (15): 2120–2152.

References

Mills, Frederick R. 1975. *Proto South Sulawesi and Proto Austronesian Phonology*. PhD diss., Department of Linguistics, University of Michigan, Ann Arbor.

Nazarov, Aleksei. 2008. *Stop voicing in Tarma Quechua.* Ms., Leiden University.

Ohala, John J. 1993. The phonetics of sound change. In *Historical Linguistics: Problems and Perspectives,* ed. by Charles Jones. 237–278. London: Longman.

Rohlfs, Gerhard. 1949. *Lautlehre*. Vol. 1 of *Historische Grammatik der Italienischen Sprache und ihrer Mundarten*. Bern: Francke.

Sirk, Ülo. 1983. *The Buginese Language*. Moscow: Nauka.

Shiraishi, Hidetoshi. *Topics in Nivkh Phonology*. PhD Dissertation, Rijksuniversiteit Groningen.

Solé, Maria-Josep. 2012. Natural and unnatural patterns of sound change? In *The Initiation of Sound Change: Perception, Production, and Social Factors*, 123–145. Amsterdam Studies in the Theory and History of Linguistic Science, Series IV, Current Issues in Linguistic Theory. Amsterdam: John Benjamins.

References

Solé, Maria-Josep, Larry M. Hyman, and Kemmonye C. Monaka. 2010. More on Post-Nasal Devoicing: The Case of Shekgalagari. *Journal of Phonetics* 38 (4): 299–319.

Xromov, Al'bert. 1972. *Jagnobskij jazyk*. Moskva: Nauka.

———. 1987. Jagnobskij jazyk. In *Osnovy iranskogo jazykoznanija. Novoiranskie jazyki: vosto?naja gruppa*, 644–701. Moskva: Nauka.

Yu, A. C. L. 2004. Explaining final obstruent voicing in Lezgian: phonetics and history. *Language*, 80(1), 73-97.

Zsiga, Elizabeth, Maria Gouskova, and One Tlale. 2006. On the Status of Voiced Stops in Tswana: Against *ND. In *Proceedings of the 36th Annual Meeting of the North East Linguistic Society (NELS)*, 721–734. Amherst, MA: GLSA.

Thank you!

* I would like to thank Kevin Ryan, Aleksei Nazarov, Adam Albright, and Donca Steriade for their useful comments and Willem Adelaar for his kind permission to analyze his recordings. Special thanks goes also to Mind Brain Behavior interfaculty initiative at Harvard University for supporting my research. All mistakes are my own.