Computational modeling of tone in Iu Mien: An inductive approach for field linguists

Ryan K. Shosted[1,2] Di Wu[1] Marissa Goldrich[1]

[1]{Department of Linguistics

[2]Beckman Institute for Advanced Science and Technology},
University of Illinois at Urbana-Champaign
rshosted@illinois.edu

Midwest Speech and Language Day, May 2, 2014

Outline

The problem

- Field linguists working on un(der)described languages are frequently confronted with the problem of classifying patterns inductively.
- This may include the number of phonemic lexical tones.
- The analysis of (phonetic) tone productions should lead to a definitive statement on the number of phonemic tonal contrasts in a given language.
- For underdescribed languages with a non-existent or little-used writing system, much of this analysis has traditionally proceeded on the basis of the researcher's impressions and resulting phonetic transcriptions.

Recent approaches

- More recently, this work involves the collection of large and comprehensive audio databases of the language's lexicon.
- Visualization and measurement of recorded production data evidently increase reliability in classification of phonemic contrasts.
- However, visualization and measurement alone may not leverage all of the data available for describing the patterns observed and may therefore fall short of a robust classification.
- We argue that computational modeling of contrasts can leverage all of the richness in a recorded database and produce results that can be quantified for their consistency and falsified by any other researcher with access to the same data.

Iu Mien

- This study deals with Iu Mien (`ium:ISO 639-3`), an underdescribed Hmong Mien language of China, Laos, Thailand, and Vietnam.
- Iu Mien has 6 lexical tones (Purnell 2012): **low level** (c; 11; ˩); **mid falling** (h; 31; ˧˩); **high mid level** (unmrk; 44; ˦); **high rise falling** (v; 453; ˦˥˧); **low rising** (x; 24; ˨˦); **low rising falling** (z; 231; ˨˧˩) ▸ Tone chart
- When high rise falling (v) syllables have a stop coda, they are realized as **high level** (v*; 55; ˥).
- Our speaker reported initially that she had only five tones but was not explicitly aware of which tonal categories were affected. This represented a nice challenge: could our model predict which tones she had merged?

Purpose

- We approach the phonological tone system of Iu Mien inductively, using computational tools to identify tone categories in the language.
- Our aim is to help field linguists characterize complex phonological systems using computational tools in order to minimize reliance on impressionistic evidence or provide falsifiable quantitative results supporting such evidence.
- What follows makes use of a **very small dataset** (245 monosyllabic items) **uncontrolled** for tonal categories or contextual effects of consonant.

Materials and pre-processing

- We recorded a database of 245 monosyllabic lexical items produced by one female speaker of Iu Mien.
- Items were of the structure V, CVC, or CVV.
- We annotated the sonorous portion of each item and extracted f0 at ten (10) normalized time points using ProsodyPro 5.3.2 (Xu 2013). See ProsodyPro details

Materials

Tonal representation in the corpus (post-hoc)

T-Letter	Descriptive	1–5	IPA	Pctg.	*n*
c	low level	11	˩	15%	22
h	mid falling	31	˧˩	6%	9
(u)	high mid level	44	˦	23%	34
v	high rise falling	453	˦˥˧	19%	28
x	low rising	24	˨˦	13%	19
z	low rising falling	231	˨˧˩	12%	18
v*	high level	55	˥	12%	17

Table: Median representation ≈ 13% ($n = 19$; $n_{exp} = 21$). $\chi^2(6) = 18.67$, $p < 0.01$. Under-representation of h-tone (mid falling ˧˩) and over-representation of unmarked-tone (high mid level ˦).

Modeling procedures

- Researchers listened to all items and classified the lexical tones without supervision or guidance.
- The canonical tone categories assigned to each word (Purnell 2012) were noted independently, with help from the speaker.
- We then performed principal components analysis (PCA) of the normalized pitch contours (Tian and Nurminen 2004); number of PCs was optimized (Jombart 2008).
  ▸ More on PC optimization
- We performed discriminant analysis (DA) on the resulting PCs using two techniques.

Discriminant analysis: Technique 1

- k-means clustering was used to identify and classify the number of tones k present in the database.
- Clusters can be compared with the categories of tones assigned by the researcher (including the dictionary-as-researcher). See a researcher-based model
- Discriminant models using $k + 1$ and $k - 1$ were also computed for reference (**not discussed here**). see the $k = 8$ model

Discriminant analysis: Technique 2

- We divided the database randomly into training (60%), testing (20%), and cross-validation (20%) sets.
- Each researcher's tone categories (along with the dictionary categories) were used as the basis of (training-based) models, which were then applied to supplementary observations (the testing and cross-validation sets).
- Each researcher's consistency was evaluated by comparing the model's predicted tone categories for the testing set with the categories actually assigned to observations in the testing set (today we will present results only for the dictionary-as-researcher). See a researcher-based model

Interpretation of model results

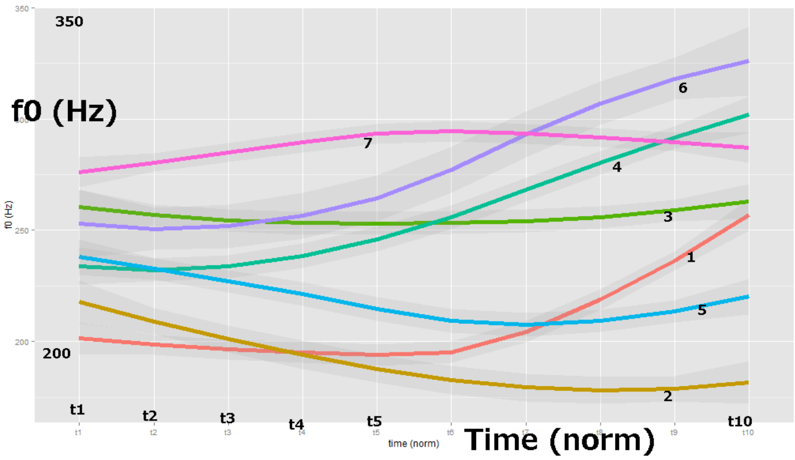
- Discrepancies between the model's predictions and the dictionary tones clarify the speaker's idiosyncratic production of some tone categories.
- Instability of group membership may lead us to conclude that some tonal categories have merged (for this speaker).
- Visualizations of tonal patterns and hierarchical clustering of the same are helpful for understanding variation in production as well as the predictive capacity of the various models under comparison.

k-means clustering

- $k = 7$
- Training model based on 80% of the data; 20% of the data reserved for testing
- Number of PCs = 2 (94.58% var. expl.)

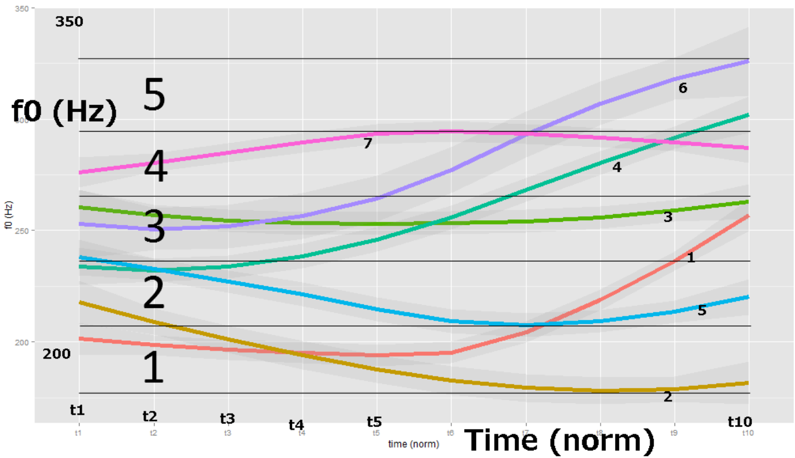
k-means clustering

Machine-learned f0 curves



k-means clustering

Machine-learned f0 curves with reference values 1–5



k-means clustering

Summary and transcription of machine-learned f0 curves

k-tone	Descriptive	IPA
1	low rising	˩˧
4	mid rising	˧˥
6	mid rising	˧˥
2	low mid falling	˨˩
5	mid falling	˧˨
3	mid level	˧
7	high mid level	˦

k-means clustering

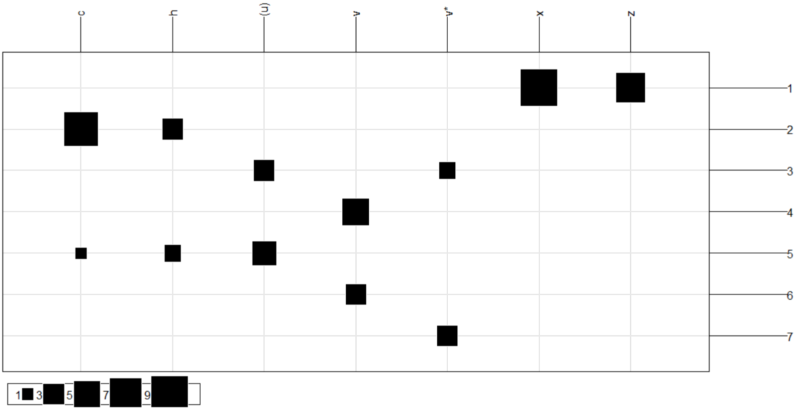


Figure: Machine-learned categories on *y*-axis, dictionary categories on *x*-axis.

k-means clustering

k-tones and comparable dictionary tones

k-tones			Dictionary tones		
k	Descriptive	IPA	Letter	Descriptive	IPA
1	low rising	˩˧	x, z	low rising, low rising falling	˨˦, ˧˦˨
4	mid rising	˨˦	v	high rise falling	˦˥˧
6	mid rising	˨˦	v	high rise falling	˦˥˧
2	low mid falling	˨˩	c, h	low level, mid falling	˩, ˧˩
5	mid falling	˧˨	(u)	high mid level	˦
3	mid level	˧	(u), v*	high mid level, high level	˦, ˥
7	high mid level	˦	v*	high level	˥

k-means clustering
Summary

- Three groups emerged: rising, falling, and level tones
- The rising group accurately corresponded to all dictionary rising tones, but failed to capture the fall at the end of the z- (low rising **falling**) and v- (high rise **falling**) tones.
- The falling group accurately corresponded to the h- (mid falling) tone but inaccurately corresponded to the c- (low level) and (u)- (high mid level) tones.
- The level group accurately corresponded to both level tones ((u)- high and and v*- high mid).

Interim problem statement

Is the model adequate? Or is there unpredicted variation in production?

- There are at least two sources of noise:
- The speaker may not be producing sufficient differentiation between all the tones (as self-reported);
- The model itself may be inadequate at capturing the distinctions (not enough data, suboptimal number of PCs, etc.).
- So, instead of asking what tone categories are present among all the speaker's productions, we'll let the machine learn about the speaker's variable productions of tone categories (as determined using the dictionary's lexical entries (Purnell 2012))...

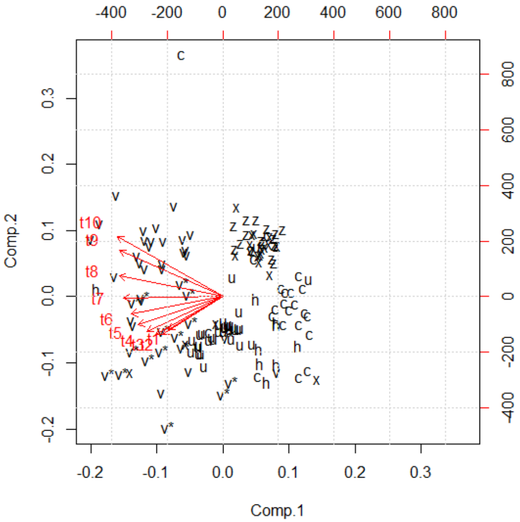
Describing clusters

Unsupervised learning of tonal contrasts based on $t_{1\ldots10}$

- Optimal number of PC's = 7 More on optimization
- Cumulative variance explained = 99.51%
- 61.23% of supplementary observations (in the 20%-testing corpus) were correctly assigned by the model based on 60% of the corpus (training).
- Leave-one-out cross-validation of the remaining 20% generated tones broadly similar in shape to those in the testing corpus.
- Shosted, Wu, and Goldrich assigned numeric values → Chao/IPA tone letters to the resulting f0 curves (training, testing, and cross-validation).

Describing clusters: Biplot

94.6% of variance explained by PC1 and PC2



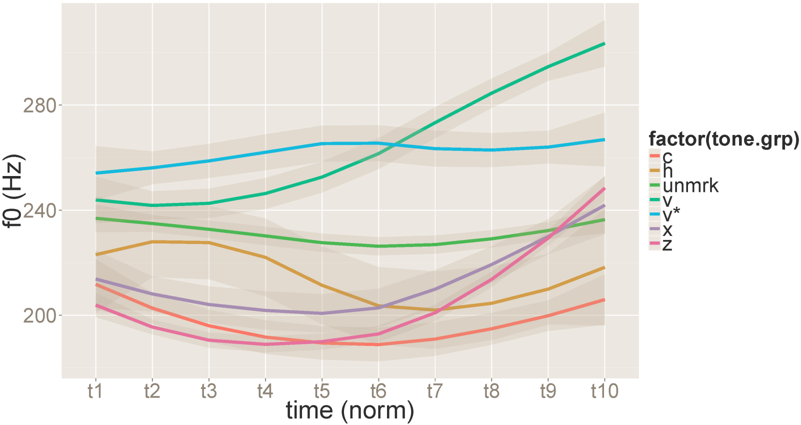
Tonal clustering

Remarks on principal component analysis

- First two PCs account for 94.6% of the variance (cumulative)
- PC1 appears strongly correlated with a point about 70% through f0 curve.
- PC2 appears strongly correlated with the endpoint of the f0 curve.

Dictionary: Training (descriptive)

147 items (=60%)



Dictionary: Training (descriptive)

IPA=Chao tone letters

	H. C. Purnell (2012)		Shosted, Wu, Goldrich	
T-Letter	Descriptive	IPA	Descriptive	IPA
c	low level	˩	low falling rising	˨˩˨
h	mid falling	˧˩	low falling rising	˨˩˨ ˧˩˨ ˨˧
(u)	high mid level	˦	high level	˧
v	high rise falling	˦˥˧	high rising	˧˥
x	low rising	˨˦	low falling rising	˨˩˧
z	low rising falling	˨˧˩	low rising	˩˩˧ ˨˩˧ ˩˧
v*	high level	˥	high level	˧

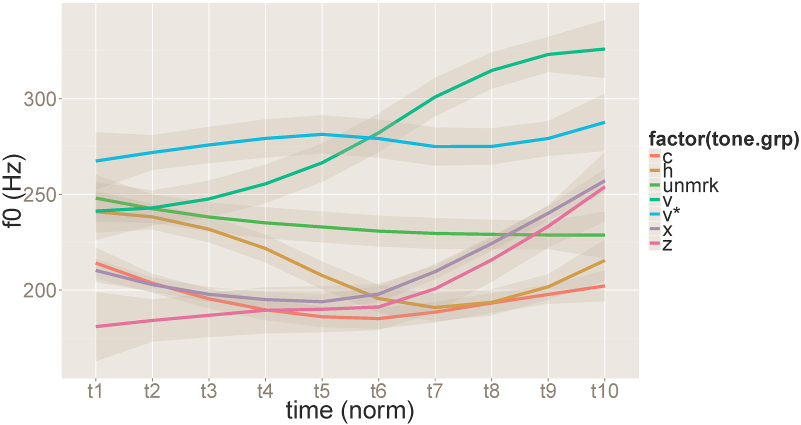
Dictionary: Training (descriptive)

Numeric

H. C. Purnell (2012)			SWG
T-Letter	Descriptive	1–5	
c	low level	11	212
h	mid falling	31	212, 312, 23
(u)	high mid level	44	33
v	high rise falling	453	35
x	low rising	24	213
z	low rising falling	231	113, 213, 123
v*	high level	55	44

Dictionary: Machine-learned categories

49 items (=20%)



Dictionary: Machine-learned categories

IPA=Chao tone letters

H. C. Purnell (2012)			Shosted, Wu, Goldrich	
T-Letter	Descriptive	IPA	Descriptive	IPA
c	low level	˩	low level	˩ ˨˩˧ ˨˩
h	mid falling	˧˩	mid falling rising	˧˩˥
(u)	high mid level	˦	high level	˧
v	high rise falling	˦˥˦	high rising	˧˥
x	low rising	˩˧	low falling rising	˨˩˥
z	low rising falling	˨˧˩	low rising	˩˥
v*	high level	˥	high mid level	˦

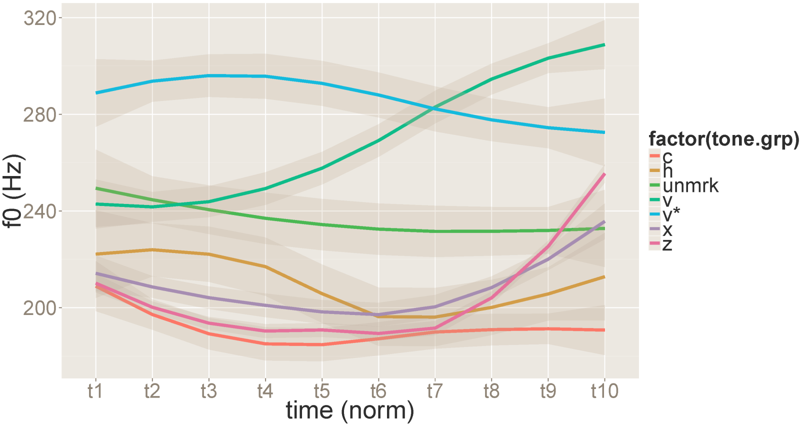
Dictionary: Machine-learned categories

Numeric

H. C. Purnell (2012)			SWG
T-Letter	Descriptive	1–5	
c	low level	11	11, 213, 21
h	mid falling	31	312
(u)	high mid level	44	33
v	high rise falling	453	35
x	low rising	24	213
z	low rising falling	231	13
v*	high level	55	44

Dictionary: Cross-validation

49 items (=20%)



Dictionary: Cross-validation

IPA=Chao tone letters

H. C. Purnell (2012)			Shosted, Wu, Goldrich	
Tone Letter	Descriptive	IPA	Descriptive	IPA
c	low level	˩	low falling	˨˩
h	mid falling	˧˩	low fall rising	˨˩˨
(u)	high mid level	˦	mid level	˧
v	high rise falling	˦˥˦	mid rising	˧˥
x	low rising	˨˦	low fall rising	˨˩˧
z	low rising falling	˨˧˩	low rising	˩˩˧ ˨˦ ˨˩˦
v*	high level	˥	high falling	˥˦

Dictionary: Cross-validation

Numeric

H. C. Purnell (2012)			SWG
Tone Letter	Descriptive	1–5	
c	low level	11	21
h	mid falling	31	212
(u)	high mid level	44	33
v	high rise falling	453	35
x	low rising	24	213
z	low rising falling	231	113, 24, 213
v*	high level	55	54

Summary
IPA

	H. C. Purnell (2012)		Shosted, Wu, Goldrich		
T-Letter	Descriptive	IPA	Train	Pred	Xval
c	low level	˩	˨˩˨	˩ ˨˩˧ ˨˩	˨˩
h	mid falling	˧˩	˨˩˨ ˧˩˨ ˨˧	˧˩˨	˨˩˨
(u)	high mid level	˦	˧	˧	˧
v	high rise falling	˦˥˧	˧˥	˧˥	˧˥
x	low rising	˨˦	˨˩˧	˨˩˧	˨˩˧
z	low rising falling	˨˧˩	˩˩˧ ˨˩˧ ˩˧	˩˧	˩˩˧ ˨˥ ˨˩˧
v*	high level	˥	˦	˦	˥˦

Summary
Numeric

H. C. Purnell (2012)			Shosted, Wu, Goldrich		
T-Letter	Descriptive	1–5	Train	Pred	Xval
c	low level	11	212	11, 213, 21	21
h	mid falling	31	212, 312, 23	312	212
(u)	high mid level	44	33	33	33
v	high rise falling	453	35	35	35
x	low rising	24	213	213	213
z	low rising falling	231	113, 213, 123	13	113, 24, 213
v*	high level	55	44	44	54

Dictionary: Machine-learned categories

Cross-category assignment (49 items)

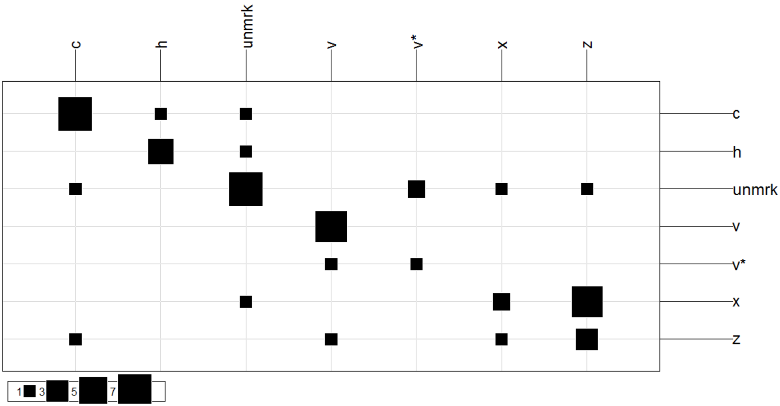


Figure: Machine-learned categories on *y*-axis, dictionary categories on *x*-axis. Note discrepancies between x- and z-tones (low rising ˩˧ and low rising falling ˩˧˩).

Dictionary: Machine-learned categories

Tonal category clusters

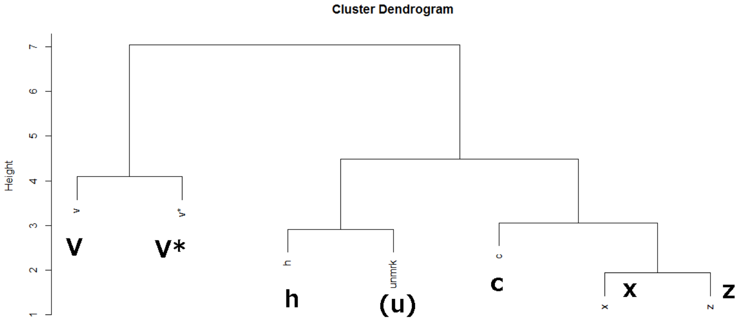


Figure: Cluster dendrogram for machine-learned categories. Note shortest distance between x- and z-tones (low rising ˩˧ and low rising falling ˩˧˩).

Discussion

- We argue that this method has a number of distinct advantages over more traditional approaches to field phonology, which are in many cases neither quantifiable nor, strictly speaking, falsifiable.
- The application of computational modeling techniques under these circumstances appears to be uncommon (cf. Tian and Nurminen 2004) – but if we have missed something, please tell us!
- Our technique offers field linguists an expanded tool set for resolving and describing the number lexical tones in a language and has application to issues of variation, dialectology, and historical sound change.

Acknowledgements

- Colleagues: Chilin Shih (Illinois), Christopher Carignan (North Carolina State University), Jonathan Amith (Gettysburg College)
- Graduate students in LING 516 Field Methods (Sp14): Staci Defibaugh, Kate Lyons, Lydia Medill, Bradley Miller

Selected references

- Jombart, T. 2008. `adegenet`: a R package for the multivariate analysis of genetic markers. *Bioinformatics 24*: 1403–1405.
- Purnell, H. C. 2012. *An Iu-Mienh—English Dictionary*. San Francisco, CA: Silkworm Books / Center for Lao Studies.
- Tian, J. and Nurminen, J. 2004. On analysis of eigenpitch in Mandarin Chinese. In *Proceedings of the 4th International Symposium on Chinese Spoken Language Processing*.

7 The Appendix

- Introduction+
- Methods+
- Results+

Iu Mien lexical tones

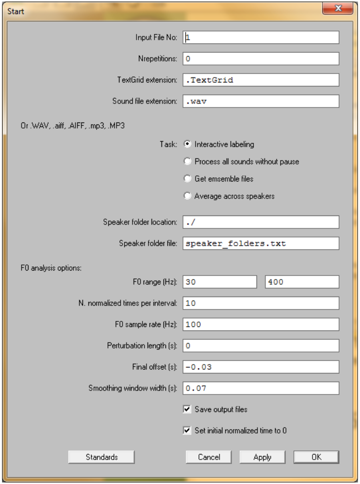
IPA=Chao tone letters

T-Letter	Descriptive	1–5	IPA
c	low level	11	˩
h	mid falling	31	˧˩
(u)	high mid level	44	˦
v	high rise falling	453	˦˥˧
x	low rising	24	˨˦
z	low rising falling	231	˨˧˩
v*	high level	55	˥

Table: v* = v- (high rise falling; 453; ˦˥˧) syllable ending in a stop consonant

◂ Back to Iu Mien

Prosody Pro 5.3.2 (Xu 2013) settings



- f0 range: 30–400 Hz
- f0 sample rate: 100 Hz
- Perturbation length: 0 s
- Final offset: -0.03 s
- Smoothing window: 0.07 s

Go back to methods

Optimized number of PCs

See Jombart (2008) for further discussion

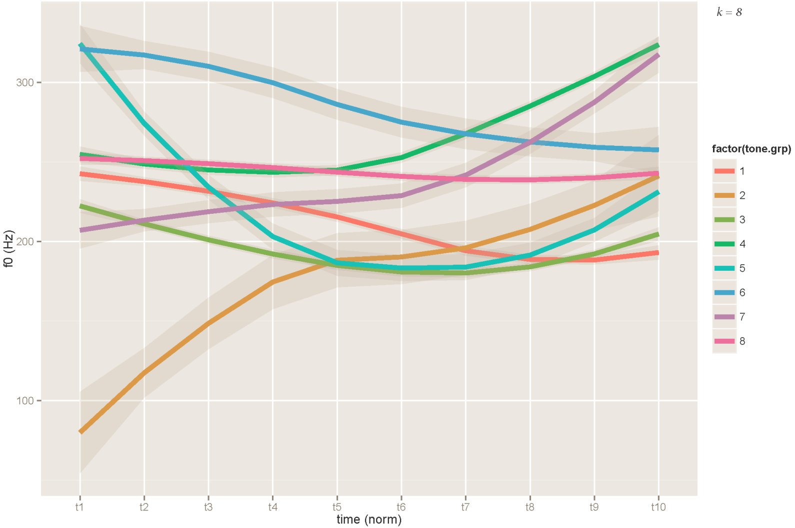
- Finds the optimum **a-score** (`adegenet` package for R, Jombart et al. (2013))
- The a-score is "the proportion of successful reassignment corrected for the number of retained PCs"
- Or more technically: "the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination)".

Back to Modeling Procedures

Back to Describing Clusters 1

k-means clustering

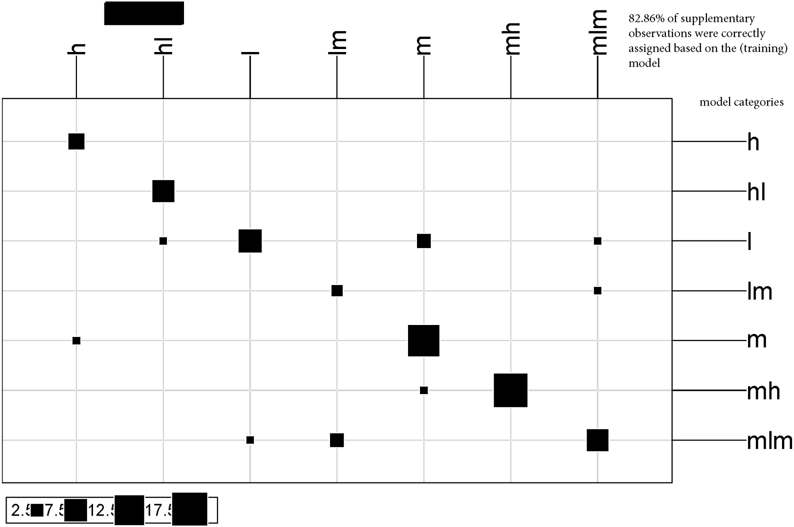
$k = 8$



Go back to Discriminant Analysis 1

Researcher-based model

Tones assigned by researcher

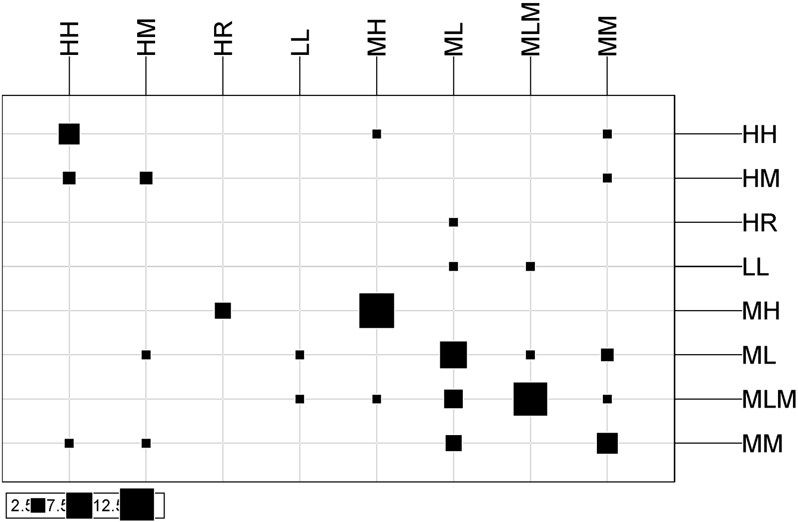


82.86% of supplementary observations were correctly assigned based on the (training) model

Go back to Discriminant Analysis 2

Researcher-based model

Tones assigned by another researcher



Go back to Discriminant Analysis 2