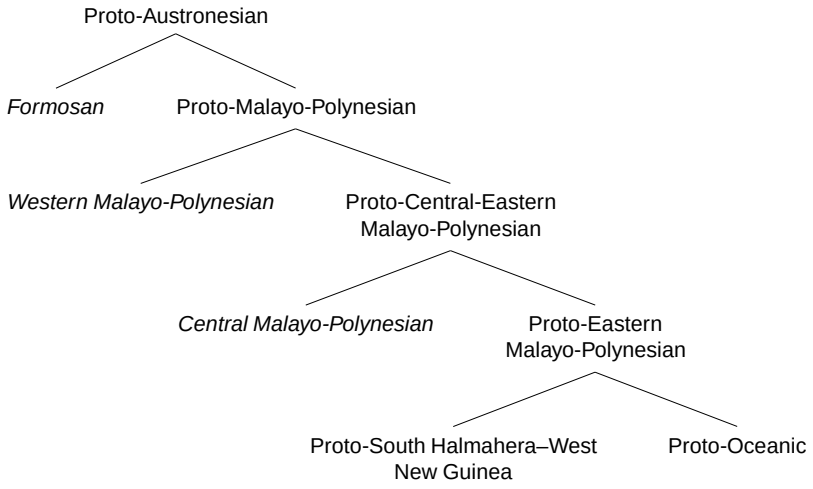
David Kamholz (UC Berkeley) linguistics.berkeley.edu/∼kamholz kamholz@berkeley.edu

## Tonogenesis in Yerisiam, an Austronesian language of West New Guinea

- Yerisiam (*ire*, also known as Iresim, Yeresiam):
  - c. 1000 speakers in three villages in Papua province, Indonesia: (i) Sima, on Cenderawasih Bay; (ii) Erega, inland near the Yamor lakes (probable Yerisiam homeland); (iii) Rurumo, on Etna Bay (Figure 1)
  - bilingualism and intermarriage: (i) Sima: with Yaur (Austronesian); (ii) Erega: with Kamoro (Asmat-Kamoro family); (iii) Rurumo: with Mer and Semimi (Mairasi family)
  - classification: Austronesian, in the South Halmahera–West New Guinea (SHWNG) subgroup (Figure 2)
- Tone in Yerisiam and SHWNG:
  - Yerisiam has lexically contrastive tone: many minimal pairs, e.g. *àká* 'ball sago', *áakà* 'four', *áàkà* 'sago palm midrib', *áaká* 'to bite'
  - other tonal SHWNG languages: Moor, Yaur (Kamholz 2012); Raja Ampat languages (Remijsen 2001)
- How and why did tone arise in Yerisiam?
- Structure of the talk:
  - §1. Overview of Yerisiam prosody
  - §2. From Proto-Malayo-Polynesian to Yerisiam
  - §3. Tonogenesis in Yerisiam
  - §4. Typological parallels and larger implications
- All Yerisiam data from my field notes, 2010–12

Figure 1: Language map of southern Cenderawasih Bay. (Adapted from SIL Indonesia map.)

Figure 2: The higher nodes of the Austronesian family tree, after Blust (2009:724–30; 736–9). Nodes in italics are not proto-languages, but rather are cover terms for multiple primary branches.




## 1 Overview of Yerisiam prosody

- Contrasts
  - five vowels *a*, *e*, *i*, *o*, *u*
  - short and long vowels: *róhé* 'nit' vs. *róohé* 'hard pith'
  - underlying tones H and L
- Permissible syllable nuclei: V, VV, V͡V, VVG, GVV (VV = long vowel, V͡V = diphthong, G = high vowel)
- TBU is the mora
  - V = $\mu$, VV/VVG/GVV = $\mu\mu$, V͡V= $\mu/\mu\mu$ (more on diphthongs below)
  - coda consonants do not count for weight
- Most prevalent lexical tone patterns:
  - (1) H throughout
  - (2) H with drop to L (H-L, HL-L, H-HL, etc.)
  - (3) L*-H (zero or more L syllables with final H syllable)
- Phonotactic restrictions
  - one long vowel per word, except if word ends with CV́V.CV́V̀ (*rónɨ́ijáà* 'gecko sp.', *báróopé* 'finger', but **báaróopé*)
  - no long vowels on final level syllables (*náà* 'dog', *rá* 'to go', but **ráa*)
  - VVG diphthongs found only on final level syllables (*róoi* 'night', *áì* 'wood', but **áaì*, **róoimé*)
  - no long vowels in type (3) words (*bàkí* 'arm', but **bàakí*)
- Type (3) words best analyzed as underlyingly toneless, based on tonal alternations when adding enclitic demonstrative *=tà*:
  - (1) *áakú* 'stone' → *áakúà=tà*
  - (2) *búubù* 'water' → *búubùa=tà*
  - (3) *hùpé* 'bottom' → *húpéà=tà* (tone overwritten as type (1))
- V͡V diphthongs
  - count for one mora if level tone: *báohé* 'turtle flipper'; *krádìa* 'flesh'; *rói* 'to sing', *dìarúa* 'sago stirrer'
  - count for two moras if contour tone: *áì* 'wood', *púkúà* 'cloud', *káèpà* 'sago chopsticks'

- Not straightforward to tie together the distribution of vowel length, VVG diphthongs, and mora assignment in V͡V diphthongs into a more general analysis
- Another complication is the skewed distribution of final *a*: rare in type (1) words, common in types (2) and (3)
- Tables below show distribution of tone patterns by syllable shape in disyllables and trisyllables, for 1440-item lexicon (percentages indicate number of items with final *a*; dashes indicate impermissible combinations)
- No apparent synchronic explanation for the skewing

| | (1) H-H | (2) H-L | H-HL | HL-L | (3) L-H | total |
|---|---|---|---|---|---|---|
| CV.CV | 21 (0%) | 32 (78%) | – | – | 42 (36%) | 95 (42%) |
| CV.CVV | – | – | 20 (35%) | – | – | 20 (35%) |
| CV.CV͡V | 6 (17%) | 21 (86%) | 14 (29%) | – | 27 (67%) | 68 (60%) |
| CV.CVVG | 9 (0%) | 0 | – | – | – | 9 (0%) |
| CVV.CV | 123 (15%) | 17 (53%) | – | 60 (68%) | – | 200 (35%) |
| CV͡V.CV | 16 (19%) | 1 (100%) | – | 16 (88%) | 2 (0%) | 35 (51%) |
| CGVV.CV | 17 (6%) | 2 (0%) | – | 1 (100%) | – | 20 (10%) |
| CVV.CVV | – | – | 17 (94%) | – | – | 17 (94%) |
| CV͡V.CVV | – | – | 5 (60%) | – | – | 5 (60%) |
| CVV.CV͡V | 1 (0%) | 1 (0%) | 4 (100%) | 3 (100%) | – | 9 (78%) |
| CV͡V.CV͡V | 2 (50%) | 0 | 2 (50%) | 0 | 4 (75%) | 8 (63%) |
| CV͡V.CVVG | 1 (0%) | 0 | – | 0 | – | 1 (0%) |
| CGVV.CVV | – | – | 3 (100%) | – | – | 3 (100%) |
| CGVV.CV͡V | 0 | 0 | 1 (100%) | 0 | – | 1 (100%) |
| total | 196 (13%) | 74 (72%) | 66 (59%) | 80 (74%) | 75 (61%) | 491 (43%) |

- Three marginal disyllabic patterns are not shown above (H-LH, HL-H, LH-L), totalling 9 items

| | (1) H-H-H | (2) H-H-L | H-L-L | H-H-HL | H-HL-L | (3) L-L-H | total |
|---|---|---|---|---|---|---|---|
| CV.CV.CV | 19 (5%) | 39 (62%) | 5 (0%) | – | – | 17 (35%) | 80 (39%) |
| CV.CV.CVV | – | – | – | 3 (100%) | – | – | 3 (100%) |
| CV.CV.CV͡V | 2 (50%) | 17 (47%) | 0 | 7 (29%) | – | 14 (93%) | 40 (60%) |
| CV.CV.CVVG | 3 (0%) | 0 | 0 | – | – | – | 3 (0%) |
| CV.CVV.CV | 86 (6%) | 25 (20%) | 6 (0%) | – | 83 (76%) | – | 200 (37%) |
| CV.CV͡V.CV | 4 (25%) | 5 (60%) | 0 | – | 14 (79%) | 1 (0%) | 24 (63%) |
| CV.CGVV.CV | 19 (11%) | 1 (0%) | 0 | – | – | – | 20 (10%) |
| CVV.CV.CV | 28 (0%) | 53 (83%) | 4 (0%) | – | – | – | 85 (52%) |
| CV͡V.CV.CV | 4 (0%) | 1 (100%) | 0 | – | – | 0 | 5 (20%) |
| CGVV.CV.CV | 2 (0%) | 4 (100%) | 1 (0%) | – | – | – | 7 (57%) |
| CV.CVV.CVV | – | – | – | 10 (90%) | – | – | 10 (90%) |
| CV.CVV.CV͡V | 0 | 3 (100%) | 0 | 2 (100%) | 2 (100%) | – | 7 (100%) |
| CV.CV͡V.CVV | – | – | – | 0 | – | – | 0 |
| CV.CGVV.CVV | – | – | – | 2 (100%) | – | – | 2 (100%) |
| CV.CV͡V.CV͡V | 0 | 1 (100%) | – | 0 | 0 | 2 (100%) | 3 (100%) |
| CVV.CV.CV͡V | 0 | 11 (100%) | 0 | 0 | – | – | 11 (100%) |
| CV͡V.CV.CV͡V | 0 | 0 | 0 | 0 | – | – | 0 |
| CVV.CV͡V.CV | 0 | 0 | 1 (0%) | – | 1 (0%) | – | 2 (0%) |
| CVV.CGVV.CV | 0 | 0 | 0 | – | 0 | – | 0 |
| CV͡V.CVV.CV | 6 (0%) | 5 (0%) | 0 | – | 5 (60%) | 0 | 16 (19%) |
| CV͡V.CV͡V.CV | 0 | 0 | 0 | – | 1 (100%) | 1 (0%) | 0 |
| CV͡V.CVV.CVV | – | – | – | 2 (100%) | – | – | 2 (100%) |
| CV͡V.CV͡V.CVV | – | – | – | 1 (100%) | – | – | 1 (100%) |
| total | 173 (6%) | 165 (63%) | 17 (0%) | 27 (78%) | 106 (75%) | 35 (60%) | 523 (45%) |

- Nine marginal trisyllabic patterns are not shown above (L-L-L, H-L-H, H-L-HL, L-L-LH, H-L-LH, H-HL-H, H-LH-L, LH-L-L, HL-H-L), totalling 22 items

## 2 From Proto-Malayo-Polynesian (PMP) to Yerisiam

- Yerisiam words with known Austronesian etymologies are listed below, grouped by tone pattern. Dashes indicate a known morpheme boundary; slashes indicate a segmentation with no independent justification.

| | | Proto-form | Yerisiam | gloss |
|---|---|---|---|---|
| (1) | 1. | PMP *batu | áakú | ‘stone’ |
| | 2. | PMP *duha | rúu-hí | ‘two’ |
| | 3. | PMP *inum | íimán/é | ‘to drink’ |
| | 4. | PMP *kaən | áan-í | ‘to eat (intr.)’ |
| | 5. | PMP *kutu | úukú | ‘head louse’ |
| | 6. | PMP *ma-bəRəqat | máak/í | ‘heavy’ |
| | 7. | PMP *ma-nipis | máníi/jáhé | ‘thin’ |
| | 8. | PMP *ma-qətaq | máaká | ‘raw’ |
| | 9. | PMP *ma-qətaq | mák-máaká | ‘green’ |
| | 10. | PMP *ma-takut | ngkák/é | ‘afraid’ |
| | 11. | PMP *pəñu | éenú | ‘turtle’ |
| | 12. | PMP *punti | píití | ‘banana’ |
| | 13. | PMP *qabaRa | áar-í | ‘to carry on shoulder’ |
| | 14. | PMP *qapuR | áau | ‘betel lime’ |
| | 15. | PCEMP *qenəp | éené | ‘to lie (down)’ |
| | 16. | PMP *qitəluR | á/kóor/é | ‘egg’ |
| | 17. | PMP *tanəm | káamán/é | ‘to plant’ |
| | 18. | PMP *taŋis | káh/é | ‘to cry’ |
| | 19. | PMP *təbuh | kóou | ‘sugarcane’ |
| | 20. | PMP *tələn | kóor-í | ‘to swallow’ |
| | 21. | PMP *təlu | kóorí-hé | ‘three’ |
| | 22. | PMP *tunu | kúun/á | ‘to burn’ |
| | 23. | PCEMP *waRəj | gwáarí | ‘rope’ |
| (2) | 24. | PMP *bulan | úùrà | ‘moon’ |
| | 25. | PMP *buku | bú-gùa | ‘knot’ |
| | 26. | PMP *dahun | ráàn/ìa | ‘leaf’ |
| | 27. | PMP *danum | ráarám-à | ‘water’ |
| | 28. | PMP *daRaq | rárà | ‘blood’ |
| | 29. | PMP *əpat | áak-à | ‘four’ |
| | 30. | PMP *kahiw | áì | ‘wood’ |
| | 31. | PCEMP *kandoRa | átóòrà | ‘cuscus’ |
| | 32. | PMP *laŋit | ráak/átè | ‘sky’ |
| | 33. | PMP *lima | rîmà | ‘five’ |
| | 34. | PMP *manuk | máan/áà | ‘bird’ |
| | 35. | PMP *ma-Ruqanay | máànà | ‘man’ |
| | 36. | PMP *m-atay | máàkà | ‘to die’ |
| | 37. | PMP *nusa | núùhà | ‘island’ |
| | 38. | PMP *niuR | núì | ‘coconut’ |
| | 39. | PMP *ŋajan | áahán-à | ‘name’ |
| | 40. | PMP *qaləjaw | óòrà | ‘sun’ |
| | 41. | PMP *qaninu | ánúunú-gùa | ‘shadow’ |
| | 42. | PMP *qasu | ógw/áahú-gùa | ‘smoke’ |
| | 43. | PMP *qatay | ákéè/nà | ‘liver’ |
| | 44. | PMP *Rumaq | rúmà | ‘ceremonial house’ |
| | 45. | PMP *susu | húuhú-gùa | ‘breast’ |
| | 46. | PMP *tasik | káhì-a | ‘salt’ |
| | 47. | PCEMP *waŋka | gwáà | ‘canoe’ |
| | 48. | PMP *zalan | jáàrà | ‘path’ |
| (3) | 49. | PMP *baRa | bà-kí | ‘arm’ |
| | 50. | PMP *buaq | ú | ‘fruit’ |
| | 51. | PMP *kaən | àn/á | ‘to eat (tr.)’ |
| | 52. | PMP *lakaw | rá | ‘to go’ |
| | 53. | PMP *ma-hiaq | mái | ‘ashamed’ |
| | 54. | PMP *ma-tuqah, PEMP *matu | màkú/i | ‘dry (coconut)’ |
| | 55. | PMP *Rambia | pí | ‘sago’ |
| | 56. | PMP *t‹in›aqi | hìná | ‘belly’ |
| | 57. | PCEMP *todan | kó | ‘to sit’ |

- Yerisiam reflexes of Proto-Malayo-Polynesian (PMP) segments:

| PMP | Yerisiam |
|---|---|
| *a | a |
| *i | i |
| *u | u, i |
| *ə (penult) | o |
| *ə (final) | a |
| *e (PCEMP) | e |
| *o (PCEMP) | o |
| *w | gw |
| *R | r |
| *l | r |
| *r | ? |
| *m | m |
| *n | n, (r) |
| *ñ | n |
| *ŋ | ∅ |

| PMP | Yerisiam |
|---|---|
| *q | ∅ |
| *h | ∅ |
| *s | h |
| *j | h |
| *z- | j |
| *-z- | ? |
| *p | ∅, p |
| *b | ∅, b |
| *t /_i | *s > h |
| *t | k |
| *d | r |
| *nt, *nd | t |
| *k | ∅ |
| *g | ? |

- Sporadic final consonant loss: **zalan* > *jáàrà* 'path' vs. **ŋajan* > *áahán-à* 'name'

## 3 Tonogenesis in Yerisiam

- Development of vowel length: penultimate vowels regularly lengthen, e.g. **batu* > *áakú* 'stone'
  - Proto-Oceanic is reconstructed with penultimate stress; Proto-SHWNG may have been similar
  - no lengthening if phonotactically impermissible: **kahiw* > *áì* 'wood', **lakaw* > *rá* 'go'
- Following vowel lengthening, some nouns are suffixed with *-a* or *-gua* (origin and meaning unknown)
- Tonogenesis:
  - type (1) pattern (H throughout) is default outcome
  - type (2) pattern (H switching to L) triggered by final *a* (not always originally present or originally final)
  - type (3) pattern (L*-H, underlyingly toneless) seems to result from:
    (i) monosyllables that would otherwise be type (1): **buaq* > *ú* 'fruit', **Rambia* > *pí* 'sago', **todan* > *kó* 'sit'
    (ii) final syllable loss: **ma-tuqah* > *màkú/i* 'dry (coconut)', **t‹in›aqi* > *hìná* 'belly'
- Location of the drop to L in type (2) words
  - if penult is short, on the ultima: **daRaq* > *rárà* 'blood', **Rumaq* > *rúmà* 'ceremonial house'
  - if *-a* or *-gua* has been added, on the added suffix: **əpat* > *áak-à* 'four', **ŋajan* > *áahán-à* 'name', **qaninu* > *ánúunú-gùa* 'shadow', **susu* > *húuhú-gùa*, etc.
  - otherwise, on the second mora of the (lengthened) penult: **bulan* > *úùrà*, **lima* > *rûmà* 'five', etc.
- Residue:
  - 7 words do not show expected vowel lengthening: **buku* > *bú-gùa* 'knot', **daRaq* > *rárà* 'blood', **Rumaq* > *rúmà* 'ceremonial house', **ma-hiaq* > *mái* 'ashamed', **ma-takut* > *ngkák/é* 'afraid', **tasik* > *káhì-a* 'salt', **taŋis* > *káh/é* 'to cry'
  - 3 words do not show expected L from final *a*: **lakaw* > *rá* 'go', **ma-qətaq* > *máaká* 'raw', **tunu* > *kúun/á* 'to burn'
  - unexpected location of L in **manuk* > *máan/áà* 'bird'
  - origin of CV́V.CV́V̀ pattern is unexplained in general
- 31 words have tone patterns other than types (1), (2), and (3), none with known Austronesian etymologies, e.g. *áìhá* 'above', *nùúsì* 'great-grandparent', *mígwàí* 'wild duck sp.', *mònìhò* 'resin', *rúmùhúa* 'wood post', *móhòràú* 'colored sky'

## 4 Typological parallels and larger implications

- Cross-linguistic universal: high vowels have intrinsically higher $F_0$ (10-15 Hz) (Whalen and Levitt 1995)
- Hombert et al. (1979:52): 'Since vowel height and prevocalic consonants cause similar perturbations, one could expect tonal development resulting from vowel merging to be as frequent as tonal development resulting from the loss of some voicing contrast in prevocalic position... however, the historical data are scanty on this point.'
- Maddieson (1978:354): 'Cases of a claimed effect of vowel height on tonal development are not generally persuasive.'
- Putative examples in Lahu and Foochow 'can be interpreted as an effect of tones on vowels. It would seem that the interaction between tones and vowel height works only in one direction: tone can affect vowel height but not vice-versa.' (Hombert et al. 1979:52)
- Is vowel quality *a* really responsible for low tone in Yerisiam?
  - some type (2) nouns have added suffix *-a* or *-gua*, which coincidentally (?) end in *a*
  - but roots ending in *a* also show low tone: **lima* > *rîmà* 'five', **bulan* > *úùrà* 'moon', etc.
- Rivierre (2001) on tonogenesis in Cèmuhî, an Oceanic language of New Caledonia with word tone:
  - high tone: *$C^hV$ > $C\acute{V}$
  - mid tone: *CV > $C\bar{V}$
  - low tone: **aqa*, **ao*, **oa* > *à*
  - short **a* > *ɛ*
- Cèmuhî versus Yerisiam
  - in Cèmuhî, tone generally preserves otherwise lost segmental contrasts: **apa* > *ā*, **aqa* > *à*
  - in Yerisiam, the triggering *a* remains
- Yerisiam and Cèmuhî are both reasonably convincing cases of *a* producing low tone: contrary to the above quotes, vowel height apparently can affect tone (but rarely)
- Tonogenesis in SHWNG
  - comparison of Yerisiam with Moor, Yaur, and Raja Ampat languages suggests that Yerisiam has no common tonal developments with other SHWNG languages
  - at least four independent tonogenesis events
- Contact and lexical borrowing may be a factor in SHWNG tonogenesis, but how can this be demonstrated?
  - Yerisiam is currently in contact with two tonal languages, Yaur and Kamoro, but no known shared features
  - time depth of Proto-SHWNG is as great as 4000 years: difficult to reconstruct ancient contact scenario with Papuan languages

## 5 Acknowledgements

This work was supported by an Individual Graduate Scholarship from the Endangered Languages Documentation Programme ('Documentation of Moor, an Austronesian language of Cenderawasih Bay, Indonesia', 2010–13). Thanks to Larry Hyman, Andrew Garrett, Claire Moyse-Faurie, and Mark Donohue, who provided crucial support and feedback on this project. Special thanks to my Yerisiam consultants, Herman Mararіampi and Hengki Akubar, without whom this work would not have been possible. *árah ráreìja hìna únduume.*

## 6 References

Blust, Robert A. 2009. *The Austronesian languages.* Canberra: Pacific Linguistics.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55(1): 37–58.

Kamholz, David. 2012. The languages of southern Cenderawasih Bay and their history. Paper presented at the 12th International Conference on Austronesian Linguistics, Denpasar, Indonesia.

Maddieson, Ian. 1978. Universals of tone. In *Universals of human language, vol. 2: Phonology*, Joseph H. Greenberg, ed., 335–365. Stanford: Stanford University Press.

Remijsen, Bert. 2001. *Word-prosodic systems of Raja Ampat languages.* No. 49 in LOT Dissertation Series, Leiden University.

Rivierre, Jean-Claude. 2001. Tonogenesis and the evolution of tonal systems in New Caledonia, the example of Cèmuhî. In *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena*, Shigeki Kaji, ed., 23–42. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Whalen, D. H. and Andrea G. Levitt. 1995. The universality of intrinsic $F_0$ of vowels. *Journal of Phonetics* 23: 349–366.