1

Language Documentation & Conservation Special Publication No. 24 (2021)
Phonetic fieldwork in Southern New Guinea
ed. by Kate L. Lindsey & Dineke Schokkin, pp. 5–12
http://nflrc.hawaii.edu/ldc/ http://hdl.handle.net/10125/24999

The phonetics of Papuan languages in Southern New Guinea: an overview

Nicholas Evans
Australian National University

The island of New Guinea and its surrounds remains at once the most linguistically diverse region on earth (with around 1,200 languages, i.e. a sixth of the world's languages in 1% of its land surface) and arguably the least studied.[1] This contrast is even more marked once one peels off the Austronesian languages wrapped around the north, east and west coastal regions and begins to investigate the 43 families and 36 isolates (Palmer 2018: 6) conventionally known as "Papuan," a term which does not entail any phylogenetic or typological unity but simply means "language which is spoken indigenously in the southwestern Pacific, and is neither Austronesian nor Australian." It is becoming increasingly clear that the extreme diversity of this region does not stop at the number of languages or genetic units, but extends out into structural characteristics at all realms of linguistic organisation.

Most relevant for this special issue, there is also great diversity in the phonetics and phonology of Papuan languages. We can use the publications in the 'Illustrations' series of JIPA (Journal of the International Phonetics Association) as a guide to how little we know about Papuan phonetics and phonology. Out of 175 published JIPA Illustrations (as at April 2021), only three are of Papuan languages: Nen (Evans & Miller 2016), Fataluku (Heston & Locke 2019) and Ende (Lindsey 2021).

This special issue focusses on one Papuan region, bringing together phonetic portraits of six languages from the Southern New Guinea (SNG) area. Until recently very little was known about any of the languages of this area (see Evans 2012, Evans et al. 2018 for surveys) but there has been a major research push over the last decade (see e.g., Carroll 2016, Döhler 2018, and Olsson 2017 for recent comprehensive grammars) and two recent JIPA illustrations have also appeared (Evans & Miller 2016 for Nen and Lindsey 2021 for Ende). The time is thus ripe to take stock of what we know of the phonetics of the region's languages.

Southern New Guinea (Figure 1) is a flat, alluvial region, about the size of Czechia and Slovakia combined, stretching from the Maro River to the west (in Indonesian Papua)

[1] For institutional and grant-agency support of my research on the languages of Southern New Guinea since 2008, I would like to thank the Australian Research Council (Grants: Languages of Southern New Guinea and The Wellsprings of Linguistic Diversity), the Volkswagen Foundation (DoBES project 'Nen and Tonda'), the Alexander von Humboldt Foundation (Anneliese Maier Forschungspreis), the Australian National University (Professorial Setup Grant) and the ARC Centre of Excellence for the Dynamics of Language (CoEDL). Most importantly, I thank the entire population of Bimadbn village for their hospitality and friendship, and especially Jimmy Nébni, Michael Binzawa, Yosang Amto and Goe Dibod, as well as speakers of other languages in their region for their patient teaching of the phonetics of Nen, Idi, Nmbo and other languages – it was the direct exposure to these languages, right from the first day of coming to the region in 2008 when Gus Iammata and Goe Dibod began teaching me Idi and Nen along the road from Dimsisi to Bimadbn, that made me realise how much phonetic diversity is present in the region. Finally I would like to thank Kate Lindsey and Dineke Schokkin for their editorial comments, and for giving me the opportunity to publish this brief overview here.

to the Fly River to the east (in Papua New Guinea), and mostly forming low-lying savannah, rainforest or swamp bounded to the north by the two aforementioned rivers and to the south by the Torres Strait. Most languages are small – ranging from a few hundred to a few thousand speakers – but on the PNG side of the region, they remain secure, still being learned by children. On the Indonesian side, they are starting to give way to Indonesian. For most groups, particularly the smaller ones, traditional "egalitarian multilingualism" prevailed (Haudricourt 1961, François 2012), with a significant proportion of marriages being linguistically exogamous and engendering bilingual households. But this has not led to any clear "linguistic area," in the sense of structural features distinctive of the region, and as we shall see below, this is as true for the phonetics as for other realms of grammar.

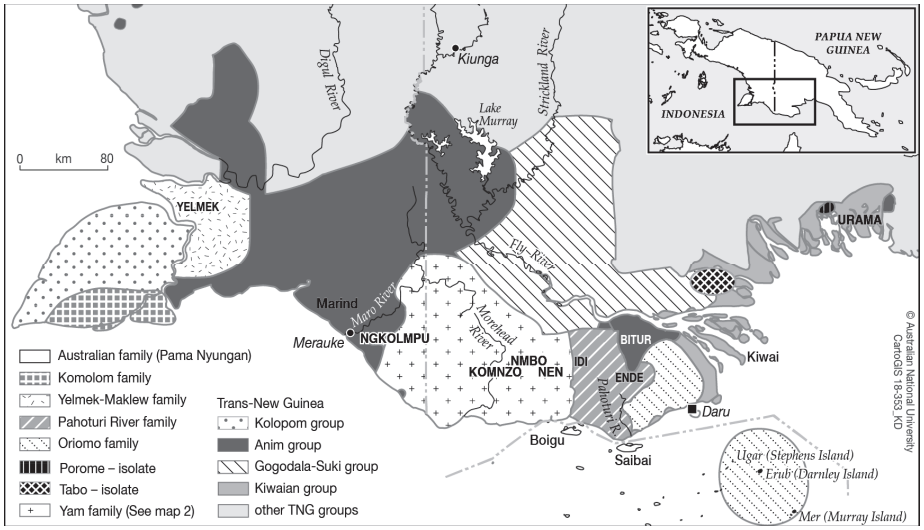

Figure 1. Map illustrating the prominent language families of Southern New Guinea

With around forty languages from seven maximal (i.e. currently unrelatable) clades, Southern New Guinea is arguably the third most diverse region of New Guinea, after the north coast and the Sepik. Four of its seven maximal clades (Evans et al. 2018) are represented in this collection – we did not manage to obtain contributions on the Komolom or Oriomo families, and the seventh clade, the Australian Pama-Nyungan family in the form of Kala Kawaw Ya, is likewise not represented here. For Yelmek-Maklew we have Tina Gregor's article on Yelmek. For Yam we have two of its three branches represented – Matthew Carroll's article on Ngkolmpu (Tonda branch), and for the Nambo branch we have Eri Kashima's article on Nmbo, complementing the recent article on the closely-related Nen in this journal (Evans & Miller 2016). For Pahoturi River we have the article on Idi by Schokkin and colleagues, offering interesting parallels and differences to the JIPA illustration of Ende by Lindsey (2021). For Trans-New Guinea we have two branches represented: Phill Rogers' chapter on Bitur, which belongs to the Lower Fly sub-branch of the Marind-Anim branch, and, for the Kiwaian branch, a description of Urama by Brown and colleagues. Note that the Trans-New Guinea family is very large, with between 300 and 500 languages (Pawley & Hammarström 2018), marking it the towering giant among Papuan language families. The spatial patterning of the Marind-Anim branch, consistent with downstream migration along the Fly and Maro rivers from origins in the Lake Murray region, suggests that they are a recent intrusion into Southern New Guinea, and as we shall see below, Bitur is rather unlike the other SNG languages in its phonological system. It is likely that the Kiwaian languages, including Urama, were also relatively recent downstream migrants into the region, but longer-established than the Marind-Anim languages, since all Kiwaian languages are

found around the Fly River mouth. Like Bitur, Urama has a simpler phonological system than the other languages represented here; it is also the only language represented that has a pitch accent system.

Reflecting on the diversity of the region on the phonological level, it is difficult to identify any typological features of the phonologies of Southern New Guinea, other than absences (e.g., no tone, no pitch accent or distinctive stress except for Urama, no ejectives etc.). The fact that all languages possess a contrast between voiced and voiceless stops[2] is somewhat unusual in the New Guinea context (where it would be more usual to contrast voiceless plain stops with voiced prenasalised stops), but hardly striking by world standards, nor is the fact that all languages in the sample have an /s/ phoneme. The number of linear positions for stops and nasals ranges from three (Yelmek, Bitur) up to five or six (Ende, Idi and Nmbo), depending on whether labial-velars and labio-velars are counted as an additional position, and whether retroflexes are treated as stops or affricates. Among nasals, only /m/ and /n/ are universally present: /ŋ/ is absent from Nmbo, Bitur and Urama (and from Nen), and /ɲ/ is absent from Bitur, Urama and Yelmek.

If we begin by looking just at the size of the consonant inventories, focussing on the six languages of this special issue of LDC plus the two recent JIPA illustrations (Table 1), we likewise see a span from what Maddieson (2005a) considers 'small' (Yelmek, Bitur, Urama) to what he considers 'moderately large' (Nmbo), with the smallest being Bitur and Urama (13 consonant phonemes each[3]) and the largest Nmbo (28). It is worth noting that the baseline for New Guinea consonant inventories is low, since New Guinea languages tend to have smaller-than-average consonant inventories (Maddieson 2005a: 149); it is also worthy of note that the two Trans-New Guinea languages in our sample (Bitur and Urama) have the smallest consonant inventories, typical for Trans-New Guinea languages which typically have consonant inventories in the 10-15 range (Pawley & Hammarström 2017: 82).

Table 1. Phoneme inventory sizes in the languages of this issue (plus Nen and Ende). The 'qualitative' measure places inventory sizes in the proposed brackets for worldwide phonological inventory sizes in Maddieson (2005a,b); note that for vowels, his brackets refer to vowel quality only, not to length, nasalisation, etc.

	C (N)	C (qualitative)	V (N)	V (qualitative)
Yelmek	13	Small	7	large
Ngkolmpu	16	Moderately small	7	large
Nmbo	28	Moderately large	8	large
Nen	23	Average	8	large
Idi	21	Average	8	large
Ende	19	Average	7	large
Bitur	13	Small	5	average
Urama	13	Small	5	average

There is more consistency within the vowel systems, with the range of qualities spanning 5 (Bitur, Urama) to 8 (Nmbo, Nen and Idi). All languages but Bitur and Urama have shorter, centralised vowels whose analytic status is complicated but for which contrastive pairs with other vowels can be found in at least some contexts. The cluster of languages

[2] Though in Ngkolmpu the contrast is not wholly symmetric: there is no contrast among coronal stops, and the voiced velar consonant /g/ is confined to loanwords. This means that the only stop position with an entrenched voicing contrast is the bilabial, /b/ vs /p/.

[3] For Urama this includes /s/, which Brown regards as marginal; omitting it would reduce the Urama C-inventory size to 12.

from the Pahoturi family and the Nambu branch of Yam contrast two short vowels whose quality is more centralised than the others (e.g., /ɐ/ vs. /ɪ/ in Nmbo; /ə/ vs. /ɪ/ in Nen and Idi), and in Idi and Nen these are integrated (especially in Idi) into the system of vowel harmony. The schwa has a marginal status in several of the Yam languages, being predictable (and hence analysable as non-phonemic and epenthetic) in Ngkolmpu and (almost) in Nen (i.e. it is only non-predictable in word-initial position); similar considerations apply in Nmbo, and in the Pahoturi languages Idi and Ende. The Idi system of vowels, described in the paper by Schokkin and colleagues, is a particularly interesting kind of cross-height vowel system, grouping "light vowels" (æ, i, u, ɪ) against "dark vowels" (a, e, o, ə). Urama is the only language in the set to have contrastive vowel length.

Turning to the more distinctive phonemes or phoneme groups of the region, particular mention needs to be made of the following six features.

1. ***Retroflexion*** in the Pahoturi languages Idi and Ende. In both these languages, the retroflex "stops" are actually postalveolar affricates or aspirates, with the point of articulation and the affrication/aspiration each playing a role as acoustic cues (seemingly in different proportions according to the speaker, at least in Idi). It is a moot point whether *place* or *manner* of articulation should be taken as the dominant feature – Schokkin and colleagues focus on place of articulation for Idi (e.g., using /ɖ/), while Lindsey focuses on both simultaneously (e.g., using /ɖʐ/ for what is structurally the equivalent phoneme). Retroflexion does not extend to nasals, but in Ende there is also a retroflex flap contrasting with the alveolar flap/tap – it appears that this contrast has been historically neutralised in Idi. More widely across the region (i.e. sampling into other Yam languages than Nen and Nmbo) retroflex segments are attested in such languages as Namo and Mblafe and probably reconstructable to proto-Yam (Evans et al. 2018, Evans 2019a).

2. ***Labial-velar stops***, voiceless, voiced and prenasalised, are found in Nmbo and Nen – in other words, this phenomenon is confined to (some) languages of the Yam family. They are not found in the other Yam language represented here, Ngkolmpu.

3. ***Rounded bilabial stops***, ***fricatives and nasals*** are found in Nmbo only among the languages of our sample. These contrast with plain bilabial stops, fricatives and nasals.

4. ***Relatively large sets of liquid phonemes*** (>2) are confined to the Pahoturi family: both Idi and Ende have three, augmenting a shared set (/r/, /l/) with a further liquid phoneme in each language (/ɽ/ in Ende and /ʎ/ in Idi); correspondence sets across these languages suggest that proto-Pahoturi had four liquids (*i.e.* all of these). Nmbo and Nen each have two (/r/ and /l/), while Yelmek (/l/ only) and Bitur (/r/ only) have but a single liquid.

5. ***Prenasalised obstruent phonemes*** present a classic difficulty of analysis – when should sequences like /nd/ be treated as unitary and when as clusters? This is particularly acute when certain word-positions, such as word-initial, neutralise the difference between prenasalised and plain obstruents in favour of the plain forms. Part of the reason for the differences in phoneme size for Ende and Idi, on the one hand, and Nen and Nmbo on the other, reflect different analytic approaches: Kashima's article on Nmbo treats these as unitary phonemes, whereas in the article on Idi by Schokkin and colleagues, they are analysed, as in Lindsey's (2021) treatment of Ende, neither as two full segments nor as a unitary segment, but instead the nasal element is treated as an underspecified feature of a particular segment that only surfaces when the segment it is associated with is a non-initial voiced obstruent. In both Idi and Ende this analysis is rendered attractive by the leftward floating of these nasal elements, within a phonological word, when there is a suitable host. In the other analytic direction, Carroll's article marshals evidence for the

unitary status of prenasalised obstruents, drawing on evidence from the general organisation of the phoneme inventory (notably prenasalised velar obstruents, but no velar nasals), phonetics (duration of prenasalised stops is only slightly longer than that of regular stops, whereas combinations of nasal plus stop are nearly double), and phonotactics, adding interesting evidence from the sonority hierarchy as it impacts Ngkolmpu phonotactics. It is clear that much scope remains for deeper investigation, both phonetic and phonological, into the behaviour of phonetic nasal+obstruent sequences across Southern New Guinea, and in our present state of knowledge, we cannot soberly claim that the differences reported on in this special issue always reflect real linguistic differences as opposed to differences of analysis.

6. ***Voiceless prenasalised obstruent phonemes*** are a rare phenomenon worldwide (prenasalised obstruents tending overwhelmingly to be voiced), but in Ngkolmpu the prenasalised obstruents are all voiceless (Carroll this issue), and in Ende there are both voiced and voiceless prenasalised obstruents (Lindsey 2021).

Apart from retroflexion (attested in some Yam languages, though not in any of the three considered here) each of the above features is confined to a single language family and so cannot be used to argue for any kind of diffusion of innovative phonological features. This is shown in tabular form (for consonants only) in Table 2, in which "x" marks the presence of a phoneme in a given language. All consonants found in any language of the sample are included, with some minor adjustments for analytic discrepancies and allophonic ranges.

From this table we can then calculate the similarity of consonant systems using an overlap measure – dividing the number of shared phonemes by some measure of the total number of phonemes. Care has to be taken here in devising the best measure: one could use the number of phonemes in the smallest inventory as the denominator, but that would exaggerate the similarity of languages with small inventories to all the others, e.g. Bitur would end up with a quotient of 1.0 with Nmbo, Nen, Idi and Ende. The phonemes from its very small and non-exotic inventory are a proper subset of the inventories of each of these languages bar Ngkolmpu and Urama – all 13 of its consonants are found in all of the other languages, except that /d/ and /z/ are not found in Ngkolmpu and /z/, /w/ and /j/ are not found in Urama. A partial way of eliminating this effect is to make the denominator the average of the consonant inventory sizes of the two languages being compared; this then lowers the overlap score between Bitur and Nmbo, Nen, Idi and Ende to the range 63.4%–76.5%. An overlap matrix for the languages in our sample is given in Table 3.

As can be seen, the highest overlap scores are between two pairs of related and neighbouring languages – Idi and Ende, which share 90.0% of their consonant inventory, and Nen and Nmbo, which share 86.3%. The third highest score, between Yelmek and Bitur, is likely due to the effects of their small and non-exotic inventories, producing a convergence in shared, ubiquitous segments. The overall range of shared consonant inventories runs from 51.3% (Idi-Urama) to 90.0% (Idi-Ende) with a mean of 68.9%.

Table 2. Matrix showing presence of phonemes across the SNG languages considered here, plus Nen and Ende. For purposes of this comparison, the Idi and Ende retroflex series are treated together, the various allophones of /z/ (fricative and affricated, dental and palatal) are lumped together, and the prenasalised obstruents in Idi and Ende are included in parentheses (though not phonemic, they are accorded a special status in these languages). Urama /s/ is also included in parentheses, though marginal in the language.

	Yelmek	Ngkolmpu	Nmbo	Nen	Idi	Ende	Bitur	Urama
p	x	x	x	x	x	x	x	x
b	x	x	x	x	x	x	x	x
^{m}p		x				(x)		
^{m}b			x	x	(x)	(x)		
b^{w}			x					
t	x	x	x	x	x	x	x	x
d	x		x	x	x	x	x	x
^{n}t		x				(x)		
^{n}d			x	x	(x)	(x)		
ʈ~ʈ^{ʂ}					x	x		
ɖ~ɖ^{ʐ}					x	x		
^{ɳ}ʈ						(x)		
^{ɳ}ɖ					(x)	(x)		
k	x	x	x	x	x	x	x	x
g	x	x	x	x	x	x	x	x
ʔ								x
^{ŋ}k		x				(x)		
^{ŋ}g			x	x	(x)	(x)		
k͡p			x	x				
g͡b			x	x				
^{ŋm}g͡b			x	x				
k^{w}					x			
g^{w}					x			
z~dʒ			x	x	x	x	x	
^{n}dz			x	x		(x)		
m	x	x	x	x	x	x	x	x
m^{w}			x					
n	x	x	x	x	x	x	x	x
ɲ			x	x	x	x		
ŋ	x				x	x		
ɸ			x					
β			x					x
ɸ^{w}			x					
β^{w}			x					
s	x	x	x	x	x	x	x	(x)
^{n}s		x				(x)		
h			x					x
r		x	x	x	x	x	x	x
ɽ						x		
l	x	x	x	x	x	x		
ʎ					x			
w	x	x	x	x	x	x	x	
j	x	x	x	x	x	x	x	

Table 3. Overlap matrix, calculated as the ratio of number of overlapping phonemes to the averaged number of phonemes between the two languages

	Yelmek	**Ngkolmpu**	**Nmbo**	**Nen**	**Idi**	**Ende**	**Bitur**
Ngkolmpu	0.759						
Nmbo	0.585	0.545					
Nen	0.667	0.615	**0.863**				
Idi	0.706	0.703	0.612	0.682			
Ende	0.750	0.743	0.638	0.714	**0.900**		
Bitur	**0.846**	0.759	0.634	0.722	0.765	0.687	
Urama	0.654	0.690	0.585	0.556	0.529	0.625	0.769

Apart from filling a significant gap in our phonetic documentation of the 860 or so Papuan languages, what this special issue does – by concentrating on a single region – is to show just how much phonetic diversity there is in languages spoken in close proximity. This is despite the many settings in Southern New Guinea where bi- or multilingual usage would be expected to have led to convergence. (But see Evans 2019b for arguments that contact can also produce divergence). Almost all the most distinctive features found in the region – retroflexion, rounded bilabials, labial-velars, richer sets of liquid phones – are quarantined within particular language families, suggesting there has been no diffusion of unusual features. The only interesting exception is the appearance of voiceless prenasalised obstruents in both Ngkolmpu, at the western extremity of the Yam languages, and Ende, far out of contact at the northern extremity of the Pahoturi languages.

It goes without saying that this small sample of languages just scratches the surface of the region's phonetic diversity – both the local region of Southern New Guinea, which certainly contains many phonological features not represented here, and the broader region of New Guinea as a whole. We hope that this small selection will whet the appetite of our readers to explore further the phonetics of this little-known and fascinating region.

References

Carroll, Matthew J. 2016. *The Ngkolmpu language, with special reference to distributed exponence*. Canberra: Australian National University. (Doctoral dissertation.)

Döhler, Christian. 2018. *A grammar of Komnzo*. Berlin: Language Science Press.

Evans, Nicholas. 2012. Even more diverse than we thought: The multiplicity of Trans-Fly languages. In Evans, Nicholas & Klamer, Marian (eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st century. Language Documentation and Conservation special publication No. 5*, 109–49.

Evans, Nicholas. 2019a. Australia and New Guinea: sundered hemi-continents of sound. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 16–19. Canberra: Australasian Speech Science and Technology Association Inc.

Evans, Nicholas. 2019b. Linguistic divergence under contact. In Cennamo, Michela & Fabrizio, Claudia (eds.), *Selected papers from the 22nd International Conference on Historical Linguistics*, 563–591. Amsterdam/Philadelphia: John Benjamins.

Evans, Nicholas & Miller, Julia C. 2016. Nen. *Journal of the International Phonetic Association.* 46(3): 331–349.

Evans, Nicholas & Carroll, Matthew J. & Döhler, Christian. 2017. Reconstructing the phonology of proto-Yam. (Paper presented at the Australian Linguistic Society conference, Sydney, December 2017.)

Evans, Nicholas & Arka, Wayan & Carroll, Matthew J. & Choi, Yun Jung & Döhler, Christian & Gast, Volker & Kashima, Eri & Mittag, Emil & Olsson, Bruno & Quinn, Kyla & Schokkin, Dineke & Tama, Philip & Van Tongeren, Charlotte & Siegel, Jeff. 2018. The languages of Southern New Guinea. In Palmer, Bill (ed.), *The languages and linguistics of New Guinea: A comprehensive guide*, 641–774. Berlin: Walter de Gruyter.

François, Alexandre. 2012. The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language* 214: 85–110.

Haudricourt, André-Georges. 1961. Richesse en phonèmes et richesse en locuteurs. *L'Homme* 1(1): 5–10.

Heston, Tyler M. & Locke, Stephanie. 2019. Fataluku. *Journal of the International Phonetic Association* 49(3): 419–425.

Lindsey, Kate L. 2021. Ende. *Journal of the International Phonetic Association*, 1-21. doi: 10.1017/S0025100320000389.

Maddieson, Ian. 2005a. Consonant inventories. In Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.), *World atlas of language structures*, 10–13. Oxford: Oxford University Press.

Maddieson, Ian. 2005b. Vowel quality inventories. In Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.), *World atlas of language structures*, 14–17. Oxford: Oxford University Press.

Olsson, Bruno. 2017. *The Coastal Marind language*. Singapore: Nanyang Technological University. (Doctoral dissertation.)

Palmer, Bill. 2018. Language families of the New Guinea area. In Palmer, Bill (ed.), *The languages and linguistics of New Guinea: A comprehensive guide*, 1–19. Berlin: Walter de Gruyter.

Pawley, Andrew & Hammarström, Harald. 2018. The Trans New Guinea family. In Palmer, Bill (ed.), *The languages and linguistics of New Guinea: A comprehensive guide*, 21–195. Berlin: Walter de Gruyter.