

Automatic categorization of prosodic contours in Bardi

Sarah Babinski & Claire Bower^{*}

Abstract. This study presents preliminary results of an automated prosodic clustering analysis of Bardi, a Nyulnyulan language from Northern Australia, using methods from Kaland (2021). Previous work on Bardi prosody identified several functions of boundary tones and two main phrase types, but stressed that findings were preliminary. Here we extend that work and show evidence for several additional phrase types, as well as confirming the overall accuracy of automated clustering. This work adds to the prosodic typology of Australian languages (cf. Fletcher et al. 2002) and provides further evidence for the functions of intonation beyond demarcation in these languages. When coupled with evaluation by a knowledgeable researcher, this automated approach can greatly expedite prosodic analysis on a large scale and expand our typology of prosodic systems.

Keywords. language science; prosody; phonetics; Australian languages

1. Introduction. Identifying types of phrasal prosody is a particularly difficult problem in language documentation work for two reasons. First, prosodic work requires specialized data collection methods and command of the prosodic literature (Gussenhoven 2004; Himmelmann 2008). For this reason prosodic documentation has not usually been a focus of general language documentation efforts; it is not until recent decades that recording equipment and analysis tools have become accessible and portable enough for this to be the case (Macaulay 2021). Second, prosodic work usually requires deep knowledge of the language's syntax and discourse structure. Intuitive analyses are subject to error based on perceptual biases of non-native speakers (Kaland 2021; Xu 2011). All of this is even more difficult when working with archival materials without access to native speakers' intuitions and judgments about the meaning that prosodic structures encode (Simard & Schultze-Berndt 2011). As a result, detailed intonational studies for endangered languages are rare (Whalen et al. 2022; Macaulay 2021).

Prosodic phenomena are as important an aspect of a language as any other linguistic phenomenon, one of many crucial topics to create a full understanding of a language's phonological system. However, prosodic description has not become a standard aspect of language documentation in the same way that segmental description is (Himmelmann 2008; Macaulay 2021). A major roadblock to describing the prosody of a language is the formation of initial hypotheses. With initial ideas about what prosodic contours to look out for, along with some knowledge of the language, it becomes easier to construct targeted sentences and texts to get more examples of a particular phrase type, and to get multiple speakers to say the same sentences in the same context to study prosodic contours more carefully (Himmelmann 2008).

One possibility to create these initial hypotheses without investing a lot of one's own research time, which is often limited due to funding constraints, is to first run a bottom-up, automatic categorization algorithm to identify phrasal contours. Besides saving time for the researcher to identify likely phrase types, an automatic categorization algorithm has the notable

^{*} The data for this paper was collected with funding from NSF grants BCS 1423711 and BCS 0651118, as well as a community AIATSIS grant awarded to Bessie Ejai and Claire Bower. Claire wishes to acknowledge the Bardi community, particularly the Wiggan, Isaac and Ejai families, as well as One Arm Point Council for support during fieldwork. This work builds on methods and discussion in Babinski (2022). Authors: Sarah Babinski, Yale University (sarah.babinski@yale.edu) & Claire Bower, Yale University (claire.bower@yale.edu).

benefit of sidestepping non-native speaker biases in identifying prosodic phenomena. It can also serve as a complement to the more traditional top-down approaches to phrasal prosody, by providing quantitative evidence to support or disprove working theories.

A small body of recent work has explored the capabilities of automatic methods in identifying prosodic phenomena such as tone and phrasal prosody. This includes Dockum (2017) and Grabowski & McPherson (2019), who have tested methods for automatic categorization of tone contrasts in a variety of languages. Cole & Shattuck-Hufnagel (2016) present Rapid Prosodic Transcription (RPT), which is a fast and easy way of identifying basic prosodic features. Kaland (2021) proposes a toolkit for automatic categorization of phrasal contours which is used in the current study.

The question remains how to correctly infer prosodic categories when there are no native speakers of the language, and how to associate function with prosodic clusters in the absence of such intuitions. Here, we present a case study in automated clustering in combination with a linguist with over twenty years' experience with the language. This work makes use of both careful fieldwork and detailed knowledge of language, as well as computational methods that behave in a largely language-neutral way, in order to approach the question of prosodic categorization. This combination of bottom-up (automatic) and top-down (traditional) analyses can serve to build up a well-rounded understanding of prosody in a language, with evidence from acoustic measurements, speaker consultations, and primary documentary materials can feed into a full analysis of phrasal prosody.

While there is no full analysis of Bardi intonation, there are some comments in the Bardi grammar (Bower 2012; 139ff) about main intonational types (see also Bower et al. 2012). In these publications, Bardi clausal intonation (at least in declarative utterances) is described as having a H* boundary tone, anchored to the first content word of the utterance. There is no work that delineates all prosodic types.

Fletcher et al. (2002) provide some discussion of Kayardild intonation in comparison to the limited amount of research conducted on the intonation systems of other Australian languages. They state (p. 1b) that “it is still not apparent, however, whether there is the same kind of relationship between the lexical stress system and the higher levels of intonational and prosodic organization that you find in languages like English.” Before such connections can be made, however, we must fill the gap in prosodic descriptions of Australian languages, and automated methods are one potential solution.

In this paper, we aim to further explore Bardi phrase-level prosody using automated clustering to group together similar contours. We then try to discover whether the clusters identified automatically have some basis in the language. Such work is hampered by the absence of prior work on both the types of melodies that characterize Bardi intonation, and typologies of Australian prosody that might help situate the Bardi automated findings. We therefore hope that this work is an impetus to further investigations of these topics.

2. Data and Methods. Bardi is a member of the Nyulnyulan family, one of the Non-Pama-Nyungan families of Northern Australia. About 2000 people identify as Bardi, though the number of fluent speakers is much smaller.

The Bardi language materials used in this study were recorded at One Arm Point Community (Ardiyooloon) between 1990 and 2008, and in the town of Derby in the early 1970s. They are narrative recordings, part of the larger set of recordings made as part of a documentation and now

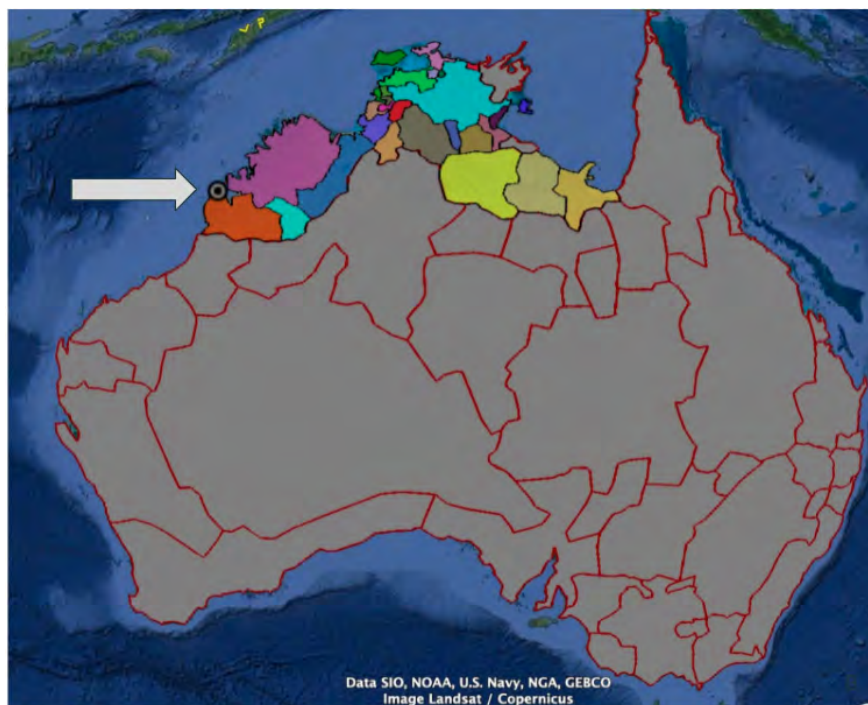


Figure 1. Location of Bardi and One Arm Point (Ardiyooloon), and the Nyulnyulan family (orange on map)

reclamation project which was started by elders in the 1990s. There is an active language reclamation program. Primary publications on the language include a reference grammar (Bower et al. 2012) and dictionary (Aklif 1999 with supplement circulated at One Arm Point in 2008).

While the full audio corpus of Bardi contains approximately 150 hours of recordings, for this project we used a sample of 1,940 utterances from 4 speakers. Each phrase was a segmented utterance (from utterances segmented at initial transcription). The original dataset was cleaned for phrases with very large pitch excursions (which are indicative of pitch tracking measurement errors, such as octave jumps). In addition, very small clusters were pruned (where $n < 6$). This resulted in omission of 4.7% of the original data (that is, 1,847 out of the original 1,940 utterances were retained). A typical example of an intonational phrase is given in Figure 2, from speaker Bessie Ejai.¹

Kaland (2021) presents case studies of the automatic prosodic categorization program along with the freely-available “Contour Clustering” toolkit². Using scripts from this toolkit, f0 measurements were taken in Praat (Boersma & Weenink 2021) at 20 intervals across the time-aligned utterance. Script settings were as follows: minimum phrase duration was set to 100ms, pitch minimum was set at 75Hz, and pitch maximum at 600Hz, time step was 0.01, and stylization resolution was 2 semitones.

F0 measurements were standardized by speaker. Each file was normalized by subtracting

¹ Examples are glossed schematically rather than with full Leipzig glossing, since the full morphological structure of Bardi verbs is complex and not at issue in this paper.

² This toolkit is available for download at <https://constantijnkaland.github.io/contourclustering/>.

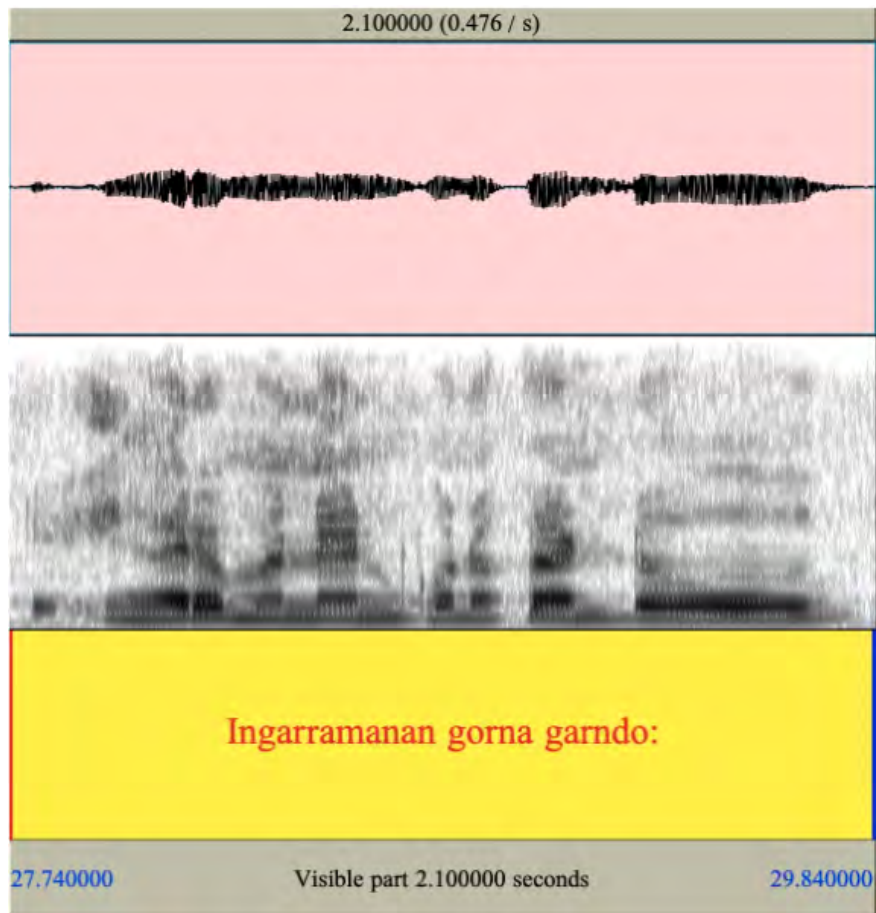


Figure 2. Example phrase

- (1) Ingarramanan gorna garndo:
they.made.it good still
“They made it so it was still good.” [<https://osf.io/u74rx/>]

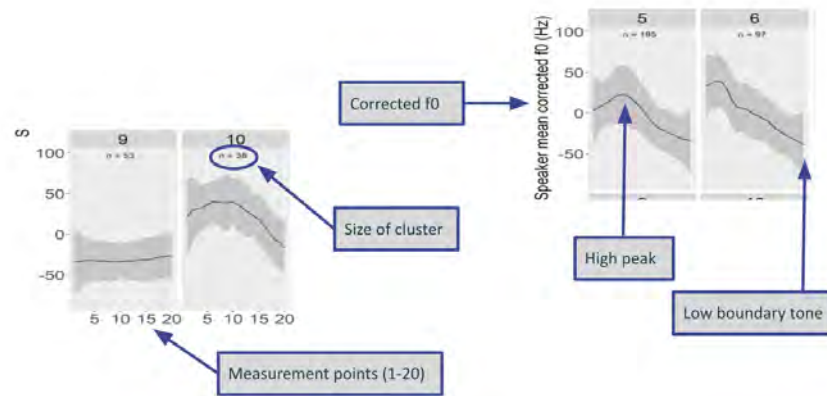


Figure 3. Components of the Kaland (2021) interface

the mean f_0 from each measurement; therefore, in the results that follow, the zero mark indicates average f_0 and positive and negative numbers are higher or lower deviations from that mean. Figure 3 provides a view of the components of the analysis, per the Kaland (2021) interface. Contours are time-normalized based on the 20 measurement points across the phrase.

Contour clustering analysis was run in R using Kaland’s (2021) graphical user interface clustering script. Data was pruned based on the automatic suggestions for subsetting when the number of clusters was set at $n = 25$, as suggested in the workflow presented in Kaland (2021). After pruning, the number of clusters was set at $n = 2$, and then investigated at each $n + 1$ until the first author (naively) judged that additional clusters were uninformative. This is the same sort of judgment recommended in Kaland (2021) to settle on the number of clusters in a relatively objective way apart from knowledge of other work on prosody of the language.

The clustering method used for this analysis is *complete-linkage clustering*. This is an agglomerative method with iterative pairwise clustering, similar to another common clustering method, UPGMA (unweighted pair group method with arithmetic mean) (Sokal & Michener 1958). Clusters are formed by grouping elements together that have the minimum distance between them. This is done in an iterative fashion to build up larger clusters, which means that a new distance value needs to be calculated for the smaller clusters in order to continue the method. For complete-linkage clustering, this new distance is calculated as the maximum distance between elements of the cluster. For example, if the first cluster has elements (a, b) and we want to determine the distance between this cluster and a new element c , we choose the larger of the distances between a and c and between b and c . This is what distinguished complete-linkage clustering from other agglomerative clustering methods; UPGMA clustering calculates this new value as the mean distance between elements of the cluster, and single-linkage clustering uses the minimum distance of the cluster elements.

3. Results.

3.1. OVERALL FINDINGS. The cluster dendrogram for Bardi is shown in Figure 4. The branching here shows a major split separating roughly one third of the data from the rest, a split that corresponds to the two major clusters in Figure 5. The larger of the two clusters (on the right in Fig. 5) corresponds to a mid tone that falls over the utterance, while the smaller cluster (left in

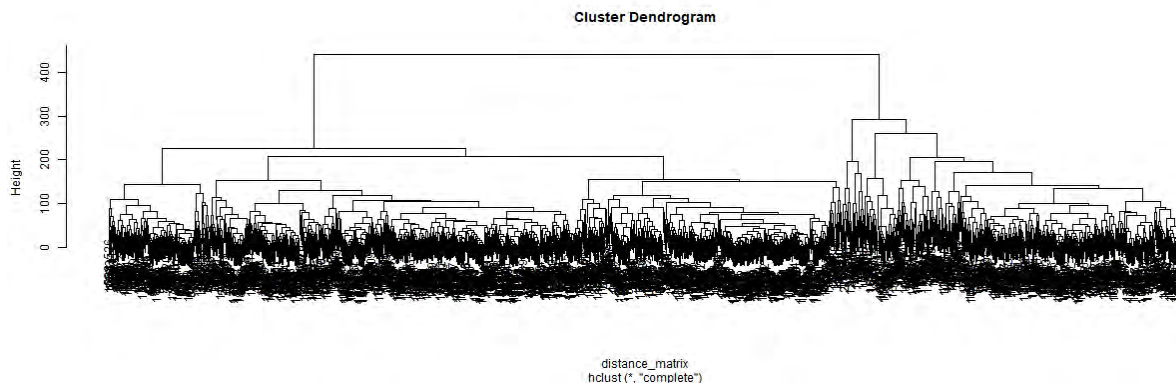


Figure 4. Cluster dendrogram for Bardi, after subsetting of data.

Fig. 5) corresponds to a higher-than-average tone that falls over the course of the utterance.

The contours in Figure 5 show the major split at the top of this dendrogram. The more common contour (Cluster 2 in the figure) shows a flattish contour, beginning at average f0 for the dataset and lowering over the course of the phrase. The smaller cluster (Cluster 1) shows a higher-than-average f0 at the beginning of the phrase, which lowers to average by the end.

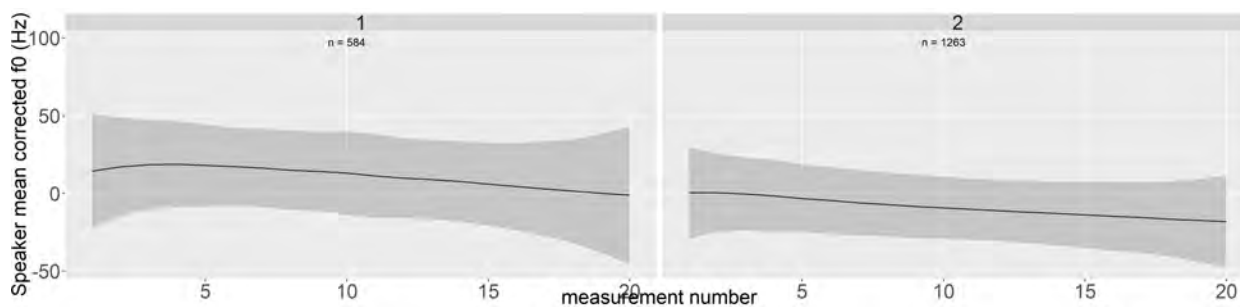


Figure 5. Two major contour clusters in Bardi.

However, looking at a two-way split in contour clusters overlooks the considerable variation present across Bardi phrases. Figure 6 shows the contour clusters detected when there are eight clusters specified, and more distinct phrase types make themselves clear. The flat contour with average f0 remains (Cluster 3 in the figure), now one of three relatively flat contours along with a low-f0 and high-f0 contour (Clusters 4 and 3 respectively). These are joined by two falling contours, one mid-to-low (Cluster 5) and one a more dramatic (but much less frequent) high-to-low contour (Cluster 6). Cluster 7 shows a high f0 gesture that is maintained throughout the phrase, followed by the low boundary tone at the end, while Cluster 8 shows an extra-high contour that seems relatively stable across the phrase, though this cluster seems to have more variation than the others. The only rising tone is seen in Cluster 2, where a slightly above average f0 rises at the end of the phrase.

It is important to note that we do not consider these eight clusters to be an exhaustive, exclusive, and correct delineation of Bardi intonational groupings. Rather, in the following section we evaluate these groupings and attempt to discover a rationale for the clustering. This is, in essence, a *post hoc* justification for the automated clusters. However, if there *are* definable cat-

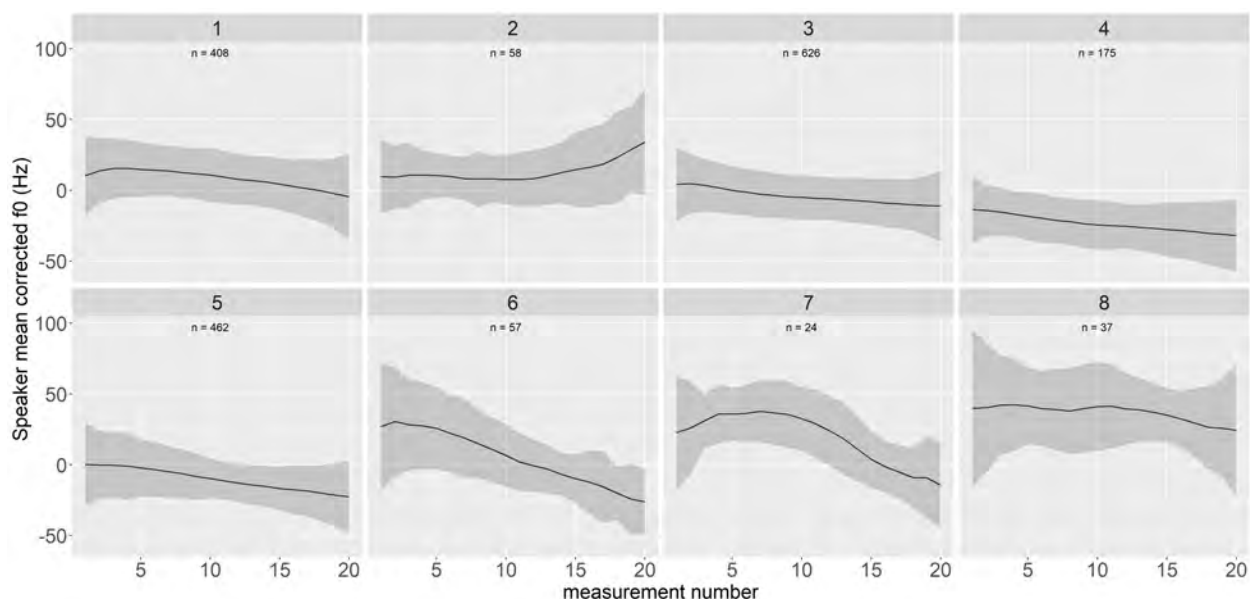


Figure 6. Eight contour clusters in Bardi.

egories that correspond to the clusters identified through the automatic contour similarity measures, this shows the utility of the approach for documentary work on underdescribed languages (even though, after further analysis, we may further refine some of these findings).

3.2. FUNCTIONS OF CLUSTERS. The eight clusters identified in Figure 6 above correspond to different functions in discourse and narrative. While the mappings are not perfect, and there are other speech act types which are not distinguished among these intonation clusters, there are nonetheless interesting mappings picked out which should be the subject of further investigation. That is, there appear to be fairly consistent associations between inflection points in Bardi narratives and the assignment to clusters given here.

Utterances are most commonly assigned to Clusters 1, 3, or 5. Together, those clusters account for about 80% of the utterances. Given that the recordings are all narratives, most of the utterances are declarative. Clusters 1, 3, and 5 have a similar shape, a shallow fall over the course of the utterance. Clusters 3 and 5 begin at the speaker's mean f_0 , while Cluster 1 begins higher (and has a slight rise).

Cluster 2 appears to be particularly associated with list intonation, though some list items also appear in Cluster 1. Cluster 1 also contains utterances which are part of a sequence of clause chaining. Consider, for example, the sentence given in (2) (and illustrated in Figure 7). In this example, Nancy Isaac is telling a story about the time her mother nearly drowned in the strong tidal current while swimming across a narrow passage between two islands. This is one utterance in a swift sequence of events.

- (2) Ginyinggi moowarnin yawoorr inangganana
 3sg hair.3poss it.pulled.it
 '...Her hair was pulling her down...' [<https://osf.io/xwhmc/>]

It is also worth noting that of the 35 examples in Cluster 2, 21 are from one speaker, and 10 are from a second speaker. This might imply that they are speaker-specific, or that there are consis-

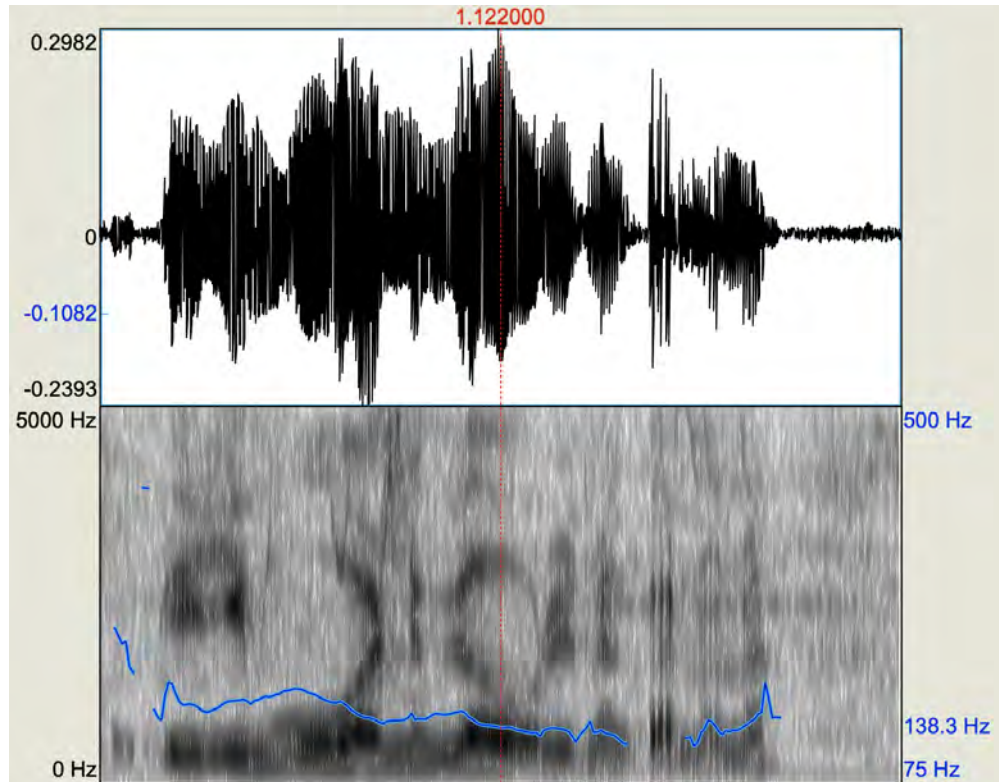


Figure 7. Pitch track and spectrogram for example (2)

tent differences in the realization of lists that show up between speakers' varieties. We have not discussed dialectal or other sociolinguistic differences between the speakers in this survey. Two genders are represented among the speakers, and speakers come from Mainland and Island clan estates. However, because most Bardi documentation has been conducted with a few speakers, it is not clear what features are associated with gendered speech or particular family groups.

Cluster 3 begins with a slight rise, and the contour falls across the intonation unit. It appears to be associated with syntactic units that are part of larger units (or intonational phrases). Note that discussion here is hampered by not yet having a full analysis of the relationship(s) between syntactic units and intonational structures. Bardi has clause chaining, as discussed in Bower et al. (2012). Tentatively, we assume that there can be a macro-intonational phrase that contains a number of sub-units. We also need some way of modeling syntactic units where there is (impressionistically) a single intonational phrase that is broken by one or more pauses.³ The precise nature of this relationship between syntax and prosody is unclear for Bardi, and is one reason for conducting a semi-automatic contour clustering. It is not clear at this point what (if anything) distinguishes Cluster 1 from Cluster 3 or 5.

- (3) Naawa inanggalanan jarroon booroo
 Noah he.lived in.this country
 '[when] Noah lived in this country... ' [<https://osf.io/9zngt/>]

³ An example of this can be found in (8) below.

- (4) gala barn injoonoojin Nyiinyinim
 well he.told.him Nyiinyi
 ‘his father Nyiinyi told him...’

Clusters 3 and 6 both have falls across the utterance. However, the number of tokens in Cluster 6 is much smaller, but the pitch range is larger (with an average of approx. 60 Hz rather than 20 Hz). Cluster 6 is used prior to longer pauses and shifts in the narrative, where Cluster 3 is part of a larger intonational unit.⁴ An example is given in (3), where the utterance (assigned to Cluster 3) is very closely linked to the following line (given here in (4)).

Three clusters—Clusters 4, 6, & 8—appear to be full intonational phrases, characterised by a final low boundary tone. Two examples from Mr. Wiggan are given below; the first in example (5) is a full utterance assigned to Cluster 8, while the second is assigned to Cluster 4, and is given in example (6).

- (5) Ginyingamba ananirr yandilybaron boogoon injoonoojin.
 That’s.why you.should.put.them in.boat he.told.him inside
 ‘That’s why you should put them in the boat,’ he told him. [<https://osf.io/7aybf/>]

- (6) Ginyinggon injoordin injoordinoo jin booroo.
 then it.dried.out it.dried.out his country.
 ‘Then his country all dried right up.’ [<https://osf.io/ehf8t/>]

One of the differences between Cluster 4 and Cluster 6 is that Cluster 4 tends to occur following a pause, while Cluster 6 is not.

Finally, Cluster 7 is mostly used for ‘‘holding the floor’’. That is, utterances in this cluster do not close a turn, but rather signal that the next utterance continues the intonational group.

- (7) agal irr baawanim
 and they child.erg
 ‘And the children...’ [<https://osf.io/z2b6h/>]

Thus in summary, about 80% of the utterances fall into Clusters 1, 3, and 5. Cluster 2 contains lists, and 7 signals the end of an utterance but not the end of a turn. Clusters 4, 6, and 8 are all full intonational phrases, differentiated by whether they are preceded by a pause. Clusters 1 and 3 appear to be part of larger intonational units.

4. Further discussion. While the phrase level cluster analysis produced results where the clusters appeared to link to function, there are additional prosodic features of Bardi narratives which did not appear in the analysis. Bower (2012:144-144) discusses a ‘‘terracing’’ feature of intonational contours. These are long stretches of narratives with two levels of flat intonation, one high in the speaker’s register, the other lower. An example can be seen in (8) below, from Bessie Ejai’s telling of a story about a time her husband was on a boat whose engine broke down, and they were nearly swamped by a whirlpool in the swift and dangerous tidal currents between Sunday Island and the mainland. In this example, the intonation is level (and higher than usual) until the word *baybirraybirr*, at which point it rapidly falls. The high level tone immediately returns, however, on the next word.

⁴ Because of the way that the utterance sample was constructed (by segmenting at pauses), we suspect that some of the utterance types identified in the Clusters below are not, in fact, full intonational phrases.

- (8) galgooriny ingirrinyan galgooriny ingirrinyan. Aamba ginyingg wangalang
swim they.did swim they.did man this young.man
baybirraybirr galgooriny galgooriny. Nyoonoomin ingirrin Boondin
one.after.the.other swimming swimming this.way-then they.were.at Boondin
“They were swimming and swimming. The men, and the young men one after another,
swimming, swimming. Then they came here to Boondin.” [Bessie Ejai: <https://osf.io/dhvx2/>]

The switches between high and low register signal inflection points in the narrative. Since this feature is found across utterances, it is not surprising that it was not detected in this work.

Note that there are intonational melodies and clause functions not identified here, thus far, that might be expected. For example, there is no Cluster exemplified by interrogatives here. This is probably because the narratives that form the bulk of the data used here contain very few questions (the few that exist are in quoted speech). Likewise, commands are not fully absent, but form a very small part of the overall dataset—again, only in quoted speech. A more naturalistic set of utterances would likely produce different sets of clusters.⁵

5. Conclusions. In conclusion, Kaland’s (2021) approach to intonation clustering appears to pick out the major phrase types in Bardi narrative speech, indicating that it is a useful tool for assisting with new prosodic analyses. Unsurprisingly, one cannot simply “read off” the phrase types from the clusters. However, the clusters returned from the automated procedure did correspond to functional intonational phrases, at least to some extent. Automatic clustering did not allow the investigation of higher order prosodic structure (though presumably one could investigate the likelihood that certain phrase types are followed by other phrase types). This method allows the researcher to sort phrases by intonational similarity, allowing for more systematic comparison. It also has the potential to uncover individual speaker differences, or differences related to social factors.

Automatic (i.e. bottom-up) and traditional (top-down) analysis methods for prosody should complement one another. Automatic methods provide the first hypotheses for identifying phrase types, which can then be followed up with perception experiments, speaker consultations, and further research with the primary documentary materials. Likewise, traditional methods can precede the automatic methods. In that case, the automatic methods can help find examples of known phrase types, or uncover categories that were missed initially, such as in stress elicitation. Using automatic methods can enhance traditional analysis by adding potentially large amounts of quantitative evidence and by potentially revealing new insights into the prosodic system such as speaker differences or social variation that may not have come out using traditional methods.

6. Supplementary materials. The sound files associated with this talk are available from the following url: <https://osf.io/a7sr5/>.

⁵ Note that the decision to focus on narratives in fieldwork was deliberate. They are the results of an oral history project begun by Bardi elders and given that in documentation projects with limited time, we chose to focus on this genre. Some other recordings of conversation have been made, along with notes about everyday language use. At the time that these Bardi recordings were made, Bardi had already ceased to be the main language of interaction for most of the Bardi community.

References

- Aklif, Gedda. 1999. *Ardiyooloon Bardi ngaanka: One Arm Point Bardi dictionary*. Halls Creek, Western Australia: Kimberley Language Resource Centre.
- Babinski, Sarah. 2022. *Archival phonetics & prosodic typology in sixteen Australian Languages*. New Haven, CT: Yale University Ph.D. dissertation.
- Boersma, Paul & David Weenink. 2021. Praat: Doing phonetics by computer. Version 6.2.01. www.praat.org.
- Bowern, Claire. 2012. *A Grammar of Bardi* (Mouton Grammar Library 57). Berlin: De Gruyter Mouton.
- Bowern, Claire, Joyce McDonough & Katherine Kelliher. 2012. Bardi. *Journal of the International Phonetic Association* 42(3). 333–351. <https://doi.org/10.1017/S0025100312000217>.
- Cole, Jennifer & Stefanie Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1).
- Dockum, Rikker. 2017. Computational modeling of tone in language documentation: Citation tones vs. running speech in Chindwin Khamti. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 43. 43–73. <https://doi.org/10.5281/ZENODO.2575294>.
- Fletcher, Janet, Nicholas Evans & Erich Round. 2002. Left-edge tonal events in Kayardild (Australian) – A typological perspective. *Speech Prosody* 2002. 295–298.
- Grabowski, Emily & Laura McPherson. 2019. DAPPr: A (semi-)automated tool for pitch annotation. *Proceedings of the International Congress of the Phonetic Sciences* 19. 1704–1708.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Himmelman, Nikolaus P. 2008. Prosody in language documentation. In Jost Gippert, Nikolaus P. Himmelman & Ulrike Mosel (eds.), *Essentials of language documentation*, 163–182. Berlin: De Gruyter.
- Kaland, Constantijn. 2021. Contourclustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association* 1–30. <https://doi.org/10.1017/S0025100321000049>.
- Macaulay, Ben. 2021. The race to document endangered languages, now that we have the technology. *Gizmodo* [online].
- Simard, Candide & Eva Schultze-Berndt. 2011. Documentary linguistics and prosodic evidence for the syntax of spoken language. In Geoffrey L. J. Haig, Nicole Nau, Stefan Schnell & Claudia Wegener (eds.), *Documenting endangered languages*, 151–176. Berlin: De Gruyter. <https://doi.org/10.1515/9783110260021.151>.
- Sokal, Robert R. & Charles D. Michener. 1958. A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin* 38. 1409–1438.
- Whalen, D. H., Christian DiCanio & Rikker Dockum. 2022. Phonetic documentation in three collections: Topics and evolution. *Journal of the International Phonetic Association* 52(1), 95–121. <https://doi.org/10.1017/S0025100320000079>.
- Xu, Yi. 2011. Speech prosody: A methodological review. *Journal of Speech Sciences* 1(1). 85–115. <https://doi.org/10.20396/joss.v1i1.15014>.