Research Article

Constantijn Kaland* and Martine Grice

Exploring and explaining variation in phrase-final f0 movements in spontaneous Papuan Malay

https://doi.org/10.1515/phon-2023-0031
Received July 10, 2023; accepted March 2, 2024; published online March 25, 2024

Abstract: This study investigates the variation in phrase-final f0 movements found in dyadic unscripted conversations in Papuan Malay, an Eastern Indonesian language. This is done by a novel combination of exploratory and confirmatory classification techniques. In particular, this study investigates the linguistic factors that potentially drive f0 contour variation in phrase-final words produced in a naturalistic interactive dialogue task. To this end, a cluster analysis, manual labelling and random forest analysis are carried out to reveal the main sources of contour variation. These are: taking conversational interaction into account; turn transition, topic continuation, information structure (givenness and contrast), and context-independent properties of words such as word class, syllable structure, voicing and intrinsic f0. Results indicate that contour variation in Papuan Malay, in particular f0 direction and target level, is best explained by turn transitions between speakers, corroborating similar findings for related languages. The applied methods provide opportunities to further lower the threshold of incorporating intonation and prosody in the early stages of language documentation.

Keywords: prosody; f0; Papuan Malay; cluster analysis; random forest

1 Introduction

Suprasegmental features of speech such as f0, duration and amplitude, henceforth *prosody*, provide crucial cues about how and what a speaker is communicating. These cues might have a function at different levels simultaneously. Among other things, prosody can signal a speaker's intention to introduce a new topic, indicate the

***Corresponding author: Constantijn Kaland**, Universität zu Köln – SFB 1252, Luxemburger Straße 299, 50939 Köln, Germany, E-mail: ckaland@uni-koeln.de. https://orcid.org/0000-0002-1813-5902
Martine Grice, Phonetics, Institute of Linguistics, University of Cologne, Köln, Germany

end of a turn, highlight a focused constituent or demarcate the end of a phrase. Because of the many different levels at which prosody is potentially meaningful, capturing the exact nature of the form-meaning relationships in a language is highly challenging (e.g. Ladd 2008). As a result, early descriptions of language often overlook prosodic aspects. This is no different for Papuan Malay, the language investigated in the current study. Although some recent studies have begun to investigate Papuan Malay prosody in more detail (e.g. Kaland 2021b; Riesberg et al. 2018), there are still many remaining aspects left for future study. This applies in particular to phrase-level intonation, that has not been described in depth.

The lack of prosodic descriptions in early stages of language documentation has been addressed in the literature. General guidelines for fieldworkers to incorporate prosody in their language descriptions are offered (Himmelmann and Ladd 2008; Jun and Fletcher 2014). To date, the intonation of many languages is described employing autosegmental-metrical approaches (e.g. Jun 2005, 2014), which are centred around the transcription of f0 movements as abstract phonological entities (i.e. pitch accents and boundary tones) that make up an intonation contour (also referred to as the *tune*). Central to this theory is the idea that tunes are compositional in nature; pitch accents and boundary tones are seen as *morphemic* in the sense that, when combined, they contribute to the overall intonational meaning in specific ways. There is however little agreement about which phonological categories should be distinguished in intonational analysis (Ladd 2008, p. 41). These challenges often result in analyses that look at a small number of phonologically relevant meanings (information structure, illocutionary force, for example), as captured by transcriptions that require manual annotations and a level of abstraction that is obtained from highly controlled (scripted) speech. This approach could introduce a researcher bias and a lack of attention to the many different functions that might be expressed by intonational variation obtained from more naturalistic speech. While we do not exclude the existence of phonological structure in Papuan Malay prosody, given the little knowledge to date, the current study focuses on the main functions of prosody at an important location in the phrase rather than on the ways in which these functions are phonologically encoded.

The aims of the current study are therefore two-fold. Primarily, we explore the main sources of phrase-final f0 variation in spontaneous Papuan Malay speech. Secondarily, we demonstrate a combination of two classification techniques that could be applied to other languages for which intonation research is in its early stages. Thus, the approach taken in this study uses time-series f0 data taken from carefully selected phrase-final Papuan Malay words produced in a collaborative task. These data are exploratively submitted to a cluster analysis using the method proposed in Kaland (2021a). Thereafter, the clustered data is analysed using random forests to investigate which potential sources of f0 variation explain phrase-final f0 variation in Papuan Malay. The latter analysis takes into account pragmatic, information structural, morphological, phonological and phonetic factors.

The remainder of this section discusses the literature on the different levels at which variation in f0 movements can be functional (1.1), an overview of what is known about the use of f0 in the prosody in Malay variants (1.2) and a brief introduction on how the methodology applied in the current paper can improve existing ones (1.3).

1.1 Sources of f0 variation at different levels

This section reviews a number of different factors that might cause variation in f0 contours. The overview is limited to the best documented linguistic factors across different languages and therefore provides a generic approach. Nevertheless, this section is meant to provide an illustration of the many sources of f0 variation at different linguistic levels that often affect the f0 contour in simultaneous fashion. Paralinguistic variation in f0, due to the emotional state of the speaker for example, is not taken into account. Nor are other acoustic cues that play a role in signalling prosodic meaning (e.g. duration, intensity, periodic energy).

It is well known that intonation is used pragmatically to organize speaking turns in conversation (e.g. Bögels and Torreira 2015; Couper-Kuhlen and Ford 2004). Although end of turns are often associated with low f0 and turn continuation with mid or high f0 across languages, there are exceptions to this generalisation (e.g. Szczepek Reed 2004). In this respect, it is crucial to note that the notion of *finality* has been associated with low f0 at other levels too. For example, at the discourse level speakers generally use final f0 movements to signal the structure between their own utterances, in particular whether they will continue talking about a certain topic or not (e.g. Swerts and Geluykens 1994; Swerts et al. 1994). It has furthermore been shown that in many languages f0 marks sentence modality. That is, questions are often signalled by a final high f0 whereas statements tend to end with a low f0, but see exceptions in e.g. Greek, Hungarian and Romanian in Grice et al. (2000), African languages in Rialland (2007), Italian varieties in Savino (2012), and Portuguese varieties in Cruz et al. (2022). However, rising intonation was found to be used in a more universal way in repair initiations (such as "huh?" in English) across 22 languages, in which the turn is given to the speaker explicitly in order to elicit a repair of what has been said before (Enfield et al. 2012; Ha and Grice 2017).

F0 is also used to highlight certain parts of the utterance. For example, semantic focus might be expressed by variations in f0 contours, either by prosodically emphasizing specific constituents or by chunking utterances into phrases (e.g. Büring 2010). Other languages set off the focused constituent by means of pausing before and/or after the focused constituent, such as in Japanese (e.g. Beckman and Pierrehumbert 1986; Jun 2014). Many languages of the world apply compression of the f0 range after the focused constituent (post-focal compression; Xu 2011). Languages might also use f0 to signal information status. For example, many West-Germanic languages use specific f0 movements to signal whether certain information is new, is in broad or narrow focus, or contrastive in the discourse context, where contrasts can relate to the previous (backward looking) or following (forward looking) constituent (e.g. Baumann and Grice 2006; Braun 2006; Dahan et al. 2002; Grice et al. 2017; Krahmer and Swerts 2001). Many languages in the Romance family use peak alignment differences in a falling movement to signal narrow and contrastive focus, e.g. Portuguese: (Frota 2016; Frota et al. 2015), Catalan (Prieto 2014), and Italian (Sbranna et al. 2023; Swerts et al. 2002).

At the lexical level, f0 variation might be caused by several factors. For example, it has been shown that tone languages, which use f0 to signal meaning differences among words, show an interaction between (phrase) intonation and the lexical tone. That is, the lexical tones are produced differently depending on the presence of boundary tones (e.g. Cantonese: Ma et al. 2006, Vietnamese: Ha and Grice 2017). When a language has word stress, f0 might be used as a cue (e.g. Thai; Potisuk et al. 1996), although the strength of f0 as a word stress correlate is often considered as weak (e.g. Gordon 2014; Gordon and Roettger 2017). Effects of stress on the f0 might also be found on specific word classes only. This is because pitch accents tend to occur on certain word classes more than others Kochanski et al. 2005. For example, Hualde (2009) reports that in Spanish content words are generally accented, whereas function words are not. Moreover, in Germanic languages, nouns are generally more often accented than verbs (e.g. Ladd 2008).

Speech segments can also affect the f0 level. It is well documented that vowels have an intrinsic f0 level in that open vowels have a lower f0 than closed vowels (e.g. Ladd and Silverman 1984; Whalen and Levitt 1995). This difference was initially found in accented syllables only (not in unaccented ones; Ladd and Silverman 1984; Steele 1986). However, later work could not replicate these results (Kingston 2007). The magnitude of the intrinsic f0 effects was also found to depend on regional dialect and socio-linguistic factors, indicating that these effects were not automatic but deliberate enhancements of phonological contrasts by the speaker (Jacewicz and Fox 2015; Kingston 2007). It is furthermore known that voicing, i.e. whether the vocal folds vibrate for the production of a certain speech segment, potentially affect the shape of f0 movements, especially adjacent to the voiceless elements. In general, voiced speech sounds are the primary carriers of f0 movements, whereby vowels play a more important role as they have more periodic energy (Albert et al. 2018; Barnes et al. 2012). Whilst f0 is often interpreted as a continuous tune, crossing the voiceless segments by means of interpolation (e.g. Mixdorff and Niebuhr

2013), more recently this has been disputed (Barnes et al. 2012). In addition there might be local effects on f0 shape by means of perturbations which are caused by consonants, in particular obstruents (e.g. Kirby and Ladd 2016; Xu and Xu 2021). It is important, therefore, to also take into account the vocalic portion of the segments when investigating the possible factors influencing the f0 shape.

To summarize the brief overview of the many sources of variation in f0, it becomes clear that when looking at a raw f0 contour, we see the product of a potentially complex interaction of many simultaneously active factors. Although this brief overview is likely to be incomplete and confined to linguistic sources of f0 variation, the number of factors listed above strongly call for reproducible, automated and more objective methods of analysis to accurately unravel them. Before exploring which methods would be suitable to do this, the next section summarizes our knowledge of the use of f0 in Papuan Malay and related Malay varieties.

1.2 F0 in Malay prosody

This section discusses the prosodic use of f0 in a number of Malay variants, in particular from the Trade Malay family to which Papuan Malay belongs.

Studies on standard varieties of Malay have found that f0 is used to mark focus. For example, in Singapore Malay (ZLM), peak alignment differences were reported to differentiate between neutral focus and subject focus (Hamzah and German 2014). In Malay as spoken in Malaysia (ZSM) f0 was not found as an acoustic correlate of word stress (Mohd Don et al. 2008; Wan 2012). However, f0 was found to cue conversational turns in a systematic way, such that a falling f0 signals an upcoming turn-transition whereas a rising or level f0 signals turn-keeping, i.e. that the speaker intends to hold the turn and continue speaking (Zuraidah and Knowles 2006).

Detailed f0 timing patterns were reported for Betawi Malay (BEW), a variety spoken in Jakarta, Indonesia (Van Heuven et al. 2008). In particular, this language was reported to lack word stress and make use of what is referred to as phrase accents. These accents were highly variable at the surface; the result of two interacting factors. That is, the default rise-fall pattern starting on the pre-final syllable in the phrase is realised on the final syllable when there is a schwa in the pre-final one or when the word is phrase-final. These two conditions work in a gradient way, such that they contribute cumulatively to the likelihood that the phrase accent is realised finally instead of pre-finally.

Standard Indonesian (IND) as spoken by Javanese was also analysed as stressless and f0 movements were reported to be mainly phrase boundary phenomena, without alignment to a particular (stressed) syllable (Goedemans and Ellen Van 2007). However, Toba Batak speakers of Standard Indonesian produced f0 movements consistently on the penultimate syllable (in the word) and accepted this pattern as a listener to a larger extent than f0 movements with a different alignment. The Toba Batak result was interpreted as an indication for word stress in this language. As this outcome contrasted with other varieties of Indonesian, it showed the importance of accounting for the language variety in prosody research in this area (Goedemans and Ellen Van 2007).

Of the Trade Malay varieties (see Paauw 2009 for an overview) Manado Malay (XMM), Ambonese Malay (ABS) and Papuan Malay (PMY) have already been investigated for f0. For Manado Malay f0 was reported to correlate with different focus contexts (Stoel 2007). That is, subject-, object- and verb-focus are indicated by an f0 movement that aligns with the focused word. Contrastive (narrow) focus was not marked by f0 movements, nor were f0 movements on discourse particles analysed as accents. Manado Malay was also reported to distinguish question types using different f0 movements phrase finally (a fall in polar questions, rise-fall in information questions and a rise-rise sequence in echo questions).

For Ambonese Malay, f0 was investigated as a potential correlate of word stress, focus and phrase prosody (Maskikit-Essed and Gussenhoven 2016). This study found that f0 distinguished between declarative (final fall) and non-declarative phrases (e.g. polar- and wh-questions; final rise), but that there was no f0 variation due to stress or focus. In addition, no stable alignment of the final f0 movements was found.

For Papuan Malay, f0 was found to be a weak correlate of word stress, whereas duration, vowel quality and spectral tilt were more strongly correlated with word stress (Kaland 2019). Contrastive focus did not show any effects on f0 contours (Kaland et al. 2023; Kaland and Himmelmann 2020). At the phrase-level, the largest f0 movements were found in final positions, which was interpreted as having a boundary marking function (Kaland and Baumann 2020; Riesberg et al. 2018, 2020). However, both in Kaland et al. (2023) and in Kaland and Baumann (2020) the pre-final (stressed) syllable in the phrase often showed a rising f0 movement, suggesting a privileged status of this syllable in this position, in particular for content words in the Kaland and Baumann (2020) study. This tendency was not found in positions earlier in the phrase, plausibly indicating an intricate interplay of word stress and phrase prosody (see also Kaland and Gordon 2022). It should be noted that despite the tendency for an f0 rise in pre-final syllables, the investigated spontaneous speech data in Papuan Malay showed more variation in phrase-final positions, a phenomenon which has not been fully understood to date (e.g. Kaland and Baumann 2020).

Summarizing the overview of Malay varieties, it becomes clear that intonation studies differ in the number of factors they investigated as potential sources of f0 variation. In addition, the Malay varieties plausibly differ in their prosodic structure. Although it should be noted that the latter conclusion largely depends on how comparable the methodologies are in the existing studies. Currently, they vary widely as to whether or not the approach was purely qualitative or (also) quantitative. In this case, it is worth recalling the different analyses on the existence of word stress (see e.g. Kaland 2021b for an elaborate discussion).

1.3 New approaches to f0 variation

As outlined above, common theoretical approaches to intonation have reached only limited consensus on which meaningful categories lie beyond the contours produced on the surface. It becomes clear that many factors, linguistic and non-linguistic, determine the contour that speakers ultimately produce. For the linguistic factors, several levels could be distinguished (Section 1.1). As shown from the discussion in Section 1.2, no Malay variety has been investigated for all of the seven factors provided in Table 1. These seven factors cover a wide range of levels at which f0 can meaningfully vary, however, they do not constitute a complete set. Thus, the prosody of the languages discussed above, and likely many other underdocumented languages, is far from being researched in a comprehensive way. This is at least partially the result of a relatively small number of studies that rely on manual intonation labels, Manual annotation is time consuming and prone to researchers' biases (Stoel 2007). It also limits the numbers of speakers that can be analysed (Maskikit-Essed and Gussenhoven 2016), may also be responsible for the number of prosodic functions analysed being reduced to make the task manageable (Hamzah and German 2014; Zuraidah and Knowles 2006) and often tends to be applied to scripted speech. There is thus a need to supplement these more traditional approaches with a methodology that makes use of non-human classification methods and does not inherently pose limits on the amount and type of speech data, nor on the number of potential sources of variation that could be investigated.

Table 1: Literature overview of factors causing f0 variation in Malay varieties (– = not reported, y = variation, p = partially/weakly, n = no variation).

Factor	ZLM	ZSM	BEW	IND	XMM	ABS	PMY
Segments	–	–	y	–	–	–	–
Word stress	–	n	n	n	–	n	p
Word class	–	–	–	–	y	–	y
Phrase boundary	–	–	y	y	–	y	y
Focus	y	–	–	–	y	p	n
Question types	–	–	–	–	y	y	–
Turn-taking	–	y	–	–	–	–	–

The current study relies on a cluster analysis approach to document f0 variation proposed in Kaland (2021a). This analysis provides an overview of which contours can be distinguished in a given speech dataset. The number of clusters can be varied according to the specific hypotheses at hand and offers the researcher an output ranging from a course-grained grouping of contours (small number of clusters) up to a fine-grained analysis for which virtually all relevant f0 differences can be distinguished (large number of clusters). Although the outcome does not a priori allow the researcher to draw conclusions on which f0 variations are meaningful, it provides an unbiased classification of contour differences with any desired level of granularity, ranging from only the largest (and most salient) ones to (potentially insignificantly) small ones. In order to usefully interpret the clustering results, the current paper proposes additional steps in the analysis to come to a more confirmatory study of intonation variation in a given language. First, we provide a way to gain a higher level of control over spontaneously produced speech. This is done by taking an informed subset of the collected data allowing for confirmatory analysis and reducing the number of required manual annotations. Second, on this subset a subsequent random forest analysis is performed to investigate the contribution of each hypothesized factor on the f0 variation as classified by the cluster analysis.

The combination of cluster analysis and random forest is used here as a novel method. It separates the exploratory stage of data analysis from the confirmatory one, which is crucially different from previous methods proposed in the literature. For example, functional principal component analysis (FPCA) directly targets the sources of numerical variation between f0 contours, which can be linked to shape features such as peak height, f0 range, steepness, etc. (e.g. Arvaniti 2019; Gubian 2011; Gubian et al. 2015). However, to understand the linguistic functions of f0 contours, it is more useful to categorise f0 contours by their overall similarity, rather than knowing which shape features underlie the categorisation. In this context, it needs to be acknowledged that f0 is just one of the prosodic features that might contribute to linguistic functionality, similar to temporal and amplitudinal features. Cluster analysis, just like FPCA, takes into account the contour as a whole, with any (combination of) shape feature(s) being a potential candidate on the basis of which clusters can be formed. The cluster analysis' output is then taken as the input for the random forest analysis, with hypothesized linguistic sources of variation being tested in a confirmatory way. In this way, the cluster analysis affects the outcomes of the random forest. Thus, the variation that is not captured by the cluster analysis will not be revealed by the random forest analysis either, although the latter would in theory be capable of doing so. This is important to realize when interpreting the results. For example, the possibility cannot be entirely excluded that the cluster analysis misses subtle, linguistically meaningful variation in f0, as it tends to cluster numerically large variation first, see also Kaland (2021a) on this issue. It is therefore important to apply a cluster analysis that is able to separate f0 contours into different clusters on the basis of small shape differences. Section 2.2 discusses how this was done in the current study and the following subsection sums up the research aims.

1.4 Research aims

The main goal of the current study is to investigate the sources of f0 variation in Papuan Malay. Research on Malay varieties has shown that many sources of f0 variation are left uninvestigated (Section 1.2). For Papuan Malay, the most relevant phrase position is the final one and the variation found there is still not fully understood (Kaland and Baumann 2020). The current study makes use of two classification techniques, one unsupervised (clustering) and one supervised (random forests). These techniques introduce reproducibility, reduce potential researcher's biases interpreting form-function relationships of f0, and crucially impose no limits on the number of forms and functions that can be investigated in a single study (1.3). In this way, the applied methods are not bound to Papuan Malay, but allow for a crosslinguistic application in early intonation research on underresearched languages.

In order to connect the unsupervised and the supervised analysis the current study makes use of manual labelling of a number of potential sources of f0 variation. This approach is applied to Papuan Malay phrase-final words, produced in a task-oriented, unscripted setting, using a subset of the factors of f0 variation discussed in Section 1.1. By performing a random forest on clustered and labelled data, the method applied in this study aims to predict which forms (clusters) are explained by which functions (labelled sources of variation). The main research question addressed in this study is therefore:

Which sources of f0 variation determine phrase-final intonation contours in Papuan Malay?

- Turn transition
- Topic continuation
- Information structure (given/new; backward/forward contrast)
- Word class
- Syllable structure
- Voicing
- Intrinsic f0 of vowels

Section 2 outlines how the different sources of f0 variation were taken into account methodologically. In Section 3 the results are given and Section 4 provides a discussion and conclusion.

2 Methodology

2.1 Recordings and annotation

An interactive task was designed in which pairs of participants (dyads) described tangram figures for each other (as used in Riesberg et al. 2018, 2020; Savino et al. 2019). Tangram figures are the outcome of the tangram dissection puzzle, consisting of seven shapes (five triangles, a square and a parallelogram), see Figure 1 for an example. Participants both had four tangram figures on a card of which one was indicated with an arrow. Their task was to find out whether the indicated figure was the same (matching) or different (mismatching). While one participant described the indicated tangram figure (describer), the other one checked whether the arrow on his/her card indicated the same picture or not (matcher). The roles switched after each completed description. The speech of both participants was recorded and then transcribed and translated into English by native Papuan Malay speakers. The transcriptions were done at the level of intonation units (Chafe 1994; Halliday 1967) by two language experts with the help of native speakers, see further details in

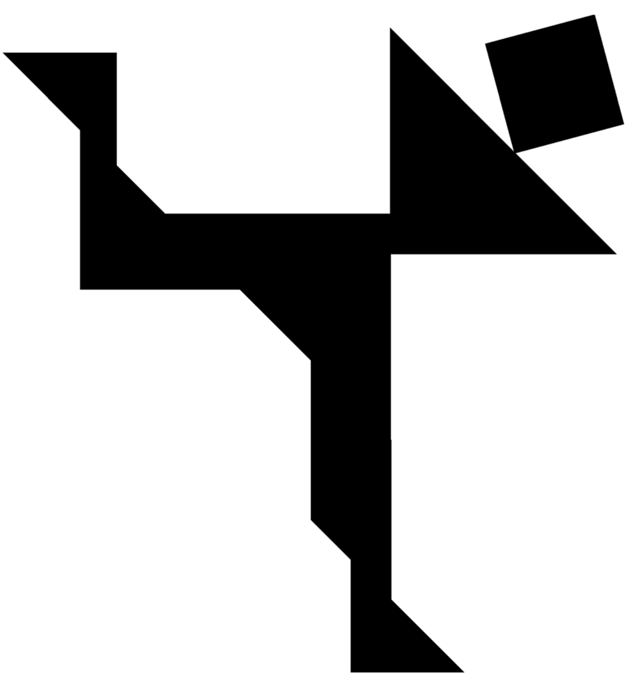

Figure 1: Example of a tangram figure.

Riesberg and Himmelmann (2012). Subsequently, the final words in all units were counted. Units which had multiple instances ending in the same disyllabic word (N = 324 word contours from 35 speakers, M duration 474.55 ms, SD 172.89 ms) were selected for further analysis (see Appendix I for the number of words per speaker). Disyllabic final words were chosen to provide comparable segmental durations for cluster analysis.

2.2 Measurements and clustering

Time-series f0 measures were taken for all final words in the selected units, following the method proposed in Kaland (2021a). Forty measures per word were taken and contours were speaker normalized to account for differences in overall f0 level and f0 range (e.g. due to gender differences) using standardisation (Rose 1987). Note that with the use of this normalization register differences within the range of the speaker are preserved. This is important as register differences have been found to be meaningful in other languages (e.g., Simard 2013). Complete linkage hierarchical clustering was carried out, based on Euclidean distances to represent the contour differences. The ideal number of clusters was evaluated using the method based on information theory (Kaland and Ellison 2023). This method seeks the minimal description length (MDL; Rissanen 1978) for a dataset, under the assumption that this is the optimal representation of that set. The MDL clustering evaluation method was developed and tested specifically for use with f0 contours. For this method, the number of clusters ranged from 2 to 15 and the *bending factor* (i.e. the degree of dependency between adjacent f0 measurement points) was set to four, following recommendations in the software application and manual (Kaland 2021a). This dependency metric optimizes the U-shape of the evaluation curve, in order to identify its lowest (MDL) point.

The optimal number of clusters according to this evaluation method was seven (Figure 2). Nevertheless, the number of clusters was set to eight, in order to allow for a small amount of additional variation captured by the cluster analysis. Thus, with 8 clusters the analysis was able to obtain course-grained differences between the contours (rises versus falls) as well a number of finer grained differences in alignment, steepness, curvature (convex/concave shapes), and f0 range. All of these are dimensions along which f0 variation is potentially meaningful. The degree of overstating the f0 variation by choosing eight instead of seven clusters is relatively small, given the shallow rise of the evaluation curve for higher numbers of clusters (7–11 in Figure 2).

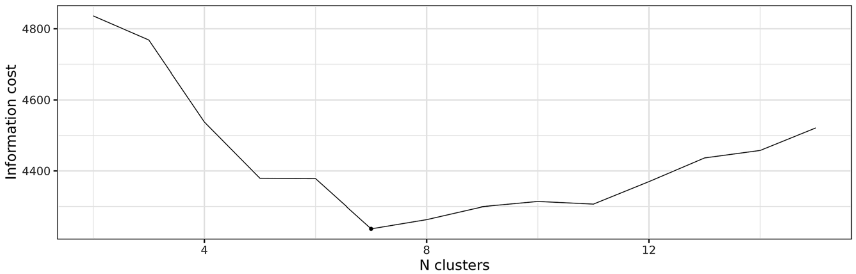

Figure 2: Evaluation of the optimal number of clusters.

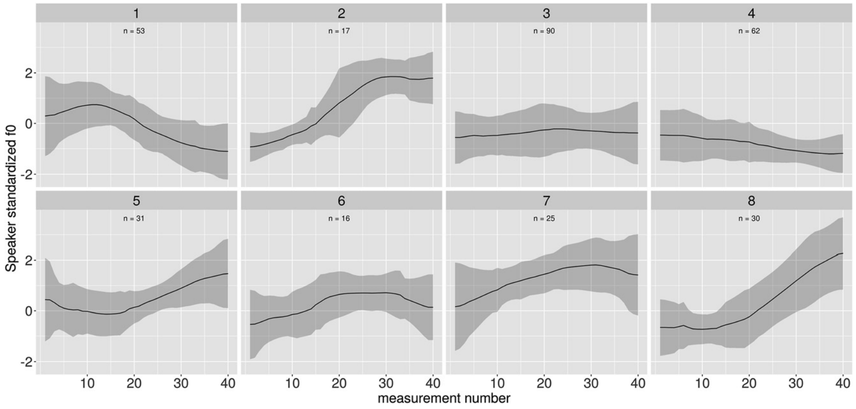

Figure 3: Visualisation of the f0 contours obtained from cluster analysis with 8 clusters assumed.

The resulting set of contours from the analysis with eight clusters (Figure 3) shows a steep fall (1), a steep early rise (2), a flat contour (3), a shallow fall (4), a shallow late rise (5), shallow early rise (6), a long rise (7) and a steep late rise (8).

2.3 Labelling

After cluster analysis contours were selected if the words to which they belonged, occurred in at least four different clusters. This was done to obtain a set with minimal segmental variation and maximal contour variation. In this way, clusters are maximally homogeneous for their segments, facilitating the comparison of different contours. The resulting subset (*N* = 133 word contours from 34 speakers, *M* duration 480.92 ms, SD 204.70 ms) was then labelled for several linguistic factors (see Appendix I for the number of words per speaker). The following provides an explanation of how the labels were assigned with glossed examples for each factor (numbers), level (letters) and speaker (roman numerals), with the target word in bold face.

For each word it was indicated whether it was turn-medial or turn-final; two levels, see (1), referred to as continuation or ending. Backchannels by the interlocutor (such as “mmm” or “yes”, glossed as INTJ – interjection) were not counted as turns and thus an utterance followed by a backchannel was still turn medial if the primary speaker continued to speak during or after the backchannel.

(1) a. Turn continuation
I Kaki eh.
leg HES
‘A leg’
Kaki sebla **angkat**
leg side lift
‘Its leg is raising.’
II Mmm.
INTJ
‘Mmm.’
I Sebla tidak.
side NEG
‘The other leg is not.’
b. Turn ending
I Kanan yang ber-diri.
right REL INTR-stand
‘The right leg which is standing.’
Trus ada.
CNJ exist.
‘Then there is.’
Ke kiri de **angkat**.
to left 3.SG lift
‘The left he is raising.’
II Di ber-diri mengadap ke mana…
3.SG INTR-stand facing to where
‘He is standing facing where …’

We also indicated whether or not the speaker continued the topic that was started in the turn unit to which the word belonged; two levels, see (2). That is, if in the next unit the speaker was still describing the same tangram figure, this would count as topic continuation. This would also be the case if the other speaker, after taking the turn, continued to describe the same tangram figure. Describing different visual aspects (i.e. of another figure or another part of the same tangram figure) or a shift from describing the visual aspects of a tangram to an explicit negotiation of whether both speakers had indeed matching or mismatching tangrams were not considered topic continuations.

(2) a. Topic continuation

I	Ini	de	sama	seperti.		
	DEM.PROX	3.SG	same	similar.to		
	'This is the same as.'					
	Kelinci	yang	ko	bilang	**jadi**.	
	rabbit	REL	2.SG	say	so	
	'The rabbit like you said.'					
	Tapi	de	punya	ini…		
	but	3.SG	POSS	DEM.PROX		
	'But you have this…'					

b. Topic ending

I	Ai,	ko	pu	ruma	tu	banyak.
	INTJ	2.SG	POSS	house	D.DIST	many
	'Hey, you have a lot of houses.'					
	Sa	punya	satu	saja	**jadi**.	
	1.SG	POSS	one	only	so	
	'Mine is only one, so.'					
II	Mmm.					
	INTJ					
	'Mmm.'					
I	Ko	lia	yang	ada	anak	panah.
	2.SG	see	REL	exist	child	arrow
	'You see there's an arrow.'					
	Itu	itu	suda			
	D.DIST	RDP	already			
	'That's it.'					

Information status and contrastivity (henceforth: information structure) were analysed for each word. That is, it was analysed whether the word was mentioned before in the description of that specific tangram (given) or not (new), or whether that word contrasted with a word in the previous unit (backward contrast) or in the next unit (forward contrast); four levels, see (3).

(3) a. New information

I Segi tiga.
angle three
'A triangle.'
Yang di depan-nya itu dia.
prep in front-POSS DEM.DIST 3.SG
'Which is in front of it.'
Lurus ke **bawa**.
straight to bottom
'Straight downwards.'
Tetapi yang di blakang itu.
but PREP in backside DEM.DIST
'But the one behind it.'

b. Given information

I Di **bawa**$_{new}$ dia punya ekor blakang.
PREP bottom 3SG POSS tail backside
'In that he has a back tail.'

II Ekor belakang ada…
tail backside exist
'In the back of its tail there is…'

I Ada persegi.
exist square
'There is a square.'

II Persegi ke **bawa**$_{given}$.
square PREP bottom
'Square downwards.'

c. Backward looking contrast

I Kotak di atas, io.
box PREP top yes
'A box is on the top, yes.
Tapi dia bentuk **bawa**.
but 3.SG shape bottom
'But its shape is at the bottom.'

d. Forward looking contrast

I Kalo di sa pu gambar begini trada segi tiga di **bawa**.
If PREP 1.SG POSS picture like.this NEG angle three PREP bottom
'Then in my picture there is no triangle at the bottom.'
De segi tiga di tenga.
3SG angle three PREP middle
'Its triangle is in the middle.'

Table 2: Overview of all words in the selected data with their English translation, annotated word class following Kluge (2017), the number of occurrences in the dataset (*N*), CV syllable structure, portion of voiced segments, and vowel sequence.

Word	Gloss	Class	*N*	CV structure	*v* Portion	V sequence
angkat	to lift	verb	8	VC.CVC	0.60	/a/–/a/
anjing	dog	noun	8	VC.CVC	1.00	/a/–/i/
atap	roof	noun	8	V.CVC	0.50	/a/–/a/
atas	top	noun	7	V.CVC	0.50	/a/–/a/
bawa	bottom	noun	10	CV.CV	1.00	/a/–/a/
duduk	to sit	verb	14	CV.CVC	0.80	/u/–/u/
ekor	tail	noun	10	V.CVC	0.75	/e/–/o/
gambar	picture	noun	13	CVC.CVC	1.00	/a/–/a/
ini	DEM.PROX	demonstrative	8	V.CV	1.00	/i/–/i/
jadi	so	conjunction	7	CV.CV	1.00	/a/–/i/
kanan	right	adverb	11	CV.CVC	0.80	/a/–/a/
miring	be sideways	verb	11	CV.CVC	1.00	/i/–/i/
ruma	house	noun	11	CV.CV	1.00	/u/–/a/
topi	hat	noun	7	CV.CV	0.50	/o/–/i/

Each word was labelled for word class, largely following the classes distinguished in Kluge (2017): adverb, conjunction, demonstrative, noun, and verb (all classes), totalling five levels. Table 2 provides an overview of all words and their classes.

The phonological composition of each word in terms of vowels and consonants was indicated using CV syllable structure annotations. This resulted in six different structures; CV.CV, CV.CVC, V.CVC, VC.CVC, CVC.CVC and V.CV (see also Table 2).

At the phonetic level, two additional factors were coded to investigate the contribution of the segments to the contour shape (see also Table 2). Vowels and voiced consonants are the main carriers of f0 movements and could therefore affect the shape of the f0 contour. This was checked by indicating the voiced portion by subtracting the number of vowels and voiced consonants from the total number of segments in each word. Although contours were partially smoothed by the time-series measures, the smoothness of the contour could still have been affected by the size of the voiced portion in the words. The obtained voiced portions were 0.50, 0.60, 0.75, 0.80 and 1.00. In addition, the vowel sequence was coded for each word to investigate the effect of intrinsic f0 (Lehiste and Peterson 1961) on the f0 contour. It is known that open vowels have a lower intrinsic f0 than closed vowels. The obtained sequences (hyphens indicate syllable boundaries) were "/a/–/a/", "/a/–/i/", "/e/–/o/", "/i/–/i/", "/o/–/i/", "/u/–/a/." and "/u/–/u/".

Note that the data did not contain questions, so that modality could not be usefully labelled and therefore not be taken into account as a potential source of f0 variation.

2.4 Statistical analysis

Random forest analysis was carried out using R (R Core Team 2019; R Studio Team 2019) and the package "ranger" (Wright and Ziegler 2017). The response variable was cluster number (eight levels, see Figure 3) and the predictors were *turn transition* (2 levels), *topic transition* (2 levels), *information status* (four levels), *word class* (5 levels), *syllable structure* (6 levels), *voiced portion* (4 levels), and *vowel sequence* (7 levels). In addition, a control-predictor *word* (the Papuan Malay word; all 14 levels) was added. *Word* is not expected to be of any predictive value and should therefore have a low variable importance. Therefore, variable importance values of other predictors that lie around or below the one of the control-predictor can be used as an additional indication of which predictors do not affect the response variable at all.

The number of trees in the analysis was increased in steps of 10,000, starting from 10,000 trees. This procedure enables an assessment of the stability of variable importance values. That is, it is important to obtain a stable ranking of the predictor variables. If repeating the random forest analysis with a certain number of trees shows different rankings across the repetitions, the number of trees is too low. The variable importance of the predictors reached a stable ranking around 50,000 trees, which was taken as the final number of trees. The number of randomly preselected predictors was set to the square root of the total number of predictors in the analysis ($\sqrt{8}$), and variable importance mode was set to "permutation". These settings are recommended for analyses with correlating predictors, following Strobl et al. (2008), and Strobl et al. (2009).

3 Results

The variable importance plot (Figure 4) shows that most predictors showed an importance value below zero. This is an indication that prediction was more accurate without these predictors. The dummy-predictor *word* was lowest ranked, indicating that this predictor worsened the accuracy of the random forest analysis the most of all predictors. The only predictors for which a positive value was found, were *turn transition* and *voiced portion*. This indicated that turn transitions predicted the f0

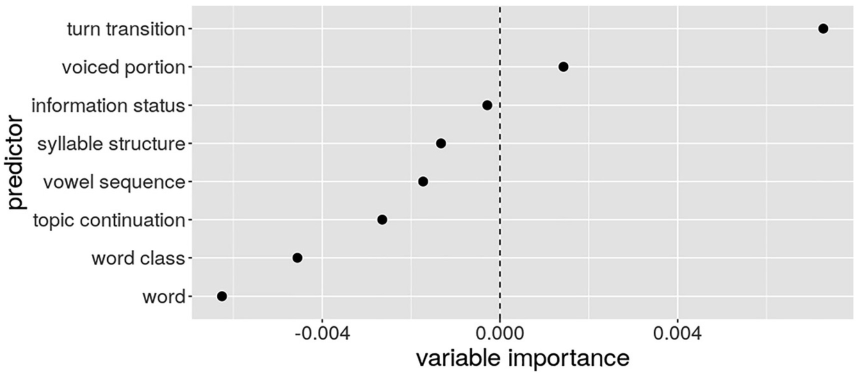

Figure 4: Variable importance plot after random forest analysis with the predictors ranked from high (top) to low (bottom).

Table 3: Proportion of turn ends per cluster.

Cluster: contour	Proportion turn ends
C1: steep fall	0.47
C2: steep early rise	0.00
C3: flat contour	0.10
C4: shallow fall	0.40
C5: shallow late rise	0.06
C6: shallow early rise	0.00
C7: long rise	0.00
C8: steep late rise	0.00

variation found in the clusters the most, followed by the size of the voiced portion. As for the former, Table 3 provides an overview of the proportions of turn ends per cluster.

The proportions in Table 3 show that turn ends were mostly found in cluster 1 and cluster 4, although they never made up more than half of all the cases in one cluster. Other clusters showed less or no turn ends, indicating that overall turn continuations were found in the majority of contours. A chi-square test on the data in Table 3 revealed that the proportions of turn ends were not randomly distributed over the clusters: [$\chi^2(7, N = 133) = 29.39$, $p < 0.001$].

Inspection of Figure 3 reveals that the contours with the highest proportions of turn ends show a falling movement. In addition, it appears that for steeper falls (cluster 1) more turn ends were found than for shallow falls (cluster 4). Vice versa, the steeper the rise, the less turn ends were found (cluster 2 and 8 in particular). To investigate whether f0 range and direction of movement indeed correlate with turn transitions, the speaker-normalized f0 data were used to compute an informed measure of f0 range. That is, for each of the 133 contours, the first measurement point (1) was subtracted from the last measurement point (40). If the outcome resulted in a negative value, the contour was labelled as “fall”. If the outcome resulted in a positive value, the contour was labelled as “rise”. Note that this is a crude measure capturing only the overall direction of the contour, not the local rises or falls. Then, the measurement point with the lowest f0 was subtracted from the measurement point with the highest f0, in order to obtain a measure of f0 direction and excursion. The outcome was then converted into a negative value if the contour was labelled as a fall and left unchanged (positive value) if the contour was labelled as a rise. In this way, the f0 excursion measure included the overall direction of the contour as well as the excursion size. Results showed that the mean f0 excursion for turn ends was –0.77 (*1.69*) and for turn continuations 0.80 (*2.02*). Pearson’s product moment correlation analysis on the proportions of turn ends and the f0 excursion values indicate a weak negative correlation: [$r(131) = -0.30$, $p < 0.001$], confirming that both direction and f0 excursion affected turn transition such that the steeper the rise the more turn continuations and the steeper the fall the more turn ends.

As for the predictive value of the voiced portion, Table 4 shows that for some clusters a voiced portion of 1 was by far the most frequent of all possible sizes (cluster 1, 3, 4, 6 and 7). However, other clusters did not show such a strong tendency (cluster 2, 5 and 8), in that the voiced portion sizes were more equally frequent. It is not immediately clear what caused the predictive value of the voiced portions on the clustering. It can be seen that more evenly distributed voiced portion sizes are all found in contours which have a rising part towards the end of the word (i.e. in the final syllable), see Figure 3. It should furthermore be noted that the variable importance of the voiced portion was much lower than the one for turn transition (Figure 4).

Table 4: Number of contours for each cluster and size of voiced portion.

Cluster: contour	**Size of voiced portion**				
	0.5	**0.6**	**0.75**	**0.8**	**1**
C1: steep fall	3	0	1	3	12
C2: steep early rise	2	0	0	1	1
C3: flat contour	5	0	2	9	14
C4: shallow fall	5	2	1	1	16
C5: shallow late rise	3	3	1	5	6
C6: shallow early rise	1	1	1	1	3
C7: long rise	1	1	3	1	11
C8: steep late rise	2	1	1	4	5

4 Discussion and conclusion

The random forest analysis on the combined results of the clustering and labelling showed that with eight different contours and seven potential sources of variation, turn transitions were able to explain a great deal of the f0 variation. Acoustic analysis confirmed that the steeper the rise, the stronger the signalling of turn continuation, and that the steeper the fall the stronger the signalling of turn ends. Figure 3 additionally reveals different f0 (target) levels at the end of the phrase. It appears that low/falling f0 levels are reached in the clusters where most turn ends are found (cluster 1 and 4). Given that these clusters were the only ones with falling movement reaching well below the speaker's mean (0 in the standardized measures), both the direction and the target level are likely to be important phonetic features of the f0 contours in Papuan Malay turn-taking. Clusters 2, 6 and 7 have a slight fall (or what Knowles (1978) and Cruttenden (1997) refer to as "slump" in a "rise-plateau-slump" contour – convex shape). Here the final f0 was not as low as in clusters 1 and 4. The smaller predictive value of voiced portion seems to be related to rises occurring word-finally (Section 3).

Although it seems a trivial observation that turn cues are realised at phrase ends, it is important to consider that the literature has shown that IP boundaries are crucial in turn transitions (Bögels and Torreira 2015). Boundary marking appeared to be the major prosodic function in previous studies on Papuan Malay monologues (Kaland and Baumann 2020; Riesberg et al. 2018, 2020) and this study extends these results to dialogues. That is, in a collaborative task such as describing and selecting a tangram figure, the prosodic marking of phrase boundaries is used to achieve smooth discourse management. It is thus not surprising that complete and clearly demarcated phrases contribute best to successful exchange of information between conversational partners (see also Ford and Thompson 1996). In this way, the pragmatic function of turn signalling is rooted in the function that chunks the speech signal into meaningful units (i.e. phrasing).

The lack of predictive value for *information status* and *word class* confirm and refine previous studies in important ways. Contrastive focus was not found to be marked by f0 in Kaland et al. (2023), however word class did show an interaction with word stress (rising f0 on the stressed syllable in content words) in Kaland and Baumann (2020). The lack of f0 variation explained by word class differences in the current study could have been the result of an asymmetric dataset. That is, most of the words analysed in this study were content words, whereas in Kaland and Baumann (2020), more function words were taken into account. In addition, the current study did not investigate the interaction with word stress. In this respect, it is worth noting that the clustering output in Figure 3 shows that rising, level or falling f0 can be found in the first half of the contour, corresponding roughly to the stressed syllable. This seems to confirm that f0 is at most a weak correlate of word stress in Papuan Malay (Kaland 2019).

It is interesting that, apart from turn transition and voiced portion (to a smaller extent), f0 variation could not be explained by any of the other predictors. This result could indicate either 1) flaws in the methodology that lead to the failure to capture relevant f0 variation or 2) that these predictors do not constitute relevant sources of f0 variation in Papuan Malay. It should be noted that the first option is unlikely, given the different dimensions along which the contours in Figure 3 vary. These dimensions are, direction, excursion size, and alignment, which could all have been relevant for one or more predictors. That is, the cluster analysis in principle allowed for f0 differences that could have been explained by any of the other predictors in the random forest analysis. The sensitivity to fine-grained shape features is confirmed by the importance of the predictor voiced portion. Given the nature of the random forest analysis, i.e. the possibility to include multiple collinear predictors, there is no reason to assume that any of the analyses prevented f0 variation from being explained by the other predictors (other than *turn transition*). It is therefore more likely that the predictors investigated in this study did not constitute meaningful categories expressed by f0 in Papuan Malay, although this conclusion is further discussed in the following.

It should be noted that rises were hardly found at turn ends, although falls were found in more than half of the cases in turn continuations. The results show, therefore, that the form-function relationship is not clearly one-to-one. In this respect, it is important to consider that the way the predictor *turn transition* was labelled (e.g. in a binary way) might not have allowed for finer grained differences in the way speakers organize their turns. The same argument holds for most of the other predictors. Although the other predictors showed negative importance values, the results could have been different if we had allowed for more category labels than applied in the current study. This observation calls for two more methodological issues to be discussed.

First, it reconfirms the major challenge in research on prosody and meaning, namely distinguishing different meaning types (e.g. Buxó-Lugo and Kurumada 2019). It might have been the case that the labelling in this study was too crude and general. This is not necessarily a problem in initial stages of prosodic research. However, the interpretability of the workflow demonstrated in this paper could improve with more (finer-grained) labels. Note that, although this option implies more elaborate manual annotations, the unsupervised and supervised classification techniques do not pose any limitations on what kind of labelling could be done. In principle it would be possible to setup an extensive and detailed framework of hypotheses using multiple (independent or correlated) predictors based on many different labels. Such an extensive approach could reveal more intricate form-meaning relationships than revealed in the current study (e.g. see the random forest analysis in Kaland et al. 2021).

Second, the number of clusters assumed in the unsupervised part of the analysis has important implications for the supervised part. In the current study, a careful choice was made to select a subset of contours that showed minimum segmental variation and maximum contour variation (Section 2.3). The number of clusters was furthermore supported by the MDL evaluation method. Note that these analyses are in principle 'blind' to the meaning of the f0 variation. They only give an indication of how the variation is most optimally captured by a certain number of clusters. Thus, for the purposes of revealing prosodic meaning in Papuan Malay phrases, it could be that *turn transition* is such a strong predictor that this would still show from a cluster analysis with only two clusters assumed. Such an analysis would, however, not captured the variation that *in theory* could be explained by other sources. The fact that these other factors did not explain the f0 variation for Papuan Malay, is a strong indication that they are indeed not relevant in the way they were modelled in this study.

Finally, it is important to consider that the researcher applying the methodology demonstrated in this paper can "configure" both the unsupervised and supervised classification in ways that help to understand the nature of the f0 variation; by varying the number of clusters and/or by elaborating the set of hypothesized sources of variation (labelling and predictor selection). The set of eight clusters and eight predictors are thus based on generic choices made for the current paper and dataset, and the purpose of demonstrating the combined methodology. It is important to realize that both the numbers (of clusters and predictors) and the kind of predictors are open to adaptation to the exact research questions.

Given the above conclusions, the results are highly comparable to the ones from Malaysian Malay in Zuraidah and Knowles (2006), in particular because that study found that turn transitions were the main source of f0 variation. The current study reconfirms what has been suggested in Zuraidah and Knowles (2006), in that the main prosodic function of Malay prosody operates at the level of conversational interaction. In this respect it should be noted that none of the predictors assessed boundary marking as a prosodic function. This was the consequence of investigating phrase-final words only. Nevertheless it can be expected that f0 movements are larger in phrase-final positions (e.g. f0 range) than in phrase-medial positions (Kaland and Baumann 2020). This finding has implications for future research into underdocumented languages. If the largest effect of intonation is only to be found when speakers interact with each other, then the use of monologic scripted speech data will have limited value. Thus, dyadic conversations should be included into a battery of recordings, possibly even as the first approach to the intonation of language.

One limitation of the study is the lack of questions elicited, since it is likely that question marking would affect f0 in Papuan Malay too. Although this was not investigated in the current study, it naturally correlates with turn transitions in the sense that statements tend to end low whereas questions (requiring the other speaker to start a turn) end high (see Section 1.1). This issue is left for future research on speech data that contains both questions and statements, using a different elicitation technique with e.g. a game requiring more questions to be asked, which could reveal similarities with modality marking in Manado Malay and Ambonese Malay (Table 1). The number of observations in this study was also limited (N = 133). Note, however, that the words were obtained from spontaneously produced speech and carefully selected from phrase-final positions to minimize the segmental variation and prioritize their f0 variation (Section 2). In this way, we believe to have reached a methodological optimum between representing the spontaneous nature of speech and simultaneously gaining control over the speech units that were analysed. It is therefore important to acknowledge that for unsupervised classification, although it does not need narrowing down the type of data (Section 1.3), careful selection has the potential of leading to more informative conclusions about the form-function relationship in prosody. We cannot exclude, however, that with more observations more types of f0 variation can be found.

Acknowledgements: The authors thank Katharina Gayler for data processing and two anonymous reviewers for valuable comments. Supplementary material for this study is available on https://osf.io/hxrjg/.

Research funding: The research for this paper has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 Prominence in Language.

Author contributions: CK: concept and design, analysis and interpretation of data, draughting and revising manuscript. MG: interpretation of data, revising manuscript.

Competing interests: The authors have no conflicts of interest to declare.

Research ethics: The experiments reported in this paper have been conducted following protocols and informed consent practices in compliance with the Helsinki Declaration. Prior approval of the protocols and informed consent procedures was granted by the Centre for Endagered Languages Documentation (CELD, Manokwari, West-Papua).

Appendix I

Number of words per participant in the respective datasets.

Participant	324 set	133 subset
1	7	1
2	11	3
3	24	9
4	2	2
5	14	5
6	2	1
7	1	1
8	24	14
9	3	1
10	10	3
11	11	6
12	17	7
13	12	5
14	12	4
15	11	4
16	9	6
17	4	2
18	7	2
19	4	3
20	6	3
21	8	4
22	2	1
23	13	3
24	5	2
25	9	3
26	13	7
27	6	3
28	17	7
29	6	0
30	7	3
31	15	6
32	14	6
33	11	4
34	2	1
35	5	1

References

Albert, Aviad, Francesco Cangemi & Martine Grice. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Speech Prosody* 2018. 804–808.

Arvaniti, Amalia. 2019. Crosslinguistic variation, phonetic variability, and the formation of categories in intonation. In Sasha Calhoun, Paola Escudero & Marija Tabain (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1–6. Canberra: Australasian Speech Science; Technology Association Inc.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383.

Baumann, Stefan & Martine Grice. 2006. The intonation of accessibility. *Journal of Pragmatics* 38(10). 1636–1657.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Bögels, Sara & Francisco Torreira. 2015. Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics* 52. 46–57.

Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4). 451–493.

Büring, Daniel. 2010. Towards a typology of focus realization. In Malte Zimmermann & Caroline Féry (eds.), *Information structure*, 177–205. Oxford: Oxford University Press.

Buxó-Lugo, Andrés & Chigusa Kurumada. 2019. *Encoding and decoding of meaning through structured variability in intonational speech prosody* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/9y7xj.

Chafe, Wallace L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Couper-Kuhlen, Elizabeth & Cecilia E. Ford (eds.). 2004. *Sound patterns in interaction: Cross-linguistic studies from conversation [OCLC: ocm56526714]*. Amsterdam: J. Benjamins Pub. Co.

Cruttenden, Alan. 1997. *Intonation*, 2nd edn. Cambirdge: Cambridge University Press.

Cruz, Marisa, Verònica Crespo-Sendra, Joelma Castelo & Sonia Frota. 2022. Asking questions across Portuguese varieties. In Marisa Cruz & Sónia Frota (eds.), *Prosodic variation (with)in languages: intonation, phrasing and segments*, 36–70. Sheffield: Equinox Publishing.

Dahan, Delphine, Michael K. Tanenhaus & Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47(2). 292–314.

Enfield, Nick, Mark Dingemanse, Julija Baranova, Joe Blythe, Penelope Brown, Tyko Dirksmeyer, Paul Drew, Simeon Floyd, Sonja Gipper, Rósa S. Gísladóttir, Gertie Hoymann, Kobin H. Kendrick, Stephen C. Levinson, Lilla Magyari, Elizabeth Manrique, Giovanni Rossi, Lila San Roque & Francisco Torreira. 2012. Huh? What? – a first survey in twenty-one languages. In Makoto Hayashi, Geoffrey Raymond & Jack Sidnell (eds.), *Conversational repair and human understanding*, 343–380. Cambridge: Cambridge University Press.

Ford, Cecilia E. & Sandra A. Thompson. 1996, December. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In Elinor Ochs, Emanuel A. Schegloff & Sandra A. Thompson (eds.), *Interaction and grammar*, 1st edn. 134–184. Cambridge: Cambridge University Press.

Frota, Sonia. 2016. *Prosody and focus in European Portuguese: Phonological phrasing and intonation [OCLC: 941440408]*. New York: Routledge.

Frota, Sonia, Marisa Cruz, Flaviane Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira & Marina Vigário. 2015, June. Intonational variation in Portuguese: European and Brazilian varieties. In Sónia Frota & Pilar Prieto (eds.), *Intonation in Romance*, 1st edn., 235–283. Oxford: Oxford University Press.

Goedemans, Rob & Ellen Van Zanten. 2007. Stress and accent in Indonesian. In Vincent J. Van Heuven & Ellen Van Zanten (eds.), *Prosody in Indonesian languages (LOT Occasional Series)*, 35–62. Utrecht: LOT, Netherlands Graduate School of Linguistics. Available at: https://dspace.library.uu.nl/handle/1874/296769.

Gordon, Matthew K. 2014. Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In Harry van der Hulst (ed.), *Word stress*, 83–118. Cambridge: Cambridge University Press.

Gordon, Matthew K. & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1). 20170007.

Grice, Martine, D. Robert Ladd & Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17(2). 143–185.

Grice, Martine, Simon Ritter, Henrik Niemann & Timo B Roettger. 2017. Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics* 64. 90–107.

Gubian, Michele. 2011. Functional data analysis for phonetic research [Publisher: University of Pennsylvania: University of Pennsylvania Type: urn:nbn:nl:ui:22-2066/94426]. *Workshop on New Tools and Methods for Very-Large-Scale Phonetics Research (VLSP 2011)*, DVD. Available at: http://hdl.handle.net/2066/94426.

Gubian, Michele, Francisco Torreira & Lou Boves. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49. 16–40.

Ha, Kieu-Phuong & Martine Grice. 2017. Tone and intonation in discourse management – how do speakers of Standard Vietnamese initiate a repair? *Journal of Pragmatics* 107. 60–83.

Halliday, Michael A. K. 1967. *Intonation and grammar in British English* [OCLC: 1165552521]. https://doi.org/10.1515/9783111357447 (accessed 21 May 2021).

Hamzah, Diyana & James Sneed German. 2014. Intonational phonology and prosodic hierarchy in Malay. In *Proceedings of Interspeech 2014*.

Himmelmann, Nikolaus P. & D. Robert Ladd. 2008. Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation* 2(2). 244–274.

Hualde, José Ignacio. 2009. Unstressed words in Spanish. *Language Sciences* 31(2–3). 199–212.

Jacewicz, Ewa & Robert Allen Fox. 2015. Intrinsic fundamental frequency of vowels is moderated by regional dialect. *The Journal of the Acoustical Society of America* 138(4). EL405–EL410.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2014. *Prosodic typology II: The phonology of intonation and phrasing*. New York: Oxford University Press.

Jun, Sun-Ah & Janet Fletcher. 2014, January. Methodology of studying intonation: From data collection to data analysis. In Sun-Ah Jun (ed.), *Prosodic typology II*, 493–519. Oxford: Oxford University Press.

Kaland, Constantijn. 2019. Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics* 74. 55–74.

Kaland, Constantijn. 2021a. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association* 53(1). 1–30.

Kaland, Constantijn. 2021b. The perception of word stress cues in Papuan Malay: A typological perspective and experimental investigation. *Laboratory Phonology* 12(1). 1–33.

Kaland, Constantijn & Stefan Baumann. 2020. Demarcating and highlighting in Papuan Malay phrase prosody. *The Journal of the Acoustical Society of America* 147(4). 2974–2988.
Kaland, Constantijn & T. Mark Ellison. 2023. Evaluating cluster analysis on f0 contours: An information theoretic approach on three languages. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 3448–3452. Prague: Guarant International.
Kaland, Constantijn & Matthew K. Gordon. 2022. The role of f0 shape and phrasal position in Papuan Malay and American English word identification. *Phonetica* 79(3). 219–245.
Kaland, Constantijn & Nikolaus P Himmelmann. 2020. Time-series analysis of F0 in Papuan Malay contrastive focus. In *Proceedings of the 10th International Conference on Speech Prosody 2020*, 230–234.
Kaland, Constantijn, Angela Kluge & Vincent J. van Heuven. 2021. Lexical analyses of the function and phonology of Papuan Malay word stress. *Phonetica* 78(2). 141–168.
Kaland, Constantijn, Marc Swerts & Nikolaus P. Himmelmann. 2023. Red and blue bananas: Time-series f0 analysis of contrastively focused noun phrases in Papuan Malay and Dutch. *Journal of Phonetics* 96. 101200.
Kingston, John. 2007. Segmental influences on F0: Automatic or controlled? [Series Title: Phonology and Phonetics]. In Aditi Lahiri, Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes: Experimental studies in word and sentence prosody*, 171–210. Berlin, New York: Mouton de Gruyter.
Kirby, James P. & D. Robert Ladd. 2016. Effects of obstruent voicing on vowel *F*0: Evidence from “true voicing” languages. *The Journal of the Acoustical Society of America* 140(4). 2400–2411.
Kluge, Angela. 2017. *A grammar of Papuan Malay*. Utrecht: Language Science Press.
Knowles, Gerry. 1978. The nature of phonological variables in Scous. In Peter Trudgill (ed.), *Sociolinguistic patterns in British English*, 80–90. London: Arnold.
Kochanski, Greg, Esther Grabe, Jerald Coleman & Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America* 118(2). 1038–1054.
Krahmer, Emiel & Marc Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* 34(4). 391–405.
Ladd, D. Robert. 2008. *Intonational phonology*, 2nd edn. Cambridge: Cambridge University Press.
Ladd, Robert & Kim E. Silverman. 1984. Vowel intrinsic pitch in connected speech. *Phonetica* 41(1). 31–40.
Lehiste, Ilse & Gordon E. Peterson. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33(4). 419–425.
Ma, Joan K.-Y., Valter Ciocca & Tara L. Whitehill. 2006. Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America* 120(6). 3978–3987.
Maskikit-Essed, Raechel & Carlos Gussenhoven. 2016. No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology* 33(2). 353–389.
Mixdorff, Hansjörg & Oliver Niebuhr. 2013. The influence of F0 contour continuity on prominence perception. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.
Mohd Don, Zuraidah, Gerry Knowles & Janet Yong. 2008. How words can be misleading: A study of syllable timing and “stress” in Malay. *The Linguistic Journal* 3. 66–81.
Paauw, Scott H. 2009. *The Malay contact varieties of eastern Indonesia: A typological comparison*. http://ubir.buffalo.edu/xmlui/handle/10477/45490 (accessed 11 July 2019).
Potisuk, Siripong, Jack Gandour & Mary P. Harper. 1996. Acoustic correlates of stress in Thai. *Phonetica* 53(4). 200–220.
Prieto, Pilar. 2014. The intonational phonology of Catalan. In Sun-Ah Jun (ed.), *Prosodic typology II*, 43–80. Oxford: Oxford University Press.

R Core Team. 2019. *R: The R project for statistical computing*. https://www.r-project.org/ (accessed 11 July 2019).

R Studio Team. 2019. *RStudio: Integrated development for R*. https://www.rstudio.com/ (accessed 11 July 2019).

Rialland, Annie. 2007. Question prosody: An African perspective [Series Title: Phonology and Phonetics]. In Aditi Lahiri, Tomas Riad & Carlos Gussenhoven (eds.), *Tones and tunes: Typological studies in word and sentence prosody*, 35–62. Berlin, New York: Mouton de Gruyter.

Riesberg, Sonja & Nikolaus P Himmelmann. 2012. The DoBeS summits-PAGE collection of Papuan Malay. https://hdl.handle.net/1839/00-0000-0000-0019-FF78-5 (accessed 11 July 2019).

Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2018. On the perception of prosodic prominences and boundaries in Papuan Malay. Zenodo. https://zenodo.org/record/1402559 (accessed 30 March 2021).

Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2020. Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 8.

Rissanen, Jorma. 1978. Modeling by shortest data description. *Automatica* 14(5). 465–471.

Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6(4). 343–352.

Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42(1). 23–48.

Savino, Michelina, Loredana Lapertosa, Alessandro Caffò & Mario Refice. 2019. Prosodic convergence in Italian game dialogues. In *7th Tutorial and Research Workshop on Experimental Linguistics*, 151–154.

Sbranna, Simona, Caterina Ventura, Aviad Albert & Martine Grice. 2023. Prosodic marking of information status in Italian. *Journal of Phonetics* 97. 101212.

Simard, Candide. 2013. Prosodic encoding of declarative, interrogative and imperative sentences in Jaminjung, a language of Australia. In *Proceedings of Interspeech 2013*, 793–797.

Steele, Shirley A. 1986. Interaction of vowel F0 and prosody. *Phonetica* 43(1–3). 92–105.

Stoel, Ruben. 2007. The intonation of Manado Malay. In Vincent J. Van Heuven & Ellen Van Zanten (eds.), *Prosody in Indonesian languages*, 117–150. Utrecht: LOT, Netherlands Graduate School of Linguistics.

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 307.

Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348.

Swerts, Marc & Ronald Geluykens. 1994. Prosody as a marker of information flow in spoken discourse. *Language and Speech* 37(1). 21–43.

Swerts, Marc, René Collier & Jacques Terken. 1994. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication* 15(1–2). 79–90.

Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4). 629–654.

Szczepek Reed, Beatrice. 2004. Turn-final intonation in English. In Elizabeth Couper-Kuhlen & Cecilia E. Ford (eds.), *Typological studies in language*, 97–117. Amsterdam: John Benjamins Publishing Company.

Van Heuven, Vincent J., Lilie Roosman & Ellen Van Zanten. 2008. Betawi Malay word prosody. *Lingua* 118(9). 1271–1287.

Wan, Aslynn. 2012. *Instrumental phonetic study of the rhythm of Malay* (Thesis) [Accepted: 2013-06-05T09:10:25Z]. Newcastle University. http://theses.ncl.ac.uk/jspui/handle/10443/1682 (accessed 5 November 2020).

Whalen, Douglas H. & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3). 349–366.

Wright, Marvin N. & Andreas Ziegler. 2017. Ranger: A fast implementation of random forests for high dimensional data in *C++* and *R*. *Journal of Statistical Software* 77(1). 1548–7660.

Xu, Yi. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences*.

Xu, Yi & Anqi Xu. 2021. Consonantal F0 perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America* 149(4). 2877–2895.

Zuraidah, Mohd D. & Gerry Knowles. 2006. Prosody and turn-taking in Malay broadcast interviews. *Journal of Pragmatics* 38(4). 490–512.