10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Time-series analysis of F0 in Papuan Malay contrastive focus

Constantijn Kaland & Nikolaus P. Himmelmann

Institute of Linguistics, University of Cologne, Germany
{ckaland, sprachwissenschaft}@uni-koeln.de

Abstract

This study reports a production experiment and acoustic analysis of Papuan Malay prosody in different contrastive focus conditions. These conditions were created by collecting descriptions of pictures with different shapes and colours. The prosody of these descriptions was examined by measures of F0, which were statistically analysed using generalized additive mixed models. These models provide a relatively novel way to analyse F0 as a contour. Results show that speakers of Papuan Malay do not use F0 to mark contrastive focus and support the idea that prosodic phenomena are confined to the final syllable(s) in a phrase. While the absence of prosodic marking provides a crucial difference with respect to some Western-Germanic languages, the boundary phenomena observed in this study rather indicate similarities.

Index Terms: prosody, contrastive focus, Papuan Malay, F0

1. Introduction

In a language such as English, speakers commonly produce "blue" with a pitch accent (in small caps) to mark contrastive information, such in the phrase (1) "Yesterday I saw a red car and today I saw a BLUE car". However, "blue" in a context without semantic contrast remains unaccented, such as (2) "I own a blue car". The acoustic prominence due to the pitch accent on "blue" in (1) therefore explicitly marks the semantic contrast with the red car. Studies have addressed the cognitive mechanisms behind the contrast marking function of prosody. Commonly, prosodic marking of contrastive information in languages such as English is reported to restrict the set of contextual alternative meanings ([1],[2]). For example, by emphasizing that the car is blue, the set of alternative meanings is restricted to colours, i.e. blue, not red (or any other colour). Compare this interpretation to (2), without prosodic marking of a semantic contrast. Thus, marking contrastive information prosodically is a useful way for speakers to single out one specific element from others when referring to the world around them.

The example above illustrates how some West-Germanic languages prosodically mark contrastive information. In these languages, contrastive focus generally leads to a change of the pitch accent location to the word that marks the contrast (to "blue" instead of the default "car"). Romance languages such as Italian [3], Romanian [4] or Spanish [5] resist such a change and the pitch accent remains in its default location. Nevertheless, the shape of the pitch accents in contrastive contexts is reported to be different from the one in non-contrastive contexts, for example in Italian varieties [6]. Other languages make use of different prosodic means to mark contrastive information. For example, in Bengali both the F0 shape and compression of the F0 range post-focally are typical in corrective or surprise information [7], a context that is semantically most similar to the type of contrastive information described in this study. Chîchewa [8] and Korean [9] insert phrase boundaries before or after the focused word.

The current body of literature on Trade Malay varieties, spoken in the Eastern part of Indonesia, lacks a systematic study of (contrastive) focus. That is, the focus contexts that have been elicited have only been described morpho-syntactically in grammars (e.g. Larantuka Malay, [10]), investigated acoustically based on a limited number of speakers and examples (e.g. Ambonese, [11]) or only investigated by visualizations of F0 contours (Manado Malay, [12]). Considerable diversity is to be expected because F0 has shown to be highly variable in closely related Betawi Malay [13], potentially masking the prosodic structure in focus contexts. In addition, conclusions drawn in studies on the prosody of Trade Malay varieties vary considerably on aspects such as word stress and pitch accents (cf. [11], [12], and [14]). The latter observation is not surprising considering the vast archipelago where these languages are spoken.

It is possible to find similarities between the Trade Malay languages discussed above. That is, prominent F0 movements are often confined to particular positions, such as phrase-final syllables in Trade Malay varieties ([11],[12]), regardless of whether they mark focus. This common observation is likely to hold for Papuan Malay as well, as the largest F0 movements occur at the right phrase edge. Particularly relevant for the question of whether Papuan Malay marks focus prosodically is the observation that F0 movements were larger in penultimate phrase position for content words than for function words [15]. That is, content words are generally considered as the domain for focus marking (e.g. [16]). While the existing research does not allow strong predictions on systematic focus marking in Papuan Malay prosody, it does hint at its possibility. Finally, it has to be noted that many languages of the world do not only use prosody to mark contrasts. Often, syntactic phrasing, lexical items or morphology can be used. In this respect, it is important to note that Papuan Malay has a repertoire of focus adverbs that in most contexts "highlight information and signal some kind of restriction, thereby adding emphasis to an utterance" ([17], p. 271). The use of focus adverbs thus eliminates the necessity to rely on prosody as the sole marker of contrastive focus.

To sum up, although much variability in the use of prosodic cues for focus marking can be expected in Trade Malay varieties, the most likely location (phrase final) and acoustic cue (F0) seem to be uncontroversial. Thus, this study investigates to what extent contrastive focus is marked prosodically in Papuan Malay, by means of a highly controlled data elicitation task. This task and the analyses are further described in the next section.

2. Methodology

A production task was carried out by presenting a sequence of minimally different picture pairs to participants, who

10.21437/SpeechProsody.2020-47

described them using specific matrix sentences. Picture pairs were selected on the basis of minimal differences in shape or colour. Shapes and colours were selected only when they occurred as Papuan Malay root in [17], to avoid the use of loanwords. All words referring to the shapes and colours consisted of two syllables to obtain a homogeneous set of noun phrases. As for the shape words, ten nouns were chosen that referred to common objects, animals or persons. As for the colour words, five of them were selected on the basis of the colours listed in [17]. Table 1 provides an overview of the words referring to the shapes and colours. For the purpose of the experiment black and white were considered colours.

Table 1. *Papuan Malay words (English gloss) for the shapes and colours used in the production experiment.*

Shapes (nouns)				Colours (adj.)	
babi	*pig*	pisang	*banana*	hitam	*black*
gunung	*mountain*	tangang	*hand/arm*	puti	*white*
kapak	*axe*	tete	*grandfather*	mera	*red*
kucing	*cat*	tangga	*ladder*	hijow	*green*
liling	*candle*	sapi	*cow*	biru	*blue*

2.1. Design and setup

The picture pairs were designed such that the only difference between the two pictures in one pair concerned either the shape or the colour. For example, if one picture showed a black pig, the other picture in the pair was either a black mountain (noun focus) or a red pig (adjective focus). In addition, pairs were created in which the pictures differed in both shape and colour (fillers). The fillers were added as a baseline condition and to avoid that there was a minimal contrast in all picture pairs, which could have revealed the purpose of the experiment. In total, 25 picture pairs were created (ten shape contrasts, ten colour contrasts and five fillers). The two pictures which formed a pair were displayed in such a way that one occurred on the left side of the screen (antecedent) and the other on the right side of the screen (target), see Figure 1. Participants were instructed to use one of two matrix sentences to describe the pictures such that the noun phrase referring to the pictures occurred in sentence final position or sentence medial position. This was done to ensure that the semantic contrast was salient and that both antecedent phrase (ANT) and target phrase (TAR) were produced within one utterance (see examples 3.1 and 3.2). Describing semantic contrasts that span two successive utterances was avoided in this way, as it could lead to the use of an invariable list intonation, in particular when the matrix sentence is repeated.

(3.1)

Di sebla kiri saya liat [ANT], dang di sebla kanang saya liat [TAR].

On the left side I see [ANT], but on the right side I see [TAR].

(3.2)

Saya liat [ANT] di sebla kiri, dang saya liat [TAR] di sebla kanang.

I see [ANT] on the left side, but I see [TAR] on the right side.

The 25 pictures pairs were presented twice; once in the first part of the experiment and once in the second part of the experiment. In each part, the pairs were presented in a randomized order, which was different for each participant. The matrix sentences were used equally often in either part of the experiment, such that one half of the participants used (3.1) in the first part and (3.2) in the second part and the other half of the participants used (3.2) in the first part and (3.1) in the second part. This was done to balance potential effects of presentation order. The words in Table 1 were used equally often to avoid word biases. That is, each noun was used five times and each adjective was used ten times.

2.2. Participants

In total 24 participants carried out the task; 13 males and 11 females (mean age: 23.6 years, age range: 18-33 years). They were all native speakers of Papuan Malay without speech problems and without colour blindness.

2.3. Procedure and data processing

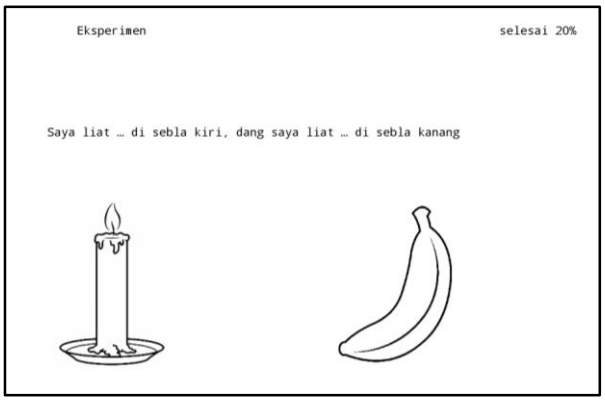

Figure 1. *Example screen capture of a picture pair in the production task (noun contrast, phrase-medial).*

The production experiment was designed using OpenSesame [18]. The experiment consisted of a script written in the programming language Python [19] and pictures displaying the combination of shapes and colours (Table 1). For each picture pair, the script generated a screen with the two pictures on either side (Figure 1). On top of the screen, a written version of the matrix sentence was displayed. Each picture pair was displayed for 7 seconds after which the participants were required to press a button to continue. Before the start of the experiment participants received oral and written instructions about the course of the task. Then, they took seat behind a computer and completed three subsequent parts of the experiment. First, participants entered their personal data. Second, they received instructions on the screen about their task. To familiarize themselves with the task, participants completed a practice round consisting of five picture pairs. At the end of the practice round participants were asked whether they felt they needed to practice more or whether they were ready to start the actual experiment. When more practice was needed, participants were presented additional stimuli. After each additional practice stimulus, participants could end the practice round. Third, when participants ended the practice session they were requested to start the actual experiment. Participants were instructed to switch off personal mobile devices during the entire experiment. Participants were instructed to take a short break after the first part of experiment, after which they were instructed to use a different matrix sentence (section 2.1). The experiment lasted approximately 20 minutes. The speech of the participants was recorded and saved on a computer as wave-file.

All references to the pictures (henceforth NPs) in the recorded wave-files were annotated on the syllable- and word-level (2 syllables per word, 4 words per stimulus, 50 stimuli per participant: N = 9600 syllables). This was done by trained

annotators who were familiar with the language and its syllabification. Syllables produced with irregularities were not taken into account for further analysis. Common irregularities included the wrong word to refer to the shape or colour (e.g. *puti* instead of *hijow*), hesitations or corrections within the word, inaudible speech or background noise. After removing these cases, 8966 syllables were left for acoustic analysis. All syllables were numbered according their position in the NP (1 to 4 for each phrase type). The F0 analysis was carried out in Praat [20]. 20 measures per syllable were taken (*M* syllable duration = 194.39 ms). The timestamps of these 20 measures were determined by maintaining equal intervals between each measure. For example, F0 measurements would be taken every 10 ms for a syllable with a duration of 200 ms, with the first measure taken at the left syllable boundary.

2.4. Statistical analysis

The F0 measures were analysed in generalized additive mixed models (GAMMs) using the packages "mgcv" [21] and "itsadug" [22]. GAMMs [23] are particularly useful for time-series data and do not assume linear relationships between response and predictor(s). This makes GAMMs the preferred method for analysing F0 as a contour (i.e. over time), which is often a complex curve that can only partially be captured by static measures such as mean F0 or F0 range.

A total of 2400 F0 contours were collected, corresponding to the F0 trajectory over the course of the four syllables (noun, adjective) in either the antecedent or the target phrase. Contour was added in a random smooth (see below) in the GAMM, which would dramatically increase the computational cost of the model if all 2400 levels would have been taken into account. In order to reduce computational cost, therefore, a random sample of 10% of the contours (240 levels) was taken into account for the random smooth.

In time-series analyses of F0, high autocorrelation in the model residuals could be problematic for significance testing with GAMMs [24]. This was accounted for in the current analysis in three ways. First, the number of F0 measures per syllable was reduced to 10 by taking every second measure of the collected 20 measurement points per syllable. This reduction also decreases the amount of local variation in the contour, for example as a result of preceding plosives. Having fewer local perturbations also makes the smoothing applied by the GAMM more accurate. With 10 measurements per syllable, however, overall rising or falling movements due to focus marking can still be accurately modelled. Second, a random smooth by contour was added to the GAMM, which accounted for the variation between the individual contours and supersedes random smooths by subject or by items due to the higher number of levels (see below). Third, an AR1 model was added to the GAMM, which included a correlation coefficient (rho = 0.37) as an estimate of the residual autocorrelation in a model without this parameter.

In the full model, F0 in semitones (ST) was the response and the interaction between the predictors focus (three levels: filler, noun, adjective), phrase position (two levels: medial, final) and phrase type (two levels: antecedent, target) was added as parametric term. In addition, a random smooth by focus was added (240 levels). To allow for significance testing using model comparisons, the smoothing parameter estimation method was set to maximum likelihood (ML, see below).

To determine whether F0 contours were significantly different between any of the levels of the predictors, two methods were used. First, model comparisons were performed between the full model (as described above) and a model in which one of the predictors was left out both as parametric term and as difference smooth (either focus, phrase position or phrase type; totalling three comparisons, see Table 2). Second, difference smooths for all minimally different combinations of all three predictors (12 for focus, 6 for phrase position, 6 for phrase type) were computed with a confidence interval (se = 1.96). In this way, the difference smooths provide an indication of the intervals at which the compared F0 contours were significantly different (i.e. where the confidence interval did not overlap an F0 difference of 0). Note that difference plots for which significance intervals were shorter than three measurement points (corresponding to 29.16 ms on average) were omitted in the results section. These short intervals correspond to the length of a segment at most and were therefore assumed of minor or no importance for the overall assessment of F0 contours as focus markers. This was the case for three difference smooths.

3. Results

Table 2. *Results of the full model comparisons, showing the ML score difference between the models, estimated degrees of freedom (Edf) and the probability (p) that the two models are different (after χ^2 testing).*

Predictor	Compared model to full model	ML diff.	Edf	*p*
Focus	phrase position * phrase type	6.50	14	n.s.
Phr. position	focus * phrase type	31.85	20	< .001
Phr. type	focus *phrase position	118.78	20	< .001

The model comparisons (GAMM) showed significant differences when either the predictor phrase position or phrase type were left out. As for the predictor focus no significant differences were found (see also Figure 2).

The difference smooths (plots are omitted in this paper to save space) showed significant differences for all predictors and are discussed accordingly. Concerning the predictor focus, target phrases with the NP in phrase final position showed that for focused nouns F0 was significantly higher compared to focused adjectives and that for focused adjectives F0 was significantly lower compared to fillers. For both effects, the intervals where the contours differed significantly were found in similar locations; i.e. around the midpoint of the NP (measure point 40) and in the final syllable (between points 60 and 80). The adjective-filler difference showed an additional significant difference in the first syllable (between point 0 and 20). As for phrase position, three difference smooths showed significant difference in the final syllable of the NP. That is, for focused nouns, the F0 in final syllable of the antecedent phrase had a higher F0 in phrase final position than in phrase medial position. A lower F0 in the final syllable of the target phrase was found for focused nouns and for focused adjectives in phrase final position compared to phrase medial position.

With regard to phrase type, three difference smooths showed significant differences. These differences always occurred in NPs in phrase final position and were mainly found in the last syllable, although some smaller intervals with significant differences were found in earlier syllables. That is, F0 in final syllables was higher in antecedent phrases than in

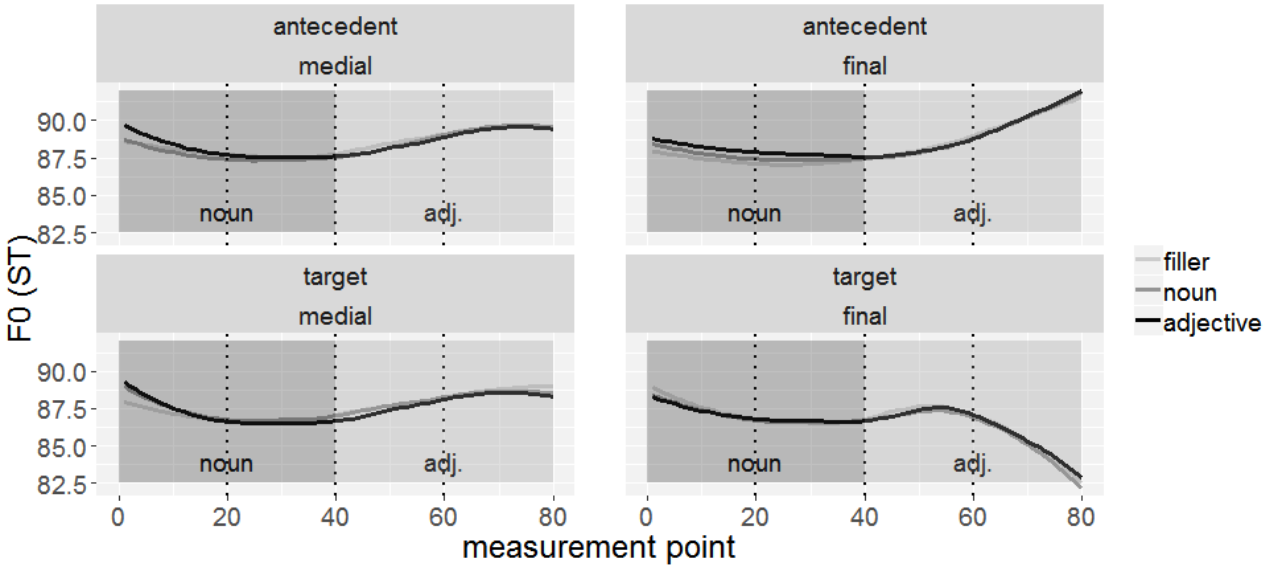

Figure 2. *Smooth plots of F0 contours (ST) of antecedent (top) and target (bottom) NPs in medial (left) and final (phrase position) in the different focus conditions. Dotted lines indicate syllable boundaries at the respective measurement points.*

target phrases, regardless of focus. For focused adjectives an additional interval in the second syllable of the noun was found where F0 was higher in the antecedent phrase than in the target phrase. For fillers an additional interval in the onset of the first syllable of the noun was found were F0 was lower in the antecedent phrase than in the target phrase.

4. Discussion and Conclusion

Apart from a couple of physiological or strictly boundary-related phenomena, the current investigation did not find evidence for contrastive focus marking by means of F0 in Papuan Malay prosody. As for the physiological phenomena, it was observed that F0 decreased over the course of the utterance, i.e. the naturally occurring declination effect [25]. Concerning the boundary phenomena, systematic marking of continuation (antecedent phrase) and finality (target phrase) on the final syllable in the phrase was found. The former was indicated by a rising F0, whereas the latter was indicated by a falling F0. The use of F0 in this way reflects the way phrases are delimited by F0 in many languages of the world, often described as boundary tones in autosegmental analyses [16].

On the basis of the results in the current study, Papuan Malay appears to differ from Manado Malay, where focus could be marked by F0 in phrase final positions [12]. Papuan Malay appears rather similar to Ambonese, where no prosodic focus marking occurred [11]. Ambonese was reported to lack pitch accents and to only make use of phrase final boundary tones with a loose temporal peak alignment (i.e. somewhere around the boundary between the pre-final and final syllable). This conclusion, however, does not necessarily hold for Papuan Malay. The use of boundary tones as known from autosegmental analyses seems restricted to the final syllable only in Papuan Malay, as discussed in the previous paragraph. From Figure 2, it can be furthermore observed that the first syllable of the adjective (points 40-60), i.e. the pre-final syllable in the phrase, shows a rising F0 in all conditions. This pre-final rise is unlikely to be an anticipation to the direction of F0 in the final syllable. That is, in most conditions F0 is rising further on the final syllable, making it likely to falsely interpret pre-final rises to belong to the final one. However, in phrase final target phrases (Figure 2 bottom right), a rise can be observed before the F0 fall on the phrase-final syllable. This observation suggests that the pre-final rises need to be interpreted as a more structural property of phrase prosody. In this respect it is important to note that penultimate syllables predominantly stand out as acoustically prominent [14], in accordance with the claim that Papuan Malay has regular penultimate word stress [17]. In the current setup, the adjectives referring to the colours are indeed marked as having penultimate stress [17]. Together, these outcomes seem to suggest a privileged status of the pre-final syllable in the phrase due to word stress patterns. Furthermore, the direction of F0 movements was found to correlate only weakly with word stress in Papuan Malay in [14], in line with the crosslinguistic observation that F0 is a better correlate of phrase prosody than it is of word prosody [26]. Thus, these results indicate that at the phrase-level the F0 rise structurally marks the pre-final syllable, which is acoustically promoted mainly by durational cues at the word level. It cannot be concluded whether this is a form of alignment of phrase accents to stressed syllables, as reported for Manado Malay, for example [12]. To this end, words with penultimate and ultimate stress need to be compared in phrase final position. Preliminary evidence suggests that there is some correlation between the location of the F0 rise and word stress [15]. However, ultimate stress is particularly rare in Papuan Malay, which challenges paradigms that aim to investigate the phenomenon empirically (e.g. [27]). Regardless of the question of (autosegmental) alignment, just by the frequency with which pre-final syllables at the phrase level are penultimate syllables at the word level, it is likely that the two prosodic levels feed into each other (see [26] for an account).

5. Acknowledgements

Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 – SFB 1252. The authors thank the staff of the Center for Endangered Languages Documentation (CELD) for help with carrying out the recording sessions and with participant recruitment, and four anonymous reviewers for their valuable comments.

6. References

[1] B. Braun and L. Tagliapietra, "The role of contrastive intonation contours in the retrieval of contextual alternatives," *Language and Cognitive Processes*, vol. 25, no. 7–9, pp. 1024–1043, Sep. 2010.

[2] E. M. Husband and F. Ferreira, "The role of selection in the comprehension of focus alternatives," *Language, Cognition and Neuroscience*, vol. 31, no. 2, pp. 217–235, Feb. 2016.

[3] M. Swerts, E. Krahmer, and C. Avesani, "Prosodic marking of information status in Dutch and Italian: a comparative analysis," *Journal of Phonetics*, vol. 30, no. 4, pp. 629–654, Oct. 2002.

[4] M. Swerts, "Contrast and accent in Dutch and Romanian," *Journal of Phonetics*, vol. 35, no. 3, pp. 380–397, Jul. 2007.

[5] A. Cruttenden, "The de-accenting and re-accenting of repeated lexical items," in *Proceedings of the ESCA workshop on prosody*, Lund, Sweden, 1993, pp. 16–19.

[6] M. Grice, M. D'Imperio, M. Savino, and C. Avesani, "Strategies for Intonation Labelling across Varieties of Italian," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, 2005, p. p.362-389.

[7] S. ud D. Khan, "The intonational phonology of Bangladeshi Standard Bengali," in *Prosodic Typology II*, S.-A. Jun, Ed. Oxford University Press, 2014, pp. 81–117.

[8] J. M. Kanerva, *Focus and Phrasing in Chichewa Phonology*. New York: Taylor & Francis, 1990.

[9] S.-A. Jun and H.-J. Lee, "Phonetic and phonological markers of contrastive focus in Korean," presented at the Proceedings of the 5th International Conference on Spoken Language Processing, 1998, vol. 4.

[10] S. H. Paauw, *The Malay contact varieties of eastern Indonesia: A typological comparison*. 2009.

[11] R. Maskikit-Essed and C. Gussenhoven, "No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay," *Phonology*, vol. 33, no. 2, pp. 353–389, Aug. 2016.

[12] R. B. Stoel, "The intonation of Manado Malay," in *Prosody in Indonesian Languages*, vol. 9, van Heuven, Vincent J. and van Zanten, E, Eds. 2007, pp. 117–150.

[13] V. J. van Heuven, L. Roosman, and E. van Zanten, "Betawi Malay word prosody," *Lingua*, vol. 118, no. 9, pp. 1271–1287, 2008.

[14] C. C. L. Kaland, "Acoustic correlates of word stress in Papuan Malay," *Journal of Phonetics*, vol. 74, pp. 55–74, May 2019.

[15] C.C.L. Kaland and S. Baumann, "Different functions of phrase-final F0 movements in spontaneous Papuan Malay," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, Eds. Melbourne, Australia, 2019, pp. 1312–1316.

[16] S.-A. Jun, Ed., *Prosodic typology II: the phonology of intonation and phrasing*. Oxford: Oxford University Press, 2014.

[17] A. Kluge, *A grammar of Papuan Malay*. Berlin, Germany: Language Science Press, 2017.

[18] S. Mathôt, D. Schreij, and J. Theeuwes, "OpenSesame: An open-source, graphical experiment builder for the social sciences," *Behav Res*, vol. 44, no. 2, pp. 314–324, Jun. 2012.

[19] G. Van Rossum and J. De Boer, "Interactively testing remote servers using the Python programming language," *CWI Quarterly*, vol. 4, no. 4, pp. 283–304, Dec. 1991.

[20] Boersma, Paul and Weenink, David, *Praat: doing Phonetics by Computer*. 2019.

[21] S. N. Wood, *Generalized Additive Models: An Introduction with R, Second Edition*, 2nd New edition. Boca Raton: Taylor & Francis Inc, 2017.

[22] J. Van Rij, M. Wieling, R. H. Baayen, and H. Van Rijn, *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. 2017.

[23] X. Lin and D. Zhang, "Inference in generalized additive mixed modelsby using smoothing splines," *J Royal Statistical Soc B*, vol. 61, no. 2, pp. 381–400, Apr. 1999.

[24] R. H. Baayen, J. van Rij, C. de Cat, and S. N. Wood, "Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models," *arXiv:1601.02043 [stat]*, Jan. 2016.

[25] J. Breckenridge, "The declination effect," *The Journal of the Acoustical Society of America*, vol. 61, no. S1, pp. S90–S90, Jun. 1977.

[26] M. Gordon, "Disentangling stress and pitch-accent: a typology of prominence at different prosodic levels," in *Word Stress*, H. van der Hulst, Ed. Cambridge, UK: Cambridge University Press, 2014, pp. 83–118.

[27] C. C. L. Kaland, "Offline and online processing of acoustic cues to word stress in Papuan Malay," The Journal of the Acoustical Society of America, vol. 147, no. 2, pp. 731–747, Feb. 2020, doi: 10.1121/10.0000578.