Phonology **33** (2016). Supplementary materials

No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay

Raechel Maskikit-Essed
Carlos Gussenhoven
Radboud University Nijmegen

Supplementary materials

On the temporal coherence and predictive power of segmental landmarks

Because correlations between segmental landmarks have a high level of multicollinearity (from $r = 0.26$ to $r = 0.96$), it is impracticable to perform an ANOVA or regression analyses on the timestamps obtained from the data reported in §3. However, the extent to which those landmarks correlate with each other may tell us something about the temporal coherence of words. A Principal Component Analysis can provide an overall summary statement by extracting the common variance from the variables entered, repeating the procedure on the residuals of the first extraction. The resulting orthogonal components can be interpreted on the basis of the extent to which the original variables contribute to it. We performed Principal Component Analyses for each language separately on the six landmarks. The Kaiser-Meyer-Olkin measure was 0.74 for Ambonese Malay and 0.85 for Dutch (respectively 'middling' and 'meritorious', according to Hutcheson & Sofroniou 1999), meaning that with 84 cases for each language we had reasonable sample sizes. After an initial analysis in which eigenvalues for the second components appeared to fall just below Kaiser's criterion of >1.0, we decided to lower the criterion so as to extract two components for each language. Table I gives component loadings (Field 2013: 706).

	Ambonese Malay		Dutch	
landmark	component 1	component 2	component 1	component 2
BegWd	0.03	**0.88**	0.19	**0.98**
BegσPen	**0.49**	**0.73**	**0.84**	**0.45**
BegRhPen	**0.68**	**0.60**	**0.93**	0.24
EndσPen	**0.91**	0.24	**0.93**	0.25
BegRhFin	**0.90**	0.31	**0.95**	0.20
EndWd	**0.81**	0.05	**0.93**	0.07

Table I

Loadings of the six landmarks on two components for Ambonese Malay and Dutch. Values above 0.40 are given in bold.

The results show that the co-variance among the six landmarks is similar in the two languages. Component 1 appears to represent the second half of the word, centring on the beginning of the final syllable, while component 2 represents the beginning of the word. The difference between the two languages lies in the degree of coherence among the landmarks. Those for Ambonese Malay are less predictable from other landmarks in the same word than those for Dutch, as shown by the percentage of variance explained by the two components. In combination, they explained 80% of the variance in the Ambonese Malay data, as against 92% in the Dutch data (see the bold values in Table II). The six landmarks thus relate to each other in similar ways in the two languages, but their common variance is less in Ambonese Malay, reflecting a lower degree of temporal integration of these landmarks than in Dutch.

	Ambonese Malay			Dutch		
	initial eigen-value	explained variance (%)	cumulative explained variance (%)	initial eigen-value	explained variance (%)	cumulative explained variance (%)
component 1	3.84	63.91	63.91	4.69	70.31	70.31
component 2	0.97	16.01	**79.92**	0.88	21.97	**92.28**

Table II

Explained and cumulative explained variance for Ambonese Malay and Dutch in six landmark variables, by two extracted components.

Stepwise linear regression analyses showed that, unsurprisingly, the two components significantly contribute to the prediction of the H-timestamps (component 1: $p < 0.0001$ for both languages; component 2: $p < 0.0001$ for Ambonese Malay, $p < 0.05$ for Dutch). Component 1 is the better predictor, explaining 58% of the variation in Ambonese Malay and 76% in Dutch. The addition of component 2 increases the explained variance by 13% in the case of Ambonese Malay, but only 1% in the case of Dutch. These data show again the diffuseness of the peak alignment in Ambonese Malay, where the combined common variance in the landmarks, i.e. the two components, leads to a better prediction than the variance of any single landmark (71% against 67% by the beginning of the final rhyme). For Dutch, the variance explained by the beginning of the penultimate rhyme is higher (81%) than that by the combined common variance (77%).

ADDITIONAL REFERENCES

Field, Andy (2013). *Discovering statistics using IBM SPSS Statistics*. 4th edn. London: Sage.

Hutcheson, Graeme D. & Nick Sofroniou (1999). *The multivariate social scientist: introductory statistics using generalized linear models*. London: Sage.