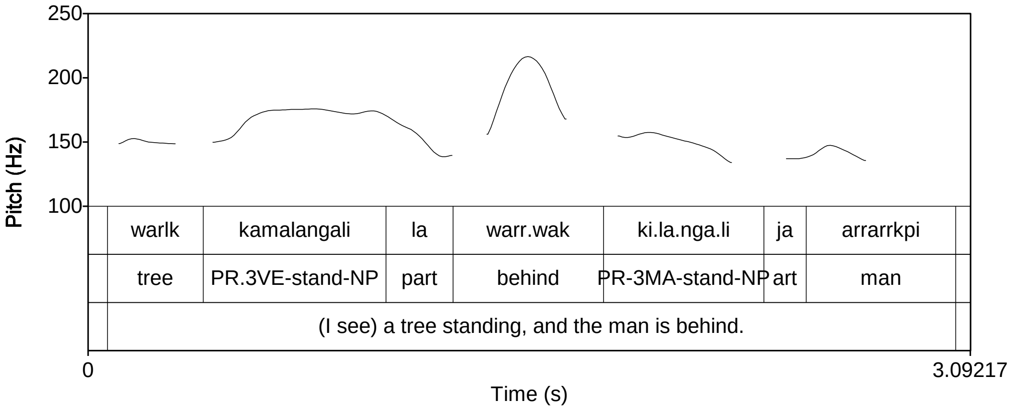
Quantitative investigation of intonation in an endangered language.[1]

SAM HELLMUTH[1], FRANK KÜGLER[2] & RUTH SINGER[3]
University of York[1], Universität Potsdam[2] & Radboud University Nijmegen[3]

1. INTRODUCTION

Mawng is a non-Pama-Njungan language of Australia (Singer 2006a) of the Iwaidjan language family. It is spoken by around 300 people and is still being acquired by children. In Mawng, the presence of an intonational pitch accent on a word does not by itself indicate information status, since, in general, all nouns and verbs bear a pitch accent. However the use of steep pitch excursions in contexts of contrastive focus, as observed by Singer (2006b), suggest that a 'special pitch accent' is available as one means of encoding focus in Mawng. This is illustrated in Figure 1 which shows a 'special focus accent' on the word *warrwak* [warwak] 'behind', with 'neutral' accents on all other content words. The focus accent usually marks a contrastive focus (as in this example which occurs in a sequence of locative descriptions in which the man was previously positioned in front of the tree), but is also observed in non-contrastive focal contexts (such as introduction of a new referent to discourse).[2]

Figure 1
Example of 'focus accent' on *warrwak* (mph-rs-1-39-9-1).



[1] We are indebted to the Mawng speakers for their contribution to this work. This research was supported by Sonderforschungsbereich 632 Project D2 'Typology of Information Structure' at the University of Potsdam (funded by the DFG), by the University of Melbourne and by the Australian Research Council grant 'Reciprocals across languages' awarded to Prof. Nicholas Evans. Ruth Singer would like to thank Bruce Birch, Janet Fletcher and Yiya Chen for useful discussions of Mawng prosody. The authors would like to thank Andreas Pankau and Kathi Moczko for their help in reformatting the data in the elicited dataset.

[2] It is not the only means available to indicate focus in Mawng, which also employs word order variation and (to some extent) an emphatic marker ('pa'), see Singer (2007 in preparation).

In this paper we present the results of a quantitative study of peak alignment and F0 excursion (or ‘scaling’) of the H peak in Mawng pitch accents. The study yields generalisations in different syllable types (open vs. closed) and in focus and non-focus variants, which enable us to propose an appropriate phonological representation for the main pitch accent and to determine whether or not the special ‘focus’ variant should be analysed as a separate phonological category.

Particular methodological issues must be faced in obtaining alignment and scaling generalisations in an unwritten, endangered language. In most European languages whose intonation has been studied in detail, the alignment and scaling of pitch accents has been investigated by means of experimental studies using carefully controlled stimuli presented to speakers in written form, and alignment properties observed in ‘lab’ speech studies of this type have been shown to be parallel to those observed in semi-spontaneous speech (Lickley *et al.* 2006). Collection of controlled data of this type is problematic in an unwritten language, although not impossible, but it is an open question whether ‘lab’ speech recordings are of sufficient value to a speech community whose language is endangered to warrant their collection over and above more naturalistic materials. The ideal solution would be to establish alignment and scaling properties on the basis of naturalistic or (semi-)spontaneous data. However, intonational phonologists have relied on carefully controlled stimuli for good reason since the scaling and alignment of intonational pitch peaks is known to be influenced by a wide range of factors in the surrounding environment (see Kügler 2007 for a summary). For example, F0 is likely to be affected by variation across target words in both intrinsic and phonological vowel length (Steele 1986, House 1989) and in duration and voicing of consonants (Rietveld & Gussenhoven 1995). At the prosodic level, variation in syllable type (Ladd *et al.* 2000, Hellmuth 2006b), proximity of a prosodic boundary (Steele 1986, Silverman & Pierrehumbert 1990, Prieto *et al.* 1995) and stress clash (Silverman & Pierrehumbert 1990) are all known to influence peak alignment.

A few studies have attempted to filter out the influence of these factors in quantitative studies of F0 alignment in (semi-) spontaneous speech (Wichmann & House 1996, Peters 1999, cf. also Kügler 2007). For example, Peters (1999) investigates inter-dialectal differences in peak alignment in German by pre-selecting tokens from a large base corpus so that they are roughly uniform (sonorant open syllables bearing a nuclear H*). Each token is categorised according to a list of factors, for use in a multiple regression analysis to determine their influence on peak alignment. We use this methodology in the present study.

We hypothesise that the ‘special focus accent’ in Mawng differs from its ‘neutral’ counterpart in either alignment or scaling (or both). A difference in peak alignment would suggest a categorical difference of phonological representation (cf. Bishop & Fletcher 2005), whereas a difference of scaling would suggest that the special focus accent is of the same phonological category as its neutral counterpart, but produced in an expanded pitch range (cf. Hellmuth 2006a). Both options are found in the literature on Bininj Gun-wok (BGW): Bishop and Fletcher (2005) mention a special L+H* tone used in the Gundjeihmi dialect of BGW to signal emphasis or narrow focus in narratives, whilst Bishop (2002) describes relative pitch scaling as the main means of encoding focus and other aspects of information structure in other varieties within the same dialect chain.

Whilst the presence (or absence) of a categorical phonological distinction cannot be established from production data alone (a perception study would be needed) this study provides important evidence in deciding the best phonological analysis of the properties of this special pitch accent. The remainder of the paper sets out how we implemented the corpus-based quantitative investigation in section 2, followed by the results of the statistical analysis in section 3, and discussion of implications of the findings in section 4.

2. METHODOLOGY

2.1. Materials

Our source dataset for the quantitative study includes two types of data: i) semi-spontaneous data elicited in response to visual stimuli (Skopeteas *et al.* 2006); ii) spontaneous data from spoken narratives (for background information on the narrative texts see Singer 2006a). Overall the data were collected with five different speakers. The semi-spontaneous data includes a relatively large number of tokens elicited in focus contexts.

Each section of the corpus (elicited/spontaneous) was analysed auditorily, with reference to F0 and spectrogram extracted using Praat 4.6.10 (Boersma & Weenink 2007), in order to select accented words which are suitable as tokens for inclusion in the alignment study. A word was included in the study only if it was in non-phrase-final position and contained mostly sonorant or voiced segments in and around the accented syllable. In practice a number of words were included which contained a stop consonant in or near the accented syllable, which reduced our ability to measure the position of the L turning point at the start of the accented syllable in a number of cases (Mawng like most Australian languages has a single series of plosive consonants which lack a voicing contrast, Evans 2003). The majority of included tokens are nominals, in which the position of word-level prominence is easier to determine; a small number of verbs in which the position of stress was clear were also included. Tokens were labelled by syllable type (V, CV and CVC) and focus status (an instance of the special pitch accent, '+ focus', or not, '-focus').

This selection process yielded 39 focus tokens and 153 'neutral' tokens. Although these token counts do not match the equal numbers of focus/non-focus tokens that one might obtain in an experimental study, this is a greater proportion of focus tokens than one might expect to find in a corpus of spontaneous speech only. We looked through approximately 800 Mawng utterances in order to obtain this set of 192 suitable tokens (540 elicited sentences + 10 texts of approximately 25 sentences each).

2.2. Analysis

Two levels of analysis were applied to the data. Firstly we wish to establish the peak alignment patterns of 'neutral' accents relative to the segmental string, in order to establish a suitable phonological representation for the Mawng pitch accent. Visually, we observe that the H peak falls either just inside or just after the end of the accented syllable, however we suspect that the distribution of peaks falling inside and just outside the accented syllable may depend on syllable type (open vs. closed), a pattern that has been observed in other languages (Ladd *et al.* 2000, Hellmuth 2007). Our specific hypothesis is that the H peak will align outside the syllable in open syllables but within the syllable in closed syllables. We use a proportional measure of peak alignment (relative peak delay, as defined below) in order to avoid the effects of greater variance among absolute peak delay values (Schepman *et al.* 2006).

Secondly, we wish to determine whether 'focus' accents differ from 'neutral' accents in alignment or scaling (or both). As dependent variables we use relative peak delay again, and a proportional measure of F0 scaling (normalised F0 excursion, as defined below). To derive the required dependent variables, segmental landmarks and pitch events (Table 1) were labelled by hand in each of the 192 tokens identified as suitable for inclusion in the study. A Praat script was used to extract durational and F0 measurements from each token and from these to calculate dependent variables as set out in Table 2.

Table 1
Segmental landmarks and pitch events labelled in each token.

O	start of the onset consonant of the accented syllable (if present)
N	start of the nucleus of the accented syllable
C	start of the coda consonant of the accented syllable (if present)
H	F0 maximum turning point in or near the accented syllable
L	F0 minimum turning point before the H peak
pH	F0 maximum of the phrase in which the token is found
pL	F0 minimum of the phrase in which the token is found

Table 2
Calculation of derived variables, with dependent variables indicated in bold.

pd	'peak delay' = distance from H peak to beginning of accented syllable
sylldur	duration of accented syllable
rpd	'relative peak delay' (pd/sylldur)
wxn	accented syllable F0 excursion in Hz ('F0 at H' – 'F0 at L')
pxn	phrasal F0 excursion in Hz ('F0 at pH' – 'F0 at pL')
zxn	normalised F0 excursion in Hz (wxn/pxn)

In order to be able to determine the degree of influence of external factors a count was made of the number of syllables from the accented syllable to the end of the word ('DistWEnd', distance to word end) and from the accented syllable to the next pitch accent ('NoIntSyll', number of intervening syllables). The position of the accented syllable within the word was also noted ('PosStress', position of stress). In addition the duration of the onset ('ODur') and coda ('CDur') were calculated (if present). All of these factors have been shown to affect the alignment of H peaks in pitch accents (see section 1).

3. RESULTS

Since our tokens were selected from a corpus we have a non-homogenous dataset, with variation in the number of tokens per speaker, per syllable type and per focus condition. In addition to this general non-homogeneity in the data, it is possible that our dependent variables (peak alignment and F0 excursion) could vary unevenly across the dataset due to factors in the prosodic environment, as discussed in section 1. In order to gauge the influence of potentially influencing factors in the non-homogenous dataset we ran two multivariate linear regression analyses with Focus, SyllableType, Speaker, DistWEnd, NoIntSyll and PosStress as potential predictors of Relative Peak Delay (rpd, our alignment variable) and Normalised F0 Excursion (zxn, our scaling variable) respectively.[3] The results of the analyses are shown in Table 3 below.

The regression analysis shows that the overall trends in the data match our hypotheses. The linear regression for Relative Peak Delay indicates that alignment is affected by a number of factors including Syllable Type, but, crucially, not by Focus. The model accounts for 56% of the variation in values of Relative Peak Delay ($R^2 = 0.558$). In fact, both DistWEnd and SyllableType have a significant effect, though these two factors are correlated (Pearson's coefficient = -.387, $p<0.01$) so it is difficult to tease apart their individual effects; Speaker and PosStress were shown to have a small effect and there was no effect at all of Focus and NoIntSyll.

In contrast F0 excursion is affected by Focus and DistWEnd only, and not by Syllable Type or any other factor. The regression model accounts for 18% of the variation in values of Normalised F0 Excursion ($R^2 = 0.179$). Both Focus and DistWEnd have a significant effect, though again these are correlated (Pearson's coefficient = -.125, $p<0.05$) so the individual effects of each cannot be distinguished.

Table 3

Multivariate linear regression: Relative Peak Delay and F0 excursion.

rpd	β	t	Sig.	zxn	β	t	Sig.
Focus	.014	.281	.779	Focus	.349	5.053	.000
Syllable Type	-.301	-5.537	.000	Syllable Type	.105	1.420	.157
Speaker	.166	3.355	.001	Speaker	-.060	-.889	.375
DistWend	.536	7.677	.000	DistWend	.306	3.209	.002
NoIntSyll	-.006	-.086	.931	NoIntSyll	-.141	-1.536	.126
PosStress	-.116	-2.358	.019	PosStress	.050	.751	.453
(Constant)		4.182	.000	(Constant)		2.431	.016

[3] ODur and CDur prove each to be highly correlated with both SyllType and Focus ($p<0.01$), co-varying with them, and were left out of the model in order to reduce the number of variables.

Figure 2

95% Confidence Intervals for values of relative peak delay (rpd), by syllable type and by focus type (reference line indicates the end of the stressed syllable).

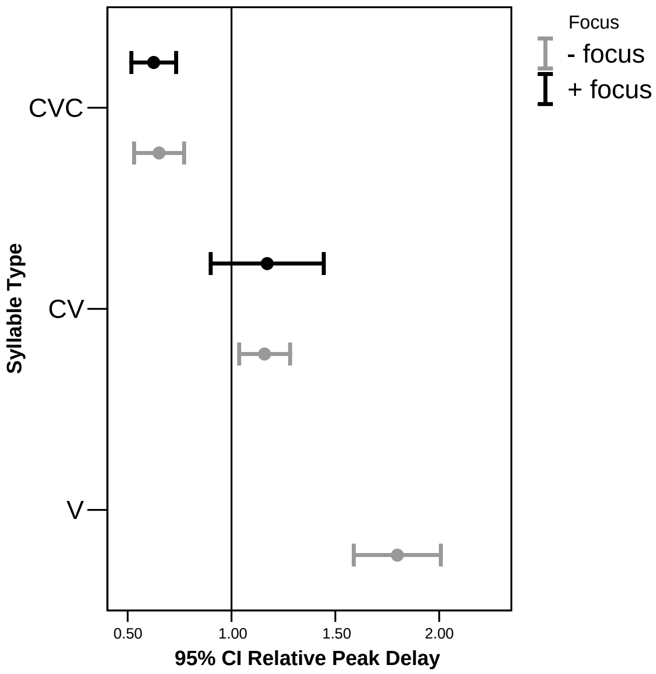


Figure 3

95% Confidence Intervals for values of normalised F0 excursion (zxn), by syllable type and by focus type.

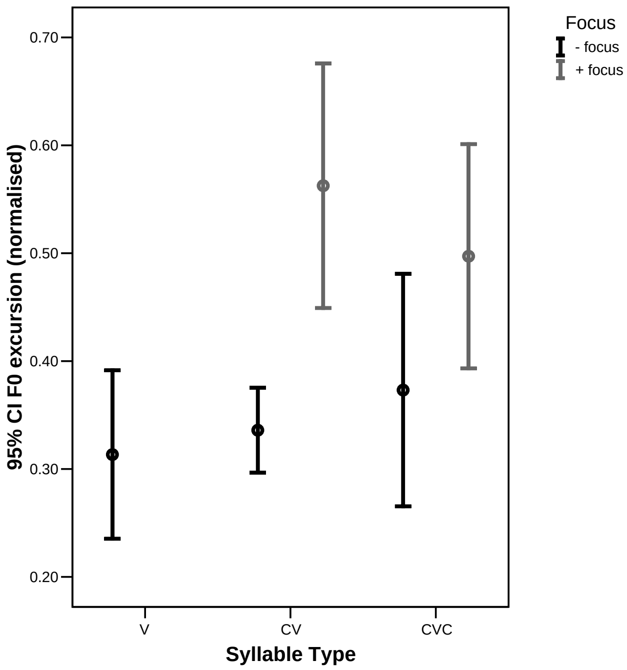


Figure 2 above shows 95% Confidence Intervals for values of rpd by syllable type and by focus type (there are only four focus tokens in the V syllable type category so they are excluded from the figure). As the regression analysis results suggest, the data are widely distributed (particularly in open syllables, V and CV), however there is nonetheless a clear difference in values of rpd dependent on syllable type, and all of these differences are significant. Our hypothesis that patterns of H peak alignment in the Mawng pitch accent would vary according to syllable type is thus confirmed. Note however that peak delay is different in each of the three syllable types observed in our corpus data, so that it is possible to distinguish different alignment patterns in open syllables with and without an onset (V vs. CV), and a different pattern again in closed syllables (CVC). We return to this point in the discussion.

Turning to the properties of the 'focus' accent, it is clear in Figure 2 that values of rpd are similar within each syllable type, regardless of focus type, and none of the slight differences are statistically significant. The difference between the neutral and focus accents is therefore not a matter of alignment. As regards F0 scaling, Figure 3 shows 95% Confidence Intervals for values of zxn (normalised F0 excursion) by syllable type and by focus type, and it is clear that values of zxn vary by focus type (and not by syllable type). The difference in values of zxn between +focus and –focus tokens is significant in CV syllables, though does not reach significance among CVC syllables (a subset of only 15 tokens). We believe that a larger pool of tokens, and thus a larger pool of CVC tokens, would yield an even clearer result. On the basis of the present dataset then, we infer that the Mawng 'focus' accent is characterised by increased F0 excursion only, and not by any significant difference in peak alignment relative to the segmental string.

4. DISCUSSION

With respect to the main Mawng pitch accent, our expectation that H peak alignment would vary in open and closed syllables was confirmed, and the regression analysis permitted us to eliminate a number of potentially influencing factors. We hope to reproduce this result, and also to eliminate the remaining co-variant factor DistWEnd (distance to word end), in an enlarged dataset with a more even distribution of tokens across syllable types.

How are we to interpret the different alignment of H in different syllable types? One possible interpretation is that the positioning of the H peak is positioned relative to some domain larger than the stressed syllable, such as the stressed foot (Hellmuth 2007). Since we did not label the end of the foot in our current dataset we are not able to test whether a measure of rpd, relative to the foot, is a more uniform predictor of H position across syllable types than rpd relative to the accented syllable.

An alternative explanation of the increasing values of rpd from CVC through CV to V syllables is that the H peak is simply aligned at a fixed duration after the onset of the stressed syllable, regardless of syllable type. This would be consistent with an analysis in which the F0 movement observed on every PWd in Mawng is not a pitch accent (associated with the stressed syllable) but a rising word-level phrase tone, anchored at the left edge of the PWd: ‘LH_w’ (constrained to occur in word-initial, and possibly also root-initial, syllables). We are not able to exclude this alternative analysis, since a test of the relationship between peak delay (pd) and syllable duration (sylldur) shows no correlation (Pearson’s correlation coefficient = 0.083; $p=0.254$; $\alpha = 0.05$). We hope to disambiguate between these two potential analyses via study of a slightly larger, more homogenous dataset. Whilst we note that a phrase-tone-only analysis is proposed for the Gunwinyguan language Dalabon (Fletcher 2007), for the present we continue to analyse the main Mawng F0 movement as a pitch accent (L+H*)[4] by analogy with analysis of regular F0 movements in both Iwaidja and BGW as pitch accents.

Under either scenario however, these quantitative generalisations enable us to incorporate into our main analysis the many instances of ‘delayed peak’ tones in which the H peak falls after the end of the stressed syllable. In their analysis of BGW, Bishop & Fletcher (2005) transcribe such cases as a separate pitch accent type (“delayed high”, ‘H*<’). Our data indicate that in Mawng this pattern is predictable from prosodic context and thus that ‘delayed peak’ tones should be seen not as a different category of pitch accent but as ‘allophonic’ variants of either the main L+H* accent or of an eventual word-level phrase-initial tone LH_w. The three ‘delayed high’ cases provided by way of illustration in Bishop & Fletcher (2005:336-339) are all found on (word-initial) open syllables (*namak*, *bayidurhdurndi* and *nyale*), suggesting that it may be possible to extend this analysis also to BGW.

We are able to make more concrete claims about the patterning of alignment and scaling in focus vs. non-focus tokens. Our quantitative study found that the Mawng ‘special focus accent’ differs from its neutral counterpart in scaling but not in alignment. We therefore analyse the ‘special focus accent’ not as a separate phonological category but instead as a hyperarticulated variant of the standard L+H* main pitch accent. We can liken this to use of contrastive modifications of the tonal space (pitch register) observed by Bishop (2002:17) in BGW. Bishop’s findings for BGW, obtained through qualitative observation, match ours obtained for Mawng through quantitative study. [5] Similar effects in English have also been known for some time, although there has been much debate as to whether the effect should be interpreted as phonologically categorical or gradient (Ladd 1994, Hayes 1994, Ladd 1996). A practical question that arises is how this accent should be transcribed, since it is clearly useful to note when a speaker uses a ‘focus accent’; we suggest that in a fully articulated transcription system for Mawng instances of the focus accent should be transcribed as such on a Misc tier.

Establishing an appropriate methodology for quantitative investigation of intonation in unwritten and/or endangered languages permits parallels to be drawn between these and ‘better-studied’ languages which in turn enable us to challenge and improve upon current versions of intonational theory, requiring them to account for a wider and more typologically varied range of languages. Improved

[4] See Hellmuth et al. (2007) for motivation of this particular phonological representation.
[5] Use of pitch range compression and expansion is also noted in Iwaidja by Birch (2003).

intonational theory can only enhance our understanding of better-studied languages; for example, the findings for Mawng could be argued to support those theories of intonational phonology which analyse interaction between pitch accent distribution and focus structure in terms of syntagmatic (rather than paradigmatic) relations (a focussed item bears *relatively* more prosodic prominence than non-focussed items, Ladd 1996). We hope that corpus-based phonetic studies of intonational phenomena in endangered languages will enhance intonational theory in the years to come.

REFERENCES

Birch, Bruce. 2003. A tour of Iwaidja intonation. (Paper presented at the Australian Linguistics Society Conference, Newcastle, 26-28 September 2003).

Bishop, Judith B. & Janet Fletcher. 2005. Intonation in six dialects of Bininj Gun-wok. In Jun, Sun-Ah (ed.) Prosodic typology: the phonology of intonation and phrasing. 331-361. Oxford: Oxford University Press.

Boersma, Paul & David Weenink. 2007. Praat: doing phonetics by computer (Version 4.6.10) [http://www.praat.org].

Evans, Nicholas. 2003. The non-Pama-Njungan languages of Australia: comparative studies of the continent's most linguistically complex region. Canberra: Pacific Linguistics.

Fletcher, Janet. 2007. Intonation in Dalabon. (Paper presented at 'Intonation of Fieldwork and Less-studied Languages' workshop, ICPhS Saarbruecken, August 5th, 2007.).

Hayes, B. 1994. "Gesture" in prosody: comments on the paper by Ladd. In Keating, P. (ed.) Phonological structure and phonetic form: Papers in Laboratory Phonology III. 64-75. Cambridge: CUP.

Hellmuth, Sam. 2006a. Focus-related pitch range manipulation (and peak alignment effects) in Egyptian Arabic. Proceedings of Speech Prosody 2006. 410-413.

Hellmuth, Sam. 2006b. Intonational pitch accent distribution in Egyptian Arabic. Unpublished PhD thesis, SOAS.

Hellmuth, Sam. 2007. The foot as the domain of tonal alignment of intonational pitch accents. Proceedings of the 16th ICPhS, Saarbruecken, Germany.

Hellmuth, Sam, Frank Kügler, & Ruth Singer. 2007. Tonal alignment and focus in Mawng. (Paper presented at ICPhS2007 satellite workshop 'The Intonation of Fieldwork Languages').

House, Jill. 1989. Syllable structure constraints on f0 timing (poster presentation at LabPhon II, Edinburgh).

Kügler, Frank. 2007. The Intonational Phonology of Swabian and Upper Saxon. Tübingen: Niemeyer.

Ladd, D. R. 1994. Constraints on the gradient variability of pitch range, or, pitch level 4 lives! In Keating, P. (ed.) Phonological structure and phonetic form: Papers in Laboratory Phonology III. 43-63. Cambridge: CUP.

Ladd, D. R. 1996. Intonational phonology. Cambridge: CUP.
Ladd, D. R., I. Mennen, & A. Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. Journal of the Acoustical Society of America. 107(5). 2685-2696.
Lickley, R., A. Schepman, & D. R. Ladd. 2006. Alignment of 'phrase accent' lows in Dutch falling-rising questions: theoretical and methodological implications. (in press). Language and Speech.
Peters, Jörg. 1999. The timing of nuclear high accents in German dialects. Proceedings of the 14th ICPhS, San Francisco, USA.
Prieto, Pilar, Jan van Santen, & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. Journal of Phonetics. 23 429-451.
Rietveld, Toni & Carlos Gussenhoven. 1995. Aligning Pitch Targets in Speech Synthesis - Effects of Syllable Structure. Journal of Phonetics. 23(4). 375-385.
Schepman, A., R. Lickley, & D. R. Ladd. 2006. Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. Journal of Phonetics. 34(1). 1-28.
Silverman, K. E. A. & Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In Kingston, J. & Mary Beckman (eds.) Papers in Laboratory Phonology. 72-106. Cambridge: CUP.
Singer, Ruth. in preparation, 2007. Information structure in Mawng. In Skopeteas, Stavros, S. Hellmuth, Gisbert Fanselow, & Caroline Féry (eds.) Information structure across languages: the interaction of syntax and phonology in comparative perspective.
Singer, Ruth. 2006a. Agreement in Mawng: productive and lexicalised uses of verbal gender agreement in an Australian language. Unpublished PhD thesis, University of Melbourne.
Singer, Ruth. 2006b. Information Structure in Mawng: intonation and focus. Proceedings of the Australian Linguistics Society Conference 2005, September 28-30, ed by K Allan Monash University, Melbourne: http:\\www als asn au.
Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry, & Manfred Krifka. 2006. Questionnaire on Information Structure (QUIS). Interdisciplinary Studies on Information Structure (ISIS): Working Papers of SFB632 University of Potsdam. 4.
Steele, S. 1986. Nuclear accent f0 peak location: effects of rate, vowel and number of following syllables. Journal of the Acoustical Society of America Supplement. 1(80). 51.
Wichmann, Anne & Jill House. 1996. Investigating peak timing in naturally-occurring speech: from segmental constraints to discourse structure. UCL Working Papers in Linguistics.