Variable cues to phrasing: finding edges in Egyptian Arabic[1]

Sam Hellmuth, University of York

1 Introduction

This paper explores variation in the tonal cues to prosodic phrasing observed in a corpus of read speech sentences in Egyptian Arabic (EA). There is variation between speakers in the cues employed to mark instances of the same type (or level) of prosodic juncture in the same position in the sentence, and there is also (what proves to be principled) variation in the types of cues employed at distinct instances of (what are expected to be) the same type/level of juncture occurring at different positions in syntactic structure. This paper seeks firstly to establish what the cues to Major Phonological Phrase (MaP)-level phrasing are in EA, amid all this variation, and, as a secondary goal, to determine whether there is MaP-level marking of 'XP-edges' in EA.

The rationale for this secondary question is to find out whether or not EA presents a challenge to edge-based mapping algorithms (Selkirk 1986, 1995, 2000, Truckenbrodt 1999), and phase-based equivalents (Kratzer & Selkirk 2007, cf. Selkirk to appear). Under edge-based mapping (Selkirk 1986, 1995, 2000), Major Phonological Phrase (MaP) boundaries are expected at the right edge of each embedded XP, whereas, under a canonical phase-based analysis, with CP and *v*P as phase heads (Chomsky 2001, 2005), a MaP boundary is predicted only at the right edge of the subject. Prior work on EA phrasing, in Hellmuth (2004), suggests that EA permits long MaPs which straddle the right edges of XPs, and these facts were there analysed in terms of interaction between prosodic minimality constraints and edge-based syntax-phonology interface constraints. An alternative analysis of this lack of sensitivity to XP-edges, to date unexplored, is that EA's long prosodic phrases could map directly to spellout domains (Adger 2006, Ishihara 2007). In order to determine whether a phase-based analysis of EA is plausible, the present corpus comprises SVO sentences in which both the subject and object position host complex noun phrases containing an embedded prepositional phrase adjunct, to probe EA phrasing for any signs of sensitivity to XP-edges.

In the present data we find conflicting evidence: boundaries are indeed observed in almost all XP-edge positions, but the cues are systematically stronger at the right edge of the subject than at the right edge of other XPs. One solution would be to treat the different cue sets as marking prosodic constituents of different levels, but this fails to take account of the fact that cues to MaP boundaries appear to be variable anyway, and that, even though the cues vary, for each speaker there is a consistent subset relation between the cues observed in different positions. We will argue that all of the variability can be understood in terms of subordination between phonetic implementation domains of adjacent MaPs (following Truckenbrodt 2002, 2004, 2007). Under this view EA proves to be a well-behaved language, with MaP boundaries at XP-edges as predicted by edge-based mapping - albeit subject to prosodic minimality constraints - and we revisit the analysis of Hellmuth (2004) accordingly.

2 Rationale for the study

2.1 Background

Egyptian Arabic (EA) is defined here as the dialect of Arabic spoken in Cairo, which also functions as a national standard (Haeri 1996, Bassiouney 2009). The segmental and metrical phonology of EA is well-described (see Watson 2002 for a summary), and syntactically EA displays SVO basic word order (Benmamoun 2000, Edwards 2009). Watson (2002) describes a number of segmental sandhi processes which are sensitive to (major) phonological phrase (MaP) level prosodic boundaries. The only segmental sandhi process investigated experimentally to date is epenthesis (which applies to break up any CCC cluster), which has been shown to apply across MaP boundaries (Hellmuth 2004, cf. Aquil 2006).

A salient feature of spoken colloquial EA intonation[2] is its unusually rich distribution of pitch accents: in EA a pitch accent is observed on almost every content word (Hellmuth 2006), and shows a very restricted range of pitch accent shapes, analysed as a single phonological category (L+H*) in Hellmuth (2006, see also Chahal & Hellmuth to appear). The twin properties of rich accent distribution and restricted accent inventory are shared with a number of other languages including Spanish and Greek (Jun 2005). For EA, this accent distribution pattern can be formalised in terms of a requirement that phonological tone, in the form of an intonational pitch accent, be associated with every Prosodic Word (PWd, Hellmuth 2007), and this pattern sets EA apart from languages in which the distribution of pitch accents is better explained relative to phrase level prosodic constituency (see Truckenbrodt, this volume, for an overview).

Prosodic phrasing in EA was investigated in Hellmuth (2004), on the basis of a corpus of read speech sentences in which the syntactic complexity and prosodic weight of arguments was systematically varied (replicating the design of Frota et al. 2007). The key finding was that phrases at the MaP level are long in EA, so that most utterances were realised within a single MaP, straddling utterance internal XP-edges. A mid-utterance boundary was only observed when the subject NP was both syntactically complex and prosodically heavy; in particular, a phrase boundary only occurred after the subject NP when it contained at least four PWds.[3] The findings were subsequently checked in a corpus of semi-spontaneous speech (narratives re-told from memory) and the same generalisation was found to hold (Hellmuth 2007): utterances were in general realised within a single MaP in EA, and straddled XP edges.

Examples from the Hellmuth (2004) dataset are reproduced below, showing the sentence in (1) realised either within a single MaP, as in Figure X.1, or in two MaPs, with a phrase break at the right edge of the subject (after [muhimm] ‘important’), as in Figure X.2. The cues to phrasing observed in Figure X.2 are pre-boundary lengthening and a local pitch reset after the boundary.

(1)	il-muˈhandis	l-maʕˈmaːri	l-muˈhimm	bijˈxumm	baˈlad-na
	[[the-[engineer	[the-architectural]$_{AP}$	[the-important]$_{AP}$]$_{NP}$]$_{DP}$	[[cheats-1ms	[our-country]$_{DP}$]$_{VP}$]$_{vP}$]$_{TP}$
	‘The important architect is cheating our country.’				

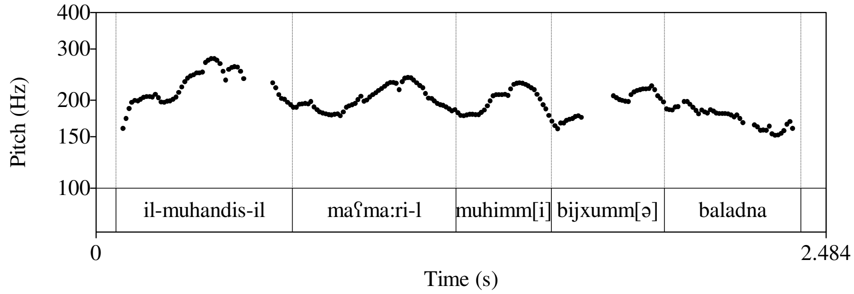

Figure X.1: Sample SVO sentence realised in a single MaP from Hellmuth (2004:102)

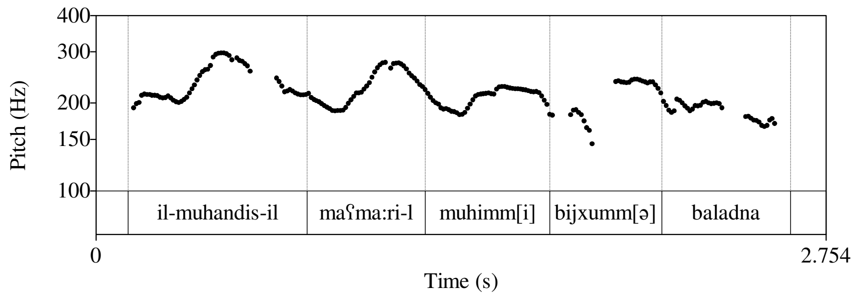

Figure X.2: Sample SVO sentence realised in two MaPs from Hellmuth (2004:102)

In an edge-based formalisation of the syntax-phonology interface (Selkirk 1986, Truckenbrodt 1999) an MaP boundary is expected at XP edges, in particular at the right edge of the subject in an SVO sentence. Following Selkirk (2000), the EA facts were formalised in an OT analysis in terms of interaction between an edge-based interface constraint (ALIGNXP) and constraints on the phonological well-formedness of prosodic constituents. Specifically, Hellmuth (2004) proposed two prosodic well-formedness constraints in EA: BINMAP requires each MaP to be comprised of at least two Minor Phonological Phrases (MiP), and BINMIP requires each MiP to be comprised of at least two PWds. In EA BINMAP outranks ALIGNXP so that a subject is only phrased alone if of sufficient prosodic weight, that is, when it contains four PWds. To account for the fact that some 3-PWd MaPs were observed in slower speech (e.g. in Figure X.2 above), it was suggested that BINMIP is reranked or redefined in slow speech, but the analysis was not fully worked out.

The cues to phrasing observed in Hellmuth (2004) were chiefly tonal. A phrase boundary was transcribed whenever two or more of the following were observed: local pitch reset, final lowering, pre-boundary lengthening, failure of epenthesis, pause or a phrase tone (H-/L-). Very few phrase boundaries were found, just 20 in a corpus of 234, and the two speakers in the study appeared to use slightly different sets of cues. Evidence to confirm that the phrase-boundaries observed were at MaP level came from additional sentences in the dataset which contained a mid-sentence parenthetical, which are expected to be phrased as independent IPs (Nespor & Vogel 1986, Kawahara this volume). The phrasing cues observed at the edges of parentheticals differed somewhat from those observed at the (few) other boundaries, in that epenthesis failed to apply across boundaries at the edges of parentheticals, but did apply across other boundaries even in the presence of the other tonal and durational cues outlined above.[4]

2.2 Aims

This paper seeks to address issues which remain outstanding or arise from the findings of Hellmuth (2004). The first of these is to pin down the cues to phrasing in EA. In many other languages the cues to prosodic phrase edges are well-established, and indeed seem to be shared across speakers (e.g. the use of phrase tones, pre-boundary lengthening or other tonal cues). The two speakers investigated in Hellmuth (2004) varied in the cues to phrasing that they used, making it difficult to make generalisations about how phrasing is cued in EA. In this paper we examine data from six speakers in an attempt to establish more clearly what the cues to phrasing are in EA, and in particular to find out whether there are any consistent cues used by all speakers.

Given the claim advanced in Hellmuth (2004) that the long phrases observed in EA arise due to a phonological constraint on the minimum size of MaPs, and that this is what causes phrases to straddle XP edges, we investigate here read speech data using stimuli in which all arguments are prosodically heavy and syntactically complex. This should trigger a greater number of phrase boundaries, facilitating the task of establishing cues to phrasing described above, but also providing a testing ground for the formal analysis proposed in Hellmuth (2004).

An alternative explanation for the lack of boundaries at some XP-edges is that the syntax-phonology mapping does not call for a boundary in that position in the first place. In a canonical phase-based formulation of the syntax-phonology interface, only CP and *v*P act as phase heads, spelling out TP and VP respectively as prosodic domains (Chomsky 2001, Adger 2006, Ishihara 2007, Kratzer & Selkirk 2007, cf. Truckenbrodt this volume). If we assume that in EA SVO sentences the subject is merged in the specifier of *v*P (cf. Ackema & Neeleman 2003), this predicts a prosodic boundary between the subject and verb (as VP is spelled out), but not at embedded XP edges within a complex NP. The current dataset comprises SVO sentences in which the object contains an NP complement plus a following (VP-internal) PP adjunct: an edge-based mapping predicts a boundary between the two XPs but a phase-based analysis does not.

The influential edge-based mapping first proposed in Selkirk (1986) has provoked fruitful work in many languages over a long period (Inkelas & Zec 1990, Truckenbrodt 1995, Truckenbrodt 1999), and a fundamental role for XPs in the syntax-phonology mapping is maintained in Match Theory (Selkirk 2009, to appear). This paper looks for evidence of a role for XPs in the syntax-phonology mapping, in a language which appears at first glance to have no need of them.

3 Methodology

The present study is designed to allow us to pick up where Hellmuth (2004) left off, eliciting read speech productions of target sentences in which all arguments are both syntactically complex and prosodically heavy. The present paper presents a fine-grained analysis of 8 SVO sentences produced 3 times each by 6 female speakers, yielding 108 tokens for input to both qualitative and quantitative analysis.[5] The materials and recording procedures are set out in §3.1, followed by a description of the qualitative and quantitative analyses in §3.2.

3.1 Materials and data collection

Eight SVO sentences were constructed in which both subject and object argument positions contained a complex NP which was both syntactically complex and prosodically heavy. The verb and object NP were the same in all eight sentences, but the subject NP varied from each other in the number of syllables, feet and prosodic words that they contained. The stimuli are provided in Table 1 and 2. The two potential boundary positions of interest here are indicated with an arrow, at the right edge of the subject ('after S' position) and at the right edge of the complement NP within the VP ('within VP' position).

Table 1: Structural template for target sentences analysed in the present study (indicating investigated potential boundary positions)

subject		*verb*	*object*	
head + complement	*adjunct PP*		*head + complement*	*adjunct PP*
il-muˈdiːr il-giˈdiːd	min juˈnaːn	bijitˈʕallim	it-tadˈriːb il-ħadiːs	fi kulˈlejat it-tarˈbeja
[[the-[manager [the-new]$_{AP}$]$_{NP}$	[from Greece]$_{PP}$]$_{DP}$	[[teaches.3ms	[the-[pedagogy [the-modern]$_{AP}$]$_{NP}$]$_{DP}$	[in [faculty the-education]$_{CS}$]$_{PP}$]$_{VP}$]$_{vP}$]$_{TP}$
	after S ➶		*within VP* ➶	

Table 2: Internal structure of the subject in the 8 target sentences analysed in the present study (the number of PWds, feet and syllables in the subject as a whole are indicated in the leftmost columns)

				subject			
target	*PWds*	*Ft*	*σ*	*head + complement*			*adjunct PP*
008	3	3	9	il-muˈdiːr	il-giˈdiːd		min juˈnaːn
				[[the-[manager	[the-new]$_{AP}$]$_{NP}$		[from Greece]$_{PP}$]$_{DP}$
010	3	4	9	il-ʔusˈtaːz	il-giˈdiːd		min juˈnaːn
				[[the-[professor	[the-new]$_{AP}$]$_{NP}$		[from Greece]$_{PP}$]$_{DP}$
012	3	4	11	il-muˈdiːr	il-mutaˈdajjin		min juˈnaːn
				[[the-[manager	[the-devout]$_{AP}$]$_{NP}$		[from Greece]$_{PP}$]$_{DP}$
014	3	5	11	il-ʔusˈtaːz	il-mutaˈdajjin		min juˈnaːn
				[[the-[professor	[the-devout]$_{AP}$]$_{NP}$		[from Greece]$_{PP}$]$_{DP}$
022	4	4	13	muˈʕallim	wiˈlaːdi	il-muˈʔaddab	min juˈnaːn
				[[[teacher	my-children]$_{CS}$	[the-polite(m.s.)]$_{AP}$]$_{NP}$	[from Greece]$_{PP}$]$_{DP}$
018	4	5	14	muˈʕallim	wiˈlaːdi	il-mutaˈdajjin	min juˈnaːn
				[[[teacher	my-children]$_{CS}$	[the-devout(m.s.)]$_{AP}$]$_{NP}$	[from Greece]$_{PP}$]$_{DP}$

The VP contains 5 PWds in all of the sentences, and is thus of sufficient weight to allow the subject to be phrased alone. If the findings of Hellmuth (2004) generalise to other speakers, then we would expect all of the 4-word subjects to be phrased alone, and at least some of the 3 word subjects (depending on speech rate). Since the sentence-final adjunct PP comprises only 2 PWds we would not expect to observe within-VP boundaries other than at slow speech rates. Note that the 4-word subject target sentences have a two word Construct State (CS) in head noun position.[6]

The 8 sentences formed part of a larger set of 44 sentences with varying lexical content and structure, which were pseudo-randomised and interspersed with distractors. The sentences were presented in Arabic typescript, using EA spelling conventions in order to elicit colloquial productions (cf. Siemund et al. 2002). Each speaker read all of the sentences through three times, performing another unrelated task between each repetition. The recordings were made in Cairo with 6 female speakers of EA. All were mother tongue speakers of EA, born and raised in Cairo, aged between 21-34 years, and none had any auditory or speech production difficulties. Speakers received a small payment as thanks for their participation. The recordings were made in a draped classroom using AKG headset condenser microphones, directly to digital .wav format in ProTools 6.0 on MBox at 44100Hz 16bit, then re-sampled at 22050Hz 16bit.

3.2 Qualitative and quantitative analysis

The full dataset of 108 tokens was submitted to both qualitative and quantitative analysis. All tokens were included in the analysis, even if they contained minor disfluencies and/or an atypical phrasing pattern. All analysis was undertaken using Praat (Boersma & Weenink 2009).

The qualitative analysis took the form of hand-labelling of information on three tiers, carried out by the author based on auditory impression and inspection of the pitch trace and spectrogram, as illustrated in Figure X.3 below. Firstly, the orthography tier segments the utterance into words and provides a broad phonetic transcription for each one; enclitics were grouped with their preceding host word. The actual transcription used a standard transliteration system for Arabic, but examples for the present paper have been re-transcribed into IPA for ease of exposition.

Secondly, on the Break Indices tier, phrase-level junctures were transcribed using the break indices listed in the Table 3 below. These are prototype break indices used here for the first time for EA, in order to test their usability. They assume the number of levels of phrasing proposed in Hellmuth (2004) and are modelled on the Break Indices proposed for Hebrew by Shaked (2007), which also assume an intermediate level of phrasing between the PWd and MaP. Break indices levels 0 and 1 were not transcribed for the present study, in order to speed up transcription.

Table 3: Prototype break indices for EA used in the present study

BI		*labelled?*
0	Morphosyntactic word boundary	no
1	PWd boundary	no
2	MiP boundary	yes
3	MaP boundary	yes
4	IP boundary	yes

Table 4: Break indices proposed for Hebrew (Shaked 2007:180)

BI	*definition*
0	Phonetically marked boundaries internal to the clitic group.
1	Clitic-group boundaries which are not prosodic word boundaries (e.g. internal to compounds and construct state nominals).
2	Word boundaries marked by pitch accent but no intonational phrase boundaries (minor/accentual phrase boundaries).
3	Intonational phrase boundaries marked by a single rising phrase tone (major/intermediate phrase boundaries).
4	Full intonation phrase boundaries, marked by a falling boundary tone.

Table 4 provides the definitions given by Shaked (2007) for Hebrew, for the boundaries of interest here. We have adopted a similar number of levels, but the definitions were not adopted a priori from Hebrew. An obvious case where the two languages appear to differ is in the distribution of pitch accents. Since in EA every PWd is expected to bear a pitch accent, the definition of BI-2 used for Hebrew by Shaked would in EA yield segmentation at every PWd boundary, rather than at MiP boundaries. Instead, we used BI-2 in EA to label MiP boundaries marked with clear phrasing cues.

Similarly, Shaked's definition of BI-1 expects a construct state (CS) nominal to be treated as a single PWd in Hebrew. CS nominals appear throughout the present dataset (e.g. utterance-final: [kullejat it-tarbeja] faculty-the-education 'the Faculty of Education') and were treated as two separate PWds by most speakers with a pitch accent on each word, as is the case in prior studies on EA (including Hellmuth 2004). We note in §4 below however that one speaker in the present dataset did produce an atypical tonal pattern on these final CS constructions, which could be seen as prosodification of the whole CS into a single PWd, and such an analysis would require an additional BI for EA defined similarly to Shaked's BI-1 for Hebrew.

The final layer of qualitative transcription was a list of the cues to phrasing observed at each labelled BI juncture point, which were transcribed using the codes listed in Table 5 below. These codes were based on the set of cues observed in Hellmuth (2004). Examples of many of the cues can be seen in the sample labelled textgrid in Figure X.3 below. Note, for example, the contrast between H (denoting rising pitch on and beyond the accented syllable) and U (denoting a peak at a higher level than the previous peak, following by a fall immediately after the accented syllable).

Table 5: Cue labels used for fine-grained description of each juncture (BI 2-4)

label	*name*	*definition*
B	boundary tone	boundary shows a full boundary tone (usually a final fall)[7]
D	downstep (final lowering)	peak of the word at the boundary is produced at a lower level than expected[8] from effects of downstep alone, relative to pitch level of the previous peak
H	phrase tone	boundary shows either a H- or L- phrase tone
L	lengthening	word at the boundary is lengthened
P	pause	boundary is followed by pause (filled, e.g. with an in-breath, or unfilled)
R	reset	following peak is produced at a higher level than the peak of the word at the boundary[9]
S	suspension of downstep	peak of the word at the boundary is produced at the same level as the previous peak
U	upstep	peak of the word at the boundary is produced at a higher level than the previous peak

Figure X.3 shows a fully labelled textgrid as created for each soundfile. The labelled tiers are as follows (from top to bottom): cues, BI, peaks, orthography, rhymes.

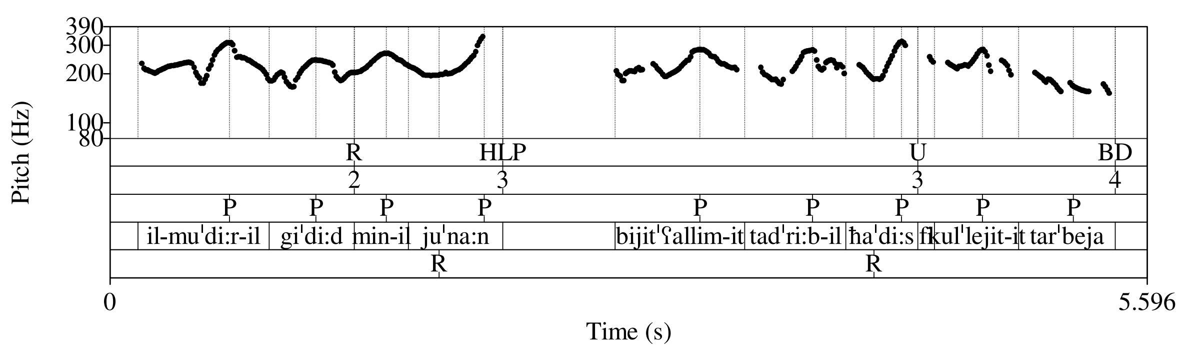

Figure X.3: Sample textgrid showing qualitative transcription and labelling for quantitative analysis.

In order to provide independent support for the qualitative analysis, two sets of quantitative measures were extracted from each sound file: f0 and duration. To facilitate accurate f0 measurements, a Pitch object was created for each sound file in Praat (floor 75Hz, ceiling 600Hz) and each Pitch object hand-corrected for pitch tracking errors, such as doubling, halving or faulty pitch tracking due to voice quality variation (for example, creaky voice was very common in the last one or two syllables of each utterance). After hand-correction of the pitch trace, a Praat script was used to identify the pitch maximum within each labelled interval on the orthographic tier, and a peak label inserted at that point. During hand-labelling of BIs and phrasing cues, the position of peak labels was also hand-checked (e.g. to delete automatically inserted peak labels on any function words which were not in fact assigned a pitch accent). A further Praat script harvested the f0 value in Hertz (measured in the hand-corrected Pitch object) at each labelled peak position in the utterance. The f0 maximum identified for a PWd at a boundary could be the reflex either of a pitch accent on that word or of a phrase tone at its right edge (e.g. if the boundary is marked with a high phrase tone). This ambiguity was tolerated so that the f0 values could act as an independent source of evidence alongside the qualitative analysis.

As a source of independent evidence for pre-boundary lengthening we measured the duration of the rhyme portion of target words at the right-edge of XP constituents that might reasonably be expected to trigger a prosodic boundary (Wightman et al. 1992). The rhyme of the final word in the subject ([junaːn] ‘Greece’) and of the final word in the complement NP of the VP ([ħadiːs] ‘modern’) was hand-labelled in each utterance. The duration of each labelled rhyme was extracted in milliseconds, then normalised by dividing each rhyme duration by the length of the whole sentence of which it was a part, and multiplying by 100, yielding a measure of each rhyme as a percentage of its host utterance.

4 Results

The transcription results show that BI-3 level MaP boundaries were inserted by speakers in after-S position in all 108 cases and in within-VP position in 91/108 cases.[10] In this section we first set out those aspects of inter-speaker behaviour which did not vary significantly in §4.1 (looking at speech rate and utterance-final cues). The main body of the section then discusses the cues to phrasing used in after-S position, which do show inter-speaker variation, in §4.2, with comparison to the cues observed at within-VP boundaries, which are different again, in §4.3.

4.1 Lack of inter-speaker variation in speech rate effects and utterance-final cues

Since phrasing in EA has been found to vary according to speech rate (Hellmuth 2004), a preliminary analysis of the speech rate used by each speaker was carried out to determine whether this affected phrasing choices in the present dataset. Figure X.4 shows mean values of speech rate (syllables per second) for each speaker, presented in order from slowest speaker to fastest. An ANOVA reveals that the speakers fall into two clear subsets ($F=19.944$, $df=107$, $p<0.001$)[11]: ‘slow’ speakers are fhg, fhm, fna; ‘fast’ speakers are fsf, faa, fhx.

Figure X.5 shows the mean number of phrase boundaries (at BI 3 or above) observed for each speaker per utterance, again presented in order from slowest to fastest speaker.[12] At first glance it would appear that the two fastest speakers (faa and fhx) do indeed produce fewer boundaries, but an ANOVA shows that the difference between speakers is only significant between the fastest speaker (fhx) and the two speakers who produced the most boundaries (fhm and fsf), who are not the slowest speakers.[13] Speech rate is thus excluded as a factor in the remainder of this paper.

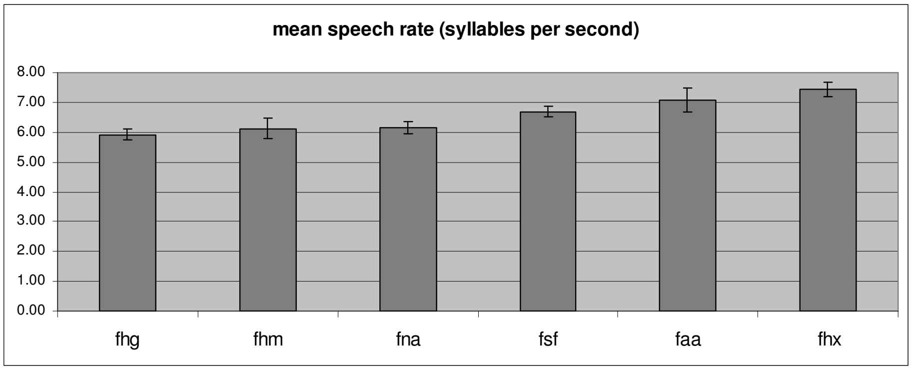

Figure X.4: Mean speech rate by speaker (slowest on the left, fastest on the right)

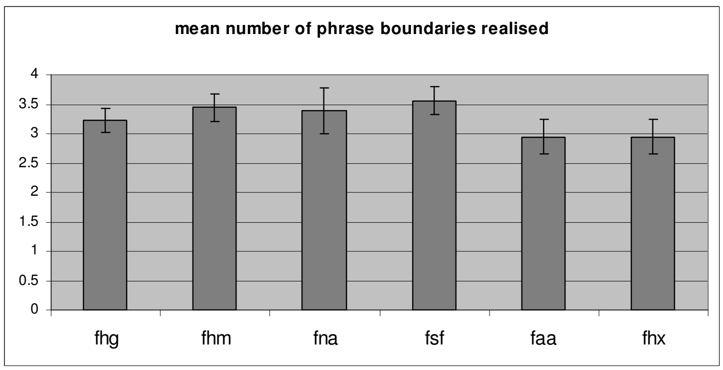

Figure X.5: Mean number of boundaries per utterance (by speaker)

Speakers also showed little variation in their use of cues in utterance-final position, as shown in Table 6 below. All speakers used a boundary tone (B), and in most cases, the height of the final peak was realised lower than expected, showing final-lowering (Liberman & Pierrehumbert 1984).[14] The only exception to this pattern was speaker fna (see Figure X.10 below) who often realised the utterance-final word with upstep relative to the peak height of the previous word.[15]

Table 6: Cue clusters transcribed in utterance-final position

fhg	*fhx*	*fhm*	*fsf*	*faa*	*fna*
BD (18)	BD (18)	BD (18)	BD (18)	BD (17)	BU (12).
				B (1)	BD (3)
					BS (1)
					B (2)

4.2 Speaker variation in cues to phrasing

There is inter-speaker variation in the types of cues used to mark prosodic constituents of the same type in EA. To demonstrate this we will first present a summary of the auditory transcription results, to identify which speakers used which sets of cues. We support the transcription results with independent quantitative evidence of two kinds: measurements of the f0 of successive H peaks through each utterance and the duration of target words at putative XP-edge positions. Finally, although pre-boundary-lengthening is used by most speakers at most boundaries, it is not observed at all of the cases labelled as boundaries during transcription: we explore evidence from the f0 of successive peak heights to determine whether it is possible to have an MaP boundary in EA which is not cued by pre-boundary lengthening.

All of the speakers realised a boundary after the subject in all of their utterances. This means that they pattern with the slow speech phrasing generalisations observed in Hellmuth (2004), in that they allow a 3-word subject to be phrased alone. The full set of cues observed in the after-S potential boundary position for each speaker are listed in Table 7 below.

All six speakers make use of pre-boundary lengthening, at most boundaries (though not all). Pauses were also used by all of the speakers but much less. With regard to their use of tonal cues, the six speakers can be divided into three groups according to their preferred cue choice in this position: speakers who tend to mark boundaries with partial reset (R), speakers who tend to mark boundaries with a phrase tone (usually H-),[16] and speakers who use a mixture of cues.

Table 7: Cue clusters transcribed in 'after-S' position, speakers grouped by cue preferences (for each speaker the most frequent cluster is underlined, most frequent cue(s) are in bold).

R-speakers		*H-speakers*		*mixed-cue speakers*	
fhg	*fhx*	*fhm*	*fsf*	*faa*	*fna*
BL**R** (1)	D**h**LP (1)	**H**L (11)	BLP (1)	HLP (3)	**H**L (1)
BP**R** (1)	D**h**LP**R** (1)	**H**LP (5)	**H**L (2)	LP (1)	**H**LP (3)
HLP**R** (1)	**h**LP (5)	**H**LR (1)	**H**LP (14)	LPRU (1)	**H**LR (1)
HP**R** (1)	HLP (1)	U (1)	LPRS (1)	LRU (2)	L (2)
LP (1)	**h**LP**R** (6)			L**S** (7)	LP**U** (1)
LP**R** (6)	L**R**S (1)			L**S**P (1)	LR**U** (1)
LP**R**S (3)	L**R**U (1)			LU (1)	L**S** (4)
L**R** (3)	LS (2)			**S** (2)	L**U** (2)
L**R**S (1)					R (1)
					S (1)
					U (1)

Speaker fhg consistently marks the after-S boundary with partial reset (R, in 17/18 cases) and speaker fhx frequently makes use of partial reset in combination with a low phrase tone on the boundary itself; they are grouped together as *R-speakers* as a result (see Figures X.6-7). Speakers fhm and fsf are *H-speakers* (see Figures X.8-9): both consistently mark after-S boundaries with a high phrase tone (H). In contrast, speakers faa and fna are *mixed-cue speakers* who vary in their use of f0 cues (see Figures X.10-11): for faa, the most common cue is to suspend downstep on the last item in the phrase (S), but phrase tones (H), upstep (U) or a following partial reset (R) are also observed; speaker fna varies between using a phrase tone (H), upstep (U) and suspension of downstep (S).

The cue labels and clusters listed in Table 7 are derived from impressionistic transcription but find independent support from quantitative analysis of the f0 properties of the utterances. The mean value of the f0 measurements for successive peaks throughout the utterance is provided in Figures X.6-11, for each speaker in turn. The f0 values for each speaker are represented scaled within her own pitch range (measured across all utterances).

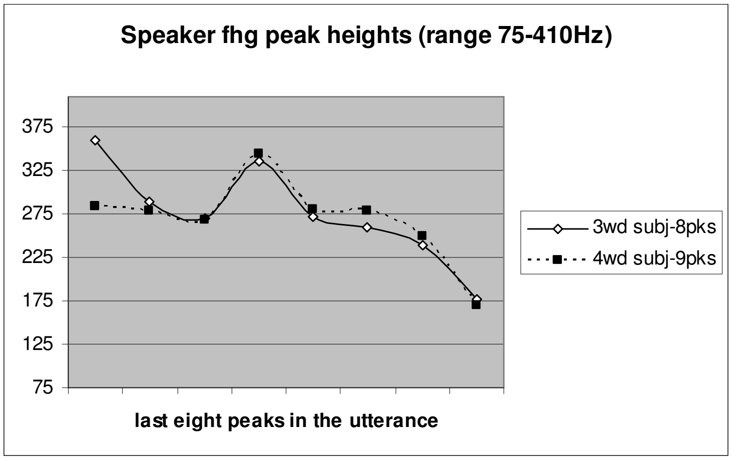

Figure X.6: Speaker fhg peak heights

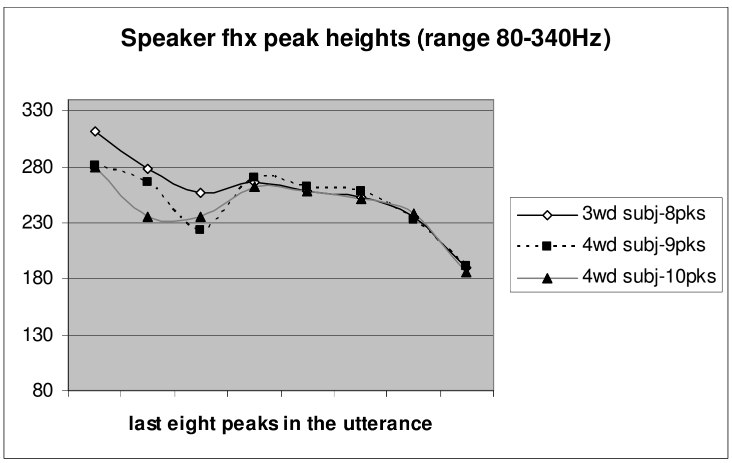

Figure X.7: Speaker fhx peak heights

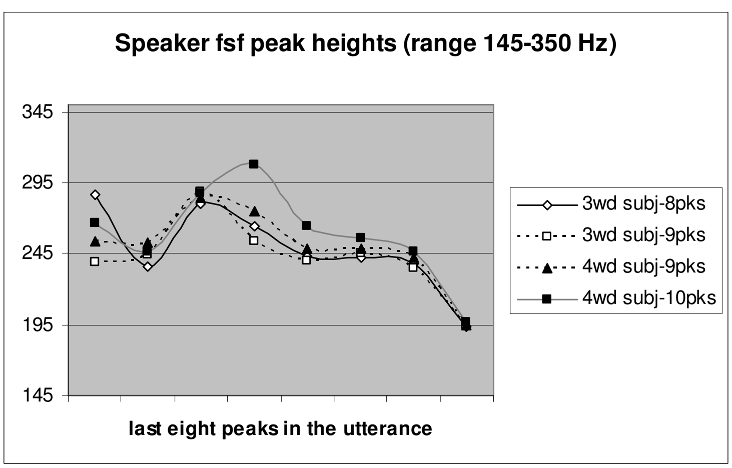

Figure X.8: Speaker fsf peak heights

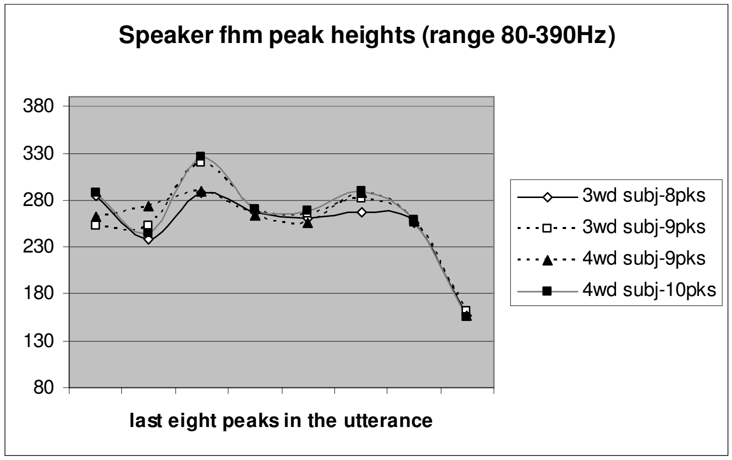

Figure X.9: Speaker fhm peak heights

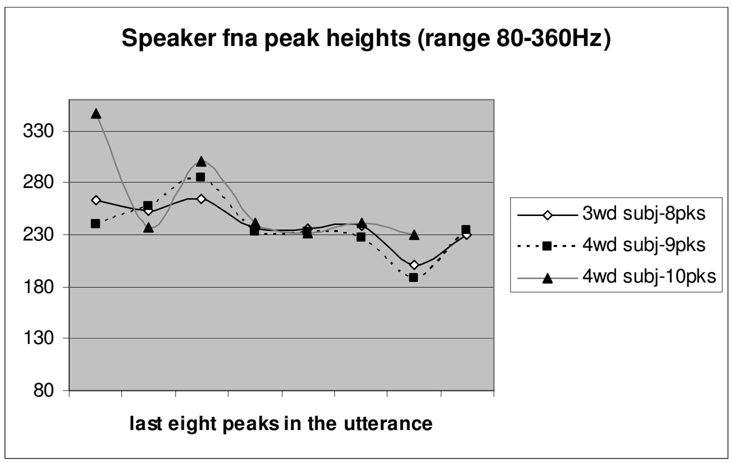

Figure X.10: Speaker fna peak heights

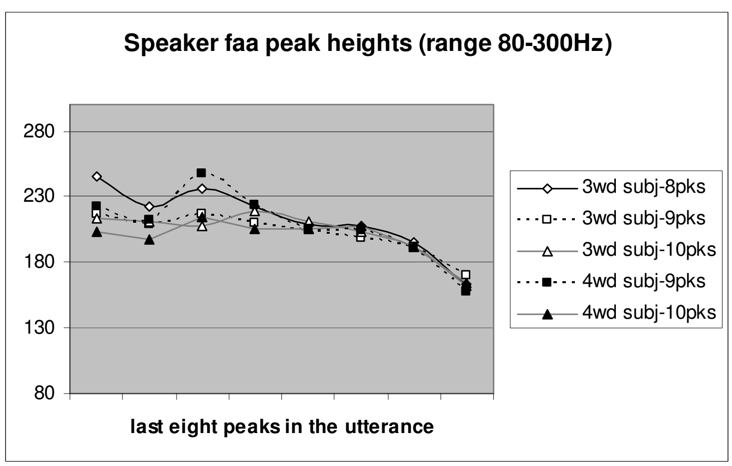

Figure X.11: Speaker faa peak heights

When interpreting Figures X.6-11 note that only the f0 values of the last eight peaks in the utterance are presented, for two reasons. The utterances vary in length, since some contain a 3-word subject and some a 4-word subject, but it is also possible for speakers to vary the number of pitch accents they produce when realising an utterance, even across productions of an identical sentence.[17] Variation in the number of accents only occurred within the subject, so in the representations below we present the last eight peaks, showing the whole utterance in 8 peak realisations but excluding the first 1-2 peaks in longer realisations. This presentation allows direct comparison across utterances with different numbers of peaks: the first three peaks in each figure represent the average f0 values in the last three peaks of the subject in every case.

In Figures X.6-7, for speakers fhg and fhx respectively, we see a clear reset of pitch at the start of the VP (on the 4th peak from the left), which matches the impressionistic transcription of their frequent use of partial reset (R) as a boundary cue. For fhx, the third peak (at the right edge of the subject) appears especially low in 4word-subject sentences due to her consistent use of a low phrase tone at the after-S boundary in these sentences. Figures X.8-9 show peak heights for the two H-speakers. In Figure X.8, for speaker fsf, we can see a clear rise in pitch for the H- phrase tone, at the right edge of the subject (on the third peak). In Figure X.9, for speaker fhm, the rise is somewhat smaller for 4-wd subject sentences with the phrase tone on peak 4 (i.e. at the end of the subject). The mixed cues speakers f0 values are shown in Figures X.10-11. Although the pattern of labels needed to describe the pattern used here is mixed, it turns out that that both speakers' peaks show a clear downward stepping shape through all utterances, so that the combination of suspension of downstep/upstep at the boundary itself, with use of following partial reset in some cases, produces a consistent overall pattern, closely resembling that observed in Figures X.8-9 (for the H-speakers).

We noted above that pre-boundary lengthening was observed for all speakers in after-S position in most cases. This labelling was also made impressionistically and thus requires independent corroboration. Measurements were taken of the rhyme portion of the final word in the subject ([juna:n] 'Greece') as a means of detecting pre-boundary lengthening (Wightman et al. 1992), then normalised for speech rate variation by calculating the duration of each rhyme as a percentage of its host utterance.

The present dataset contains only utterances with prosodically heavy subjects, so we do not have any tokens without a boundary in after-S position to use as controls. Nevertheless we can test the reliability of the impressionistic use of the 'L' label in after-S position in the present data by comparing these cases with those which were transcribed as having a boundary in after-S position but without the use of lengthening (L) as a cue. Comparison of mean normalised rhyme durations (nr1dur) in the final word of the subject, grouped by labelling with/without lengthening, shows that the cases transcribed impressionistically as +L (n=100) are indeed longer than those transcribed as -L (n=8), as illustrated in Figure X.12 below. The difference in mean values between the two groups is statistically significant (oneway ANOVA: $F=9.110$, $df=106$, $p=0.003$).

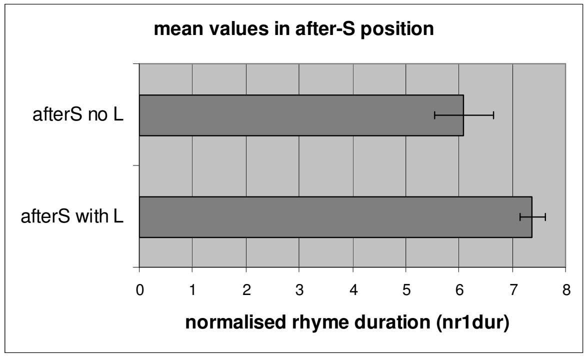

Figure X.12: Mean normalised rhyme durations in last word of the subject, labelled with/without an L cue.

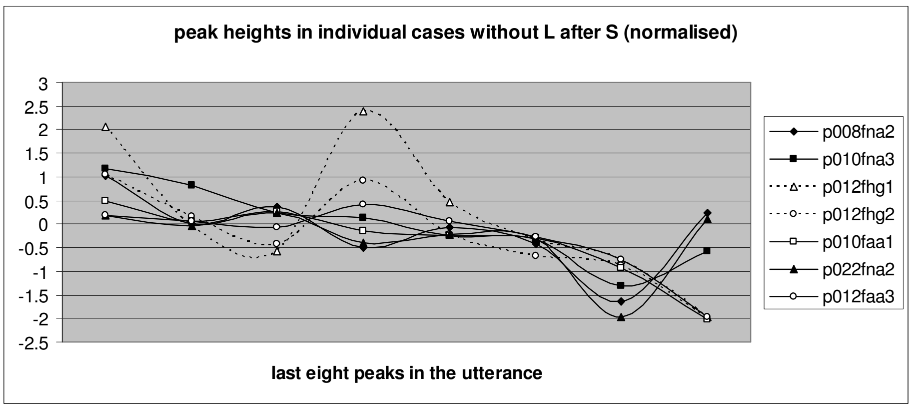

Figure X.13: Normalised f0 values in cases labelled as a boundary without an L cue, in after-S position.

The durational evidence confirms that boundaries labelled as lengthened were indeed lengthened. What evidence is there to support the claim that boundaries marked without an L cue are indeed boundaries? Peak f0 measurements for the 7 tokens labelled with exclusively tonal cues in after-S position are shown in Figure X.13.[18] Speaker fhg (an 'R' speaker) shows partial reset on peak 4, whereas peaks 3 and 4 are similar in height for the other two 'mixed cue' speakers (faa and fna) who use a combination of suspension of downstep at the edge of the subject and following partial reset. This suggests that it is justified to treat these tokens as having a boundary, even without pre-boundary lengthening as a cue, though we note that lengthening is observed in most cases.

In summary then, all six speakers produced a boundary at the right edge of the subject XP, but they vary in the cues to phrasing used at these boundaries. Pre-boundary lengthening is almost always used, but does not appear to be absolutely required, and although a phrase tone (H-) was frequently used, again, it does not appear to be a necessary cue. The most consistent cues appear to be those that are the reflexes of phonetic implementation of downstep register domains: suspension of downstep, partial rest and upstep. In the next section we explore the cues observed at boundaries in within-VP position.

4.2 'Positional' variation in cues to phrasing

In this section we motivate the claim that there is 'positional' variation in the types of cues used to mark XP edges in EA. To do this we detail the cue clusters observed at BI-3 boundaries in the within-VP position, and argue that there is a subset relation between the types of cue clusters observed in the two key predicted boundary positions: cues in after-S position are stronger than those in within-VP position, for all speakers.

Overall there were slightly fewer boundaries observed in within-VP position (91/108) than in after-S position (108/108). This confirms that the speakers in the current study are patterning with the slow speech generalisations of Hellmuth (2004). Table 8 lists the cue clusters observed at the within-VP boundary for all speakers (grouped by their after-S cue choice preferences).

Table 8: Cue clusters transcribed in within-VP position, speakers grouped by after-S cue preferences (for each speaker the most frequent cluster is underlined, most frequent cue(s) are in bold).

after-S 'R-speakers'		*after-S 'H-speakers'*		*after-S 'mixed-cues'*	
fhg (n=18)	*fhx (n=11)*	*fhm (n=18)*	*fsf (n=18)*	*faa (n=11)*	*fna (n=15)*
H**L** (1)	LP**S** (1)	H (1)	H**L**R (1)	**L** (5)	HLP (2)
H**L**P (1)	L**S** (1)	L**U** (1)	**L** (2)	**L**S (3)	L**S** (2)
L (9)	LU (1)	R (2)	**L**PS (1)	S (3)	LU (1)
LR (1)	R (1)	RS (1)	**L**R (4)		**R** (5)
LS (5)	**S** (5)	R**U** (1)	**L**S (9)		**RS** (2)
LU (1)	U (2)	S (3)	S (1)		**S** (3)
		U (9)			

Here, again, the speakers differ from each other in their preferred cue clusters, but if we were to propose groupings among them based on within-VP cue preferences these would require different groupings: cue preferences at after-S position do not directly predict cue preferences at within-VP position. Overall there is less use of pre-boundary lengthening (L), but it is still the primary within-VP cue for speakers fhg, fsf and faa. The other three speakers mostly rely on tonal cues: fhx mostly uses suspension of downstep (S), as does fna in combination with partial reset (R); fhm mostly uses upstep (U).

In order to confirm the impressionistic labelling of 'L' at within-VP position measurements were taken of the rhyme portion of the final word in the complement NP ([ħadiːs] 'modern'), then normalised by calculating the duration of each rhyme as a percentage of its host utterance. Since there are cases both with and without a boundary in within-VP position, as well as instances of boundaries labelled with/without L, we make a three-way comparison. Comparison of mean normalised rhyme durations (nr2dur) in the final word of the complement NP to the verb, grouped by presence of a boundary and labelling with/without lengthening, illustrated in Figure X.14 below, shows firstly that the boundary cases transcribed impressionistically as +L (n=54) are all longer than the boundary cases transcribed as -L (n=38) and than the cases transcribed without a boundary (n=16). Secondly, there is no difference in mean values of rhyme duration between non-boundary cases and those transcribed with a boundary but without lengthening. A one-way ANOVA confirms that the differences are significant ($F=34.512$; $df=105$; $p<0.001$).[19]

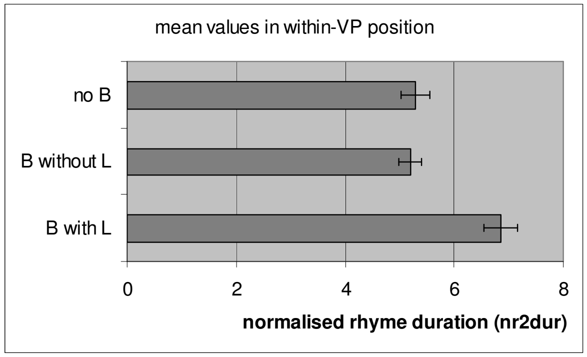

Figure X.14: Mean values for normalised rhyme durations in the final word of the complement NP in the VP, labelled with/without a boundary and with/without an L cue.

Is there any evidence that the cases labelled as having boundaries at within-VP position, but without pre-boundary lengthening (L) as a cue, are indeed boundaries? Figure X.15 shows f0 peak heights averaged for each speaker in the last five peaks of the utterance (that is, in the entire VP), in the 39 cases labelled as a boundary but without pre-boundary lengthening at the within-VP boundary. For comparison, Figure X.16 shows the parallel peak measurements in 17 tokens labelled as having no boundary at all.

In interpreting Figures X.15-16 recall that a within-VP boundary is expected to pick out the middle peak in the VP (at the right edge of the complement NP), which should show signs of upstep or suspension of downstep, or a following reset. The f0 trends in Figure X.15 do all show some kind of suspension of the downstep pattern on or after the middle peak, though the effect is smaller for some speakers (e.g. fna) than others (e.g. fhm). Examination of the f0 trends in Figure X.16 suggests that a within-VP boundary should have been transcribed in at least some of these cases for speaker fhx (non-boundary cases for fhx, n=7), whose peaks (on average) show a reset before the final PP adjunct phrase. In contrast, there was indeed no boundary here for speakers faa and fna (compare level pitch between peaks 2 and 3 in the VP in Figure X.15 with falling pitch between peaks 2 and 3 in Figure X.16, for these speakers).

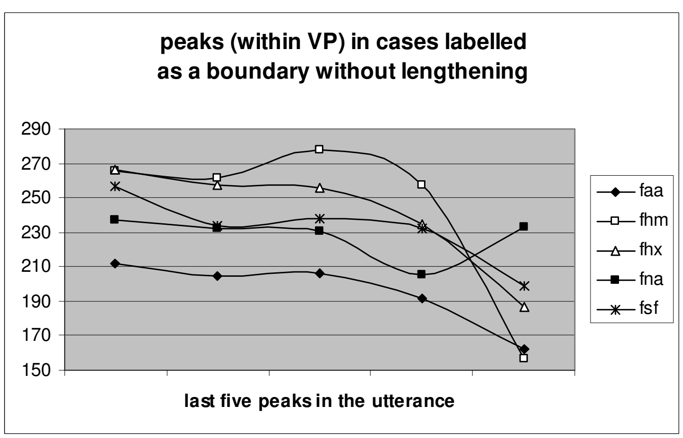

Figure X.15: Average f0 of the last five peaks in the utterance (the VP) by speaker, for cases labelled as a boundary in within-VP position but without an L cue.

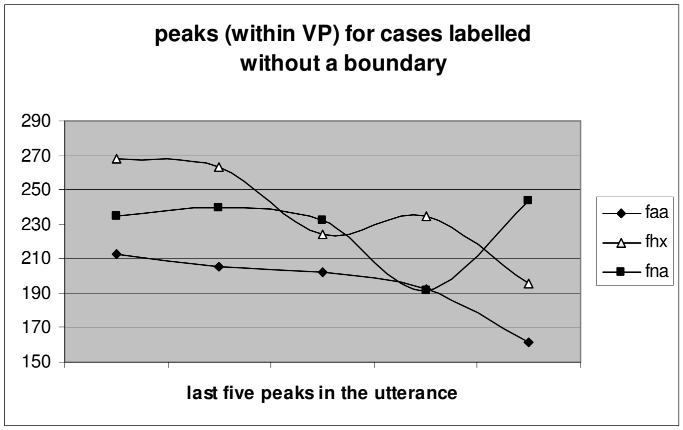

Figure X.16: Average f0 of the last five peaks in the utterance (the VP) by speaker, for cases labelled as having no boundary in within-VP position.

The impressionistic transcription and the cue clusters that arise from it (in Tables 7-8) are thus substantially confirmed by the supporting evidence from durational and f0 measurements. Our final task in this section is to explore the differences between the cues observed in those tables, that is, in after-S and within-VP positions. Recall that speakers vary among themselves in their preferred cue choices, and also in their own choices of cues in different positions. We suggest however that the 'positional' variation in cue choices between these two potential phrase boundaries is in fact principled, in that the cues used in within-VP position are always a subset of those used in after-S position, for any given speaker.

Figure X.17 summarises percentage use of each cue type in after-S and within-VP positions for all speakers and shows that use of each cue varies between the two positions. Pre-boundary lengthening is a purely durational cue, and is used consistently (though not universally) in after S position (93%), and less, though still roughly half of the time, in the subordinate within-VP position (48%). The use of a phrase tone can be seen as a phonological cue, which we might expect to be a required cue to a phrase edge (for example following Beckman & Pierrehumbert 1986, Pierrehumbert & Beckman 1988). In fact in the present corpus it is used only roughly half the time in after-S position (53%), and is somewhat rare in within-VP position (5%). Pauses are known to be a very variable cue to phrasing (Cruttenden 1997) and are used here roughly half the time in after-S position (54%), but are rare in within-VP position (5%).

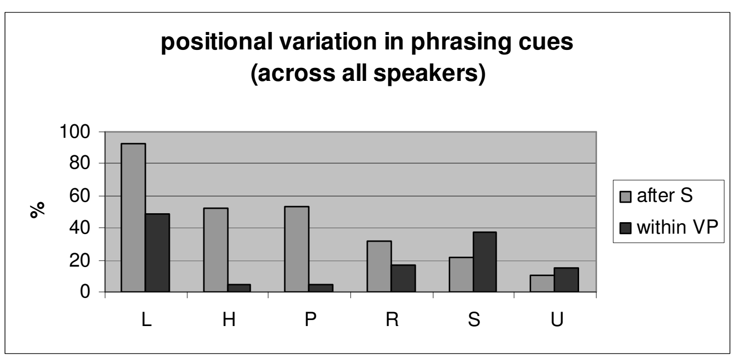

Figure X.17: Percentage use of phrasing cues in different positions (across all speakers)

The remaining cues, partial reset (R), suspension of downstep (S) and upstep (U), are arguably all related to phonetic implementation, and can be seen as the reflexes of the register domains of successive prosodic constituents (cf. Truckenbrodt 2002, 2004, 2007). As with the other cues, the proportion of R use is lower in within-VP position (18%) than in after-S position (34%). In contrast however, use of both S and U increases in within-VP position (37% and 15% respectively), compared to after-S position (21%, 10%). We suggest therefore that increased use of these 'phonetic' cues to register domains is in fact the norm in the subordinate within-VP boundary position, and that use of R was under-reported in the present analysis due to the definition of reset used during the labelling process. A reset was defined here relative to the immediately previous peak, so a reset would not be transcribed in cases where the preceding (phrase-final) peak bears a high phrase tone or is subject to upstep. We expect that transcription using a definition of reset relative to the level of the preceding domain (rather than the preceding peak) would yield a higher proportion of R cases in within-VP position than in after-S position.

Finally, as already mentioned above, we note that the cues used by an individual speaker in after-S position are not a clear predictor of the cues that she will use in within-VP position. The two speakers who mostly use R after-S (fhg and fhx) diverge in VP position: fhg only retains L while fhx switches to use of S (a register domain cue). Similarly, the two after-S H-speakers (fhm and fsf) both dispense with the phonological phrase tone cue: fsf only retains L while fhx switches to use of upstep (U). The after-S mixed-cue speaker faa retains only lengthening in within-VP position. The only speaker to use the same (mixed) set of cues in both positions is fna (note that this speaker is also distinct from all the others in her use of utterance-final upstep, as discussed above, and also has some instances of boundaries inserted within the subject).

In summary then, somewhat different sets of phrasing cues are used by speakers in within-VP and after-S positions, and the within-VP cues are a 'weaker' subset of those observed in after-S position. In the next section we set out arguments for and against choosing to treat these boundaries as junctures at the same level of prosodic phrasing, and explore the theoretical implications of that choice.

5 Discussion

In this section we seek to interpret the inter-speaker and positional variation in cues to BI-3 in EA described in §4 above. We first explore the possibility that some of the boundaries labelled as BI-3 should instead be treated as boundaries at some other level (that is, either IP or MiP, in §5.1, then explore how and why boundaries at the same level might be realised with different cue sets (in §5.2). Section §5.3 revisits the formal analysis of Hellmuth (2004) in the light of the findings of the present study.

5.1 Should 'different cues' imply 'different phrasing level'?

An obvious question that arises is whether the different cues that we observe here are really all cues to the same level of phrasing in the prosodic hierarchy. Perhaps some speakers are breaking the utterance up into constituents at some higher or lower level, such as IP or MiP, and consequently use different cues. We treat this question by examining the degree of reset at boundaries for each speaker (as a diagnostic for IP level boundaries) and by looking at the cues to phrasing observed at boundaries labelled BI-2 (MiP) in the present dataset for comparison with those labelled as BI-3 (MaP).

The most likely candidates for reanalysis as full IP boundaries are those in after-S position. In order to determine whether an after-S boundary is an IP or a MaP level boundary we rely on the observation made in Truckenbrodt (2002, 2004, 2007) for German, which we take to hold cross-linguistically, that boundaries between two MaP level constituents within a single IP will show partial reset only, due to global downstep between the two MaP register domains. In contrast, a full IP boundary is expected to show a full reset to the level of the initial peak in the utterance. We exclude from this investigation three tokens in the dataset which were transcribed with an IP boundary after the subject.[20] Figure X.18 shows mean peak heights by speaker in three positions: on the first and second peaks in the utterance, and on the verb (the peak immediately following the after-S boundary). All speakers show partial reset on the verb, either to the same height as the second peak (faa, fhx, fna) or to a level intermediate between that of the first and second peaks (fhg, fhm, fsf). Crucially however, none of the speakers show reset after S to the initial level of the utterance. We conclude therefore that it is indeed appropriate to treat these boundaries as instances of MaP level boundaries[21], even though they are cued differently by different speakers.

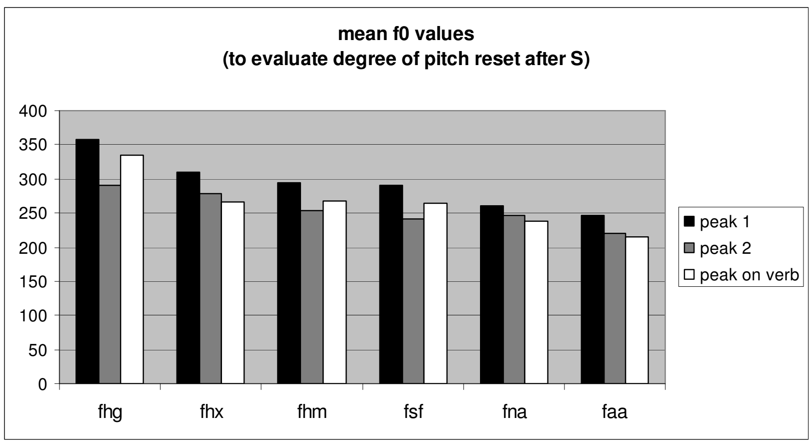

Figure X.18: Mean f0 values on the first and second peak of the utterance and on the verb (by speaker).

A possible reanalysis of the weaker cues in within-VP position would be to suggest that some of the boundaries labelled here as BI-3 are in fact BI-2 MiP level boundaries. To explore whether this is a plausible analysis, we look at the cues observed at boundaries which were labelled BI-2 in the dataset, and where they occurred. In total, just 28 MiP level boundaries were labelled in the dataset. The cues observed at these boundaries for different speakers are detailed in Table 9 below. The most common cue used was a partial reset (relative to the previous peak), which was also the cue to MiP boundaries observed in Hellmuth (2004). There were however a number of more complex cue clusters observed, which resemble those observed at MaP boundaries for the relevant speaker (e.g. 'HL' was the most common cluster for speaker fhm in after S position, and 'LS' was the most common cluster for speaker fhg in within-VP position).

Table 9: Cue clusters observed at boundaries labelled BI-2 (MiP) by speaker

fhg (n=4)	*fhx (n=6)*	*fhm (n=5)*	*fsf (n=7)*	*faa (n=4)*	*fna (n=2)*
LS (1)	L (2)	HL (1)	L (1)	LS (1)	LR (1)
R (1)	R (4)	R (4)	R (6)	R (1)	R (1)
S (2)				S (2)	

All of the 28 boundaries labelled MiP were observed within the subject, and of these 13 were observed in 3-word subject sentences and 15 in 4-word subject sentences. In 3-word subject sentences the MiP boundary was always found between the complement NP and the adjunct PP (that is, before [min(-il) juna:n] 'from Greece'). The 4-word subject sentences allow us to test whether the subject-internal boundaries labelled as BI-2 are MiP or MaP boundaries: we would expect a MaP boundary to occur at the right-edge of a syntactic XP, whereas an MiP constituent, which appears in EA to be purely rhythmic in nature, should enhance eurhythmy by forming binary MiP domains (containing two PWds each).

The syntactic structure of the 4-word subject is nested in such as way that the right edge of the complement XP is after the third word in the subject, not at the mid-point of the four word sequence. We suggest that boundaries observed after the third word are in fact MaP boundaries, and indeed they are the ones that prove to bear the 'stronger' cue clusters: LS for faa and fhg, HL for fhm, L for fhx, LR for fna and L for fsf. All of the MiP boundaries falling at the mid-point of the 4-word subject (after [wila:di-l] 'my children', n=8) are marked by partial reset (R) only, suggesting that this is indeed the best diagnostic for an MiP boundary. This in turn implies that at least some of the boundaries labelled as BI-3 (MaP) in within-VP position which bear only partial reset (R) as a cue, can plausibly be reanalysed as MiP boundaries. There are however only eight such boundaries (for speakers fhx, fhm and fna; see Table 8 above). All of the other boundaries labelled BI-3 at within VP position (n=83) can indeed be considered MaP level boundaries, even though they bear a variety of cues.

We have set out evidence in support of treating the boundaries that are variably cued in EA as instances of the same level of prosodic phrasing. In the next section we explore why the cues are so variable.

5.2 How can 'different cues' imply 'same phrasing level'?

How can it be that boundaries at the same level of phrasing are marked with different cue sets by different speakers? We explore this question by comparing the degree of pitch reset observed in within-VP position and the degree of lengthening in after-S and within-VP positions.

The cues most consistently used *across all speakers* in after-S position are lengthening (L) and tonal register cues (R, S, U). These are reflexes of phonetic implementation, rather than phonological cues (such as a phrase tone), and are also the cues most likely to be retained, or made use of in within-VP position.

For the tonal register cues, we have a principled way of understanding why f0 effects are smaller at the within-VP boundary than at the after-S boundary: following Truckenbrodt (2002, 2004, 2007), the degree of reset/upstep will be relatively less at the start of more embedded domain (such as within-VP position) than at the start of a less embedded domain (such as after-S). Figure X.19 shows mean f0 by speaker in the three phrase-initial positions (initial, on the verb and on the penultimate PWd), for 8/9 peak utterances containing a within-VP MaP boundary (n=72).[22] We can see that there is a clear downstep relationship between phrases and that the degree of downstep among phrases is remarkably even (except for speaker fhg who shows reduced downstep between the first two phrases). For comparison, Figure X.20 shows the downstep relation that holds for each speaker among the first three peaks of the utterance, within the first phrase.[23] According to Truckenbrodt (2002, 2004, 2007) we should expect the degree of downstep among phrases to resemble that observed between the first two peaks in the first phrase. The degree of downstep among phrases (in Figure X.19) closely resembles that between the first two peaks (in Figure X.20) for speakers fhx and faa, but the norm in this dataset is for the degree of downstep to be somewhat greater within phrases than between them.

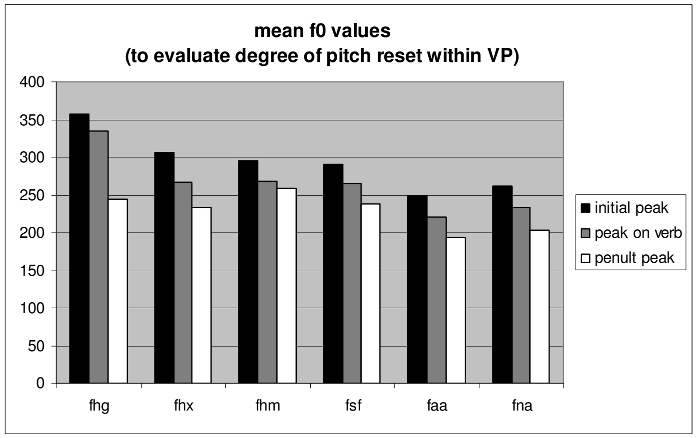

Figure X.19: Mean f0 values on phrase-initial peaks, by speaker
(for 8 or 9 peak utterances containing a within-VP boundary, n=72).

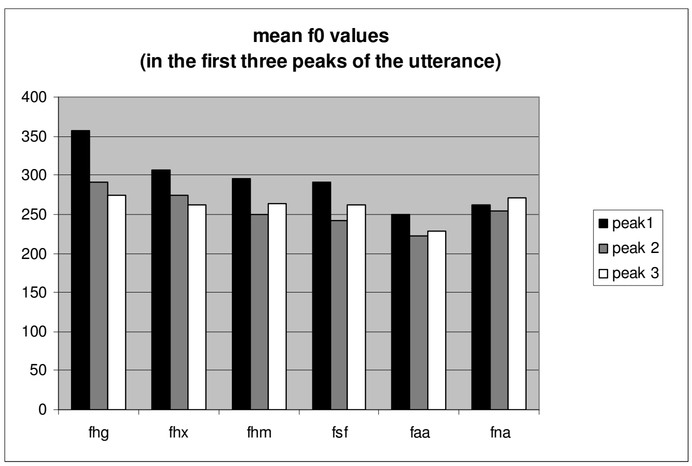

Figure X.20: Mean f0 values on the first three peaks in the utterance, by speaker
(for 8 or 9 peak utterances containing a within-VP boundary, n=72).

The phonetic implementation theory of Truckenbrodt (2002, 2004, 2007) allows us to accommodate variant f0 cues at boundaries of the same type, due to the notion of embedding of register domains. Is there any evidence that the degree of lengthening observed at MaP boundaries in the two positions is similarly reduced? Figure X.21 shows mean values of normalised rhyme duration for target words i) at the right edge of the subject (nr1dur) iff a boundary was labelled L in that position (n=100) and ii) at the right edge of the complement NP of the verb (nr2dur) iff a boundary was labelled L in that position (n=54). The difference in mean duration in the two positions is statistically significant (ANOVA: F= 7.097, df = 152, p=0.009).

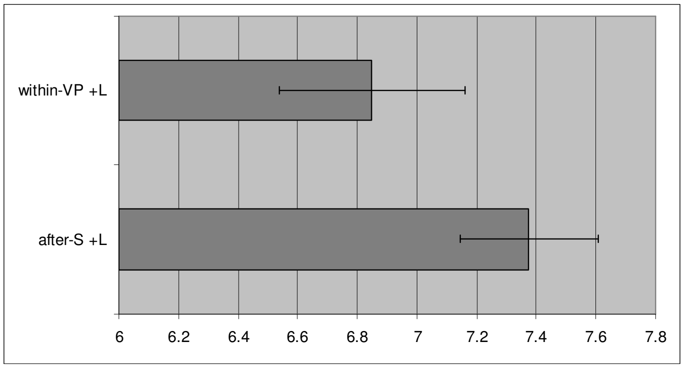

Figure X.21: Mean values of normalised rhyme duration for target words labelled L in their respective position, after-S or within-VP, across all speakers.

It could be argued that a statistically significant difference in the degree of lengthening in the two positions is evidence for a different level of phrasing (cf. Chahal 2001, 2003), but we suggest that to do so would require treatment of varying degrees of f0 excursion in tonal cues to phrasing as evidence of different levels also (contra Truckenbrodt 2002, 2004, 2007). We therefore analyse this durational difference not as evidence for different levels of phrasing but as evidence that our adjacent MaP domains are embedded within a single IP. Since pre-boundary lengthening is another reflex of phonetic implementation it is to be expected that we find a degree of durational difference similar to that seen in f0 values between adjacent register domains (cf. Ladd 2008:299).

5.3 Re-analysis

The formal analysis proposed in Hellmuth (2004) was based on double-embedding of MiPs within MaP in order to arrive at a requirement that MaPs contain minimally 4 PWds. The empirical phrasing generalisations in the present dataset appear to be somewhat less restrictive, since a large number of MaPs containing less than 4 PWds are observed.

Why might the facts be different here? One reason might be that the speech dataset analysed in Hellmuth (2004) comprised just two speakers who both used a somewhat fast speech rate, whereas the present dataset contains a greater mix of fast/slow speakers. Furthermore, Hellmuth (2004) employed a prosodic transcription protocol in which only boundaries bearing two or more cues to phrasing were labelled as boundaries, on the a priori assumption that fixed cues to phrasing would be identified. The results of the present study suggest that even reduced tonal cues to phrasing may still indicate the presence of a prosodic boundary and that the most reliable cues to phrasing are the relative heights of successive peaks in the utterance, which provide information about where downstep register domains begin and end.

Given these new empirical insights, the implications for the analysis proposed in Hellmuth (2004) are twofold. We reconsider the status of MiP and sketch how the analysis could be reformulated in a phase-based approach.

5.3.1 Re-visiting MiP in EA

There were two reasons for proposing an MiP level of phrasing in Hellmuth (2004): i) as a theoretical means to generate a requirement for a 4 PWd MaP, by requiring doubly-embedded binary domains, and ii) to account for instances of apparent ‘rhythmic boost’ occurring in some adjacent PWd pairs (cf. the consistent ‘R’ cue to a MiP boundary observed here). The present study appears to eliminate the first of these two reasons, but provides further confirmation that the second reason still holds. How are we to analyse these facts?

One could maintain the analysis of Hellmuth (2004) with the caveat that BINMIP is a feature of fast speech only. Alternatively, one could dispense with BINMIP and instead analyse the phrasing generalisations of EA with BINMAP only (formulated as a requirement that each MaP contain at least two PWds), but lose much of the empirical coverage, including the key fact that short SVO sentences are routinely phrased into a single MaP, such as the 3-word subject sentence reproduced from Hellmuth (2004) in Figure X.1 above.

A more promising approach would be to also revisit the status of MiP. We propose that EA has only a single level of phrasing between the IP and the PWd, namely the MaP, and that the rhythmic boost effect observed both here and in Hellmuth (2004) arises due to compounding of PWds, in violation of the Strict Layer Hypothesis (cf. Ladd 2008: 297ff., Itô & Mester this volume). If we propose, plausibly, that compounding is a feature of fast and/or rhythmic speech, then faster speech will yield somewhat longer MaPs on the assumption that in EA BINMAP (formulated to count PWds) pays attention to the maximal PWd (Itô & Mester this volume).[24]

5.3.2 Re-visiting the role of XP in the syntax-phonology mapping in EA?

If we accept the foregoing arguments that cues to phrasing are variable in EA, and that the boundaries observed within-VP are of the same order as those observed after-S, then we have shown that EA does display sensitivity to XP constituency, given sufficient prosodic material.

Could the facts of EA nonetheless be amenable to re-analysis in a multiple-spell-out approach? One option would be to adopt the Highest Phrase Condition (HPC, Kratzer & Selkirk 2007), which predicts a within-VP boundary as well as an after-S boundary in our target sentences, as shown in (2) below: the HPC causes the highest phrase in VP (the spellout domain of vP), to be mapped to a MaP. The highest phrase in VP is the head + complement, shown in bold. This produces an MaP boundary at the right edge of the complement within the VP just as we observe in many cases in the present dataset.

(2) [[[head [complement]$_{AP}$]$_{NP}$ [adjunct]$_{PP}$]$_{DP}$ [*v* [verb **[[head [complement]$_{AP}$]$_{NP}$** [adjunct]$_{PP}$]]$_{VP}$]$_{vP}$]$_{CP}$

spell out domain of vP phase: <-->

Highest Phrase in spellout domain: <------------------------------------->

In order to account for the variability in our dataset (the fact that not all speakers realise a within-VP boundary) however, we must however retain a role for BINMAP (in its revised form, that is, counting PWds). To maintain the HPC analysis, we would have to propose, contra Kratzer & Selkirk (2007), that the final adjunct PP is phrased as an MaP under an exhaustive parse. This is because, in cases where the speaker fails to realise a within-VP boundary, we infer that the adjunct PP (formed of two PWds) has formed a compound PP and that it is BINMAP that prevents the adjunct from being phrased into a separate MaP. Unfortunately however, the proposal that the parsing of syntactic material into prosodic structure is exhaustive would prevent the HPC from accounting for the original facts of German for which it was proposed (Kratzer & Selkirk 2007), since the analysis for German depends on an inexhaustive parse. An alternative would be to adopt a prominence-driven phase-based account (such as Stress-XP, Truckenbrodt this volume), but this is not readily adapted to analysis of a language in which accentual prominences are distributed at the PWd level, rather than the MaP level.

The only remaining option for explaining the many within-VP boundaries observed in the current dataset in a phase-based approach would be to argue for a different syntactic analysis of the target utterances. We have assumed thus far that the utterance-final adjunct PP is VP-internal as in (3); if however it were analysed as a vP-adjunct, as in (4), spellout of VP as the complement to vP would yield a prosodic break in the position that we have here termed within-VP position.

(3) [[the- [manager [the-new]$_{AP}$]$_{NP}$ [from Greece]$_{PP}$]$_{DP}$
[[teaches$_{3ms}$ [the-[pedagogy [the-modern]$_{AP}$]$_{NP}$]$_{DP}$ [in [faculty-the-education]$_{CS}$]$_{PP}$]$_{VP}$]$_{vP}$]$_{TP}$

(4) [[the- [manager [the-new]$_{AP}$]$_{NP}$ [from Greece]$_{PP}$]$_{DP}$
[teaches$_{3ms}$ [[the- [pedagogy [the-modern]$_{AP}$]$_{NP}$]$_{DP}$]$_{VP}$ [in [faculty-the-education]$_{CS}$]$_{PP}$]$_{vP}$]$_{TP}$

We note therefore that what is needed to fully rule out a role for XPs in the syntax-phonology mapping in EA is evidence that a sequence of unambiguously VP-internal XPs (e.g. in a ditransitive) also trigger within-VP boundaries of the sort observed here, and we leave this question to future research.

6 Conclusion

This paper has explored in detail the cues to phrasing observed in a corpus of read speech sentences, using quantitative measurements of f0 and rhyme durations to support the results of a fine-grained qualitative transcription. It appears that there is no required phonological cue to the edges of MaP boundaries in EA, though phrase tones are sometimes observed. The most common and consistent cues to MaP edges are phonetic (cf. Kawahara & Shinya 2008): pre-boundary lengthening and peak height manipulation reflecting downstep register domains, with speakers free to interpret these in a number of ways. This is parallel to Truckenbrodt (2002, 2004, 2007)'s findings for Southern German in which speakers also varied in how phonological tones (of pitch accents or phrase tones) were associated to prosodic domains (and thus varied in their phonetic implementation). It is known that cues to phrasing also vary in e.g. English, hence the adoption of BI-2 for indeterminate cases in the MAE-ToBI labelling system (Beckman et al. 2005, Brugos et al. 2008).

A question that arises is what governs the distribution of phrase tones (such as H- and L-) in EA, if they are not obligatory MaP edge markers. An obvious answer is that their distribution is likely to reflect pragmatic meaning and/or information structure. One hypothesis would be that low phrase tones (L-) act as a focus marker and that high phrase tones (H-) act as a marker of continuation, and thus of topics, as has been suggested by El Zarka (2008). The evidence presented here indicates that if what is inserted is indeed a phrase tone, which forces a prosodic boundary, then this would obligatorily condition boundary cues in the form of phonetic implementation of downstep register reset, and optionally also other cues such as pause or lengthening. The status of these tones is testable in that if a boundary is indeed inserted it should trigger a register domain reset of the type observed here; if no reset is found then the inserted tone might be better analysed as a free-floating tone which functions as a focus/topic marker. We leave investigation of this possibility to future research.

As for the syntax-phonology mapping in EA, we have suggested that the analysis of Hellmuth (2004) can be amended by dispensing with the MiP level and instead allowing for compounding of PWds in fast or rhythmic speech. Under this analysis EA has only one level of phrasing between the IP and the PWd, the MaP, which is consistent with the claim that there are universal restrictions on the set of prosodic constituents (Itô & Mester, this volume; cf. Selkirk to appear).

The edge-based analysis of Hellmuth (2004) could be re-cast in a phase-based approach by appeal to the HPC, but requires exhaustive parsing of the utterance into prosodic structure (contra Kratzer & Selkirk 2007) if it is to account for inter-speaker variability in the occurrence of within-VP boundaries (a final MaP must be present for BINMAP to prevent its formation if composed of too few PWds). To work for EA then, the HPC requires a revision which means it no longer captures the data for which it was originally proposed (the distribution of stressless XPs in German, cf. Truckenbrodt this volume). Since the distribution of pitch accents is PWd level in EA, rather than MaP level, phase-based approaches that govern the distribution of MaP prominences (rather than MaP domains) such as StressXP (Truckenbrodt this volume) won't work for EA either. An alternative syntactic analysis of the target sentences, with the final PP as a vP-adjunct rather than a VP-adjunct, could be consistent with a spellout-domain approach though, suggesting that further investigation of e.g. ditransitives will be fruitful.

Nonetheless, it turns out that EA, which appeared to be a language which did not mark XP edges, does display sensitivity to XP constituents, *iff* sufficient prosodic material is available (that is, in long utterances such as those analysed here). So, even if a spellout-domain approach can be shown to explain the occurrence of within-VP boundaries in long and complex sentences, the fact that MaPs in EA are nonetheless subject to prosodic minimality constraints, calls for an intervening prosodic representation between the syntax and phonology, rather than a purely syntactic solution, as Selkirk has long argued.

References

Adger, David. 2006. Stress and phasal syntax. (Ms.). Queen Mary, University of London.

Ackema, P., and A. Neeleman. 2003. Context-Sensitive Spell-Out. *Natural Language and Linguistic Theory* 21:681–735.

Aquil, Rajaa M. 2006. *The segmentation/parsing unit in Cairene Arabic.* PhD, Georgetown University.

Bassiouney, Reem. 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.

Beckman, Mary & G. A. Elam. 1993. Guidelines for TOBI Labelling (version 3.0 1997). The Ohio State University Research Foundation.

Beckman, Mary, Julia Hirschberg, & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, Sun-Ah (ed.) *Prosodic Typology: The phonology of intonation and phrasing.* 9-54. Oxford: Oxford University Press.

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

Benmamoun, Elabbas. 2000. *The feature structure of functional categories.* Oxford: OUP.

Boersma, Paul & David Weenink. 2009. Praat: doing phonetics by computer (Version 5.1.19) [http://www.praat.org].

Borer, Hagit. 1996. The construct in review. In Lecarme, J Lowenstamm, & Ur Shlonsky (eds.) *Studies in Afroasiatic Grammar.* 30-61. The Hague: HAG.

Brugos, Alejna, Nanette Veilleux, Mara Breen, & Stefanie Shattuck-Hufnagel. 2008. The alternatives (Alt) tier for ToBI: advantages of capturing prosodic ambiguity. *Speech Prosody 2008 (Campinas, Brazil).* 273-276.

Chahal, Dana. 2001. *Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English.* PhD, University of Melbourne.

Chahal, Dana. 2003. Phonetic cues to prominence in Arabic. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2067-2070.

Chahal, Dana & Sam Hellmuth. to appear. Comparing the Intonational Phonology of Lebanese and Egyptian Arabic. *Prosodic Typology. Volume 2.* Oxford: Oxford University Press.

Chomsky, Noam. 2001. On derivation by phase. In Kenstowicz, M. (ed.) *Ken Hale: a life in language.* 1-52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2005. On phases. Cambridge, MA.: MIT.

Cruttenden, Alan. 1997. *Intonation*. Cambridge: CUP.

Edwards, Malcolm. 2009. Word order in Egyptian Arabic: form and function. In Owens, Jonathan & Alaa Elgibali (eds.) *Information Structure in Spoken Arabic* Oxford: Routledge.

El Zarka, Dina. 2008. Leading, linking and closing tones and tunes in Egyptian Arabic - what a simple intonation system tells us about the nature of intonation. (Paper presented at the annual meeting of the Arabic Linguistics Society, Michigan, March 2008.).

El Zarka, Dina & Sam Hellmuth. 2009. Variation in the intonation of Egyptian Formal and Colloquial Arabic. *Langues et Linguistique.* 22 73-92.

Ferguson, C. A. 1957. Two problems in Arabic phonology. *WORD: Journal of the International Linguistic Association.* 13(3). 461-479.

Frota, Sonia, Mariapaola D'Imperio, Gorka Elordieta, Pilar Prieto, & Marina Vigario. 2007. The phonetics and phonology of intonational phrasing in Romance. *Amsterdam Studies in the Theory and History of Linguistic Science IV.* 282 131-154.

Haeri, Nilofaar. 1996. *The sociolinguistic market of Cairo: gender, class and education.* London: Keegan Paul International.

Hellmuth, Sam. 2004. Prosodic weight and phonological phrasing in Cairene Arabic. *Proceedings of Annual Meeting of Chicago Linguistic Society.* 40.

Hellmuth, Sam. 2006. *Intonational pitch accent distribution in Egyptian Arabic.* PhD, SOAS.

Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. *The Linguistic Review.* 24(2). 289-314.

Inkelas, S. & D. Zec. 1990. *The phonology-syntax connection.* Chicago: University of Chicago Press.

Ishihara, Shinichiro. 2007. Major phrase, focus intonation and multiple spellout. *The Linguistic Review.* 24(2). 137-167.

Itô, Junko & Ralf-Armin Mester. 2010. Trimming the prosodic hierarchy. In Borowsky, Toni, Shigeto Kawahara, Tahahito Shinya, & Mariko Sugahara (eds.) *Prosody matters: Essays in honor of Elisabeth Selkirk.* London: Equinox.

Jun, Sun-Ah. 2005. Prosodic typology. In Jun, Sun-Ah (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing.* 430-458. Oxford: OUP.

Kawahara, Shigeto. 2010. The intonation of nominal parentheticals in Japanese. In Borowsky, Toni, Shigeto Kawahara, Takahito Shinya, & Mariko Sugahara (eds.) *Prosody matters: essays in honor of Lisa Selkirk.* London: Equinox.

Kawahara, Shigeto & Tahahito Shinya. 2008. The intonation of gapping and coordination in Japanese: evidence for Intonational Phrase and Utterance. *Phonetica.* 65 62-105.

Kratzer, Angelika & Elisabeth O. Selkirk. 2007. Default phrase stress, prosodic phrasing and the spellout edge: The case of verbs. *The Linguistic Review.* 24(2). 93-135.
Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.
Liberman, M. & Janet Pierrehumbert. 1984. Intonational variance under changes in pitch range and length. In Aronoff, M. & R. T. Oehrle (eds.) *Language Sound Structure: Studies in Phonology present to Morris Hallé.* 157-233. Cambridge, M.A.: MIT Press.
Nespor, M & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
Rastegar-El Zarka, Dina. 1997. *Prosodische Phonologie des Arabischen.* PhD, Karl-Franzens-Universität Graz.
Rifaat, Khalid. 1991. *The intonation of Arabic: an experimental study.* PhD, University of Alexandria.
Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook.* 3 371-405.
Selkirk, Elisabeth O. 1995. Sentence prosody: intonation, stress and phrasing. In Goldsmith, J (ed.) *The Handbook of Phonological Theory.* 550-569. Cambridge, Mass.: Blackwell.
Selkirk, Elisabeth O. 2000. The interaction of constraints on prosodic phrasing. In Horne, Merle (ed.) *Prosody: theory and experiment.* 231-262. Dordrecht: Kluwer.
Selkirk, Elisabeth O. 2009. On Clause and Intonational Phrase in Japanese: The Syntactic Grounding of Prosodic Constituent Structure. *Gengo Kenkyu: Journal of the Linguistic Society of Japan.* 136 35-73.
Shaked, Amit. 2007. Competing syntactic and prosodic constraints in Hebrew prosodic phrasing. *The Linguistic Review.* 24(2). 169-199.
Selkirk, Elisabeth O. to appear. The Syntax-Phonology Interface. In Goldsmith, John, Jason Riggle, & Alan Yu (eds.) *The Handbook of Phonological Theory*. 2nd edition.
Siemund, R, B Heuft, K Choukri, O Emam, E Maragoudakis, H Tropf, O Gedge, S Shammass, A Moreno, A Nogueiras Rodriguez, I Zitouni & D Iskra. 2002. OrienTel - Arabic speech resources for the IT market. LREC 2002 (Arabic workshop).
Truckenbrodt, Hubert. 1995. *Phonological phrases: their relation to syntax, focus and prominence*. PhD, MIT.
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry.* 30(2). 219-255.
Truckenbrodt, H. 2002. Upstep and embedded register levels. *Phonology.* 19 77-120.
Truckenbrodt, H. 2004. Final lowering in non-final position. *Journal of Phonetics.* 32(3).313-348.
Truckenbrodt, Hubert. 2007. Upstep on edge tones and on nuclear accents. In Riad, Tomas & Carlos Gussenhoven (eds.) *Tones and Tunes: Experimental studies in word and sentence prosody. Volume 2: Phonology and Phonetics.* 349-386. Berlin: Mouton de Gruyter.
Truckenbrodt, Hubert. 2010. Stress and VP-shells in German. In Borowsky, Toni, Shigeto Kawahara, & Mariko Sugahara (eds.) *Prosody matters: essays in honor of Lisa Selkirk* London: Equinox.
Watson, Janet C. E. 2002. *The phonology and morphology of Arabic*. Oxford: OUP.
Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America.* 91(3). 1707-1717.

[1] The author thanks the Egyptian participants for their time, and Laura Downing, Shigeto Kawahara, Bernadette Plunkett, Hubert Truckenbrodt and the audience of the ZAS Syntax-Phonology Colloquium in Berlin for comments.
[2] Most early descriptions of intonation and prosodic phrasing in EA (Rifaat 1991, Rastegar-El Zarka 1997) treat the Egyptian pronunciation of the formal register of the language, Modern Standard Arabic (MSA), which exists alongside the spoken variety in the well-known diglossia of the Arab world (Ferguson 1957). The two varieties are however similar in many intonational properties (see El Zarka & Hellmuth 2009 for discussion).
[3] Some 3-word subjects were phrased as separate MaPs when speakers were asked to slow their speech rate, as in the realisation illustrated in Figure X.2.
[4] The stimuli were constructed so that a CCC cluster straddled all potential XP edge boundary positions. See for example the subject-final adjective [muhimm] 'important' followed by the verb [bijxumm] 'he cheats' in (1) above.
[5] This subset is part of a larger database of 44 SVO and VOO (double object constructions) sentences recorded with 12 speakers of EA (6 male, 6 female).
[6] A Construct State (CS) phrase is usually analysed as a single morphosyntactic word (Borer 1996) but functions prosodically as two PWds in EA (Hellmuth 2006). We analyse it here as a 'CS' sequence, without further discussion of its internal syntactic structure, for ease of exposition, but assume that it does constitute a lexical XP projection.
[7] A small number of some mid-utterance complex boundary configurations were observed which required analysis using a phrase tone + boundary tone sequence (e.g. H-L%).
[8] The relative height of peaks was judged by eye during labelling, for later corroboration with quantitative results.
[9] Note that reset was defined relative to the height of the preceding peak, rather than with reference to domain register levels.
[10] Cues observed at junctures labeled as BI-2 (MiP) are discussed in §4 below.
[11] Levene's test shows that the variances among speakers are not equal (p=0.001), so Tamhane's test was used to determine for each speaker pair whether the two speakers fall into the same or a different subset.
[12] In this and all other figures, errors bars denote 95% Confidence Intervals.
[13] Overall the ANOVA shows a significant difference among speakers (F=4.074, df=107, p=0.002) but Tamhane's test shows that the only significant differences are between fhx and fhm (p=0.13) and between fhx and fsf (p=0.25); Tamhane's test was used because Levene's statistic shows that the variances are not equal (p<0.001).
[14] This can be observed in the low f0 values seen on the final peak in Figures X.6-11 below.
[15] The final adjunct PP of the utterances is a construct state. Speaker fna produced these with a qualitatively different pitch accent on the first word than that observed on other words; the pitch rises throughout the first word to a peak late in the stressed syllable of the second word. One analysis would be that she is treating the CS nominal as a single PWd, with a single rising pitch accent realised across the both words. Further investigation of this atypical realisation is beyond the scope of this paper and is left for future research.
[16] All phrase tones were labelled 'H' during transcription, whether low or high. The only speaker who used low phrase tones was fhx, and this is shown here by presenting the phrase tone in lower case.
[17] This is because it is possible in EA to assign PWd status to a function word if it is bimoraic. In the current dataset the function word most often realised with an accent was [min] 'from', with encliticisation of the following definite article [il] 'the', yielding bimoraic [min-il] 'from the'. The planned stimuli sought to avoid this; the printed text was [min junaa:n] 'from Greece'; but many speakers inserted the definite article yielding [min-il-junaa:n] 'from (the-) Greece'.
[18] One further token was labeled with an after-S boundary but without L, but the whole utterance was phrased atypically (with a full IP boundary within the subject) and is not discussed further here.
[19] By Levene's statistic the variances are not equal (p=0.002), so Tamhane's test was used for post-hoc comparisons; the difference is significant for 'no B' vs. 'B with L' (p<0.001) and 'B no L' vs. 'B with L' vs. (p<0.001) only.
[20] Two from speaker fhg and one from speaker fsf.
[21] A check on individual tokens reveals that the differential between the peak on the verb and the first peak is positive (i.e. the verb peak is realised at a higher f0 level than the initial peak) in just three cases (p022fhg1, p022fhg3, p08fna1). These cases should probably therefore be reclassified as IP level boundaries.
[22] Within-VP boundaries which were labelled with R only, and which should probably be reanalysed as MiP boundaries, are excluded.
[23] The third peak in the utterance is somewhat raised relative to the second for the four non-'R' speakers (fhm/fsf/faa/fna), reflecting their use of a high phrase tone and/or upstep on the third peak in 8 peak utterances.
[24] This proposal would require BI-2 to be redefined for EA, either along the lines of Shaked (2007) for Hebrew, or by leaving BI-2 free for use in indeterminate cases as in MAE-ToBI (Beckman & Elam 1993).