Title: Towards a prototype intonational transcription system for Egyptian Arabic: testing the local f0 contour properties of intonational pitch accents in spontaneous speech.

Author: Dr Sam Hellmuth

Affiliation: University of York

Address: Department of Language & Linguistic Science

University of York, Heslington, York, UK YO10 5DD

Tel: +44 1904 432657

Fax: +44 1904 432673

Email: sh581@york.ac.uk

Abstract:

This paper sets out the properties of a prototype notational system for the transcription of Egyptian Arabic (EA) intonation, and tests the system by comparing the results of transcription of a small corpus of spontaneous conversational speech with known facts about EA intonation from experimental studies on 'laboratory' (read) speech. In particular the transcription system is used here to test the claim that a single pitch accent type is observed in EA. Specifically, a local phonetic contour annotation tier, adapted from that used in the IViE notation system for English, is used to establish the shape of the local f0 contour in the vicinity of the stressed syllable of a subset of target words identified in a small corpus of spontaneous speech. The results indicate that all of the variation in the local f0 contour can be explained from properties of the local prosodic context (target syllable type, prosodic context) and thus support the claim that a single accent type is sufficient for the description of spontaneous speech. The paper discusses potential future development and applications of the proposed transcription system for EA intonation.

Keywords: intonation, transcription, Egyptian Arabic, alignment, pitch accents

Running head: Towards a prototype intonational transcription system for Egyptian Arabic.

## I Introduction

This paper sets out the properties of a first prototype notational system for the transcription of Egyptian Arabic (EA) intonation, and tests the system by comparing the results of transcription of a small corpus of spontaneous conversational speech with known facts about EA intonation from experimental studies on 'laboratory' (read) speech. In particular the transcription system is used here to test the claim that a single pitch accent type is observed in EA. This is achieved by using a 'local phonetic contour' tier, adapted from that used in the IViE [1,2] and IVTS [3] notation systems, proposed for English and French respectively. The paper explores one way in which this annotation tier might be implemented for useful transcription of the intonation of spontaneous EA speech. The background to the study is set out in section II; section III describes the properties of the proposed transcription system; section IV describes the results of the transcription and compares them to known facts from experimental studies; the paper concludes with a discussion in section V.

## II Background to the study

Egyptian Arabic (EA) is defined here as the colloquial dialect of Arabic spoken in Cairo, Egypt, and by educated middle class Egyptians throughout Egypt. The segmental and metrical phonology of EA is well-described, and has been the subject of much research (see [4] for a summary). In contrast, EA intonation has received comparatively less attention, and this situation is paralleled across most spoken dialects of Arabic. Recent studies have however established some key properties of EA intonation, including the typical shape of global contours in different utterance types[5], the lack of complete deaccentuation after a focus[6,7], the alignment patterns of pitch peaks occurring on pre-nuclear (non-phrase-final) accented words[8] and the frequent distribution of these pitch peaks, such that one occurs on every content word [9]. This array of facts have led the author to propose (in other work[10,11]) that

only a single pitch accent category is necessary for the phonological description of EA pitch accents (L+H*). Other authors have reached a similar conclusion for non-phrase-final accents in the Egyptian pronunciation of Modern Standard Arabic[12,13] (although they assign a different phonological label to the single accent in their studies). In general, these claims are based on analysis of either scripted or broadcast speech, which might be expected to be uniform in character. The aim of the small pilot study reported here was to test whether a single accent category is sufficient for the description of non-phrase-final accents as they are realised in fully spontaneous colloquial EA speech. It is however difficult to obtain quantitative generalisations from spontaneous speech due to the inevitable variation in segmental and prosodic contexts of potential target accented words (though not impossible[14]). In order to solve this problem the present paper proposes a prototype notation system for fine-grained prosodic transcription of EA intonation. Specifically, a local phonetic contour annotation tier is used to establish the shape of the local f0 contour in and around the stressed syllables of a subset of target words identified in a small corpus of spontaneous speech. The properties of the local f0 contour in these words is compared to known facts about EA intonation established experimentally, and the range of variation in the observed shape of pitch accents is discussed.

## III Methodology

The Intonational Variation in English (IViE) labelling system was designed to facilitate the creation of a corpus of "directly comparable transcriptions of several varieties of English in a single labelling system"[15: p1]. It is similar to the Tones and Break Indices (ToBI [16]) labelling system in that phonological pitch targets ('tones') are labelled on a tier separate from other aspects of the transcription. ToBI comprises: a tone tier, an orthographic tier, a break index tier and a miscellaneous tier. In ToBI, "the tone and break index tiers represent the core

prosodic analysis" [16: p8]. The innovation in the IViE labelling system was the addition of a tier for the labelling of 'acoustic-phonetic structure'; this tier comprises a labelling of the "shape and alignment of f0 patterns relative to the location of (accented) strong syllables" [17: p2]. With the advent of technology such that labelled transcriptions are almost invariably presented alongside a spectrogram and f0 pitch trace of the speech fragment in question, one could argue, as Wightman[18] has done, that the local f0 contour does not need to be labelled, since users of the transcription can themselves see and interpret the f0 contour. However, the local f0 contour tier in IViE can play an important role whilst a fully fledged phonological analysis of the language or varieties concerned is being developed. It is this strength of the IViE approach that is exploited here, in order to test the (phonological) claim that a single accent type is sufficient for the description of EA accents in prenuclear position, in spontaneous as well as scripted speech. It is thus the necessary properties of a local f0 contour tier for EA that are explored in the present paper; future work will explore the required properties of other tiers for accurate transcription of EA intonation.

The dataset transcribed for the present study is a set of 100 target words from a spontaneous speech telephone conversation (between two female speakers, from the Call Home corpus[19]). The relevant portions of the conversation were transcribed by the author with reference to the f0 contour and spectrogram using Praat version 4.6.10[20], on three tiers: i) *words*, broad phonetic transcription, on separate tiers for each speaker, ii) *tones* (phonological labels: pitch accents and boundary tones), and iii) *local f0 contour* (phonetic-acoustic structure tier). The criterion for selection of target words was the presence of little or no perturbation of the f0 contour during the word. The position of the pitch peak (or valley) within each target word was identified automatically using the pitch maxima (minima) function within Praat, and a dummy "H*" or "L*" label assigned to the peak on the tones tier.

The shape of the f0 contour immediately before and after the pitch peak/valley was then annotated on the local f0 tier, as follows. Firstly, a capital letter (*L*, *M* or *H*) was assigned to denote the height of the f0 contour during the accented syllable, according to whether the height of the pitch peak/valley was low, mid or high in the speaker's pitch range. Next, the relative height of the f0 contour on unaccented syllables occurring immediately before and after the peak/valley was described: adjacent pitch low in the speaker's pitch range was transcribed '*l*' , adjacent pitch in middle of the speaker's pitch range was transcribed '*m*' and adjacent pitch high in the pitch range (and thus in most cases level with the pitch on the accented syllable itself) was transcribed '*h*'. The speakers' pitch range throughout the whole conversation had previously been calculated for each speaker independently (speaker A: 120-520Hz; speaker B: 80-520Hz). Finally, the position of the pitch peak/valley within the accent syllable was checked; if the peak/valley occurred very early or very late in the accented syllable this was noted, by insertion of a line 'l' immediately before the capital letter (for an early peak/valley) or immediately after the capital letter (for a late peak/valley). This additional labelling convention is proposed here and is not part of the original IV notation systems. The set of 100 labels annotated on the local f0 contour was then harvested, along with relevant information about the syllable structure, stress position and prosodic context of each target word. A sample annotation grid is provided in Figure 1.

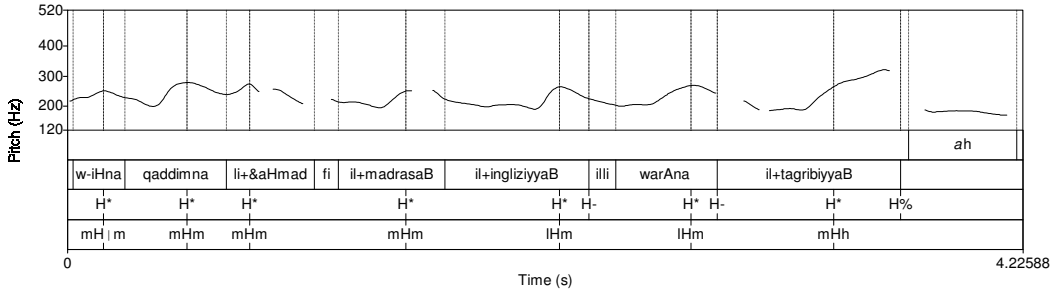


Figure 1: Sample of annotated text (*tiers top to bottom: words B, words A, tones, local f0*).

| w-iHna | qaddimna | li+&aHmad | fi | il+madrasaB | il+ingliziyyaB | illi | warAna | il+tagribiyyaB |
|---|---|---|---|---|---|---|---|---|
| ˈwiħna | ʔadˈdimna | ˈlaħmad | fi | l-madˈrasa | l-ingiliˈzijja | ˈilli | waˈraːna | t-tagriˈbijja |
| and we | applied | for-Ahmed | in | the-school | the-English | REL | behind-us | the-near |

## IV Results

All of the accents were transcribed with an 'H' on the accented syllable (and none with L), that is, no pitch valleys were found, only pitch peaks (note that the conversation does include declarative questions). This finding suggests that no 'L*' type phonological labels are required for the description of pitch accents in EA. The next two sections explore the results of the transcription with respect to the alignment of H peaks within the accented syllable (section IV.1) and the shape of the preceding and following f0 contour (section IV.2).

### IV.1 Position of the H peak within the accented syllable

In the great majority of cases (80%) the peak was observed to occur within the accented syllable. Consistent alignment of the high peak within the accented syllable is a good reason to propose that the H tone is the phonologically strong tone (the 'starred tone') of the EA pitch accent. The exact position of the H peak within the accented syllable was found to co-vary with the position of the stressed syllable within the word and with the position of the word within the phrase, factors known to affect peak alignment, particularly in pre-nuclear rising pitch accents [21,22]: the nearer the accented syllable is to a prosodic boundary (at the word or phrase level) the earlier the peak will be aligned. For example, there were just two cases where the pitch peak was observed to be very early in the accented syllable (denoted by 'lH'); both were words occurring at the end of an intonational phrase and with stress on the final syllable: [inʃaʔalˈlaah], [tuˈreːl]. In addition, peak alignment was observed to vary with the type of syllable bearing stress in the word (CV, CVV or CVC) [cf. 8]. In phrase-initial words the H peak of the accent was observed to be at the very end of the stressed syllable in all cases ('Hl'), regardless of the position of stress in the word, however in phrase-medial words, the H peak was at the end of the syllable only in words bearing initial-stress, e.g.

[ˈɣaːja]. In sum, there is no evidence in the present (small) corpus for variation in the position of the H peak within the accented syllable other than that caused by local prosodic factors.

IV.2 Shape of the local f0 contour before and after the H peak

Turning to the local pitch contour immediately before and after the H peak, the most common transcription used in this pilot transcription set was '*mHm*', observed on 55% of target words. Overall, the level of preceding pitch was very consistently observed to be '*m*' (on 82% of target words). There are three types of context in which preceding pitch was labelled 'l', indicating that the f0 contour rose to the H peak from a somewhat lower level in the speakers pitch range. One such context was on words which were the first accented word in the utterance, preceded only by an unaccented function word (e.g. [wi hijja…] 'and she…'), and in these cases the rise in pitch starts already at the onset of the utterance (suggesting a possible %L initial boundary tone). Preceding local pitch also seems to be somewhat lower when there are relatively large number of unstressed syllables between successive accents. This can result in a short 'low plateau' between two accents [cf. 23], and was observed in polysyllabic words such as [ingiliˈzijja] 'English' which contains three unstressed syllables before the stressed syllable, as in Figure 1. Finally, preceding local pitch seemed also to be lower when the word is followed by relatively high pitch, such as in words followed by a high phrase tone (H-), as observed on the word [waˈraːna], also in Figure 1. There were no instances at all of preceding high pitch ('*h*'). In sum then, the level of preceding pitch seems to be stable, and any variation can be ascribed to factors in the local prosodic context.

Turning to the level of following pitch, this was also most often observed to be '*m*' but was more variable, with following 'm' observed in 66% of target words only. Occurrence of a

following ‘h’ or ‘l’ was however also found to be dependent on local prosodic factors such as other tonal events following the accented word. For example, all words followed by a H% boundary tone showed a continuous rise throughout the accented syllable (thus ‘*h*’ following the H peak), as in the word [tagri'bijja] in Figure 1. In contrast, words followed by a L-phrase tone or L% boundary tone all show either mid ‘*m*’ or low ‘*l*’ following pitch. A word was observed to be more likely to have a ‘*mHl*’ shaped accent (than ‘*mHm*’) if the next stressed syllable was relatively distant.

### IV.3 Summary

This detailed transcription of target words found in a small corpus of naturally occurring speech suggests that observed variation in the local pitch contour both preceding and following the accented syllable in EA can be described as a function of the surrounding tonal environs. Such variation is arguably therefore predictable from the prosodic context, and does not constitute evidence for additional pitch accent types in EA of any kind. Nonetheless it is interesting to note that the relative height of the following local pitch contour was observed to be more variable than that of the preceding local pitch contour. This matches findings regarding the alignment properties of L turning points preceding and following the H peak in EA, whereby it is the preceding L tone that is stably aligned at the onset of the stressed syllable [8,24].

## V Discussion and conclusion

This paper has made use of the labelling conventions of a local f0 contour tier, in order to test a claim regarding the best phonological representation of the range of phonetic realisations of accented syllables in a limited corpus of target words found in spontaneous colloquial EA speech. The labelling system makes it possible to categorise potential candidates for

membership of a single phonological category according to slight variations in the actual local f0 contour, and thus to determine whether any of these variation should in fact be considered as evidence of membership of some other phonological category. In the present study, it was found that all of the observed variations in the local f0 contour could be attributed to factors in the local prosodic environment, and it is thus argued that all of the 100 accented syllables should be assigned membership of a single phonological category. This was labelled with a dummy 'H*' in the annotation. We propose that the correct phonological designation is L+H*, for two reasons, both of which have previously been established experimentally [8] but which also become evident from the current study: firstly, the accented syllable is invariably marked by a pitch peak (H) which is positioned consistently within the stressed syllable, and any exceptions can be explained with reference to factors in the prosodic context; secondly, the level of pitch preceding the peak is more stable than the level of pitch following the peak, suggesting the presence of a leading L tone. Use of a prosodic annotation tier at the local f0 level is thus shown to be a useful tool in establishing phonological categories in spoken EA.

The labelling conventions of the local f0 contour tier represent only a small part of the range of tools available within the IV transcription systems; other tools include the use of a rhythmic tier (and corresponding Implementation Domains) to capture the relationship between rhythmic structure and intonational pitch events[15] and a global f0 contour tier proposed (in [3]) for transcription of declination and downstep. In addition we believe that it would be profitable to define and test a set of Break Indices for EA intonation[cf.16], in order better to determine the prosodic phrasing properties of spontaneous speech, and implementation of all these additional tools is planned. A crucial test of any successful labelling system however is the extent to which independent transcribers are able to use it to reach consensus about how a particular stretch of speech should be labelled. A study of inter-

transcriber agreement using the proposed labelling system is thus also necessary, to test the robustness of the prototype annotation guidelines for the local f0 contour tier demonstrated here, and to develop robust transcription norms for EA on rhythmic, global f0 contour and break indices tiers.

Establishment of an agreed intonational transcription system for EA will facilitate future interdisciplinary research between phonological and speech technology research teams. In addition a working transcription system will permit better descriptions of aspects of EA intonational phonology which remain as yet undescribed as well as paving the way for future documentation of intonational variation in Egypt, and comparison of EA intonation to that of other varieties of Arabic.

References


1. E. Grabe, F. Nolan, & K. Farrar. (1998) IViE - a comparative transcription system for intonational variation in English. Proceedings of the 5th International Conference on Spoken Language Processing (Sydney, Australia)., 1259-1262.
2. E. Grabe. (2001) The IViE Labelling Guide. (Version 3.0) http://www.phon.ox.ac.uk/IViE/guide.html.
3. B. Post & E. Delais-Roussarie. (2006) Transcribing intonational variation at different levels of analysis. *In* Proceedings of Speech Prosody 2006. R.Hoffmann and H.Mixdorff, eds. Dresden, TUD Press Verlag der Wissenschaften GmbH.
4. J.C.E. Watson. (2002) The phonology and morphology of Arabic. Oxford, OUP.
5. O.A.G. Ibrahim, S.H. El-Ramly, & N.S. Abdel-Kader. (2001) A model of F0 contour for Arabic affirmative and interrogative sentences. 517-524. Mansoura University, Egypt. 18th National Radio Science Conference.

6. K. Norlin. (1989) A preliminary description of Cairo Arabic intonation of statements and questions. Speech Transmission Quarterly Progress and Status Report, 1, 47-49.
7. S. Hellmuth. (2006) Focus-related pitch range manipulation (and peak alignment effects) in Egyptian Arabic. *In* Proceedings of Speech Prosody 2006. R.Hoffmann and H.Mixdorff, eds. pp. 410-413. Dresden, TUD Press Verlag der Wissenschaften GmbH.
8. S. Hellmuth. (2007) The foot as the domain of tonal alignment of intonational pitch accents. Proceedings of the 16th ICPhS, Saarbruecken, Germany.
9. S. Hellmuth. (2007) The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. The Linguistic Review, 24, 289-314.
10. S. Hellmuth. (2006) Intonational pitch accent distribution in Egyptian Arabic., Unpublished PhD thesis, SOAS. PhD.
11. D. Chahal & S. Hellmuth. (2009) Comparing the Intonational Phonology of Lebanese and Egyptian Arabic. *In* Prosodic Typology. Volume 2. Oxford, Oxford University Press.
12. D. Rastegar-El Zarka. (1997) Prosodische Phonologie des Arabischen., Unpublished PhD thesis, Karl-Franzens-Universität Graz. PhD.
13. K. Rifaat. (2005) The structure of Arabic intonation: a preliminary investigation. *In* Perspectives on Arabic Linguistics XVII-XVIII: Papers from the seventeenth and eighteenth annual symposia on Arabic linguistics. M.T.Alhawary and E.Benmamoun, eds. pp. 49-67. Amsterdam/Philadelphia, John Benjamins.
14. S. Hellmuth, F. Kügler, & R. Singer. (2007) Quantitative investigation of intonation in an endangered language (and how this methodology could help us understand the intonation of better-studied languages). Proceedings of the Language

Documentation and Linguistic Theory Conference (SOAS, London, December 2007).

15. E. Grabe. (2001) The IViE Labelling Guide. (Version 3.0). University of Cambridge.
16. M. Beckman & G.A. Elam. (1993) Guidelines for TOBI Labelling (version 3.0 1997). The Ohio State University Research Foundation.
17. E. Grabe. (2001) The IViE Labelling Guide. (Version 3.0) [http://www.phon.ox.ac.uk/IViE/guide.html].
18. C.W. Wightman. (2002) ToBI or not ToBI? Proceedings of Speech Prosody 2002.
19. K. Karins, M. Liberman, C. McLemore, & E. Rowson. (2002) CallHome Egyptian Arabic Speech Supplement LDC2002S37. Philadelphia, Linguistic Data Consortium.
20. P. Boersma & D. Weenink. (2007) Praat: doing phonetics by computer (Version 4.6.10) [http://www.praat.org].
21. P. Prieto, J. van Santen, & J. Hirschberg. (1995) Tonal alignment patterns in Spanish. Journal of Phonetics, 23, 429-451.
22. D. Chahal. (2001) Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: a comparison with English., Unpublished PhD thesis, University of Melbourne. PhD.
23. D. El Zarka & S. Hellmuth. (2008) Variation in the intonation of Egyptian Formal and Colloquial Arabic. Langues et Linguistique.
24. S. Hellmuth & D. El Zarka. (2007) Variation in phonetic realisation or in phonological categories? Intonational pitch accents in Egyptian Colloquial Arabic and Egyptian Formal Arabic. Proceedings of the 16th ICPhS, Saarbruecken, Germany.