

Homepage of the [IViE project](#)



The IViE Labelling Guide

Version 3, Copyright©2001 Esther Grabe



A [snapshot](#) of an IViE transcription.

0. Table of Contents

[1. Introduction](#)

[2. Technical Information](#)

[3. The Structure of IViE](#)

[4. More on the Prominence Tier and the Target Tier](#)

[5. Summary](#)

[6. References](#)

1. Introduction

This labelling guide describes the structure and application of the IViE system for prosodic labelling. IViE stands for 'Intonational Variation in English', and is pronounced like the woman's name 'Ivy'. IViE is based on ToBI, the current standard for prosodic labelling of English intonation (Silverman et al. 1992, Beckman and Ayers, 1997), but unlike the original ToBi, IViE allows for directly comparable transcriptions of several varieties of English in a single labelling system. Additionally, IViE transcriptions capture rhythmic differences between varieties, and differences in phonetic realisation.

In the IViE system, prosody is transcribed on three levels:

- (a) rhythmic structure
- (b) acoustic-phonetic structure
- (c) phonological structure

The three levels allow us to transcribe rhythmic variation, variation in pitch accent realisation, and variation in tune structure. Cambridge English and Bradford Punjabi English, for instance, differ in their rhythmic structure (Grabe et al., to appear). Leeds and Newcastle English differ in the phonetic realisation of pitch accents (Grabe, Post, Nolan and Farrar, 2000). Finally, Cambridge and Belfast English differ in their phonological structure. Belfast English speakers produce pitch accents which are not part of a Cambridge speaker's inventory (Grabe et al., to appear).

[Back to the top](#)

1.1 How IViE transcriptions are made

IViE transcriptions are made step-by-step. Labellers begin by deciding on the location of rhythmically strong syllables. Rhythmically strong syllables can be accented or unaccented and are labelled 'P' (prominent). The label is placed in the middle of the strong vowel.

Next, labellers transcribe the pitch movement surrounding syllables labelled as prominent. This transcription is made on the phonetic tier or target tier. Clearly, there are many acoustic-phonetic aspects of f₀ one could label (e.g. pitch range, declination, register, alignment). In the IViE system, the phonetic transcription is chiefly about alignment. Labellers transcribe the shape and alignment of f₀ patterns relative to the location of strong (accented) syllables in the text.

The domain for an alignment transcription is the *Pitch Accent Implementation Domain* or ID. More information on the ID is given in section 3.3.1 below.

Finally, labellers come up with phonological classifications. These are made on the tone tier. Note that phonological analysis in intonation is difficult and controversial, and there are no widely accepted tests (whether experimental or introspective) for phonological category membership. Therefore, on the IViE tone tier, transcribers are provided with a toolkit for phonological analysis. The labels given on this tier do not constitute a phonological analysis of any particular variety of English. Instead, users work with a pool of labels from which they choose different subsets for different varieties of English. In other words, IViE provides the labels, and labellers may use these to draw up phonological systems for different varieties of English. The resultant variety-specific systems are directly comparable because we have imposed one constraint on the possible form labels take: all tonal morphemes in the pool are left-headed, and because we have added an option that many other two-tone systems do not have: intonation phrase boundaries do not have to be associated with a high or a low tone, but can be left unspecified (Grabe, 1998b, boundary specifications are H%, L% or % (no change))

[Back to the top](#)

1.2 Comparative approach to phonological analysis

In our work on varieties of British English, we take a comparative approach to intonation analysis. Our corpus contains directly comparable data from twelve speakers of each variety of English recorded, in a range of speaking styles. Therefore, we do not need to label utterances from a particular speaker in isolation. We can compare the intonational structure of utterances produced in similar or identical context and produced with similar speaker intent. All IViE labels are assigned in this way. We start our work as follows:

- (1) We draw up a first set of hypotheses about variety-specific intonation systems on the basis of read speech (controlled sentences, read passage). Our read speech data provide the starting point because these data contain directly comparable intonation contours produced by different speakers in identical contexts.
- (2) Then we compare the contours produced in read speech with semi-spontaneous speech data (a retold version of our reading passage). In the semi-spontaneous data, speakers produce lexical items which have also been produced in the reading passage. We can compare the intonation contours produced around these items directly across the two speaking styles and across speakers.
- (3) Next, we transcribe Map Task data (goal-directed interaction). These data allow for a comparison of contours on the same lexical items across speakers and an investigation of deaccenting and reaccenting. We can also check whether speakers produce contours in the Map Task which they do not produce in other speaking styles.
- (4) Finally, we label conversational data, and this is the most difficult set of data to label (greatly increased range of delivery, speaker overlap, two overlapping fundamental frequency traces). The information about variety-specific intonation structures that we have collected from the other speaking styles is helpful here.
- (5) On the basis of our comparisons, we draw up an language-specific inventory of pitch accents and boundary tones.

[Back to the top](#)

2. Technical information

Like ToBI, IViE works in conjunction with xwaves(TM), a commercial software package which used to be available from Entropic. IViE labelling can also be carried out using [PitchWorks](#), or any other signal processing package which allows the user to add text labels (e.g. [PRAAT](#) or [wavesurfer](#)) or with pen and paper.

Xwaves runs under UNIX, and the IViE labelling tool displays the speech pressure wave, a labelling template and the fundamental frequency trace. If you place the cursor into one of the labelling tiers, and click the right mouse button, you get a menu with labels for that tier. The xwaves IViE labelling tool is similar to the ToBI labelling tool and displays a wave form together with the corresponding F₀ trace and 5 empty labelling templates. Within each template, a menu can be called up via the right mouse button, and this menu contains the relevant prosodic labels which can be inserted, deleted or shifted where appropriate. The other mouse-clicks work as in ToBI: a left mouse-click on a word labelled on the orthographic tier plays the word, and a middle mouse-click plays the stretch of speech from the end of the previous word up to the cursor.

The IViE labeller, menus, instructions and examples can be downloaded from our server, or ordered from us on a CD (free of charge). If you'd like to see some IViE labelled data, please send me message. Comments and suggestions are very welcome.

[Back to the top](#)

3. The Structure of IViE

IViE has five levels of transcription (two orthographic, three prosodic), arranged as follows:

5	Comment Tier	Alternative transcriptions and notes
4	Phonological Tier	Formal linguistic representations of speakers' intonational choices
3	Target Tier	Phonetic transcriptions; syllable-based; allow transcribers to draw up a first set of hypotheses about accent alignment
2	Prominence Tier	Location of prominent syllables (stressed and accented)
1	Orthographic Tier	Transcriptions of the words spoken

3.1 THE ORTHOGRAPHIC TIER

If you put the cursor in the lowest tier (the orthographic tier), and hold down the right mouse-button, you will see a menu that allows you to insert, delete, replace or move around words. Type the words spoken by the speaker, one by one, and align them with the end of the words in the speech wave using the MOVE button (right-mouse-menu). After choosing MOVE, click the middle mouse button. The word which is closest to your cursor will jump to the location of the cursor. If the words have been inserted correctly, you can hear each word by clicking on the word with the left mouse-button.

3.2 THE PROMINENCE TIER

The second lowest tier is intended for the transcription of stressed and accented prominences, rhythmic boundaries and hesitations.

The right-mouse menu offers three symbols:

'P' = prominence,

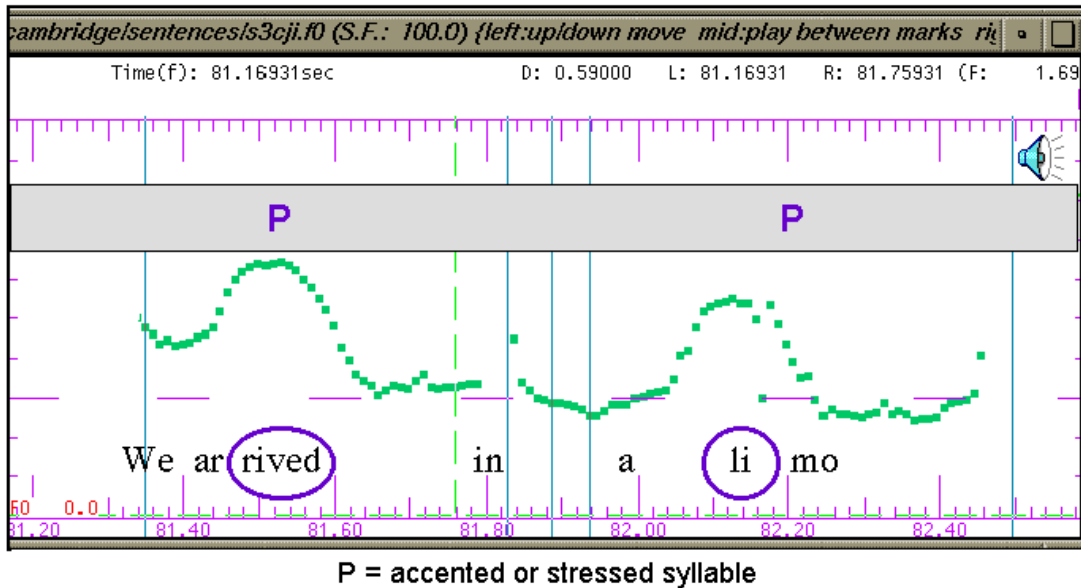
'%' = rhythmic boundary,

hash sign = hesitation, or speech error.

Insert P in the middle of a prominent syllable, % at the end of a word that is followed by a rhythmic boundary, and hash at the location of the hesitation or error. An example is shown in the PowerPoint slide shown below (NB: I haven't put the % symbol in - it would be right at the end of utterance, lined up with the end of the last word *limo*).

[Click here to listen to the utterance \(.wav\)](#)

Locations of prominent syllables: P



Slide 1. Location of two accented syllables in *We arrived in a limo*. Note that the same pattern is possible on e.g. *We hesitated in a limo*, i.e. we argue that *arrived* is not followed by an intermediate phrase boundary L-.

3.3 THE TARGET TIER or PHONETIC TIER

On the target tier, transcribers enter a syllable-based transcription of the alignment of f0 patterns surrounding prominent syllables. Transcriptions are made within the 'ID' or accent *Implementation Domain* (see 3.3.1 below).

NB: As the phonetic transcriptions are syllable-based, it is clear that they do not (and are not intended to) provide quantitative information about f0 alignment (this is what the f0 trace is for). Instead, the phonetic transcriptions provide the researcher with a first set of hypotheses about f0 alignment in his/her data. These hypotheses then require investigation via acoustic measurements.

3.3.1 The Implementation Domain (ID)

An ID corresponds to an accent foot plus the preaccentual syllable. Each ID contains

- (1) the preaccentual syllable,
- (2) the accented syllable,
- (3) all following syllables (if any) up to the next accented syllable.

NB. As the preaccentual syllable is included in the ID, there is a one-syllable overlap between IDs (for a related concept, see Gussenhoven's 1984 tonal association domain. ADs correspond to accent feet, but do not include the preaccentual syllable).

3.3.2 Phonetic labels

Six target labels are available for phonetic transcriptions: H, M, L are used for pitch levels or glides on accented syllables (e.g. L = level, LH = glide) and h, m, l for the unstressed syllable preceding the strong syllable, and for any unstressed syllables following the strong syllable, up to the next strong syllable.

'%' is used on the target tier to indicate the end of an ID which co-incides with a rhythmic boundary.

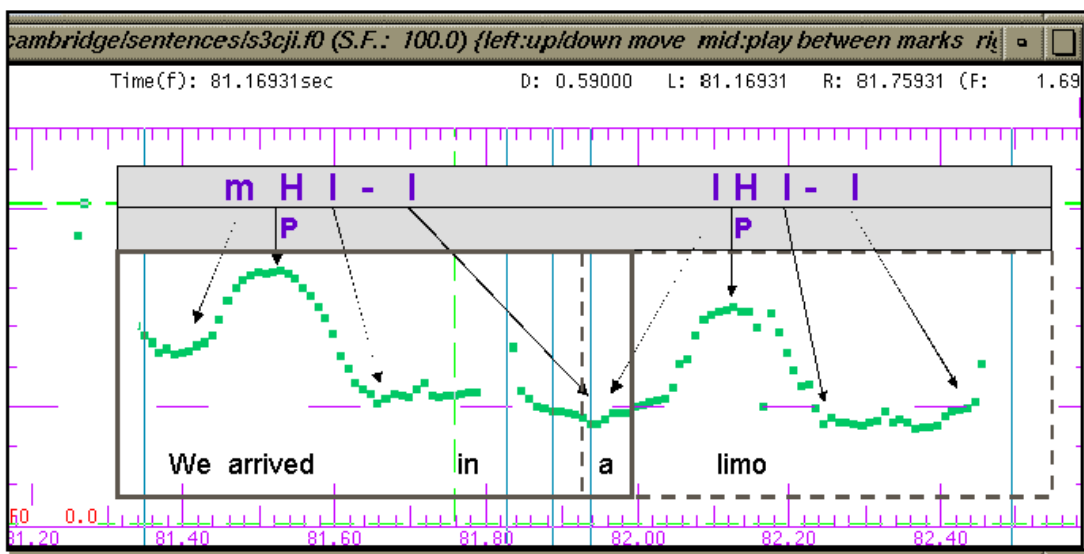
'-' stands for *interpolation*, and is used to connect the final pair of phonetic labels in an ID (i.e. the penultimate label and the one that transcribes the pitch level at the ID boundary). Please note that this hyphen is not like the '-' diacritic used to mark a phrase-accent or an intermediate phrase boundary). The combination 'mH-l-', for instance, indicates that the preaccentual syllable is mid, the accented syllable is high, and after that, there is an interpolation in the f0 trace to a low target at the ID boundary. If the ID ends with an accented syllable, no '-' is used. E.g. H*L % produced on the text 'the cat' could be IHL (no '-' between any of the target labels) and H* % would be IH.

Each of the following syllables can be given a label, and minimally two labels must be assigned (accented syllable and pre- or postaccentual syllable):

- (1) the preaccentual syllable
- (2) the accented syllable
- (3) the postaccentual syllable unless it's on the transition path to the final syllable
- (4) the final syllable in the ID

An example of a phonetic transcription is shown in the slide below. The grey boxes indicate the locations of the two IDs in the example. The arrows show the approximate location of the labelled targets in the relevant syllables.

Transcriptions of phonetic realisation: L H M



Slide 2. Phonetic labels describing the pitch pattern surrounding two stressed and accented syllables in *We arrived in a limo*.

Phonetic labels are assigned separately for each ID - relationships between pitch levels across IPs are not taken into account (such relationships are modelled on the phonological tier).

Finally, the target tier contains a selection of phonetic labels, but in general, transcribers make up their own and type them into the tier using the 'insert' command.

3.3.3. Parsimony of phonetic transcriptions

Generally, the idea is that labellers should use the smallest number of labels which transcribe the pitch pattern in an ID adequately, i.e. transcriptions should be parsimonious. Each syllable in the ID *can* be given a separate label, but most of the time, this is unnecessary. H*L, for instance, can be realised as 'l' on the preaccentual syllable, 'H' on the accented syllable followed by a transition to a low target 'l' at the ID boundary. In such a case, three levels are sufficient although the ID may contain more than three syllables.

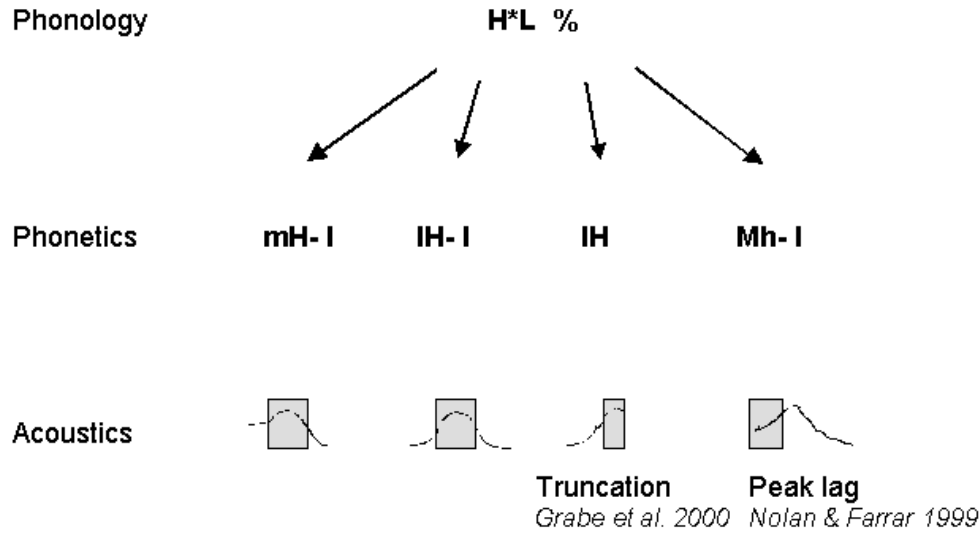
3.3.4. Motivation for a phonetic level of transcription

Data from the IViE corpus show that the phonetic implementation of intonation is extremely variable. At first sight, there is so much variation that one might be tempted to think that the variation is chaotic. But the variation is not chaotic - if one compares realisations of an utterance produced in identical contexts within and across varieties, one can see that much of the variation is systematic. But one can arrive at this conclusion only if one has a way of making a record of the phonetic variation. The transcriptions on the target tier provide such a record. They provide a set of testable hypotheses, for instance about the alignment of pitch accents and tunes in different varieties of English. Intonation researchers can then test these hypotheses via acoustic measurements.

In the same vein, the phonetic labels allow us to compile an index of phonology-phonetics mappings in different varieties of English. H*L, for instance, can be realised as hM-l, as mH-l, as hH-l, as Mh-l (peak lag) or IH (truncation). In Cambridge English, hM-l appears to map onto !H*L, but in Newcastle, hM-l is frequently the consequence of an early peak alignment and does not point towards a downstepped

production of H*L. The slide below provides further examples of phonetic/phonology mappings in British English.

Examples of phonetic variation



Slide 3. Examples of a number of possible phonetics phonology mappings.

Finally, phonetic intonation labels are of pedagogical value. When labellers or students are trained, they learn to arrive at phonological classifications by careful listening, and by examining the F0 trace. Problems may arise if labellers are not sure which stretches of speech they should listen out for, or if they rely too heavily on F0. The target transcriptions allow teachers to work out why a student may have come to a certain decision, and offer a starting point for discussion. But experienced transcribers can also disagree on phonological categorisations of pitch patterns - again, the phonetic transcriptions offer a starting point for discussion.

[Back to the top](#)

3.3.5 Examples

The following selection of labelling examples is sorted by direction of f0 movement from stressed syllable to immediately following syllable; this may be up, down, or level (IP initial options not given, self-explanatory). Note that the following examples represent a small selection of possible ID labels.

A. Levels

Stressed syllable in IP initial ID

up	also up
L-h ○ ● and also ○ ● ○ ● ● ● ● ●	Lh-h ○ ● ●
Down	also down
H-l ○ ● ● ●	Hl-l ○ ● ● ●
Level	n.a.
level ○ ●	n.a.

Stressed syllable in IP final ID

up	down	level
----	------	-------

IH ● ○	hL ● ○	level ● ○
--------	--------	-----------

Stressed syllable in IP medial ID

up
IL-h ● ○	ILh-h ● ○	mL-h ● ○	mLh-h ● ○	hL-h ● ○	hLh-h ● ○	lM-h ● ○	lMh-h ● ○
down
hH-l ● ○	mHl-l ● ○	mH-l ● ○	etc.	etc.	etc.	etc.	etc.

B. Glides

If the stressed syllable is produced with a pitch glide, transcribers use two capital letters.

For instance, on a monosyllabic word in IP-final position: IHL, hLH, mHL, mLH, hMl or lMH.

A monosyllable produced in isolation can be HL or LH. Unstressed syllables which are produced with a pitch glide receive two small letters, e.g. mHlh.

C. Pitch accents which are immediately adjacent

Label IDs as before from left to right, the second accent will have an accented syllable as the first pitch label, therefore capitalise and follow with left bracket to indicate that this is not a glide, i.e. a case where one stressed syllable is associated with two tones

For instance:

H*	!H*L
lH	H[M-l
P	P
the MEAL	EARly

NB: in the xwaves template, P is aligned with the middle of the stressed syllables 'meal' and 'ear', and the other labels are aligned with the P labels.

Labels are aligned roughly in the middle of the vowel in the stressed syllables, and in the middle of the ID if there is no stressed syllable. Note that pitch labels should also be aligned with phonological labels on the tone tier.

The pitch movement transcription is given separately for each syllable which is marked as rhythmically strong and accompanied by moving pitch (i.e. each accented syllable). In other words, each ID is transcribed in isolation, and relationships between successive accents are NOT taken into account (relationships between accents such as downstep are transcribed on the phonological level).

[Back to the top](#)

3.4 THE PHONOLOGICAL TIER

On the second highest tier, the intonational structure is labelled. Intonational phonology is a controversial area of research, and there are no widely accepted tests for phonological category membership of pitch patterns. In the IViE project, we draw up intonational phonological systems for different varieties of British English on the basis of comparisons of contours produced by different speakers in comparable contexts and produced with comparable speaker intent. The resulting transcriptions are comparable across varieties because

(a) all pitch accents specifications are taken from a single pool of labels (but not all labels or label combinations are used for every variety),

(b) all pitch accents are left-headed

(c) the system offers three rather than two boundary specifications (some varieties of British English make use of two boundary tones, but some have three different types of boundaries).

The following table shows the IViE tone labels:

IViE option	Contour can look like this (description and possible tone target labels)
H*L	High target on prominent syllable followed by low target in same ID, e.g. H-l, mH-l or mHl-l
H*	High target, common in initial position in so-called flat hats, e.g. lH-h
!H*L	Downstepped high target, low target, e.g. hM-l
L*HL	IP internal or IP final rise-fall: Low target on prominent syllable, high target on next syllable followed by low target, e.g. lLh-l
L*H	Low target on prominent syllable followed by high target, e.g. mLh-h, mL-h, or lL-h
L*	Low target
H*LH	IP internal or IP final fall-rise: high target on strong syllable, low, high, e.g. mHl-h

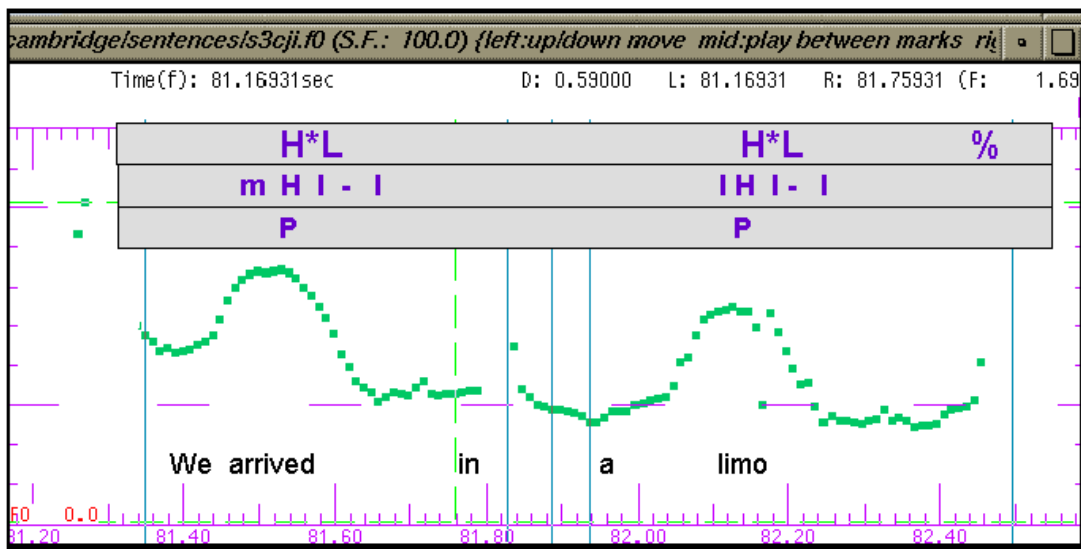
Intonation phrase boundary specifications:

Phrase-initial	Phrase-final	Transcribes:
%H	H%	high target
%	%	no pitch movement at boundary
%L	L%	low target

Extra Symbol	Transcribes:
#	Hesitation, interruption

Slide 4 below shows the complete IViE transcription for *We arrived in a limo*.

Example: phonological transcriptions



Slide 4. Complete IViE transcriptions.

NB: The % boundary specification in IViE has been taken from Grabe (1998a NB. transcribed as 0% in Grabe 1998a, but the 0 has been omitted). A % boundary symbol means that the tonal specification on the last syllable in the intonation phrase does not differ from the immediately preceding tone. In practise, a % specification says: here we have a relevant landmark in the contour, i.e. an intonation phrase boundary, and we know that, e.g. because there is a rhythmic continuity, and phrase-final lengthening, but nothing is happening in the tonal domain. The pitch level reached at the end of the IP-final accent continues at the same level. As nothing has changed, no tone is specified.

% boundary specifications offer transparent transcriptions of the rise-plateau patterns found in Northern Irish English:




 L*H H%	 L*H %	 L*H L%
...the right of the lamb. (.wav)	d'you know where the alleyway is? (.wav)	...there is a library! The relevant bit is right at the end of the pitch trace; the rising-levelish-falling section. (.wav)

Figure 1. Three boundary options in Northern Irish English.

The data in Figure 1 are from the IViE map task. In this task, the participants have to find their way round a small town, and the 'lamb' is a pub. The speech files listed in the table provide the context for the sections shown in Figure 1. The figure shows that in Northern Irish English, after L*H, pitch may either rise at the boundary, remain level, or fall (note that in General Southern British English, pitch can either remain level, or rise, but not fall). The IViE boundary specifications capture the three Northern Irish options and the two General Southern British (GSB) options comparably and transparently (i.e. the only difference is that we do not find L% in GSB).

[Back to the top](#)

3.5 THE COMMENT TIER

On the comment tier, transcribers can add alternative transcriptions and make notes.

4. More on the Prominence Tier and the Target Tier

In IViE, two tiers have been added to the original ToBI system, the prominence tier and the target tier. The new tiers are intended to increase the transparency and replicability of the labels on the tone tier. In essence, they permit a step-by-step breakdown of the process which leads to a specific tonal transcription. In English, this process begins with the identification of rhythmically prominent (stressed) syllables because the pitch movements transcribed on the tone tier are anchored to these syllables. In the IViE system, we have made this identification process overt, rather than implicit, i.e. everyone who labels English in a two-tone system makes decisions about the location of anchor points for accents.

The decomposition of the tonal transcription into a phonological and a phonetic tier is motivated as follows:

(1) Certain intonational distinctions are phonetic in one variety of English, and phonological in another (Grabe et al. 2000 and to appear [Prosody2000 proceedings]), and some involve the locations of prominent syllables in utterances. Accordingly, the IViE system was designed to allow for the transcription of prosodic differences between dialects at different levels of representation. Phonological differences are transcribed on the phonological tier, phonetic differences can be noted on the target tier, and differences relating to the distribution of prosodic prominences are labelled on the prominence tier.

(2) Labellers can disagree about the phonological classification of a particular accent pattern, but they are less likely to disagree about the phonetic realisation that pitch movement (i.e. people usually agree that IH-l is IH-l, but some may classify this pattern as H*L %, some as L+H*L-L% and yet others as H*L-L%).

The target tier offers an opportunity for consensus - a level at which transcribers can, and are likely to agree about the nature of a particular prosodic pattern. On the target tier, that consensus can be recorded.

[Back to the top](#)

5. Summary

IViE has five tiers: an ORTHOGRAPHIC tier for the words, a PROMINENCE tier for transcriptions of the location of stressed and accented syllables, a PHONETIC TIER for pitch levels or glides surrounding rhythmically prominent syllables, a PHONOLOGICAL TIER for phonological classification and generalisation, and a COMMENT tier for alternative transcriptions and notes.

6. References

- Beckman, M. and Elam, G. A. (1997). [Guidelines for ToBI labeling](#), version 3. Linguistics Department, Ohio State University.
- Cruttenden, A. (1996). *Intonation*. Cambridge, CUP.
- Fletcher, J., Grabe, E., and Warren, P. ([to appear, preprint in .doc format](#)). Intonational variation in four dialects of English: the high rising tune. In Sun-Ah Jun (ed) *Prosodic typology and transcription - a unified approach*. To be published by OUP.
- Grabe, E. (1997). Comparative intonation analysis: English and German. In A. Botinis, G. Kouroupetroglou and G. Carayannis (eds.) *Proceedings of the ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications*. Athens, Greece.
- Grabe, E. (1998a). *Comparative Intonational Phonology: English and German*. MPI Series 7, Nijmegen, The Netherlands.
- Grabe, E. (1998b). Pitch accent realisation in English and German. *Journal of Phonetics* 26, 129-144.
- Grabe, E., Post, B. and Nolan, F. ([to appear, preprint in .doc format](#)). Modelling intonational Variation in English. The IViE System. In *Proceedings of Prosody 2000*, 2-5 October, Krakow, Poland.
- Grabe, E., Nolan, F., and Farrar, K. (1998). IViE - a Comparative transcription system for intonational variation in English. *Proceedings of the 5th Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia.
- Grabe, E., Post, B., Nolan, F., and Farrar, K. (2000). Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28, 161-186.
- Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: CUP.
- Nolan, F. and Grabe, E. (1997). Can ToBI transcribe intonational variation in the British Isles? In A. Botinis, G. Kouroupetroglou and G. Carayannis (eds.) *Proceedings of the ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications*. Athens, Greece.
- Nolan, F. and Farrar, K. (1999). Timing of f0 Peaks and Peak Lag. *Proceedings of the International Congress of Phonetic Sciences*, 961-967.
- Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). ToBI: a standard for labeling English prosody. In *Proceedings of the Second International Conference on Spoken Language Processing (ICSLP)*, 2: 867 - 70. Banff, Canada.

[Back to the top](#)



Comments and suggestions welcome

- Last modified on 3/09/2001 -