Guidelines for ToBI Labelling

(version 2.0, February 1994)

by Mary E. Beckman and Gayle M. Ayers

Contents

0 Preface **2**
- 0.1 What are the "Guidelines for ToBI Labelling"? . . . 2
- 0.2 Getting and using the digitized utterances and f0 tracks . . . 2
  - 0.2.1 Getting the digitized utterances and f0 tracks . . . 2
  - 0.2.2 Using the digitized utterances and f0 tracks . . . 5
- 0.3 Getting and using the audio recording and paper f0 records . . . 6
- 0.4 Future editions and a disclaimer . . . 7

1 Overview and some basics **8**
- 1.1 The basic parts of ToBI . . . 8
- 1.2 Guiding principles . . . 9
- 1.3 The marking of stress — Pitch accents and prominence . . . 10
- 1.4 The marking of stress — Intonational phrasing and prominence . . . 11
- 1.5 What lines up with what? . . . 12

2 More on the tone tier **12**
- 2.1 Tones and fundamental frequency . . . 12
- 2.2 Some familiar contours, and the contrast between H* and L+H* . . . 14
- 2.3 The timing of the phrase accent and 'upstep' . . . 17
- 2.4 Difficult combinations of nuclear pitch accent and following phrase accent . . . 19
- 2.5 The initial %H boundary tone . . . 20
- 2.6 Pitch accent timing, and the L*+H pitch accent . . . 22
- 2.7 The H+!H* pitch accent . . . 24
- 2.8 Downstep . . . 24
- 2.9 Uncertainty about accent placement and accent type . . . 27
- 2.10 When something is accented that you would not expect . . . 29
- 2.11 The point of highest f0 . . . 31

3 More on the break index tier **31**
- 3.1 The break index tier relative to other tiers . . . 31
- 3.2 Break indices 0 and 1 . . . 32
- 3.3 Break indices 3 and 4 . . . 33
- 3.4 Break index 2 . . . 34
- 3.5 The p diacritic (and the %r tone label) . . . 35
- 3.6 Ordinary uncertainty. . . . 37

4 The miscellaneous tier (and other aspects of the marking of disfluencies) **38**
- 4.1 The miscellaneous tier defined . . . 38
- 4.2 Suggested guidelines for marking disfluencies . . . 39

References **42**

Index **44**

0 Preface

0.1 What are the "Guidelines for ToBI Labelling"?

ToBI (for Tones and Break Indices) is a system for transcribing the intonation patterns and other aspects of the prosody of English utterances. It was devised by a group of speech scientists from various different disciplines (electrical engineering, psychology, linguistics, etc.) who wanted a common standard for transcribing an agreed-upon set of prosodic elements, in order to be able to share prosodically transcribed databases across research sites in the pursuit of diverse research purposes and varied technological goals. Silverman et al. 1992 describe the motivation for and development of the ToBI system. If you ask for this handbook in hard copy, that paper will be appended as Appendix B. Appendix A (which is included both in the hard copy and in the ASCII file version of this labelling guide) is "The ToBI Annotation Conventions", the definitive summary statement of the symbols and marks used in ToBI transcriptions, and of the conventions that we have agreed upon for their use. The rest of this labelling guide is a more detailed description of the system, with reference to accompanying utterances of two types: example utterances to illustrate points made in the text and exercise utterances to give labellers practice on the points made in the text. These utterances are set off in the text of the labelling guide using the following typographic conventions.

`EXAMPLE <<basename>>:` orthographic transcription
tonal transcription and/or break index values

`EXERCISE <<basename>>:` orthographic transcription

Each example utterance is also referred to in the text by its basename within pairs of angle brackets — e.g., the first example utterance is ≪jam1≫. We have chosen the examples and arranged the exercises with the aim of leading new users through the system in a self-taught training course, trying to choose utterances in each of the six practice sets that show only phenomena that have been introduced up to that point.

The utterances that accompany this labelling guide can be obtained in two formats: as digitized computer files (with electronic record of the f0 contour) or as an audio tape (with paper record of the f0 contour). These utterances and this labelling guide were developed at the Ohio State University with partial support from the National Science Foundation, and New Mexico Institute of Mining and Technology has generously provided a distribution site for the electronic records (see Section 0.2). They are available to any interested user, but only for non-commercial use. The National Science Foundation, the Ohio State University, and New Mexico Institute of Mining and Technology make no warranty and accept no liability associated with the use of these materials. These materials may be obtained only as described in Sections 0.2 and 0.3, and are not to be redistributed by other user sites. Users may not redistribute these materials from their own sites, but should instead tell interested people how to obtain their own copy from the distribution sites.

0.2 Getting and using the digitized utterances and f0 tracks

If you have *waves*^{TM} (an Entropic Research Laboratory product) or a similar computer display system, obtain the speech files, electronic record of the f0 contour, and label files by ftp from Colin Wightman at the New Mexico Tech ditribution site.

0.2.1 Getting the digitized utterances and f0 tracks

There are two options for obtaining the ToBI materials depending upon how much disk space users have available. For those with sufficient disk space there is a single large tarfile for convenience. This option requires about 40 MB available during the installation process; the full materials occupy about 20 MB once the installation is completed and the tarfile is removed. If you do not have enough space to have both the complete tarfile and all the installed files at the same time, use the second option. There are four smaller tarfiles which together contain all the materials contained in the single large tarfile. The

"Guidelines for ToBI Labelling", scripts, and tools are contained in the "essential" tarfile, and the speech files, f0 records, and label files are divided into three parts by order of occurrence in the Guidelines.

On your home system, enter the command: `ftp kiwi.nmt.edu` (or use the internet address for kiwi, which is 129.138.1.82) from the directory where you would like to have the materials (the installation process will create a directory called `TOBI-TRAINING` where all the files will be put). When prompted, type the login name `ftp` and your user name on your home system as a password. If you now list the files available (`ls`), you will see several directories as listed below. Users should feel free to explore these directories.

`docs` – contains documentation files such as this labelling guide.
`pub` – used for temporary storage of files uploaded by users.
`tarfiles` – contains compressed sets of files for easy retrieval.
`tools` – contain shellfiles and tools for transcription.
`utts` – contains utterances and transcriptions for reference.

Further descriptions of the ToBI ftp site at kiwi.nmt.edu and some guidelines can be found in the README file. If you run into serious problems send email to Colin Wightman at cww@kiwi.nmt.edu.

All the files for the training materials have been placed in one large tarfile and in four smaller files as described above. To get it/them, change to the tarfile directory (`cd tarfiles`). Since tarfiles are binary files, enter the command `binary`. You can transfer the complete materials by typing `get complete_tobi_release_2.0.tar.Z`. When the transfer is complete, return to your home system with the command `quit`. Alternatively, to get the smaller files with part of the training materials only, type `get tobi_release_2.0_essentials.tar.Z`. This will give you the Guidelines for ToBI Labelling, and the scripts and tools that you will need. Then to get the actual examples and exercises speech files, f0 contours, and label files, get one or more of the "part" files. You can take as much as you have room for and then delete things after you are through with them to make room for the next set of examples and exercises to work through. Part 1 has the files for Section 1 to Practice 2, Part 2 has the files for Section 2.6 to Practice 4, and Part 3 has the files for Section 3.2 to Practice 6. Transfer these by typing `get tobi_release_2.0_part1.tar.Z`, `get tobi_release_2.0_part2.tar.Z`, and `get tobi_release_2.0_part3.tar.Z` respectively. When you have all the files you want, return to your home system with the command `quit`.

Back on your home machine, make sure the tarfile(s) is/are in whatever directory you would like the ToBI files to reside under. The installation process will create a directory underneath the one in which you put the tarfile(s) and all the ToBI material will go in the new directory. To install the files, first enter the command `uncompress complete_tobi_release_2.0.tar.Z` (or uncompress the "essentials" and "parts" files). Once the file(s) is/are uncompressed (it will take a few minutes), enter the command `tar -xvf complete_tobi_release_2.0.tar` (or tar -xvf the other .tar files to extract all of the subdirectories and files). Don't forget to delete the tarfile once you're happy that everything got installed correctly.

Summary instructions are listed below.

1) `ftp kiwi.nmt.edu` (or `ftp 129.138.1.82`)
   Name: `ftp`
   Password: `username` on home system
2) `binary`
3) `cd tarfiles`
4a) `get complete_tobi_release_2.0.tar.Z`
   or, if disk space is a concern, chose a relevant subset of
4b) `get tobi_release_2.0_essentials.tar.Z` (Guidelines for ToBI Labelling, scripts, tools)
   `get tobi_release_2.0_part1.tar.Z` (Section 1 to Practice 2)
   `get tobi_release_2.0_part2.tar.Z` (Section 2.6 to Practice 4)

`get tobi_release_2.0_part3.tar.Z` (Section 3.2 to Practice 6)

5) `quit`

6a) `uncompress complete_tobi_release_2.0.tar.Z`
or relevant subset of

6b) `uncompress tobi_release_2.0_part1.tar.Z`
`uncompress tobi_release_2.0_part2.tar.Z`
`uncompress tobi_release_2.0_part3.tar.Z`

7) `tar -xvf complete_tobi_release_2.0.tar`
(or relevant tobi_release_2.0 parts)

8) `rm complete_tobi_release_2.0.tar`
(or relevant tobi_release_2.0 parts)

The directory structure that you should recover by the untarring all the files is described below. The top level directory is `TOBI-TRAINING`. It includes an ASCII version of this labelling guide (called `labelling_guide_v2.ASCII`), two subdirectories called `EXAMPLES` and `PRACTICE` (which hold the speech, f0, and label files), and two script files called `examples` and `exercises`. The last two are *waves*^{TM} scripts for displaying the examples and exercises and for labelling the exercises. These two scripts assume this directory structure. Non-*waves*^{TM} transcriptions of the examples are given in the ASCII file `Nonwaves-transcriptions`, and more detailed instructions about how to use the scripts and useful shortcuts for the mechanics of labelling using the *waves*^{TM} scripts are given in `README-transcriber`.

`TOBI-TRAINING`
- `EXAMPLES` (where speech, f0, and "answers" are kept)
  - `AND1.breaks` (break index label file)
  - `AND1.d` (speech)
  - `AND1.f0` (f0)
  - `AND1.misc` (misc label file)
  - `AND1.tones` (tones label file)
  - `AND1.words` (words label file)
  - . . .
- `PRACTICE` (where user transcriptions are kept)
  - `I-mean.breaks`
  - `I-mean.words`
  - `I-mean.tones`
  - . . .
- `examples` (script for displaying examples and "answers")
- `exercises` (script for practice labelling)
- `labelling_guide_v2.ASCII` (ASCII version of Guidelines and Conventions)
- `Nonwaves-transcriptions` (ASCII version of non-waves(tm) labelling)
- `README-transcriber` (instructions and shortcuts for scripts)

After you have untarred the files, you will have to type a few commands to use the utterances as intended. To make the scripts executable and to protect the speech files and "answer" label files which are kept in the directory `EXAMPLES` from being overwritten by mistake, type the following three commands at the unix command line from within the `TOBI-TRAINING` directory.

```
chmod +x examples
chmod +x exercises
chmod -w EXAMPLES
```

There are also a few other tools included in the complete ToBI release and the "essentials" file. These are compressed tarfiles which must be uncompressed and untarred (as above) to be installed. They are not strictly necessary for working through the "Guidelines for ToBI Labelling", but are helpful tools to have. `cardinals.tar.Z` contains the files necessary for displaying cardinal examples of ToBI label categories, `transcriber.tar.Z` contains the files necessary for transcribers to transcribe their own data,

and `checker.tar.Z` contains the files necessary to invoke the John Pitrelli's checking program, which checks transcriptions and reports errors.

`cardinals.tar.Z` allows the user to display and play cardinal examples of ToBI label categories by pushing buttons in a menu display. If this tool is installed, the button menu with these examples is called up automatically each time the `examples`, `exercises`, or `transcriber` (see next tool description) scripts are invoked. Read the `README-transcriber` file for more information. Install by uncompressing and untarring as described above.

`cardinals.tar.Z` (for displaying cardinal examples of ToBI)

When uncompressed and untarred yields:

`README-transcriber`
`aux_examples/` (additional examples of ToBI labelling)
`cardinals`
`docard`

Labellers who wish to label their own data should use the script `transcriber` which is included in `transcriber.tar.Z`. This uses the same format as the training materials. Cardinal examples are available if they are installed. Read the `README-transcriber` file for more information. Install by uncompressing and untarring as described above.

`transcriber.tar.Z` (for transcribing your own examples)

When uncompressed and untarred yields:

`breakindexmenu`
`miscmenu`
`tonemenu`
`transcriber` (script for doing transcriptions)
`wordsmenu`

Labellers should check their transcriptions of their own data to see that they have a "legal" ToBI transcription. The script `check-transcription` checks the label files for consistency and adherence to the ToBI conventions for labelling. Read the `README-checker` file for more information. Install by uncompressing and untarring as described above.

`checker.tar.Z` (for checking transcriptions)

When uncompressed and untarred yields:

`README-checker`
`check-and-behead-breaks.awk`
`check-and-behead-misc.awk`
`check-and-behead-tones.awk`
`check-and-behead-words.awk`
`check-transcription` (script for checking transcriptions)
`check-transcription.awk`

0.2.2 Using the digitized utterances and f0 tracks

Now you are ready to use the two $waves^{TM}$ scripts. Both of them take as their argument(s) the base-name(s) of the utterance(s) you want to display. For example, as you are reading about example utterance ≪jam1≫ in the labelling guide, you can listen to the speech and look at the associated transcription by typing:

```
examples jam1
```

This will call up xwaves with two data windows to display the speech waveform and f0 trace, and a third window for the ToBI labels of our "answer" transcriptions. This script is set up to just display the information and play the speech; it does not allow the user to change the labels.

In order to practice transcribing one of the example utterances (say the first exercise in PRACTICE ONE), type:

```
exercises amelia-p2
```

This will display the speech waveform and f0 with only the word labels and placeholders for the break index labels in the labelling window. You can then use the labelling menus to add the tones and substitute break index values for the place holders in the break index section of the label window.

Both of these scripts can be used to display several utterances in a series. For example, to queue up the first four examples in the Guidelines, type:

```
examples jam1 cough made1 made2
```

After you have finished looking at ≪jam1≫, push the `CONTINUE` button in the *waves*^{TM} control panel. The example ≪cough≫ will then be the displayed.

A note on where the label files are stored: The "answer" label files (`basename.tones`, `basename.breaks`, `basename.misc`) are kept in the directory `EXAMPLES` along with the speech, f0, and word labels files (`basename.d`, `basename.f0`, `basename.words`). The label files that the user creates when labelling with the script `exercises` are stored in the directory `PRACTICE`.

A note on multi-transcriber sites: Setting up a site to be a multi-transcriber site is fairly straightforward. The main idea is that instead of all labellers working with the `exercises` script and having their labels saved in the directory `PRACTICE`, each labeller will have a personal `exercises` script and directory where the labels will be stored. Take the name `USER` as our demonstration (where any name can substitute for `USER`).

For each user, make a separate directory within the top level directory `TOBI-TRAINING`, copy the `exercises` script to a personal copy for the user, edit the `user` script, and then invoke the script `user` exactly as one would invoke the `exercises` script (make sure the `user` script is executable — `chmod +x user` if necessary.) Additionally, one may want to copy the break index placeholders into the `USER` directory.

1) `mkdir USER`
2) `cp exercises user`
3) Change line in script `user` which says:
   `PRACTICEDIR=PRACTICE`
   to say:
   `PRACTICEDIR=USER`
4) `cp PRACTICE/*.breaks USER/`
5) `chmod +x user` (if necessary)
6) E.g., start Practice 1 by typing:
   `user amelia-p2`

0.3 Getting and using the audio recording and paper f0 records

To get an audio tape of the utterances and a printed paper copy of the labelling guide and f0 tracks, send your request along with a check for $25.00 made out to The Ohio State University to:

ToBI Labelling Guide, c/o Mary Beckman
Ohio State University, Linguistics Dept.
222 Oxley Hall, 1712 Neil Ave.
Columbus, OH 43210-1298 USA

(The $25.00 just barely covers the cost of making copies of the tape and booklet and of the mailing to a

North American location.)

On the audio tape each utterance is played twice in a row. The utterances occur in the order that they are listed in the Labelling Guide text, with two more repetitions if an utterance is mentioned again later in another section. However, it is far easier to use the utterances if you can play each one as many times as you want and if you can zero in on some section of an utterance at will, and we recommend that you find some way of doing so. For example, you might use a tape recorder with a recordable tape loop device (and loops of several lengths). Or if you have a Kay DSP 5500 or some other kind of computer system with fast A/D capabilities, digitize each utterance from the tape into a buffer where you can play repeatedly while looking at the paper record of the f0 contour and the accompanying labels.

To transcribe the example utterances, you will need to use the non-*waves*^{TM} conventions described in Section 9 of The ToBI Annotation Conventions. The last section (Appendix C) of the booklet is a listing of an ASCII file containing the non-*waves*^{TM} format labelling of each example and exercise utterance corresponding to the *waves*^{TM}-format labelling displayed on the sheet with the f0 contour. They are given in alphabetical order by basename. This ASCII file and another ASCII file containing the orthographic labels and field placeholders of all of the exercise utterances file can be obtained by anonymous ftp from Colin Wightman by doing the following:

On your home system, enter the command: `ftp kiwi.nmt.edu` (or use the internet address for kiwi, which is 129.138.1.82). When prompted, type the login name `ftp` and your user name on your home system as a password. If you now list the files available (`ls`), you will see several directories:

`docs` – contains documentation files such as this labelling guide.
`pub` – used for temporary storage of files uploaded by users.
`tarfiles` – contains compressed sets of files for easy retrieval.
`tools` – contain shellfiles and tools for transcription.
`utts` – contains utterances and transcriptions for reference.

To get them, change to the docs directory (`cd docs`) and transfer the files by entering the commands `get Nonwaves-transcriptions` and `get Nonwaves-exercises-templates`. When the transfers are complete, return to your home system with the command `quit`.

0.4 Future editions and a disclaimer

If you have comments on this Labelling Guide — particularly, if you have suggestions for improvements or better example utterances you would like to give to us — we would be very grateful if you would direct the comments to us at:

e-mail: tobi@ling.ohio-state.edu

other-mail: ToBI Labelling Guide, c/o Mary Beckman
Ohio State University, Linguistics Dept.
222 Oxley Hall, 1712 Neil Ave.
Columbus, OH 43210-1298 USA

This e-mail address is also the place to send us your e-mail address if you want to be added to our list of "subscribers" to be notified of any future editions of the Labelling Guide.

The ToBI labelling system was originally developed to cover the three most widely used varieties of spoken English — namely, general American, standard Australian, and southern British English. We do not claim to cover other varieties. Indeed, we have already determined that ToBI proper does not adequately cover many other British varieties such as the Glasgow dialect, and modified variants need to be developed by users who want to use it in transcribing utterances in these other dialects. By the same token, we must stress that ToBI was not intended to cover any language other than English, although we endorse the adoption of the basic principles in developing transcriptions systems for other languages, particularly languages that are typologically similar to English. More general comments about using the ToBI system for other dialects of English or about adapting ToBI labelling principles to develop comparable systems for the transcription of other languages may also be addressed to the ToBI e-mail

address listed above for forwarding to appropriate interested members of the larger ToBI group.

1 Overview and some basics

1.1 The basic parts of ToBI

A ToBI transcription for an utterance consists minimally of a recording of the speech, an associated electronic or paper record of the fundamental frequency contour, and (the transcription proper) symbolic labels for events arranged in four parallel tiers. (Other tiers can be added for the needs of particular sites — see Section 4.) The four tiers of labels, arranged in the order that they appear in the default labels window for the examples and exercises programs, are:

1. an orthographic tier
2. a tone tier
3. a break index tier
4. a miscellaneous tier

The tone and break index tiers represent the core prosodic analysis. The tone tier is the part of the transcription that corresponds most closely to a phonological analysis of the utterance's intonation pattern. It consists of labels for distinctive pitch events, transcribed as a sequence of high (H) and low (L) tones marked with diacritics indicating their intonational function as parts of pitch accents or as phrase tones marking the edges of two types of intonationally marked prosodic units. The inventory of pitch events and their definitions are based on autosegmental analyses, in particular the analysis of Pierrehumbert and her colleagues (see Pierrehumbert and Hirschberg 1990, and the references cited in it) with some modifications toward such alternative analyses as that of Ladd 1983. In example utterance ≪jam1≫, there is a production of the question *Will you have marmalade, or jam?* with two pitch accents (the L* tones), two phrase accents (the H- tones), and a H% boundary tone.

```
EXAMPLE <<jam1>>: Will you have marmalade, or jam?
                                 L*      H-    L* H-H%
```

The break index tier marks the prosodic grouping of the words in an utterance by labelling the end of each word for the subjective strength of its association with the next word, on a scale from 0 (for the strongest perceived conjoining) to 4 (for the most disjoint). These categories of association strength, or 'break indices' are based on work by Mari Ostendorf, Patti Price, Stefanie Shattuck-Hufnagel, and their associates (see, e.g., Price et al. 1991). We equate the two highest break indices with prosodic groupings that are marked intonationally. For example, break index 3 after the word *marmalade* in utterance ≪jam1≫ corresponds to the end of the intermediate phrase indicated by the H- phrase accent.

```
EXAMPLE <<jam1>>: Will you have marmalade, or jam?
                       1   1    1          3  1   4
```

The orthographic tier is arguably not part of any core prosodic analysis, except inasmuch as the labels on this tier can be used to interface the transcription to dictionary entries which do indicate such things as which syllable is likely to be most stressed in each word, prosodic information which is not otherwise included in the ToBI system. The orthographic tier is a straightforward transcription of all of the words in the utterance, in ordinary English orthography. When using $waves^{TM}$ and a transcriber script, or any similar computer labelling system, the convention is to align each orthographic label to the end of the word.

The miscellaneous tier, like the orthographic tier, can include many events that are arguably not part of prosody per se. However, many events that are typically marked on this tier are important for interpreting the analyses on the tone tier and break index tier, because they disrupt the smooth rhythm of the utterance or interrupt the intonation contour. This tier is essentially a 'comment' tier that can be

used to mark events such as the cough in example utterance ≪cough≫. Except for very few exceptions (most notably, the label 'disfl' often stands alone to flag the occurrence of a perceived disfluency of some type), labels on this tier come in pairs, to mark the beginning and end of each event interval. If it were not for the disruption of the cough labelled on the miscellaneous tier here, the tone transcription would have to be parsed as either unfinished or ill-formed.

```
EXAMPLE <<cough>>: Will you have marmalade ...
                        L*        L*
                       1   1    1         1p
                                   cough<          cough>
```

1.2 Guiding principles

As should be obvious from the preceding examples, ToBI does not try to transcribe all aspects of prosody, or even all aspects that are amenable to symbolic transcription. In deciding what to include and what to leave out, we were guided by three principles. First, we wanted to be able to distinguish in our transcription all of the categorically distinct intonation patterns and prosodic units of the language (or rather of the three intonationally similar dialects that we claim to cover — see Section 0.4 above). Second, we felt we should not transcribe aspects of prosody which are more amenable to quantitative measures than to the categorical divisions of a symbolic transcription. Finally, we did not want to squander the user's energies in transcribing even categorical aspects of prosody which are predictable from other parts of the transcription or from auxiliary tools such as dictionaries.

The categorical aspects of prosody which we try to capture completely (by the first principle) are of two types. The first is the prosodic structure — the rhythm of more and less stressed words alternating with each other, and the grouping of words into prosodic constituents of various sizes — and the second is the intonation pattern — the sequence of contrastive pitch events that we call pitch accents, phrase accents, and boundary tones.

An example of the noncategorical aspects of prosody which we leave out (in accordance with the second principle) is the local tempo of each word in the utterance, which we feel could be more accurately and directly captured by some quantitative measure such as normalized segment duration (e.g., Campbell 1992) than by any symbolic transcription such as an arbitrary division into, say, categories '1', '2', and '3' (for 'slow', 'medium', and 'fast' tempi). An exception to this principle is the marking for each phrase of the point of highest fundamental frequency associated with an accent (HiF0), which we use as a measure of pitch range in order to facilitate research on the relationship between pitch range and discourse structure (see, e.g., Grosz and Hirschberg 1992, and references therein). We anticipate being able to do away with this marking when we have developed automatic tools for detecting accent-related peaks directly from the fundamental frequency contour in conjunction with the tone tier transcription.

A categorical aspect of prosody which we leave out (in accordance with the third principle) because it should be fairly predictable is the marking of the stressed and unstressed syllables within each word. By this level of stress we mean the word-internal alternation between more and less stressed syllables where the relative prominence of any pair of syllables is fairly fixed and can be thought of as inherent to the word's dictionary entry. For example, if the first and third syllables in the word *marmalade* are not pronounced with more prominence than the second, native speakers will judge the vowels in these two syllable to be mispronounced. (That is the first and third syllables should not have reduced vowels, whereas the second one should.) Since such word-internal rhythms are thus a fixed part of the word's pronunciation, we leave this specification out. That is, for example, in the transcription of utterances ≪jam1≫ and ≪cough≫, we have not marked the first and third syllables as relatively more stressed than the second syllable, since this aspect of the prosodic structure would be marked in any dictionary entry for the word, so that users of ToBI-transcribed databases could interface the orthographic tier with an online dictionary to fill in this information.

1.3 The marking of stress — Pitch accents and prominence

If the stress patterns within words are largely predictable from the dictionary entries for the word, what about other levels of stress? It has been recognized for some time now (e.g. Bolinger 1972) that other aspects of the stress pattern cannot be predicted from the grammar with anything like the confidence with which we can predict the more stressed syllables within a word. Indeed the factors predicting the prominence of a word relative to other words in the same sentence is a matter of much current debate (see e.g. Hirschberg 1993), and is one of the issues which we hope ToBI transcribed databases will be most useful in helping to resolve.

Example utterance ≪made1≫ illustrates the unpredictability of prominences above the word, with three different productions of the same sentence — *Marianna made the marmalade* — each of which has a different stress pattern. In the first production, there are two syllables that are relatively more prominent than any other, the accented syllables in the words *Marianna* and *marmalade*. In the second production of the sentence, on the other hand, there is only the one relatively more prominent syllable in *Marianna*, and *marmalade* has been 'deaccented'. This level of stress is marked in the ToBI system by directly transcribing the pitch accent on the tone tier. Thus, in the transcription of the first production in the example, there are H* accents marked for both *Marianna* and *marmalade*, whereas in the second production there is only the L+H* accent marked on *Marianna*. (The third production, like the first, also has accents on *Marianna* and *marmalade*, but it has a different stress pattern because both of these accents are nuclear stresses, whereas in the first production only *marmalade* has a nuclear accent. We will describe this higher level of stress in more detail in the next subsection.)

```
EXAMPLE <<made1>>:      Marianna made the marmalade.
in three productions 1)    H*             H*    L-L%
                     2)  L+H*                    L-L%
                     3)  L+H*L-H% L*       H*    L-L%
```

Note that there is another difference between the first production and the last two: the second and third productions begin at a much lower fundamental frequency than the first. This is due to the distinction, marked on the tone tier, between a single-tone H* pitch accent and a bitonal L+H* pitch accent. This contrast is independent of the difference in stress pattern, which depends on the pattern of pitch accent PLACEMENT and not on the type of pitch accent. To see this, compare the first two productions of the sentence in ≪made2≫ with the second two productions. (These first two sentences are the same as productions (1) and (2) in ≪made1≫.)

```
EXAMPLE <<made2>>:       Marianna made the marmalade.
in four productions  1)     H*              H*    L-L%
                     2)   L+H*                    L-L%
                     3)   L+H*             !H*    L-L%
                     4)     H*                    L-L%
```

The stress patterns are the same, but the choice of H* versus L+H* pitch accent type is the opposite. (For the relationship between the second pitch accent and the first in production (3) and the diacritic '!' that marks this relationship, see Section 2.8 below. The somewhat less low beginning in the third production is also dicussed in Section 2.2.) The same stress patterns are illustrated again in the third and fourth productions in ≪made3≫ with yet another pitch accent type, this time a L* pitch accent (with a following rise into H- phrase accent and H% boundary tone).

```
EXAMPLE <<made3>>:     Marianna made the marmalade.
in four productions  1)  L+H*           !H*    L-L%
                     2)    H*                   L-L%
                     3)    L*             L*    H-H%
                     4)    L*                   H-H%
```

In transcriptions using $waves^{TM}$ label files (or any similar computer labelling system), the stress that comes from associated pitch accents can be parsed from reading the tone tier, since the waveform is used

to place the mark for a pitch accent somewhere in the syllable that is phonologically associated to the accent. In the non-*waves*^{TM} transcription conventions, the stress is marked even more explicitly in the symbolic string, by putting an asterisk in the orthographic transcription just before the vowel of each accented syllable.

1.4 The marking of stress — Intonational phrasing and prominence

Above the level of contrast between pitch-accented versus unaccented words, native speakers of English can distinguish another level of stress contrast, that between the last accented word of a phrase and any preceding accent. In the first production in utterance ≪made1≫, for example, the word *marmalade* feels more prominent than *Marianna*. In the last production of the sentence, on the other hand, *marmalade* does not feel necessarily more prominent than *Marianna*. The sentence has been divided into two intonational phrases, so that each of these words is the last accented word in its own phrase. (This level of prominence is often called the 'nuclear stress' or 'nuclear accent' of the phrase.) Note that the level of prominence need not be marked explicitly, since the word with nuclear stress is defined positionally; it is the last accented word, or the accented word (if there is only one in the phrase). Thus the prominence contrast between a nuclear accent and a mere (prenuclear) accent can be read from the transcription of the accents on the tone tier relative to the boundaries marked between the phrases.

```
EXAMPLE <<made1>>:       Marianna made the marmalade.
in three productions 1)     H*              H*    L-L%
                                  1    1   1          4
                     2)   L+H*                       L-L%
                                  1    1   1          4
                     3)   L+H*L-H% L*       H*    L-L%
                                  4    1   1          4
```

There are two separate markings indicating the boundaries of an intonation phrase; one is the sequence of phrase accent and boundary tone on the tone tier, and the other is the 4 on the break index tier. The break indices are numbered from 0 (for least disjuncture) to 4 (for most pronounced disjuncture). The numbering captures the hierarchical nature of these prosodic groupings. At the highest level of the break index hierarchy and at the next lower level, the sense of disjuncture between adjacent words is connected closely to the intonation pattern. The boundary after *Marianna* in the third production of the sentence in ≪made1≫ is one at the highest level in the break index hierarchy transcribed in ToBI. This level is marked tonally by a boundary tone (H% or L%) at its end (and sometimes at its beginning, too, in which case it is %H). The next lower level (break index 3) is marked by a phrase accent (H- or L-) at its end.

An intonation phrase contains one or more intermediate phrases, and the end of an intonation phrase is by definition also the end of an intermediate phrase (break index 3). This fact is reflected on the tone tier in the requirement that there be a sequence of phrase accent (for the last intermediate phrase) followed by a boundary tone at the end of every intonation phrase. The last production of the sentence in ≪made1≫ illustrates this nicely with clear reflexes of the tone string in the fundamental frequency contour. Note first the fall from the peak for the L+H* nuclear pitch accent to the L- phrase accent for the first intermediate phrase, followed by the small rise in fundamental frequency to the H% boundary tone at the intonation phrase boundary.

Utterance ≪insert≫ illustrates the next lower level of disjuncture, that between two intermediate phrases that are grouped into one intonation phrase. In the second production of the sentence *'I' means insert*, there is a fall from a H* nuclear accent into a L- phrase accent, but there is no subsequent boundary tone, since this in not an intonation phrase boundary.

```
EXAMPLE <<insert>>:   'I' means insert.
in two productions 1) H*          H* L-L%
                        1     1       4
                   2) H* L-       H* L-L%
                        3     1       4
```

Note that the first production of the sentence in ≪insert≫ contrasts with this second production in its stress pattern in the same way as the first and third productions of ≪made1≫. The notion of nuclear accent is defined relative to the intermediate phrase. The contrasting productions in ≪made4≫ illustrate the same contrast in one versus two nuclear accents with L* pitch accents and a H- phrase accent at the boundary between the two intermediate phrases in the production with two nuclear accents. (The *? on the *made* in the first production illustrates a very common type of ambiguity about accent placement that is discussed below in Section 2.9.)

```
EXAMPLE <<made4>>:     Marianna made the marmalade.
in two productions 1)      L*   *?       L*    H-H%
                               1    1   1          4
                   2)      L* H-          L*    H-H%
                               3    1   1          4
```

1.5 What lines up with what?

The conventions for placing labels when using the *waves*^{TM} labelling system are prescribed in the ToBI Annotation Conventions so that labellers can use tools such as John Pitrelli's checker program to check for inadvertent omissions and grammatical errors. To quickly summarize, the break index label is placed at or just after the word label. Phrase accent and boundary tone labels are placed on or just before the corresponding 3 or 4 break index label. Pitch accents are placed somewhere within the accented syllable, preferably within the interval that can be identified with the syllable's vowel.

In the non-*waves*^{TM} transcription conventions, the orthographic, tone, and break index labels are ordered within each line so that such a transcription could be generated fairly quickly by merging and sorting a set of *waves*^{TM}-format label files.

2 More on the tone tier

2.1 Tones and fundamental frequency

As noted above, one of the basic parts of a ToBI transcription for an utterance is an electronic or paper record of the fundamental frequency contour. The transcription of events on the tone tier is closely linked to this record. In the case of pitch accents, the labeller can make this link explicit by choosing to place the label for the pitch accent specifically at the f0 maximum or minimum that realizes the starred tone of the accent, if this f0 event is within the interval of the accented syllable nucleus. (If the maximum or minimum does not actually occur within the syllable nucleus, there are optional conventions for marking the maximum or minimum as well as the accented syllable using the symbols '<' and '>', for a late or early f0 event, respectively — see Section 4.2 in "The ToBI Annotation Conventions".) There is a more practical connection, as well, inasmuch as most transcribers find the fundamental frequency contour an invaluable aid in making the analysis of the intonation pattern that is embodied by the transcription on the tone tier.

In interpreting the f0 contour to make the tonal transcription, it is important to keep in mind that several non-tonal aspects of an utterance can also strongly influence the fundamental frequency pattern. One of the most ubiquitous of these influences is the way in which consonant segments in the utterance interrupt the smooth course of the f0. Voiceless stops such as [p] and [t] and voiceless fricatives such as [f] and [s] create 'holes' in the f0 contour just by being voiceless. Moreover, it is not possible usually to read the intended pitch during a voiceless consonant by interpolating from the last f0 value before voice offset to the first f0 after voice onset because obstruent consonants (stops, fricatives, affricates) all cause dramatic perturbations in the fundamental frequency contour over and above any interruption of voicelessness per se. As an 'intrinsic' characteristic of its voiceless specification, a voiceless obstruent is usually associated with a dip into the consonant constriction and a dramatic fall starting from a much higher frequency just after the consonant release. Even voiced obstruents disturb the f0 contour; a voiced stop or fricative can be associated with a fall into and rise out of an often quite-deep valley during the

consonants constriction. Utterance ≪blond-baby1≫ illustrates some of these effects. There is a dip in the f0 around 1.9 s into the file for the [d] at the beginning of *difference* and the sharp fall around 5.29 s right after the [p] in *pink*. (To be sure, the perturbation caused by the [p] here is very small compared to many cases of voiceless obstruents that we have seen.)

```
EXAMPLE <<blond-baby1>>: what's the difference among my long memory
                           H*      !H*     L-L%    L+H*     !H* H-H%
                         your blond baby    and the pink carpeting
                        L+H*     *? !H* L-H% L*      L*   H*   L-L%
```

In interpreting such 'intrinsic' segmental effects, it is important to note the actual voicing of the consonant, and not simply its phonemic status. For example, phonemically voiced obstruents in stressed syllable initial position for many speakers are not always really voiced. Note, for example the /b/ of *blond* at about 3.95 to 4.0 s in ≪blond-baby1≫, which is voiceless unaspirated and has f0 perturbing characteristics more like those of the /p/ of pink. Also, the consonant /t/ in American and Australian English is usually a voiced flap (a short [d]-like segment) when it begins an unstressed syllable, as in the /t/ of *carpeting* in example utterance ≪flap2≫. Similarly, /h/ is often voiced between vowels. Thus, the perturbation caused by these two phonologically voiceless consonants is often like that for a /d/ or a /v/, rather than like a true [t], as shown by the /h/ in example utterance ≪voiced-h≫ (at around 3.04s).

```
EXAMPLE <<flap2>>:      The pink carpeting.
                            H*   H*    L-L%
EXAMPLE <<voiced-h>>:   Give him a hand with that.
                                  H*         L-L%
```

Example utterance ≪flap≫ gives another environment where flapping is common; see the flapped /t/ across the word boundary at around 1.34s. The flapping here is also important for transcribing break indices (see Section 3.2).

```
EXAMPLE <<flap>>:       Don't hit it to Joey.
                        H*              L*+!H L-L%
```

Another kind of problem in interpreting the f0 contour comes from shifts into voice qualities other than normal modal phonation. For example, for most speakers, subglottal pressure falls very sharply at the very end of an utterance. If the cross-glottal pressure difference becomes very weak, there may no longer be good glottal closings — i.e. the phonation may become quite breathy — so that even fairly robust pitch-tracking algorithms can easily fail. For some speakers this switch to breathy voice might happen even earlier if the utterance has a long low-pitched stretch corresponding to a L- phrase accent. Or, a speaker might break into creaky voice in such a region. In fact, many speakers break into creaky voice in almost any region with very low fundamental frequency. Since creaky voice is typically characterized by very irregular glottal periods (i.e., the fundamental frequency is physically not well-defined), pitch-tracking algorithms often do not do well during these portions of the utterance, creating a messy 'spattering' of values, like that seen in the f0 trace between 4.95 and 5.08 seconds in ≪jam2≫. Here the creak is due to the L*.

```
EXAMPLE <<jam2>>:       Will you have marmalade, or jam?
                                    L*     H-     L* H-H%
```

The pitch tracker can also completely fail, and give no f0 values at all, as in the region of the L- in the second production of ≪made1≫ after about 3.4s.

```
EXAMPLE <<made1>>:     Marianna made the marmalade.
second production    2) L+H*                   L-L%
```

In these two examples (as in many other occurrences of the same tone types in many of the example utterances in this labelling guide), the creaky voice is reliably interpreted by native speakers as a very low pitch value for some low tone. However, creaky voice does not automatically mean a very low L tone. Creaky voice can also occur as one common manifestation of a glottal stop, a segment which in English often occurs phonetically as a way to set off a word beginning with a stressed syllable that has no onset consonant. For example, the word *airline* in ≪glottal-stop≫ begins phonetically with a glottal stop realized as creaky voice.

```
EXAMPLE <<glottal-stop>>: And set training and experience standards
                              H*   H*    H-      H*        H*    L-L%
                          for airline inspectors and mechanics.
                              H*         H*   L-        H* L-L%
```

Nor are breathy voice and creaky voice the only source of pitch-tracking errors. Even in parts of the utterance with normal modal voicing, pitch-tracking algorithms can sometimes go wrong because of fluctuations of amplitude or because of the vowel's resonance characteristics. A perfectly ordinary period-to-period oscillation in amplitude can cause a halving of the estimated fundamental frequency value, as illustrated in the region between 4.8 and 4.93 seconds and again between 5.3 and 5.45 seconds in example utterance ≪pitch-halving≫. (Compare this to ≪no-pitch-halving≫, which is exactly the same utterance, pitch-tracked with somewhat different assumptions about the signal parameters which the pitch-tracking program uses in its consistency-checking algorithm.) Or, if the first formant is much higher than the fundamental, the pitch tracking program might take the amplitude of harmonics that it amplifies as an intervening glottal pulse, effectively doubling the pitch, as in the region between 14.07 and 14.18 seconds in example utterance ≪pitch-doubling≫. Transcribers must therefore learn when to trust their ears to catch such misparsings in the fundamental frequency track (or to use an alternate record of the fundamental frequency contour, such as the narrow-band spectrogram). When all of these perturbing effects are taken into account, however, the fundamental frequency contour becomes a valuable aid in transcribing the events on the tone tier.

```
EXAMPLE <<pitch-halving>>:      Jim builds a big daisy-chain.
                                    H*           H*        L-L%
EXAMPLE <<no-pitch-halving>>:   Jim builds a big daisy-chain.
                                    H*           H*        L-L%
EXAMPLE <<pitch-doubling>>:     Then I don't know if I can explain
                                      H*                    L+H*
                                it to you.
                                       L-L%
```

2.2 Some familiar contours, and the contrast between H* and L+H*

The inventory of events that are transcribed on the tone tier are five pitch accents, two phrase accents, and two boundary tones (plus downstepped counterparts of pitch accents and phrase accents with H tones). The summary statement in Appendix A lists the symbols for all of these tones and defines their use. In the previous sections, we already illustrated several familiar intonation patterns involving these tones. For example, the first productions of the sentences in example utterances ≪made1≫ and ≪insert≫ were instances of the 'declarative contour' which is an intonation phrase containing one or more H* pitch accents and ending in a sequence of L- phrase accent and L% boundary tone — i.e. (H*) H* L- L%. (When there is more than one accent, particularly when there is one relatively early and one relatively late accent, this contour is often called the 'hat pattern'.) The last production in utterance ≪made1≫ illustrated a sequence of L- phrase accent followed by H% boundary tone that is sometimes called the 'continuation rise'. The first production in utterance ≪made4≫ was an example of the 'yes-no question contour', consisting of one or more L* accents followed by a H- phrase accent and H% boundary tone — i.e. (L*) L* H- H%.

The productions in ≪made1≫ and the first two productions in ≪made3≫ also illustrate one of the more difficult contrasts in pitch accent type — that between the two types of 'peak accent' in which the peak is timed to occur on the accented syllable (H* versus L+H*). These two pitch accents are alike in that both have high fundamental frequency targets timed to occur on the accented syllable. They are alike also in that the actual timing of the f0 peak that realizes the high tone can vary depending on the phonetic length of the syllable and on the neighboring tones. In longer syllables just before a L- phrase accent, the peak tends to come fairly early in the syllable, whereas in short syllables with no immediately following tone target, the peak for the high tone can be quite late, sometimes after the actual acoustic end of the syllable. This is illustrated in the hat pattern utterance in ≪word1≫. The peak for the high tone of the first H* on *word* comes rather late (in the last third of the syllable), whereas the peak for the high tone of the second H* comes very early in *word* before the L- low tone target (during the first quarter of the syllable). How then do the two pitch accents differ?

```
EXAMPLE <<word1>>:      Your word is your word.
                             H*           H* L-L%
```

The essential difference is what happens before the high tone. The leading L tone in L+H* is meant to transcribe a rise from a fundamental frequency value low in the pitch range that cannot be attributed to a L* pitch accent on the preceding syllable or to a L- phrase accent or L% boundary tone at a preceding intermediate-phrase or intonation-phrase boundary. For H*, by contrast, there is at most a small rise from the middle of the speaker's voice range (unless, of course, the H* follows soon after some low tone such as a L* pitch accent or L- phrase accent). Example utterance ≪won≫ is a minimal pair illustrating this contrast.

```
EXAMPLE <<won>>:           Marianna won it.
in two productions      1)    H*      L- L%
                        2)  L+H*      L- L%
```

In the English intonation system as described by Pierrehumbert and Hirschberg 1990, H* and L+H* have distinct meanings, which make the latter more likely to occur in a contrastive context such as the one evoked by the second production of the sentence in ≪made1≫. In theory, this contrast between H* and L+H* can occur anywhere within a phrase. However, the distinction is difficult to make when the accented syllable is the first in the utterance, as in the second production of the sentence in ≪anna≫. These three productions are examples of almost exactly the same patterns as exemplified by the three productions in ≪made1≫. However, because the word *Anna* has no unstressed syllables before the main stressed one, it is difficult to realize the low tone for the nuclear accent on the first word in the second production. In cases such as this, where the evidence for L+H* comes from (theory-dependent) intuitions about meaning rather than from any clear low pitched region in the fundamental frequency contour, the ToBI Annotation Conventions prescribe H* instead. (The *? on the *married* in the first production illustrates a very common type of ambiguity about accent placement that is discussed below in Section 2.9.)

```
EXAMPLE <<anna>>:         Anna married Lenny.
in three productions 1) H*     *?     H* L-L%
                     2) H*               L-L%
                     3) H* L-H%     L+H* L-L%
```

Even when there is a long enough stretch between the beginning of the utterance and the accent, L+H* can be difficult to distinguish from H* because the categorical distinction in meaning is not always matched by a categorical distinction in the f0 level of the low tone. (The mapping of phonetic continua onto discrete oppositions is a well-known problem in segmental phonology as well.) Utterance ≪made2≫ above illustrates this. The L tone of the L+H* in the third production is not so low as that in the second production. When such utterances are taken out of context, it is possible for even intonational experts to be confused, and in fact, another transcriber with long experience in transcribing English pitch accents questioned our transcription of this as L+H*. (We are confident in the transcription, and did not mark it as X*? — see Section 2.9 below — but only because we know the context.)

The last productions in ≪made1≫ and ≪anna≫ are very similar to another type of contour where one needs to be especially careful in choosing between H* and L+H*. In both of these sentences, the nuclear stress for the second intonation phrase occurs late enough that the low-pitched region of the L+H* (nuclear) pitch accent could be distinguished even if there were no H% boundary tone intervening between the L+H* pitch accent and the L- phrase accent for the preceding phrase. In the very similar contours of example utterances ≪noone≫ and ≪for-marianna≫, on the other hand, there is no H% boundary tone, and one must play close attention to the timing in order to decide whether the accent in the second phrase should be transcribed as H* or L+H*. (Note that the first utterance in ≪for-marianna≫ also probably illustrates grouping at the level of the intermediate phrase and not a full intonation phrase; see Section 2.4 for the difficulty of telling these levels apart in this context).

```
EXAMPLE <<noone>>:          But Marianna knows noone.
                                 L+H* L-L%   L+H* L-L%

EXAMPLE <<for-marianna>>: 1) That one's for Marianna.
                               H*     L-       L+H* L-L%

                          2) Give me the brown one for Marianna.
                             H*           H*  L-L%        H* L-L%
```

The first response alternative in example utterance ≪mother4≫ illustrates another idiomatic intonation contour which might be confused with L+H*. This is the 'surprise-redundancy' contour described by Sag and Liberman 1975. Here the preceding low pitched region comes from a L* pitch accent on a prenuclear accented word. The second response alternative shows the subtle way in which this rising sequence differs from L+H*. The simple interpolation from the L* to the H* is more gradual than the steep rise within the L+H* accent, although the difference can be very subtle when there are only a few syllables between the two accents in the L* H* sequence, as it is here.

```
EXAMPLE <<mother4>>:  Who's it for?     Mary's mother.  It's for Mary's mother.
                       L*      H* L-L%  L*     H* L-L%           *?   L+H* L-L%
```

PRACTICE ONE

H* versus L+H*, L* H* L- L%, L* H- H%, and other accents in familiar contours

Transcribe these exercises using the `exercises` script.

EASY:

EXERCISE ≪amelia-p2≫:	*Amelia.* (two productions)
EXERCISE ≪mother1≫:	*Marianna's mother.*
EXERCISE ≪mole1≫:	*A new mole.*
EXERCISE ≪anna1≫:	*Anna married Lenny.* [Compare to last production in ≪anna≫]
EXERCISE ≪thought≫:	*That's what I thought.*
EXERCISE ≪lazy≫:	*He's lazy and crazy and stupid.*

INTERMEDIATE:

EXERCISE ≪memphis1≫: *Are you going to visit your mother when you're in Nashville?*

EXERCISE ≪memphis2≫: *My mother lives in Memphis.*

EXERCISE ≪heavy-rain≫: *Heavy rain possible. High around 70.*
[Transcribe only the second sentence for now, concentrating on the *seventy*.]

EXERCISE ≪wellies1≫: *Are you gonna wear your wellingtons?*
[Concentrate on the nuclear accent and following tones. (We know that there must be a prenuclear accent of the same type as the nuclear one, but we're not sure where it is.)]

EXERCISE ≪eileen-leaving≫: *Eileen is leaving.* (two productions)

EXERCISE ≪tree1house≫: *My classmate who lives in a treehouse was written up in Atlantic.*

DIFFICULT:

EXERCISE ≪dream≫: *So, what did you dream?*
[Don't worry about the tune on *So* for now.]

EXERCISE ≪thermometer≫: *Keep the thermometer under your tongue.*
[Transcribe only the *under your tongue* for now.]

EXERCISE ≪fail1≫: *So, a lotta times they fail.*
[Concentrate on transcribing the pitch accents here, and don't worry about transcribing the *So*.]

EXERCISE ≪happens≫: *And what happens is, when you...*
[Concentrate only on the first clause (before *when*).]

EXERCISE ≪I-mean≫: *You know what I mean?*

EXERCISE ≪anyway≫: *But anyway, if you can't see that then I don't know if I can explain it to you.*
[Note that the f0 tracker has doubled the pitch on the word *can*, and that your transcription of the tones nearby should take this into account.]

2.3 The timing of the phrase accent and 'upstep'

The examples so far also illustrate two important points about the phrase accent. First the phrase accent is unlike the boundary tone in that it is not necessarily localized at the phrase edge. Rather, when the nuclear accent is far from the end of the intermediate phrase, the phrase accent fills in the space in between it and the phrase edge, creating a long flat valley for L- realized over a long stretch, as in the first speaker's production in example utterance ≪names≫ (the region between 11.06 and 11.64 seconds), or a long plateau-like region for H- realized over a long stretch, as in the second speaker's production in example utterance ≪names≫ (the region between 13.3 and 13.9). Second, the H- phrase accent triggers an 'upstep' (a local raising of the pitch range to the end of the phrase), so that a following H% boundary tone is realized as a second rise at the end of the plateau-like region. This second point can also be seen clearly in example utterance ≪names≫ starting at 13.9 seconds. Compare the much lower f0 target for the H% boundary tone in example utterance ≪names≫ at 11.78, where the H% occurs after a L- phrase accent and therefore is not realized in such an upstepped pitch range.

```
EXAMPLE <<names>>:
        Anna may know my name, and yours too.      Anna may know our names?
       H*                 L-H%     H*    H* L-L%  L*                   H-H%
```

(Some experienced transcribers may want to call the pitch accent on *Anna* in the first sentence a L+H*. We remind them of the annotation guidelines that say in effect "Whenever there might be any doubt whatsoever, such as on the absolute utterance initial syllable, choose H* rather than L+H*.")

The summary statement on ToBI conventions prescribes that, in a *waves*^{TM} label file, the phrase accent (or phrase accent and following boundary tone) should be marked at a point at or just before the end of the last segment in the word ending the intermediate phrase (or full intonational phrase) and always before the related break index mark. The conventions say that the phrase accent should be placed here even when the nuclear accent occurs quite early and the phrase tone is realized over a long period of time, as in these two example utterances.

Note that when the nuclear accent is close to the end of the intonation phrase, it is impossible to discern any inflection point between the high f0 target for the H- phrase accent and the even higher f0 target for the upstepped H% boundary tone. The upstepped boundary tone after *jam* in example utterance ≪jam1≫ illustrated the smooth single rise that results in this case.

Example utterance ≪money≫ illustrates the full paradigm of combinations of phrase accent and following boundary tone that can occur at the end of an intonation phrase. Note that because of the upstep of the pitch range after the H- phrase accent, the L% boundary tone of a H- L% sequence does not have an absolutely low f0 target, just a lower one than that of the upstepped H% boundary tone. The contrast between a H- L% and a H- H% sequence is particularly salient when the preceding nuclear pitch accent is H*, as in the two sentences in example utterance ≪name1≫.

```
EXAMPLE <<money>>:  1) Is that Marianna's money?
                          H*     H*         L-H%
                    2) That's Marianna's money.
                             H*          L-L%
                    3) That's Marianna's money.
                             H*          H-L%
                    4) Is that Marianna's money?
                          L*     L*         H-H%

EXAMPLE <<name1>>:  My name is Marianna.
in two productions  1)            H* H-H%
                    2)            H* H-L%
```

The productions in ≪money≫ of the two combinations with final L% boundary tone illustrate another potential difficulty, and highlight the importance of listening to the speech and not just looking at the f0 record when doing the intonational analysis for the tone tier transcription. When an intonation phrase is not the last one in an uninterrupted stretch of speech and it ends with a L% boundary tone, it is difficult to distinguish from an intermediate phrase ending with the corresponding phrase accent just by examining the f0 contour. That is, the pitch differences between L- L% sequence and a mere L-, or between a H- L% sequence and a mere H-, are very subtle at best. Here the transcriber must rely on the subjective sense of degree of disjuncture, which is probably cued by such other things as the amount of preboundary lengthening or the degree of final lowering in the case of L- L% versus L-. (Note that the difference here must be transcribed also on the break index tier — Section 3.) Example utterances ≪park2≫ and ≪oregano≫ illustrate this difficulty. In ≪park2≫, looking just at the f0 contour, we might have L- L% and a full intonation phrase boundary or just L- and a mere intermediate phrase boundary between the nuclear H* pitch accents on *probably* and *pleasantest*. The ambiguous durational cues (two experienced transcribers argued even over whether the boundary should be before or after the *the*) supports the notion that this must be the latter, a mere intermediate phrase boundary. By contrast, the strong sense of pause (caused by the lengthening on *and*?) in the tonally identical stretch between the accents on *shortest* and *probably* support a full intonation phrase. In ≪oregano≫ the two productions contrast the L* H- intermediate phrases typical of a list in the first production with the ambiguous H* H- or H* H- L% plateaus of the second.

```
EXAMPLE <<park2>>:      Definitely the shortest and probably the pleasantest
                        H*           L-  H*    L-L%  H*       L-  H*
                        way to go is through the park.
                               L-              L+H* L-L%
```

```
EXAMPLE <<oregano>>:    1) Let's see   I need oregano 'n marjoram 'n some
                           H*   H* L-L%        L*  H-    L*    H-
                           fresh basil okay?
                          L+H*  !H* L-  H* H-H%
                        2) Oh  I don't know    it's got oregano 'n marjoram
                          H* !H*      !H* L-L%           H*  H-    H*    H-
                           'n some fresh basil.
                                       H* H-L%
```

The f0 patterns on *oregano* and *marjoram* illustrate also another difference between the H- and L- phrase accent, particularly in the contexts of unlike tones on the preceding nuclear accent — i.e. in the context of L* H- versus, say, H* L-. When the nuclear accent is on an early syllable in the last word in the phrase, the L- of a H* L- sequence seems to kick in very immediately with a sharp fall that typically begins during or just after the accented syllable. In the analogous situation, the rise from a L* to a H- begins as early, but the f0 change is much more gradual. Here, for example, the f0 seems to be rising continuously from about a third of the way into the accented syllable all the way to the end of the phrase. In this case, there is no real inflection point leveling out into a plateau.

The last clause of the first production in ≪oregano≫ also illustrates anew the difficulty mentioned above in connection with utterance ≪mother4≫ in Section 2.2. What is the best analysis of the fall to a low level immediately after *marjoram* and subsequent rise to a high f0 on *fresh*? How can we distinguish, say, a sequence of L* H* from the L+H* that we have transcribed? One thing to note is that, since accented syllables must be stressed, other characteristics of a syllable must be compatible with a tonal analysis that puts a pitch accent on it. The words *and* (*'n*) and *some* here do not sound stressed at all. Both have been reduced to the point that they have syllabic nasals as their nuclei. This supports the analysis of L+H* on *fresh* over an analysis of L* H*, even though the fall from *marjoram* looks so much steeper than the gradual rise from *and* back up to the H tone on *fresh* that the f0 pattern may seem more compatible with a L* on *and*. Note, however, that there may be mistracking due to breathy voice on *and*. Also, the *some* shows a strong perturbation from the initial voiceless [s] that obscures how low the intended f0 is later in the syllable.

2.4 Difficult combinations of nuclear pitch accent and following phrase accent

In most of the examples so far, nuclear H* has occurred before L-, where the following fall in pitch makes it easy to discern the pitch accent, and nuclear L* has occurred only before H-, where it was easy to spot from the immediately following rise in pitch. But the choice of pitch accent type is independent of the choice of following phrase accent, and there is nothing to preclude H* from occurring before H- or L* before L-. The second production in ≪oregano≫ illustrated the first case of this 'stylized high-rise' contour (Ladd 1980), which is becoming more and more familiar to contemporary American English speakers.

The combination of L* and following L- is also not rare. There are two situations where this sequence is typically encountered. The first is illustrated in ≪nose≫, and the first sentence in ≪tags≫. This L* L- H% pattern is typical of such vocative tags. The second sentence in ≪tags≫ shows that tag questions can have this contour too. However, tag questions can also take a H* L- L% intonation pattern (the third sentence in ≪tags≫), which seems to be precluded on the vocative tag for pragmatic reasons (see Beckman and Pierrehumbert 1986).

```
EXAMPLE <<nose>>: Oh   don't nuzzle me you marmalade-nose.
                 X*? L- H*   !H*   L-      L*        L-H%
```

(Section 2.8 will explain the '!' diacritic in the second pitch accent, and Section 2.9 will explain the X*? accent on *Oh*.)

```
EXAMPLE <<tags>>:       1) Where are you going, Willy?
                                        H*  L- L* L-H%
```

```
2) He won't be going, will he?
      H*       H*  L- L*   L-H%
3) He won't be going, will he?
      H*       H*  L- H*   L-L%
```

The f0 contour for the L* L- H% vocative tag contour can be confused with a longish postnuclear stretch in the sequence H* L- L%, as shown in example utterance ≪vocative1≫. As with medial L- intermediate phrase boundary versus L- L% intonation phrase boundary discussed above, the transcriber may have to rely entirely on the subjective impression of greater versus lesser disjuncture to capture this difference between an intermediate phrase boundary at a vocative tag and no boundary. (Note again that the difference here must be accompanied by different symbols on the break index tier.)

```
EXAMPLE <<vocative1>>:  1) Anna will win, Manny.
                              H* L- L* L-H%
                        2) Anna will win Manny. (She won't lose him).
                           H*        H*    L-L%
```

The other situation in which one often sees a L* nuclear accent and following L- phrase accent is in the 'contradiction contour', an intonational idiom illustrated in ≪gloria≫ and ≪elephant3≫. This contour is discussed at length in Sag and Liberman 1975 and chapter 3 of Ladd 1980. The L* L- H% sequence starting at the nuclear syllable is like the contour in the vocative tag, but this is not the only essential component of this intonational idiom. Crucially, there must be a fall from an early prenuclear H* pitch accent (or from an initial %H boundary tone — see next section) onto a nearby L* accent. If the L* nuclear accented syllable is far from the beginning of the utterance (as is the case with the nuclear accent on *incurable* in ≪elephant3≫), there might be another L* on some prenuclear syllable with relatively prominent secondary stress (e.g., the fourth syllable of *elephantiasis* in ≪elephant3≫). Note that ≪elephant3≫ also illustrates the possibility of having two pitch accents on one word when there is more than one full stressed syllable (see Section 2.9 for more examples).

```
EXAMPLE <<gloria>>:     Ah Gloria  you're not ugly.
                        H*  L* L-L% H*        L* L-H%

EXAMPLE <<elephant3>>:  Elephantiasis isn't incurable.
                       H*      L*             L* L-H%
```

2.5 The initial %H boundary tone

The contradiction contour also illustrates another phenomenon that we have not discussed so far — namely, the possibility of a boundary tone marking the initial as well as the final boundary of an intonation phrase. Utterance ≪bananas≫ is an example. Here the event that provides the high pitch for the early fall onto the L* tone cannot be an accent, since the first syllable of *bananas* is reduced (i.e. completely unstressed and hence unaccentable).

```
EXAMPLE <<bananas>>: Bananas aren't poisonous.
                    %H L*           L*     L-H%
```

In the intonational analysis assumed in the ToBI system, the final boundary tone is mandatory, whereas an initial one is not. The initial boundary tone differs from the final ones also in that it seems to be limited to absolute utterance-initial position, and in that it is always high. Thus, unlike the final boundary, where there is a paradigmatic choice between L% and H%, the phrase-initial boundary tone contrasts merely with the absence of a boundary tone. That is, %H contrasts with the default (unmarked) initial pattern, which in absolute utterance-initial position tends to start in the middle part of the speaker's pitch range (as opposed to beginning of utterance-medial intonation phrases, where

the pitch simply continues from the value at which the previous phrase ended). This utterance-initial midrange pitch value is illustrated in ≪loan1≫, where the first and second productions show a rise from the mid value to H* and a fall from the same default mid value to L*, respectively. The third production then contrasts with the second in that it has an initial %H boundary tone. These two examples also illustrate the typical effect of having an initial %H boundary tone in the surprise-redundancy contour. The one with the initial %H has a greater vividness, conveying either more surprise or more insistence that this is the information that the hearer should really already know.

```
EXAMPLE <<loan1>>:       You need a loan.
In three productions  1)     H*     H* L-L%
                      2)     L*     H* L-L%
                      3)  %H L*     H* L-L%
```

The ToBI conventions prescribe that %H be an analysis of last resort. That is, like L+H*, which is used instead of H* only when there is no other possible explanation for the low pitch before the peak (see Section 2.2, above), %H is used only when there is no other plausible explanation for an initial high pitch. It should be marked only when a high-pitched beginning for an utterance cannot be attributed to a H* accent on the first few syllables in the utterance — i.e., when the first word itself does not appear to be accented or when its accented syllable occurs too far into the word to account for the initial high target. Thus it should not be used in ≪gloria≫, where the high pitch at the beginning of the phrase *You're not ugly* is attributed to a H* accent on the first syllable *You're*. In ≪elephant3≫, similarly, although the main stress of the word *elephantiasis* clearly is at the L* accent on the fourth syllable, we have the option of analyzing the earlier high pitch as another pitch accent earlier in the word, since the first syllable has a lexical 'secondary stress' (i.e. is rhythmically more prominent than the surrounding syllables).

PRACTICE TWO

Phrase accent and boundary tone contrasts

The following examples are practice utterances for the phrase accent and boundary tone contrasts discussed in the last few sections. Transcribe these exercises using the **exercises** script.

EASY:

EXERCISE ≪manitowoc≫:	*Does Manitowoc have a bowling alley?*
EXERCISE ≪mother2≫:	*For Marianna's mother.*
EXERCISE ≪cream≫:	*Would you like some cream?*
EXERCISE ≪wellies2≫:	*No, I think I'll wear my hiking boots.* [Don't worry about transcribing the tune on *No* for now.]
EXERCISE ≪voice≫:	*You lost your voice.*
EXERCISE ≪flour2≫:	*Oh nothing special, you know flour and butter and sugar.* [Transcribe just the second part, after the *you know*.]
EXERCISE ≪audience1≫:	*Good evening radio audience.* [Don't worry about the transcription of *radio*.]
EXERCISE ≪good2≫:	*I thought it was good.*
EXERCISE ≪legumes1≫:	*Legumes are a good source of vitamins, and of protein as well.*
EXERCISE ≪legumes2≫:	*Legumes are a good source of vitamins, but not the best.* [Transcribe only the first part, up through *vitamins*.]
EXERCISE ≪legumes3≫:	*Legumes are a good source of vitamins, and so are greens.* [Transcribe only the first part, up through *vitamins*.]
EXERCISE ≪stalin≫:	*I was wrong, and Stalin was right. I was wrong.*

INTERMEDIATE:

EXERCISE ≪friend1≫: *A friend of mine um works for NASA.*

EXERCISE ≪good1≫: *I thought it was good?*
[Play ≪good2≫ for contrast.]

EXERCISE ≪pigs≫: *They've eaten the pigs.* (two productions)

EXERCISE ≪flour1≫: *I need flour and sugar and butter and oh I don't know.*
[Transcribe only the part up through *butter* for now.]

EXERCISE ≪atlanta≫: *Yes I would uh like the information on the flight leaving from uh Philadelphia to Atlanta.*
[Concentrate just on the parts *like the information* and *Philadelphia to Atlanta.*]

EXERCISE ≪good-aft≫: *Good afternoon. Information Services.*

EXERCISE ≪knock-stuff≫: *Mostly they just sat around and knocked stuff. You know, the school, other people.*
[Concentrate for now on the second sentence, starting at *You know...*]

EXERCISE ≪drive≫: *I'm not going to drive to school today.*

EXERCISE ≪spoon1≫: *There's a spoon in here.*

DIFFICULT:

EXERCISE ≪mother3≫: *I've told you a million times! It's for Mary's mother.*
[Compare ≪mother4≫; don't agonize too much over the tones around *for* in the second sentence.]

EXERCISE ≪experience1≫: *Well I mean, would you hire somebody that doesn't have no experience?*

EXERCISE ≪trafficlight≫: *That's right at the traffic light.* (two productions)

2.6 Pitch accent timing, and the L*+H pitch accent

The examples so far have illustrated both possible phrase accents, both boundary tones, and three of the five types of pitch accent — the low accent (L*), the plain 'peak accent' (H*), and the 'rising peak accent' (L+H*). We have also discussed the timing of the f0 peak in the two types of peak accent, pointing out that it is somewhat variable; in particular, that it occurs somewhat later relative to the segments of the accented syllable when it is the accent at the beginning of a 'hat pattern' contour and relatively earlier before L- (see Section 2.2). Such differences in timing are not distinctive, and seem to be related to the phonetics of pre-boundary lengthening. For example, we might think of the relatively earlier placement of the peak in the latter case as a matter of lengthening the part of the syllable after the nuclear pitch accent peak in order to accommodate the L- phrase accent within the intermediate phrase (see, e.g., Silverman and Pierrehumbert 1989, for a discussion of such phonetic accounts). These phonetic differences in timing can be ignored in the transcription on the tone tier.

There is another difference in the timing of apparent peak accents, however, that must not be ignored, because it is distinctive. Both the small rise from mid pitch that is usually seen with an utterance-initial H* accent and the definitive rise from low pitch that is necessarily seen to transcribe a L+H* accent contrast phonologically with another accent type that involves a rise from low pitch into a peak that occurs much later, making the low tone align with the accented syllable. This is the 'scooped' accent L*+H, illustrated in the first production of ≪millionaire≫. The second production in this example utterance is of the contrasting 'rising peak' accent L+H*. These two pitch accents have very different meanings, as described by Ladd 1980 and Ward and Hirschberg 1985, and the difference in timing here is a phonological difference that is represented in the ToBI system by the contrasting specifications of L*+H versus L+H*. That is, phonologically, both of these accents are a L plus a H, but in the 'scooped' accent, the L is the starred tone (associated to the accented syllable) rather than the H. The associated phonetic

difference is that the rise is much later in the 'scooped' accent, and it is the timing of the minimum f0 relative to the segments of the associated syllable that is salient.

```
EXAMPLE <<millionaire>>:    Only a millionaire.
in two productions      1) H*      L*+H    L-H%
                        2) H*    L+H*      L-H%
```

Because it is the L target in the 'scooped' accent that is associated to the stressed syllable, and not the H, the high pitch target is specified only as occurring somewhat later than the L, and the timing of the peak f0 relative to the segments is not controlled. If the stressed syllable is long, the rise to the peak might be accomplished entirely within the accented syllable. But if the stressed syllable is short, the peak may occur one or more syllables later. This is illustrated in example utterance ≪stein≫, which shows a relatively fixed rise relative to the low f0, which makes the peak occur within the last part of the long syllable *Stein* but two syllables later relative to the short accented first syllable in *rigamarole*.

```
EXAMPLE <<stein>>: Stein's not a bad man.
                   L*+H             L-H%
                   Rigamarole is monomorphemic.
                   L*+H                   L-H%
```

Note that although the crucial difference between L+H* and L*+H is the timing of the low pitched portion, some speakers produce a secondary difference, whereby the L of L*+H is consistently somewhat lower in the pitch than the L of L+H*. This is particularly apparent in the first accents of the two productions in ≪bloomingdales≫. This also means that L*+H is not nearly so confusable with H* as is L+H*. There is also considerable interspeaker differences. Some speakers have rather mid-level L tones even in L*+H. (This fact will be relevant when you transcribe ≪noodle1≫ and ≪noodle2≫ in PRACTICE THREE.) Other speakers have very low L tones even in L+H*. This does not affect the relative heights of the L's in L*+H versus L+H*.

```
EXAMPLE <<bloomingdales>>: There's a lovely one in Bloomingdale's.
in two productions:        1)        L*+H           L*+!H     L-H%
                           2)      L+H*          L+!H*        L-L%
```

Another thing to note in the sequence of accents in these two productions is the introduction of another set of symbols for the second (the nuclear) accents. These new symbols are actually the same accent types as the first accent in their respective utterances. The extra '!' in the symbols for the nuclear accents is a diacritic to denote the way in which the second accent peak is lower than the preceding peak. This lowering of the second peak is due to a process called 'downstep', which is defined as a categorical compression of the pitch range that reduces the f0 targets for any H tones subsequent to the specification of the downstep — i.e. the counterpart of the 'upstep' triggered by the H-. We will describe downstep and what triggers it in more detail in Section 2.8 below after introducing the last remaining pitch accent type in the ToBI analysis, transcribed as H+!H*.

Finally, as in deciding how low the f0 must be to count as L+H* rather than H*, transcribers should be aware of slight interspeaker differences in the timing of the L tone in differentiating L+H* from L*+H. Our impression is that American speakers (such as the speaker of ≪bloomingdales≫) do not always make L*+H rise as late as most RP British speakers do. The second (downstepped) L*+!H on the word *Bloomingdale's* in the first production, in particular, might seem quite early to a British transcriber. Note, however, that there is a very low pitch level throughout the [b] and the [l], and the f0 does not begin to rise until the voicing begins in the [u], making the peak occur considerably after the [m] release. This is quite late for a nuclear L+H* before a L- (cf. our comments above in Section 2.2.), as can be seen by comparing this rise to the rise in the comparably downstepped nuclear L+H* in the second production. In the second production, the rise begins before the [b] and is completed well before the release of the [m].

2.7 The H+!H* pitch accent

The nuclear accent in the second production in example utterance ≪theresa≫ illustrates this pitch accent type. It is characterized by a fall from a preceding higher pitch onto a lower pitch level on the accented syllable. This accent type corresponds to the type called H+L* in Pierrehumbert's original system. The substitution of the letters '!H' for 'L' in the name of the pitch accent reflects the fact that the pitch target on the accented syllable is only somewhat lower than the preceding H tone target; it is not so low as the f0 target for the plain 'low' accent (L*) or for the L tone of the 'scooped' accent (L*+H), or even for the L of the 'rising peak' accent (L+H*). The renaming of this pitch accent type was intended to make the analysis somewhat more concrete and intuitive for the transcriber.

```
EXAMPLE <<theresa>>:  You want an example? How about Mother Theresa?
                           H*       H* H-H% H*          *?       H* L-L%
                                  You want an example? Mother Theresa.
                                     H*       H* H-H%        H+!H* L-L%
```

(The *? in the first production indicates uncertainty about whether that word is accented — see Section 2.9.)

2.8 Downstep

Downstep is a phonologically triggered compression of the pitch range that lowers the f0 targets for any H tones subsequent to a downstep trigger. In Pierrehumbert's model of intonation, downstep is said to be triggered by any bitonal pitch accent. In example utterance ≪bloomingdales≫ discussed above, for example, the progressive reduction of the second L+H* or L*+H peak relative to the preceding one would be analyzed as an automatic consequence of the fact that these two pitch accent types are composed of two tones, L plus H.

In the ToBI system, this compression of the pitch range is marked by having alternative names for accents which are used for the first downstepped high tone target after the downstep trigger. Thus in the first production in example utterance ≪bloomingdales≫, the second 'scooped' accent is transcribed with L*+!H rather than L*+H to denote that a downstep has occurred. And similarly in the second production in this example utterance, the second 'rising peak' accent is transcribed with L+!H* rather than L+H*. When there are more than two such bitonal accents in a row, each accent triggers another instance of downstep, so that each subsequent accent peak is reduced yet again relative to the immediately preceding one. This is illustrated in example utterance ≪yellow2≫. Example utterance ≪calling≫ shows that it is not just pitch accents which are affected by downstep. The !H- phrase accent here is reduced to a mid level by the downstep triggered by the preceding L+H* nuclear pitch accent. (Note the characteristic mid-tone tail, as the downstepped !H- phrase accent triggers a subsequent upstep of the L% boundary tone.)

```
EXAMPLE <<yellow2>>: There's a lovely yellowish old one.
                        H*     L+H*  L+!H*    L+!H*   L-L%
EXAMPLE <<calling>>:    Marianna.
                          L+H* !H-L%
```

Transcribers who are familiar with Pierrehumbert's system will recognize that ToBI differs in this explicit marking of the reduced pitch range directly on the first H tone affected by the downstep. In Pierrehumbert's system, downstep is not explicitly marked because it is redundant to the specification of the trigger in the preceding bitonal accent.

Pierrehumbert's system differs from ToBI in yet another way; it includes a sixth pitch accent type, H*+L, which bears the same relationship to H+L* (ToBI's H+!H*) as L*+H does to L+H*. That is, the fall to a slightly lower pitch target occurs after the accented syllable instead of into the accented syllable. Typically, the endpoint of this fall is no lower than the pitch target of a subsequent downstepped H tone, and the contrast between H* and H*+L thus hinges on recognizing the downstep triggered by the H*+L.

Many first time transcribers find this comparatively abstract analysis unintuitive and therefore difficult. In the ToBI system, therefore, we have eliminated H*+L in favor of marking the downstep directly

on the first reduced H tone. Thus H* in ToBI corresponds to both plain H* and the downstep triggering H*+L. Users of databases transcribed with the ToBI system who need to analyze the data in terms of the intonational categories in Pierrehumbert's system, can recover each H*+L tone by searching for a downstepped !H* or !H- marked immediately after a H* (or !H*) accent. For example, in utterance ≪really1≫ the second production is a plain 'hat pattern' (H* H* L- L%) whereas the first is a 'downstepped hat', which would be transcribed as H*+L H* L- L% in Pierrehumbert's system. The second production in utterance ≪calling2≫ illustrates another very familiar intonation pattern, the 'calling contour', which in Pierrehumbert's system would be transcribed with H*+L H- L%.

```
EXAMPLE <<really1>>:      That's really illuminating.
in three productions    1)        H*     !H*    L-L%
                        2)        H*      H*    L-L%
                        3) Transcribe this one in PRACTICE THREE

EXAMPLE <<calling2>>:       Anna.
in two productions      1) L*  H-H%
                        2) H* !H-L%
```

A fact to note about downstep is that it is local to an intermediate phrase. Each new intermediate phrase represents a new paradigmatic choice of pitch range, at which downstep can be reset. This is illustrated in ≪yellow3≫, where the intermediate phrase boundary after the *yellowish* allows a new choice of pitch range, so that the peak on *old* is not downstepped relative to that on *yellowish*, unlike in ≪yellow2≫. (See Section 2.9 to read about the X*? symbol marking the peak on *old*. It indicates that there is an accent on *old* but we are not completely certain which type of accent it is. It could be simply H* after an unexpectedly steep rise from the preceding L-. Or it could be L+H*, with a less steep rise than expected.)

```
EXAMPLE <<yellow3>>: It's lovely and yellowish, and it's an old one.
                          L+H*      L+!H*    L-              X*?  L-L%
```

Note, however, that the peak on *yellowish* looks downstepped relative to that on *lovely*. This is due to the relationship between the pitch ranges chosen for the two intermediate phrases. The topic structure of a discourse is marked in part by the choice of pitch range for the succession of intermediate phrases; large topics begin with expanded pitch range and end with very reduced pitch range (see, e.g., Brown et al. 1980; Hirschberg and Pierrehumbert 1986). This often creates an effect of 'paragraph intonation' in which the relationship among successive phrasal pitch ranges mimics the phrase-internal relationship between preceding peaks and subsequent following lower downstepped peaks. The successive pitch ranges in utterances ≪park1≫ through ≪park5≫ in PRACTICE TWO above illustrate 'paragraph intonation' over a longer time frame.

Sometimes it is not easy to tell the difference between two phrase-internal accents with the second downstepped relative to the first and two intermediate phrases with the second phrase in a lower pitch range relative to the first. Example utterance ≪levels≫ illustrates such a difficult case.

```
EXAMPLE <<levels>>: There are many intermediate levels.
                            L+H*     L+!H*   L+!H* L-L%
```

PRACTICE THREE

L*+H, H+!H*, and downstep

Transcribe these exercises using the **exercises** script.

EASY:

EXERCISE ≪yellow1≫: *There's a lovely yellowish old one.*
[Compare to ≪yellow3≫.]

EXERCISE ≪really1≫: *That's really illuminating.*
[Transcribe third production now; first two are examples from 2.8.]

EXERCISE ≪eileen1≫: *Eileen's pro-English.*

EXERCISE ≪eileen2≫: *Eileen's pro-English.*
[Compare to ≪eileen1≫.]

EXERCISE ≪calling3≫: *Marianna.* (two productions)

EXERCISE ≪windy≫: *Becoming windy.*

EXERCISE ≪park1≫: *Okay to get from home to the station.*

EXERCISE ≪park4≫: *But uh in fact I have to go along the main road for a little ways it's probably about three hundred yards.*

EXERCISE ≪legumes2≫: *Legumes are a good source of vitamins, but not the best.*
[Repeated exercise from PRACTICE TWO. Now transcribe the second part, after *vitamins*.]

INTERMEDIATE:

EXERCISE ≪friend2≫: *A friend of mine works for NASA.*
[Compare to ≪friend1≫.]

EXERCISE ≪mile≫: *You give him an inch, he takes a mile.*

EXERCISE ≪really2≫: *That's really illuminating.* (three productions)

EXERCISE ≪park3≫: *It would be nice to be able to go right out the back door and into the park cause it's actually right behind the house.*

EXERCISE ≪flour1≫: *I need flour and sugar and butter and oh I don't know.*
[Repeated exercise from PRACTICE TWO. Transcribe only the part after *butter*.]

EXERCISE ≪sununu≫: *... and denies speculation that Chief of Staff, John Sununu, is meddling in the region's environmental affairs.*

EXERCISE ≪noodle1≫: *We have a lean mini-noodle with beans. Well, we have a lean mini-noodle dish.*

EXERCISE ≪noodle2≫: *We have a lean mini-noodle with beans. We have a lean mini-noodle dish.*

EXERCISE ≪romanelli≫: *John Romanelli, John Romanelli, please return to the ticket counter.*

EXERCISE ≪thatone≫: *Do you really think it's that one?* (two productions)

EXERCISE ≪word≫: *Your word is your word.*
[Compare to ≪word1≫.]

DIFFICULT:

EXERCISE ≪thatone2≫: *Do you really think it's that one?* (two more productions)
[Don't agonize too much over the tones around *Do you* in the second production.]

EXERCISE ≪heavy-rain≫: *Heavy rain possible. High around 70.*
[Repeated exercise from PRACTICE ONE. Transcribe the first sentence now, concentrating on the *rain*.]

EXERCISE ≪tree2house≫: *My classmate who lives in a treehouse was written up in Atlantic.*
[Compare ≪tree1house≫ in PRACTICE ONE.]

EXERCISE ≪knock-stuff≫: *Mostly they just sat around and knocked stuff. You know, the school, other people.*
[Repeated exercise from PRACTICE TWO. Concentrate now on the first sentence, the part before *You know...*]

EXERCISE ≪argument≫: *If he can then there's no argument about it.* (two productions)

EXERCISE ≪sublime1≫: *Sublime mnemonic rhyme and free meter.*

EXERCISE ≪sublime2≫: *Sublime mnemonic rhyme and free meter.*

EXERCISE ≪fail≫: *And what happens is: when you... when you buy my business, and you try to run my business, it's really hard for you to run my business. So a lotta times they fail.*
[Concentrate particularly on the *When you buy my business*, and don't worry about the preceding interrupted *when you...*]

EXERCISE ≪business≫: *A lot of people have done this; they sell their business, and they have... If something goes wrong, and they have the first rights to buy it back.* [Interviewer: *Oh, really?*]
[You've already transcribed parts of this in earlier practice sets. Here concentrate on filling in the missing pieces up through *something goes wrong*, leaving for now the *and they have...*]

2.9 Uncertainty about accent placement and accent type

In addition to conveying topic structure, pitch range variation is used for many discourse purposes. For example, a lower (or higher) pitch range than surrounding phrases can be used to set off a stretch of speech as an aside or a parenthetical. This is illustrated in example utterance ≪capote≫. Also, expanded pitch range can convey extra liveliness or involvement, as illustrated in the much larger pitch range on the phrase *Now be careful* in ≪onions≫.

```
EXAMPLE <<capote>>: Capote died Saturday at the Bellaire home of
                    L+H*       !H*    L-        H* H*   L+H*  L-
                     Joanne Carson    (estranged wife of talkshow host
                          L+H*   L-L%          L+H*   L- H*       *?
                     Johnny Carson), and she was among those who
                      H*         L-L%   L+H*     !H*   !H*
                     eulogised him.
                   H+!H*          L-L%
EXAMPLE <<onions>>:  Okay now chop the onions... Now be careful.
                    H* L- H*        H+!H*   H-   L*    L+H* L-H%
                     Okay,  chop the onions, and put them into that bowl.
                    L+H* L-L%     H+!H*   H-              H*      H+!H* L-L%
```

Because speakers can vary their pitch ranges seemingly without limit to convey discourse organization or degree of involvement, and because downstep can happen many times within a single phrase, sometimes

it is difficult to tell whether a tone is H or L, even when one is sure a tone is there. The accents on *smoke* and *yeah* in the asides in example utterance ≪smoke≫ illustrates this. Or, in very reduced pitch ranges, it can be difficult to tell whether a syllable is accented or not. Utterance ≪sold3≫ illustrates this. In the very reduced pitch range after the second downstep on *else*, it is impossible to know how many accents there are.

```
EXAMPLE <<smoke>>: Can I smoke?  <<interviewer says "You can smoke.">>
                          X*? H-H%
                    Yeah?  <<interviewer: "Does this door have to stay
                   X*? X-? X%?
                    open?" If it>> No, it doesn't have to be; you can close it.
EXAMPLE <<sold3>>: He sold it to somebody else, they bought the
                      H*          !H*      !H*          *?
                    whole company, and he made lots of money on
                            *?  -X?              *?     *?
                    the business...
                        H*     L-L%
```

In the first type of uncertainty, ToBI prescribes that the transcriber use the notation X*? to simply mark the clear presence of an accent, without forcing an arbitrary commitment to the accent's type. Thus, the accent on *smoke* should be transcribed as X*? rather than as L* or H*. (X*? should not be used to mark uncertainty between the L- and H- phrase accents or between the L% and H% boundary tones. There the transcriber should instead mark X-? for the phrase tones and X%? for the boundary tones.) In the second type of uncertainty, when the transcriber is not certain even that there is a pitch accent (as, for example, on the *bought* in utterance ≪sold3≫), the mark *? should be used instead of X*?.

In addition to very compressed pitch ranges, there are several particular tone sequences which are prone to inducing uncertainty about the presence of accent. One such case is the downstepped H* !H* !H* ... sequence just illustrated. In many cases, words after the first H* in such sequences are ambiguous between being accented with !H* and being 'deaccented' (i.e. being in the postnuclear low stretch in a H* L- L% sequence). This is not always ambiguous, however. Utterance ≪anna2≫ illustrates a clear contrast between downstepped and deaccented.

```
EXAMPLE <<anna2>>:     Anna married Lenny.
in two productions 1) H*                 L-L%
                   2) H*     !H*    !H* L-L%
```

Another case of inherently ambiguous tone sequences which occurs very commonly is when there is a long stretch of speech in a 'hat pattern' contour. Utterance ≪peel≫ illustrates this. The word *off* sounds very prominent, giving a strong subjective impression of accent. However, because the word lies in the plateau between the first H* on *peeled* and the nuclear H* pitch accent on *Hawaii*, it is difficult to tell whether *off* also bears a H* accent, or just the preceding *peeled*. The word *host* in the phrase *talkshow host Johnny Carson* in ≪capote≫, the word *married* in the first production of ≪anna≫, and the word *Mother* in the first production in ≪theresa≫ in Section 2.7 above are three more illustrations of this very common ambiguity. The first production in example utterance ≪made4≫ (given in Section 1.4 above) illustrates the analogous situation with a L*; *Marianna* probably has a L* accent (note the dip down into it from the mid pitch level that begins the sentence) and *marmalade* clearly has a L* accent, but what about the *made* in between?

```
EXAMPLE <<peel>>: [Ever since the roof of a 19-year old Aloha Airlines
                  Boeing 737] peeled off over Hawaii last April, ...
                               H*    *?         H* L-    H* L-L%
```

(See example utterance ≪older-aircraft≫ in the next PRACTICE for the whole context of this phrase.)

In cases such as these it is better to err on the side of conservatism and mark the word with *? or nothing. In particular, the transcriber should take care not to let grammatical expectations guide the

marking of accents. If we find ourselves giving in to such thoughts as "This is a content word and therefore probably is accented", we preclude the use of our transcriptions to test whether content words are indeed likely to be accented.

A final source of uncertainty is particularly true of transcribing sentences in isolation extracted from the context in which they originally occurred. This is uncertainty about accent type due to unfamiliarity with a particular speaker's normal speaking range for a particular style of speech. For example in utterance ≪hurt≫, the nuclear pitch accent on *hurt* is probably L*; the pitch is lower than the "neutral" value at the beginning of the utterance. However, 200 Hz is very high for a low, and unless one knows from experience that this speaker has a very high-pitched voice, one might be tempted to transcribe this utterance with a H* nuclear accent.

```
EXAMPLE <<hurt>>: But would it hurt you?
                          *?      X*?    H-H%
```

Utterance ≪beef≫ also illustrates this point. Here we have transcribed each of the accents in the second speaker's response as L*, since we know from many other examples in this labelling guide that this speaker's normal range is higher than the first speaker's voice. This utterance also exemplifies an intonation pattern we have not shown before: a sequence of all low tones, for all of the accents, the phrase accent, and the boundary tone.

```
EAMPLE <<beef>>: Here's your Chateaubriand, ma'am.
                  H*                 L+H* L-  L* L-H%
                 I don't eat beef.
                    L*    L*   L* L-L%
```

2.10 When something is accented that you would not expect

It is important also to not fall into the obverse reasoning and hesitate to mark an accent simply because the accented word is not a content word. Example utterance ≪AND1≫ is a nice illustration of this point. Here, the pitch pattern unequivocally supports a nuclear 'peak' accent on the *and*. There is a suggestion of accent on the same word in ≪hennessy≫, but less clearly; here the glottalization at the beginning of the word is the main cue that it is accented (recall that stressed vowel-initial syllables are set off by glottal stops — Section 2.1).

```
EXAMPLE <<AND1>>: ...design improvements, and a schedule...
                        H*      H*    L-H% H* L-   H*  L-L%

EXAMPLE <<hennessy>>: Hennessy is widely respected   for his legal
                       H*    L-    H*      !H*  L-L%       L+H* !H-
                      scholarship and his administrative abilities.
                         H*     H- H*             *?      H*     L-L%
```

Another thing to watch out for is that in very emphatic speech, a word with two fairly strong syllables can bear two pitch accents. Utterance ≪understand≫ is such a case, where our normal expectation is that only the most stressed final syllable in *understand* will be accented. Here, however, the first syllable is also accented, so that the phrase *to understand* can realize the 'surprise-redundancy' contour L* H* L-L%.

```
EXAMPLE <<understand>>: I'm simply trying to get you to understand.
                        H*   L* H-  L*     H- L*      H- L*   H* L-L%
```

Example ≪philadelphia≫ gives another case of such double accents, apparently without the impetus of realizing any particular intonational idiom.

```
EXAMPLE <<philadelphia>>: from Philadelphia to Dallas
                             L+H* !H*   L-    H*   L-L%
```

This phenomenon of double accents, and the apparently related phenomenon of 'stress shift', have been examined extensively by Stefanie Shattuck-Hufnagel and her colleagues (see Ross et al. 1992) in a corpus of newscasts which they have transcribed using something like the ToBI system. Some of the exercises involving this phenomenon are from this study. It is apparently fairly common.

PRACTICE FOUR

uncertainty about pitch accent placement or type

Transcribe these exercises using the **exercises** script.

EASY:

EXERCISE ≪legumes3≫: *Legumes are a good source of vitamins, and so are greens.*
[Repeated exercise from PRACTICE TWO. Transcribe the second part, after *vitamins*.]

EXERCISE ≪sold1≫: *He sold the business to somebody else*

EXERCISE ≪artwork≫: *State law now requires public construction projects to set aside 1% of their budgets for artwork.*
[Concentrate particularly on the part from *to set aside* on.]

EXERCISE ≪howto≫: *I know we've gotta do it but I don't know how to do it.*
[There isn't an intermediate phrase break between *how* and *to*. You'll learn how to transcribe sequences such as this in Section 3.]

EXERCISE ≪noodle4≫: *Do you have a lean mini-noodle dish?*

EXERCISE ≪noodle5≫: *D'you have a lean mini-noodle dish?*

INTERMEDIATE:

EXERCISE ≪butcher≫: *How'd your operation go? Don't talk to me about it; I'd like to strangle the butchers.*

EXERCISE ≪physicist≫: *He's a physicist 'n works at NASA.*

EXERCISE ≪environ1≫: *And Ballaga seems determined to stay the environmental course.*

EXERCISE ≪environ2≫: *plenty of room to flex environmental muscles*
[We had trouble on the accents around *environmental* too, so don't agonize over their type.]

EXERCISE ≪older-aircraft≫: *Ever since the roof of a 19-year old Aloha Airlines Boeing 737 peeled off over Hawaii last April, sweeping a flight attendant to her death, attention has been focused on the older aircraft.*
[Don't agonize too much over the *Ever since the roof* part. We found the tonal analysis really hard there too.]

DIFFICULT:

EXERCISE ≪spanish≫: *And I had registered for Spanish, simply because I'd taken it for five years in high school.*

EXERCISE ≪two-million≫: *I'll buy it back from you for like two million, because ya done ran it into the ground, you're having problems, like you... you're not gonna make it, and you go bankrupt, so I'm gonna buy it back from you for like, for next to nothing.*

EXERCISE ≪sold2≫: *'coz he was like, a millionaire*
[Don't worry about the type of the boundary (if there is one) after *like*.]

EXERCISE ≪massachusetts2≫: *Hewlett-Packard has announced it's buying Massachusetts-based Apollo computer.*

2.11 The point of highest f0

The only other label on the tone tier which we have not discussed is the transcription of HiF0, the point of highest fundamental frequency associated with a pitch accent within an intermediate phrase. HiF0 is used currently as a rough measure of the phrase's pitch range. It is transcribed only for intermediate phrases in which there is an accent with a H component — i.e. a H*, L+H*, L*+H, or H+!H* accent. Thus, for example, in ≪made4≫ (given above) one should not transcribe HiF0.

The summary statement in the Annotation Conventions (Appendix A) offers the following advice about HiF0:

> Transcribers should take reasonable care to choose a point in time that reflects the target of the H for the accent. In several cases this will mean choosing some point other than the actual f0 maximum. For example, sometimes the highest f0 value in an accented syllable reflects the 'intrinsic' effect of a voiceless consonant and will thus be a poor estimate of the speaker's choice of pitch range. More seriously, in a phrase where the highest accent-related f0 occurs in a H* H- H% sequence, choosing the absolutely highest value for HiF0 will artifactually inflate the pitch range estimate by the amount of the upstep on the H%. In such cases, we recommend that the syllable's amplitude contour be used to pinpoint HiF0 within the candidate region.

For an example, see ≪good1≫ from PRACTICE TWO. HiF0 would be at the amplitude peak for *good* at about time 2.61.

3 More on the break index tier

3.1 The break index tier relative to other tiers

The other core part of the prosodic transcription proper is the break index tier. If we think of the tone tier as a marking of the speech signal mediated primarily by our interpretation of the analysis of the f0 contour, the analogous way to think of the break index tier is as a marking of the speech signal as mediated primarily by the rhythmic and segmental analysis implicit in the orthographic tier. The summary statement of ToBI conventions describes this relationship as follows:

> Break indices represent a rating for the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance. They are to be marked after all words that have been transcribed in the orthographic tier. All junctures — including those after fragments and filled pauses — must be assigned an explicit break index value; there is no default juncture type.

Thus, the events on the break index tier are labels of the utterance's prosodic grouping — that is, each label denotes a boundary of some kind of constituent which ends at the word that the transcriber has marked on the orthographic tier. The convention for placing break index marks in a *waves*^{TM} label file is that the number should be associated with a point in time at the end of the marked word as indicated by the label in the orthographic tier. It should be located exactly at, or slightly to the right, of this word marker, so that break indices can be unambiguously associated with other tiers.

There are 5 break indices, numbered 0 through 4, roughly in order of lesser to greater degree of perceived separation between the marked word and following material. The break indices are meant to be a label of the SUBJECTIVE strength of the boundary. However, this does not mean that there are no objective criteria for marking the boundaries, or that the five labels form a uniform five-point scale. For example, the lowest-level break index (0) is defined in terms of connected speech processes, such as the flapping of word-final /t/ and /d/ before a following vowel-initial word in many American and Australian dialects, processes that prosodically group words together into 'clitic groups' — larger compound-word-like constituents above the level of the word (see Section 3.2). At the other end of the

scale, the two highest break indices (3 and 4) are defined in relationship to the prosodic constituents (intermediate phrases and intonation phrases) that are assumed by the marking of phrase accents and boundary tones on the tone tier (see Section 3.3).

Mainstream phonological theory might lead us to expect that these intonational constituents and the lower-level clitic group constituents will form a strict hierarchy (see Selkirk 1981; Nespor and Vogel 1986). The numerical scale of break index values reflects a mild bias in favor of such strictly hierarchical models (see the discussion in Price et al. 1991). Rather than building the expectation rigidly into its transcriptions, however, ToBI provides two regular mechanisms for denoting mismatches between different cues to subjective boundary strength. First, break index 2 denotes a mismatch between the constituency prescribed by the tonal transcription and the sense of disjuncture due to pauses and pause-like phenomena (see Section 3.4). Second, there is a diacritic 'p' that can be appended to break indices 1, 2, and 3 to convey some sort of prosodic disfluency — for example, an abrupt cutoff after a false start or a perceptible prolongation or pause which sounds as if the speaker were hesitating while searching for the next word (see Section 3.5). These two provisions should allow transcribers to avoid the circularity of basing a theory about the nature of the prosodic hierarchy upon the transcription of databases that might be used to explore such issues as the relationship between intonational constituents and pause (see, e.g., Woodbury 1993 who proposed that pauses can be placed independently of intonational boundaries when the discourse structure requires the indication of competing segmentation strategies for topic structure versus rhetorical structure).

3.2 Break indices 0 and 1

Except in more deliberate speech styles, such as the information-packed style of radio news announcers, the break index value that will be encountered most frequently is probably 1. The ToBI conventions define break index 1 negatively, as the label to be used for "most phrase-medial word boundaries", as contrasted with the marked phrase-medial cases transcribed by break index 0. Break index 0, conversely, is defined with positive criteria as the value "for cases of clear phonetic marks of clitic groups; e.g. the medial affricate in contractions of `did you` or a flap as in `got it`." Since the other break indices are also defined by positive criteria (markings on the tone tier — see Sections 3.3 and 3.4), we can think of break index 1 as the 'default' (although, of course, there is no real default index in the sense of having a value that need not be marked because it is understood). We have already seen many examples of break index 0 in previous example utterances. For example, in example ≪understand≫ in Section 2.10 above, there are three cases of 0 break index: the flapped /t/ on the two instances of the word *to* after *trying* and *you* and the palatalization of the /t/ at the juncture between *get* and *you* all are examples of connected-speech processes that we take as criteria for break level 0.

```
EXAMPLE <<understand>>: I'm simply trying to get you to understand.
                            1      3      0  3   0   0  3          4
```

Example utterance ≪kinds-v≫ illustrates yet another such connected-speech process: the apparent deletion of the vowel in *of* after *kinds*, to make a phonotactically impermissible /zv/ word-final cluster.

```
EXAMPLE <<kinds-v>>: What kinds of planes...
                          1     0  1      4
```

Note that in some cases the phenomena denoted by break index 0 are so frequently encountered in particular types of sequences, that orthographic conventions have developed for marking them. For example, the flapping of the /t/ and consequent cliticization of the word *to* onto a preceding auxilliary verb *got* is sometimes indicated in writing by *gotta*. Or the deletion of the initial /h/ and vowel of *have* in sequences such as *would have* can be indicated by spelling it *would've*. In such cases, the transcriber has the alternative of marking the prosodic grouping by the choice of label on the orthographic tier instead. For example, by labelling the word as *gotta* rather than *got to* the transcriber has eliminated the word boundary where a 0 label might be placed on the break index tier.

3.3 Break indices 3 and 4

Break indices 0 and 1 form a natural progression with indices 3 and 4. These two break index strengths are equated with the intonational categories of intermediate (intonation) phrase and (full) intonation phrase. Thus, whenever the tonal analysis indicates a L- or H- phrase accent, the transcriber should decide where the end of the intermediate phrase marked by this tone label is and place a 3 on the break index tier to align with the orthographic label for the last word in the intermediate phrase. Similarly, whenever the tonal analysis indicates a L% or H% boundary tone, the transcriber should place a 4 on the break index tier at the end of the last word in the intonation phrase. In actuality, the ordering of these two analyses is sometimes reversed. This is particularly the case with the L% boundary tone; the transcriber might be convinced of the percept of a 4 versus a 3 level boundary before deciding that there must be a L-L% or H-L% sequence as opposed to merely a L- or H- to be marked on the tone tier. Recall from the discussion in Section 2.3 that there may be little or no difference in f0 values between the end of a mere L- and a L-L% sequence or between a mere H- and a H-L% sequence; A L% following a L- is in the bottom of the speaker's pitch range just as a L-, whereas a L% following a H- is upstepped to the same level as the preceding phrase accent. In such cases, the analysis is necessarily more subjective; the transcriber must rely on the percept of degree of disjuncture with less help from the f0 contour. Some pertinent examples from earlier sections are repeated here.

```
EXAMPLE <<names>>:
    Anna may know my name, and yours too.     Anna may know our names?
    H*                L-H%      H*    H* L-L% L*                   H-H%
        1   1    1  1    4    1     1   4         1   1    1   1     4


EXAMPLE <<park2>>: Definitely the shortest and  probably the pleasantest
                   H*           L-  H*      L-L% H*         L- H*
                              1   3        1   4         1   3           1
                         way to go is through the park.
                                  L-             L+H* L-L%
                            0  1  3  1       1   1    4
EXAMPLE <<oregano>>:  1) Let's see   I need oregano 'n marjoram 'n some
                           H*   H* L-L%        L*  H-    L*     H-
                                1   4   1    1       3  0        3  0  1
                           fresh basil okay?
                          L+H*  !H* L-  H* H-H%
                                1     3    4
                        2) Oh I don't know    it's got oregano 'n marjoram
                          H* !H*      !H* L-L%          H*  H-    H*   H-
                             1 1     1    4       1   1       3  0     3
                           'n some fresh basil.
                                        H* H-L%
                             0    1     1     4

EXAMPLE <<nose>>: Oh    don't nuzzle me you marmalade-nose.
                  X*? L- H*   !H*     L-    L*           L-H%
                      3      1      1  3   1         1    4
```

When using *waves*^{TM} label files, a 3 or 4 break index label and the corresponding phrase accent or boundary tone are placed together at the orthographic label, with the break index label coming last if the labels on the three tiers cannot be absolutely synchronized.

PRACTICE FIVE

break indices 0, 1, 3, and 4

You have already transcribed the tones on the following. Now transcribe the break indices.

EASY:

EXERCISE ≪manitowoc≫: *Does Manitowoc have a bowling alley?*
[See PRACTICE TWO for tones.]

EXERCISE ≪butcher≫: *How'd your operation go? Don't talk to me about it; I'd like to strangle the butchers.*
[See PRACTICE FOUR for tones.]

EXERCISE ≪stalin≫: *I was wrong, and Stalin was right. I was wrong.*
[See PRACTICE TWO for tones.]

EXERCISE ≪flour2≫: *Oh nothing special, you know flour and butter and sugar.*
[See PRACTICE TWO for tones. Transcribe just the second part, after the *you know*.]

EXERCISE ≪thought≫: *That's what I thought.*
[See PRACTICE ONE for tones.]

INTERMEDIATE:

EXERCISE ≪I-mean≫: *You know what I mean?*
[See PRACTICE ONE for tones.]

EXERCISE ≪noodle1≫: *We have a lean mini-noodle with beans. Well, we have a lean mini-noodle dish.*
[See PRACTICE THREE for tones.]

EXERCISE ≪knock-stuff≫: *Mostly they just sat around and knocked stuff. You know, the school, other people.*
[See PRACTICE TWO for tones.]

DIFFICULT:

EXERCISE ≪argument≫: *If he can then there's no argument about it.* (two productions)
[See PRACTICE THREE for tones.]

EXERCISE ≪artwork≫: *State law now requires public construction projects to set aside 1% of their budgets for artwork.*
[See PRACTICE FOUR for tones.]

EXERCISE ≪anyway≫: *But anyway, if you can't see that then I don't know if I can explain it to you.*
[See PRACTICE ONE for tones.]

3.4 Break index 2

As noted in the previous section, each 3 on the break index tier must correspond to the marking of a phrase accent for the intermediate phrase on the tone tier, and each 4 must correspond to the marking of a boundary tone. The implication is that any other interword juncture will be something that can be transcribed on the break index tier with either a 0 or a 1. However, the subjective impression of boundary strength does not always allow such a neat correspondence. In the course of developing the ToBI transcription system, we encountered several utterances in which we felt a strong sense of disjuncture at a boundary between two words where the pitch pattern showed no evidence of the necessary tonal events for either of these two levels of intonational constituency. We also encountered the converse case:

utterances in which the pitch pattern at a boundary between two words clearly indicated an intermediate or intonation phrase boundary with none of the preboundary lengthening or other cues that support the subjective sense of a strong disjuncture. Break index 2 was devised to mark cases of these two types of 'mismatch' between the subjective boundary strength and the intonational constituency. These two types are described in the ToBI Annotation Conventions as follows:

> a strong disjuncture marked by a pause or virtual pause, but with no tonal marks; i.e. a well-formed tune continues across the juncture.
>
> OR
>
> a disjuncture that is weaker than expected at what is tonally a clear intermediate or full intonation phrase boundary.

Example utterance ≪iraqi≫ illustrates the first type of mismatch, and example utterance ≪quincy≫ illustrates the second. In ≪iraqi≫, the smooth sequence of apparent downstepped peak accents with no clear intervening phrase accent suggests that the words *six*, *southern*, *iraqi*, and *cities* all belong to the same intermediate phrase, yet there is an intonation phrase sized pausing between each adjacent pair of these words. In ≪quincy≫, the clear tonal markings for at least an intermediate phrase boundary are unaccompanied by any clear preboundary lengthening, making some transcribers uncomfortable in labelling this juncture with a 3.

```
EXAMPLE <<iraqi>>: The Pentagon reports fighting in six southern
                     L+H*            L- L*  H-H%    H* !H*
                     1        1       3        4  2   2        2
                   iraqi cities.
                   !H*   X*? L-L%
                        2      4

EXAMPLE <<quincy>>: uh Quincy.  Could I have the number to uh
                       H*  L-        H*          !H*       L-L%
                     4      2       1 1    1   1      1  1  4
                   Shore Cab.
                    *?   H* H-L%
                        1   4
```

Break index 2 was devised for cases where the mismatch between the tonal marking and the disjuncture is not accompanied by any sense of hesitancy or disfluency. When 2 is used in the first way (to indicate a stronger sense of disjuncture than 1 even while producing a coherent contour for an uninterrupted intermediate phrase), it can have the rhetorical effect of careful deliberation, as in the ≪iraqi≫ example. In the opposite case (when 2 is used to mark intermediate phrase boundaries which do not have a very strong sense of disjuncture) the speaker may be speaking quickly to hold the floor or to convey a sense of urgency, while using the tonal marks necessary to convey attentional focus on several closely placed words. We suspect that both types of 2 will be explained ultimately by a better understanding of the complexities of discourse structure, an understanding that can best be achieved by the transcription and analysis of many occurrences in natural dialogue.

3.5 The p diacritic (and the %r tone label)

[Christine Nakatani and Elizabeth Shriberg contributed greatly to the preparation of this and the following sections.]

There are other cases of mismatch between tone tier and segmental rhythm, however, where break index 2 does not seem to be appropriate. For example, in utterance ≪display≫, the pauses after *Baltimore*, *which*, and *leave* do not have the feel of a speaker striving for an effect of judicious deliberation, as in the *six southern Iraqi cities* phrase of the ≪iraqi≫ example, but rather sound disfluent, as if the speaker were hesitating as he searches for the next word. Such cases can be distinguished from fluent cases of 2 by the use of the p diacritic.

```
EXAMPLE <<display>>: Display all the flights from Baltimore to Dallas
                            1   1   1       1    3-        2p 0      4
                     which leave after 4:00 p.m.
                          2p    2p    3p   2p   4
```

The p diacritic is used in conjunction with a break index 1, 2, or 3, to indicate a disfluency in the timing or separation of words across a break. The notation 'p' was chosen initially to denote the prolongation of the hesitation pause with break indices 2 and 3, but we have since extended the diacritic's usage to cover also abrupt cutoffs before restarts and repairs, which are often but not necessarily separated from the disfluent stop by a pause. In this case, the appropriate break index is 1. Thus the inventory of combinations of break index and p diacritic is:

- 1p — an abrupt cutoff before an actual repair, or as if stopping to permit a repair or restart of some kind
- 2p — a hesitation pause or prolongation of segmental material where there is no phrase accent perceived in the intonation contour
- 3p — a hesitation pause or a pause-like prolongation where there is a phrase accent in the tone tier.

The p diacritic is not used with break index 4, because it is difficult to reliably identify hesitations between two full intonational phrases. Example utterances ≪amazing≫ and ≪cheapest≫ illustrate the use of the diacritic with break indices 1 and 3. Example ≪display≫ also had an occurrence of 3p. Note the presence of the phrase accent distinguishing this interword juncture from the surrounding cases of 2p.

```
EXAMPLE <<amazing>>:    um But I had I mean the stuff he knows is kind of
                          0   0 1p  3 1    4   1     1  1     4  1    1 1
                        amazing 'coz he does a lot of um environmental
                               3    1p 1    1 1   0  1  4             1
                        impact stuff
                              2p    4

EXAMPLE <<cheapest>>: I want to see the cheapest flight from Atlanta
                       1    1  1   1   3p       1      1    1       3
                      to Baltimore
                        1         4
```

In general the p diacritic should be used conservatively, and should not become a substitute for 2. A good test for appropriateness is to imagine whether the break would have been the same if the speaker were asked to repeat the utterance with the same intonation, but more 'fluently'. If the break were the same upon repetition, it should probably not get the p.

Note also that the prolongation of segmental material for a 2p label can physically occur at the beginning of a word rather than at the end, as in example ≪least≫, where the hesitation lengthens the [l] of *least* rather than the vowel of *the*.

```
EXAMPLE <<least>>: Between Boston and Denver I'd like to a flight that
                            3      1   1      4   1    1  3p 1     1    1
                     takes the least amount of stops to get to Boston
                          3p  2p    1      0  1     4  1   1  0      4
```

Closely associated with these definitions of 1p, 2p, and 3p in the break index tier is the tone tier label %r, for restarting with a brand new intonation contour when the the last contour was interrupted without being finished by some disfluency. This is most common at a 'repair', where the speaker abruptly stops and begins again with the intended or 'repaired' material, as in example utterance ≪amazing≫, already cited above, and in example ≪connections≫, below.

```
EXAMPLE <<amazing>>:    um But I had    I mean     the stuff he knows
                          0   0 1p   3   1     4      1     1  1     4
                               H* H* L- H* !H* L-L%      H*       !H* L-L%
                        is kind of amazing 'coz    he does a lot of um
                          1    1  1       3    1p    1    1 1   0  1  4
                                   L+H*  L-      %r H* !H*            L-L%
                        environmental impact stuff
                                     1      2p    4
                          H*          H*         H-L%

EXAMPLE <<connections>>:  What are the plane sizes   for these flights and
                              1   1   1     1     4     1     1       4  1
                            H*           L*      H-H%      H*    H*  H-L%
                          do they ha(ve)- do  are there any other flights
                            2p   1       1p  1p   1     1   1     1       1
                                             %r          H*        !H*
                          that have s-    connections
                              1    1  1p             4
                                      %r    H*    L-L%
```

As with the use of the p diacritic, one should be conservative in using %r. It is needed only if there is good evidence that a new intonational phrase has begun after disfluent pause, evidence such as a notable change in f0 range or amplitude. It should not be used in cases such as the *had* after the first 1p in ≪amazing≫, which continues with a fluent H* accent in the same pitch range (unlike the H* !H* on *he does* after the second 1p in this utterance, which is in a new pitch range). Nor should %r be used in example utterance ≪abbreviation≫, where after the speaker stumbles and pauses momentarily around the end of *what is the*, the intonation on *abbreviation* continues as if there had been no interruption.

```
EXAMPLE <<abbreviation>>: What is the b-  abbreviation      n     under
                              0  1   1  1p            3-      3       3p
                                          H*          H- L+H* L- H* !H-
                          the category d c  mean
                             1         1 1  1    4
                               H*      H* H*    L-L%
```

Especially, %r should not be used after a 3p, where the (re)start of a new intonation contour is already implicit in the break index for the intermediate phrase.

3.6 Ordinary uncertainty.

In addition to these two well-defined types of 'uncertainty' due either to conflicting evidence about boundary strength (break index 2) or to the interruption of fluent prosodic production at repairs and hesitations (the 'p' diacritic), there will be cases of ordinary garden-variety uncertainty for other reasons. For example, (as we have already discussed above in Section 2.3) the f0 contour for an utterance-medial intonation phrase that ends with a L% boundary tone is often difficult to distinguish from a mere intermediate phrase. In such cases, where the transcriber cannot decide from other cues whether the tonal analysis should be L- versus L-L% (or H- versus H-L%), the break index marking is also necessarily ambiguous. The ToBI conventions prescribe that in such cases of transcriber uncertainty, the higher-level boundary should be chosen, and uncertainty marked by appending the '-' diacritic. Thus, in ≪park2≫ given above in 2.3, if no decision can be made between L- and L-L%, the correct break index marking is '4-'.

The same convention applies at lower levels of the hierarchy. For example, if the transcriber thinks that a word-final /d/ has been pronounced as a flap, joining the word it ends into a close prosodic unit with the following word, but is not certain that it is a flap and not just a rather short [d], then the correct break index marking is '1-'. A similar case involving /t/ is given in example utterance ≪democrat≫. Here it is not clear whether the /t/ at the end of the word *democrat* has been flapped, or not released.

```
EXAMPLE <<democrat>>: The chairman,  Wendell Ford,  democrat of Kentucky...
                        L+H*   L- L+H*    !H*  L- H*           L+H* L-L%
                       1        3        1     3          1- 1        4
```

Examples ≪rewarding≫, ≪noodle2≫, and ≪noodle3≫ illustrate cases where tonal sequences evident in the pitch contour might seem compatible with several alternative analyses, some with and some without a medial intermediate phrase break. When such utterances are transcribed outside of their larger discourses, these contours might be highly ambiguous.

```
EXAMPLE <<rewarding>>:  A really rewarding day.
                          L+H* L-          H* L-L%
EXAMPLE <<noodle3>>:    We have a lean mini-noodle dish.
                              L+H* L-    L+H*     L-L%
```

(compare ≪noodle2≫ given above in PRACTICE THREE)

The minus symbol associated with uncertainty in break index value cannot be used in conjunction with the p diacritic. Uncertainty about whether or not to use the p should be conveyed by using 'p?'.

4 The miscellaneous tier (and other aspects of the marking of disfluencies)

4.1 The miscellaneous tier defined

The miscellaneous tier is in essence a 'comment' tier for the optional marking of events of any kind other than the standard words, tones, and disjunctures marked on the orthographic tier, the tone tier, and the break index tier. Many of the events labelled on the miscellaneous tiers are things that span longish intervals. In this, miscellaneous events are like the word events labelled on the orthographic tier. However, the two types of events are very different, in that a strict succession of miscellanous events is not essential to speech, whereas speech must be a succession of produced words (or pieces of words). Therefore, whereas the ToBI convention is to mark each event on the orthographic tier only at the end of the interval that the event spans, it prescribes that an event on the miscellaneous tier should in general be marked for both its end and its beginning, using the diacritics '>' and '<', respectively. Thus labels on the miscellaneous tier usually come in pairs, such as:

```
\hspace*{.5in}breath< breath>  laugh< laugh>  cough< cough>
```

Example ≪cough≫ in Section 1.1 illustrated the use of the miscellaneous tier to mark the cough that interrupts the utterance.

```
EXAMPLE <<cough>>: Will you have marmalade ...
                        L*        L*
                       1   1    1         1p
                                 cough<            cough>
```

Another similar example is the laughter that interrupts the pitch contour in utterance ≪laugh≫.

```
EXAMPLE <<laugh>>: To me  it this seems very obvious;   to make it on
                     1  3   1    1     1    1       4     1    1  1  3
                      H* L-               L+H*    L-L%     H*     !H* L-
                                       laugh<     >laugh laugh<
                   to make it by hand is much more fun than to make it on
                     1    1  1  1    3  1    1    1   1    1  1    1  0  1
                       H*       L+H* L-   H*
                              >laugh
                   a computer.
                    1        4
                             L-L%
```

Since such markings are useful for parsing the disruption of otherwise tonally well-formed intonation contours, we can think of them as a source of 'disfluency'. Indeed, the ToBI Annotation Conventions encourage the marking of disfluencies, and suggest the use of 'disfl<...disfl>' (or 'disfl') as a general flag for them:

> In general, it is the assumption of the participants in the common transcription group that silences should be automatically detectable, at least to a first approximation, and that transcriber time should not be spent marking these by hand. Disfluencies, by contrast, are not automatically detectable, and the absence of markings for them makes it difficult to parse the tone and break index tiers. For these reasons, transcribers are urged to mark disfluencies on the miscellaneous tier using 'disfl<' and 'disfl>' (or 'disfl' if the disfluency is extremely localized), and to provide these marks in the miscellaneous tier menu when using $waves^{TM}$.

However, it is often easier to determine that something is disfluent in some region than it is to determine exactly where the disfluency begins and ends. For this reason, the ToBI Annotation conventions specify that the marks can be used more like a disfluency flag rather than the demarcation of a precise region:

> ...the marks 'disfl<' and 'disfl>' (or simple 'disfl') should be interpreted as rough pointers to the disfluent region and transcribers should not agonize over placing them precisely.

Note that here the ToBI Annotation Conventions explicitly mention the use of a single mark, rather than a pair of marks for the beginning and end of a region. However, they specifically recommend this usage only for disfluencies, to encourage the marking of something that is typically very difficult to locate precisely in time. Transcribers should be careful about using a single (unpaired) label on the misc tier for anything other than marking the general location of a perceived disfluency, since in any other circumstance, the usual interpretation must be that the event is so localized that its beginning is virtually the same point as its end.

Example utterance ≪fare≫ is an example of a disfluency marked in this way.

```
EXAMPLE <<fare>>: show me the cheapest fare    from Da(llas)- from
                      1  1   1        1    4       1        1p    1
                    H*        L+H*     !H* L-L%        %r
                                                disfl< disfl>
                   Philadelphia to Dallas     excluding    restriction
                               3  1      4             4              4
                  L+H* !H*     L-  H*    L-L%  L+H*   L-L%      H* L-L%
                   v   u    slash one
                     1  3        1   4
                   H* !H* L- H*   H* L-L%
```

Although the miscellaneous tier is a general-purpose 'comment' tier, we recommend that when transcribers at a particular site find themselves often adding comments that fit some particular pattern other than these, they consider defining another extra tier for that purpose.

Christine Nakatani and Elizabeth Shriberg, both of whom have worked extensively on disfluencies in naturally spoken utterances, differentiate more finely, and suggest guidelines for other transcribers who wish to differentiate types of disfluencies in the same way. The following section is adapted from their suggested guidelines, and uses many of their examples.

4.2 Suggested guidelines for marking disfluencies

Nakatani and Shriberg have identified several different types of events that they feel should count as disfluencies in ToBI. Not all of these need be marked on the misc tier in order to be recovered. In particular, mere hesitation pauses can be recovered from the use of the 2p or 3p marks on the break index tier and (in the case of many filled pauses) from the transcription of the filler material on the orthographic

tier. Phenomena that might be flagged as disfluencies on the misc tier include such phenomena as stumbling over a word, or abruptly cutting off a word or phrase in midstream to make a fragment, as in ≪fare≫ cited above, or ≪transport≫ below. These are examples of the first of the major classes of disfluency which Nakatani and Shriberg identify, including what they call 'phonetic error'.

```
EXAMPLE <<transport>>: show ground transpor- ground transportation
                           1      1        1p      1              4
                               disfl:repair<
                                 disfl:repair>
                       at  atlanta
                         2p       4
```

The second of the three major classes is the hesitation pause. This includes both silent pauses as in the examples transcribed with 2p above, and filled pauses — that is hesitation intervals during which the speaker holds the floor by producing hesitation noises or other material, as in ≪weight≫.

```
EXAMPLE <<weight>>: The weight on a six on a seven sixty seven is
                       1      0  1 1   2p 1 1     2p    1     3- 2p
                    three thousand uh three hundred and twelve
                         1        2p 4     1       1   1      1
                    thousand pounds uh is that including passengers
                            2p     2p 4  1    1         3          4
```

Nakatani and Shriberg recommend that the spelling of hesitation noises be standardized so that later users of a ToBI transcribed database need search only for a limited set of 'words' in recovering the disfluency. In particular, for standard American English, they recommend the use of only *um*, *uh*, or *mm*. That is, transcribers should not invent other spellings such as *ah* or *uhhhh* to reflect differences in the quality of the reduced vowel or the duration of the syllable. With this stipulation, filled pauses of this sort would not need to be flagged on the misc tier, since they would be recoverable from the orthographic tier.

A filled pause may be perceived as unaccented, and yet as constituting its own intermediate or intonational phrase. Normally each intonation phrase is required to have at least one pitch accent. In the case of filled pauses this criterion is relaxed; an unaccented filled pause in its own phrase can be labelled with the phrase accent (chosen from the full inventory) without the requirement that a pitch accent be marked on the filled pause.

The last major class of disfluency is the class of repairs and fresh starts, which Nakatani and Shriberg define as "lexical self-corrections of parts of sentences and whole sentences, respectively". They give us utterance ≪fare≫ as an example of a repair, and ≪connections≫ as an example of a fresh start. (Here we have used the misc tier to mark these interpretations of the disfluencies.) These two examples also illustrate abrupt cutoffs resulting in word fragments.

```
EXAMPLE <<fare>>: show me the cheapest fare from Da(llas)- from
                     1  1   1        1    4    1        1p    1
                                          repair< repair>
                  Philadelphia to Dallas excluding restriction
                              3  1      4         4           4
                  v u slash one
                   1 3     1   4
```

```
EXAMPLE <<connections>>: What are the plane sizes for these flights and
                             1  1   1     1     4   1     1       4   1
                         do they ha(ve)- do are there any other flights
                           2p   1      1p  1p  1     1   1     1      1
                             restart<
                                      restart>
                         that have s-  connections
                             1    1  1p           4
```

More detailed suggestions about how to flag repairs can be obtained by writing directly to Christine Nakatani (chn@das.harvard.edu) or Elizabeth Shriberg (ees@speech.sri.com).

PRACTICE SIX

break index 2, the p diacritic, disfluencies

Transcribe these exercises using the **exercises** script.

EASY:

EXERCISE ≪park5≫: *Uh and then I go under a footbridge and into the park.*

EXERCISE ≪business≫: *A lot of people have done this; they sell their business, and they have... If something goes wrong, and they have the first rights to buy it back.*
[Repeated exercise from PRACTICE THREE. Transcribe the phrase *and they have,...*]

EXERCISE ≪howto≫: *I know we've gotta do it but I don't know how to do it.*
[Repeated exercise from PRACTICE FOUR.]

INTERMEDIATE:

EXERCISE ≪mean≫: *Because I I mean, to make a map on computer is not n- nearly as much fun.*

EXERCISE ≪semester≫: *The advisor to f- fill out my schedule for the first semester said "Why don't you take Introduction... Intro... Introductory Linguistics."*

EXERCISE ≪spoon2≫: *There's a spoon in here.*
[Compare ≪spoon1≫ in PRACTICE TWO.]

EXERCISE ≪author≫: *The author of more than eight hundred state supreme court opinions (Hennessy is widely respected for his legal scholarship and his administrative abilities.)*
[This is the first part of ≪hennessy≫ in Section 2.10.]

EXERCISE ≪usually-not≫: *Usually not, no. Nah. Usually they won't give you chances.*

DIFFICULT:

EXAMPLE ≪tuition≫: *My learning experiences are on the job, so when I screw something up instead of s- spending all this money to go to college... When I screw up a job, that's my tuition for college. That's exactly, exactly how it works, there's no difference at all.*

EXERCISE ≪fail≫: *And what happens is: when you... when you buy my business, and you try to run my business, it's really hard for you to run my business. So a lotta times they fail.*
[Repeated exercise from PRACTICE THREE. You've transcribed most of the tones already. Now you're ready to worry about the break indices, particularly those around the first *when you...* and *when you try to run my business.*]

EXERCISE ≪figureout≫: *Half the job is accomplished by just starting it.*
[Interviewer: *Mm-hmm*] *So just start doing it, and you'll figure it out.* [Interviewer: *Yeah*] *You know what I mean?*

References

Beckman, M. E., and J. B. Pierrehumbert. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook* 3:255–309.

Bolinger, D. 1972. Accent is predictable (if you're a mind reader). *Language* 48:633–644.

Brown, G., K. Currie, and J. Kenworthy. 1980. *Questions of Intonation.* Croom Helm.

Campbell, W. 1992. Prosodic encoding of English speech. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, 663–666. Banff, Canada.

Grosz, B., and J. Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, 429–432. Banff, Canada.

Hirschberg, J. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence* 63((1-2)).

Hirschberg, J., and J. Pierrehumbert. 1986. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 136–144. New York, NY.

Ladd, D. R. 1980. *The Structure of Intonational Meaning: Evidence from English.* Indiana University Press.

Ladd, D. R. 1983. Phonological features of intonational peaks. *Language* 59:721–759.

Nespor, M., and I. Vogel. 1986. *Prosodic Phonology.* Foris Publications.

Pierrehumbert, J. B., and J. Hirschberg. 1990. The meaning of intonation contours in the interpretation of discourse. In *Plans and Intentions in Communication and Discourse*, ed. P. R. Cohen, J. Morgan, and M. E. Pollack, 271–311. (SDF Benchmark Series in Computational Linguistics). MIT Press.

Price, P., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1991. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of the Acoustical Society of America* 90:2956–2970.

Ross, K., M. Ostendorf, and S. Shattuck-Hufnagel. 1992. Factors affecting pitch accent placement. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Vol. 1, 365–368. Banff, Canada.

Sag, I., and M. Liberman. 1975. The intonational disambiguation of indirect speech acts. In *Papers from the 11th Regional Meeting of the Chicago Linguistics Society*, 487–497.

Selkirk, E. 1981. On the nature of phonological representation. In *The Cognitive Representation of Speech*, ed. T. Myers, J. Laver, and J. Anderson, 379–378. North-Holland.

Silverman, K. E. A., M. Beckman, J. F. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 1992. TOBI: A standard for Labeling English Prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Vol. 2, 867–870. Banff, Canada.

Silverman, K. E. A., and J. B. Pierrehumbert. 1989. The timing of pre-nuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, ed. J. Kingston and M. E. Beckman, 72–106. Cambridge University Press.

Ward, G., and J. Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61:747–776.

Woodbury, A. 1993. Against intonational phrases in Central Alaskan Yupik Eskimo. Presented at Linguistics Society of America Annual Meeting.

Index

≪AND1≫, 29
≪I-mean≫, 17, 34
≪abbreviation≫, 37
≪amazing≫, 36, 37
≪amelia-p2≫, 16
≪anna≫, 15, 16, 28
≪anna1≫, 16
≪anna2≫, 28
≪anyway≫, 17, 34
≪argument≫, 27, 34
≪artwork≫, 30, 34
≪atlanta≫, 22
≪audience1≫, 21
≪author≫, 41
≪bananas≫, 20
≪beef≫, 29
≪blond-baby1≫, 13
≪bloomingdales≫, 23, 24
≪business≫, 27, 41
≪butcher≫, 30, 34
≪calling≫, 24
≪calling2≫, 25
≪calling3≫, 26
≪capote≫, 27, 28
≪cheapest≫, 36
≪connections≫, 36, 37, 40, 41
≪cough≫, 9, 38
≪cream≫, 21
≪democrat≫, 37, 38
≪display≫, 35, 36
≪dream≫, 17
≪drive≫, 22
≪eileen-leaving≫, 17
≪eileen1≫, 26
≪eileen2≫, 26
≪elephant3≫, 20, 21
≪environ1≫, 30
≪environ2≫, 30
≪experience1≫, 22
≪fail≫, 27, 42
≪fail1≫, 17
≪fare≫, 39–41
≪figureout≫, 42
≪flap≫, 13
≪flap2≫, 13
≪flour1≫, 22, 26
≪flour2≫, 21, 34
≪for-marianna≫, 16
≪friend1≫, 22, 26
≪friend2≫, 26
≪gloria≫, 20, 21
≪glottal-stop≫, 14
≪good-aft≫, 22
≪good1≫, 22, 31
≪good2≫, 21, 22
≪happens≫, 17
≪heavy-rain≫, 17, 27
≪hennessy≫, 29, 41
≪howto≫, 30, 41
≪hurt≫, 29
≪insert≫, 11, 12, 14
≪iraqi≫, 35
≪jam1≫, 8, 9, 18
≪jam2≫, 13
≪kinds-v≫, 32
≪knock-stuff≫, 22, 27, 34
≪laugh≫, 38, 39
≪lazy≫, 16
≪least≫, 36
≪legumes1≫, 21
≪legumes2≫, 21, 26
≪legumes3≫, 21, 30
≪levels≫, 25
≪loan1≫, 21
≪made1≫, 10–16
≪made2≫, 10, 15
≪made3≫, 10, 15
≪made4≫, 12, 14, 28, 31
≪manitowoc≫, 21, 34
≪massachusetts2≫, 30
≪mean≫, 41
≪memphis1≫, 17
≪memphis2≫, 17
≪mile≫, 26
≪millionaire≫, 22, 23
≪mole1≫, 16
≪money≫, 18
≪mother1≫, 16
≪mother2≫, 21
≪mother3≫, 22
≪mother4≫, 16, 19, 22
≪name1≫, 18
≪names≫, 17, 33
≪no-pitch-halving≫, 14
≪noodle1≫, 23, 26, 34
≪noodle2≫, 23, 26, 38
≪noodle3≫, 38
≪noodle4≫, 30
≪noodle5≫, 30
≪noone≫, 16
≪nose≫, 19, 33
≪older-aircraft≫, 28, 30
≪onions≫, 27
≪oregano≫, 18, 19, 33
≪park1≫, 25, 26
≪park2≫, 18, 19, 33, 37
≪park3≫, 26
≪park4≫, 26
≪park5≫, 25, 41
≪peel≫, 28

≪philadelphia≫, 29, 30
≪physicist≫, 30
≪pigs≫, 22
≪pitch-doubling≫, 14
≪pitch-halving≫, 14
≪quincy≫, 35
≪really1≫, 25, 26
≪really2≫, 26
≪rewarding≫, 38
≪romanelli≫, 26
≪semester≫, 41
≪smoke≫, 28
≪sold1≫, 30
≪sold2≫, 30
≪sold3≫, 28
≪spanish≫, 30
≪spoon1≫, 22, 41
≪spoon2≫, 41
≪stalin≫, 21, 34
≪stein≫, 23
≪sublime1≫, 27
≪sublime2≫, 27
≪sununu≫, 26
≪tags≫, 19, 20
≪thatone≫, 26
≪thatone2≫, 27
≪theresa≫, 24, 28
≪thermometer≫, 17
≪thought≫, 16, 34
≪trafficlight≫, 22
≪transport≫, 40
≪tree1house≫, 17, 27
≪tree2house≫, 27
≪tuition≫, 42
≪two-million≫, 30
≪understand≫, 29, 32
≪usually-not≫, 41
≪vocative1≫, 20
≪voice≫, 21
≪voiced-h≫, 13
≪weight≫, 40
≪wellies1≫, 17
≪wellies2≫, 21
≪windy≫, 26
≪won≫, 15
≪word≫, 26
≪word1≫, 15, 26
≪yellow1≫, 26
≪yellow2≫, 24, 25
≪yellow3≫, 25, 26
waves^{TM}, 2, 4–8, 10, 12, 18, 31, 33, 39
'!' diacritic, *see* pitch range, downstep, 10, 19, 23

accented syllable, 10–12, 15, 19, 22, 23, 31
amplitude, 14, 31, 37
audio recording, 2, 6, 7

basename, 2, 5–7
boundary effects
    lengthening, 18, 22, 35
    lowering, 18
boundary tone, 8–12, 14–17, 20–22, 24, 28, 29, 32–34, 37
    initial high, 11, 20, 21
break indices, 2, 4, 6, 8, 11–13, 31–33, 36–38, 41, 42
    break index 0, 8, 11, 31–34
    break index 1, 32–36
    break index 2, 32, 35, 36
    break index 3, 8, 12, 33, 34
    break index 4, 8, 11, 12, 33, 34
    clitic groups, 31, 32
    subjective boundary strength, 8, 18, 20, 31–35

calling contour, *see* contour types, calling contour
cardinal examples, 4, 5
clitics, *see* break indices, clitic groups
connected speech processes, 31, 32
    break index 0, *see* break indices, break index 0
    deletion, 32
    flapping, *see* flapping
    palatalization, 32
content words, 29
context, 16, 29
continuation rise, *see* contour types, continuation rise
contour types
    calling contour, 25
    continuation rise, 14
    contradiction contour, 20
    declarative, 14
    downstepped hat, 25
    hat pattern, 15, 22, 25, 28
    stylized high-rise, 19
    surprise-redundancy, 16, 21, 29
    yes-no question, 14
contradiction contour, *see* contour types, contradiction contour
core prosodic analysis, 8, 31

deaccenting, 10, 28
declarative contour, *see* contour types, declarative
dictionary entries, 8–10
digitized utterances, 2, 5
discourse/topic structure, 9, 25, 27, 32, 35
    paragraph intonation, 25
disfluencies, 9, 35, 39, 40
    'p' diacritic, 32, 35–38, 41
    'phonetic error', 40
    false start, 32
    flags, *see* tier, miscellaneous
    fragments, 31, 40
    hesitation, 36, 37, 39, 40
    repairs, 36, 37, 40, 41
    restarts, 36
disjuncture, 18, 20, 32–35, 38
    degrees of, *see* juncture, degrees of
downstep, *see* pitch range, downstep

downstepped hat, *see* contour types, downstepped hat

e-mail address, 7, 8
emphatic speech, 27, 29

flags, *see* tier, miscellaneous
flapping, 13, 31, 32, 37
ftp, 2, 3, 7
fundamental frequency, 8–15, 31
    maximum, 12, 31
    minimum, 12, 23

glottal stop, 14, 29

hat pattern, *see* contour types, hat pattern
holding the floor, 35

installation, 2–5
interspeaker differences, 23
intonation pattern, 2, 9, 11, 12, 14, 19, 25, 29

juncture
    degrees of, 31
    disjuncture, *see* disjuncture

meaning contrast, 15
mid-level tones, 21–24, 28
multi-transcriber sites, 6

non-*waves*^{TM}, 4, 7, 11, 12
nuclear accent, 10–12, 15, 17–20, 23, 24, 29

other languages, 7
overview, 8

pauses, 18, 31, 32, 36, 37
    filled, 39, 40
    hesitation, 36, 39, 40
    silent, 40
phrase
    intermediate, 11, 12, 16–18, 20, 22, 25, 30–33, 35, 37, 38
    intonation, 11, 14, 16, 18, 20, 32, 33, 35, 37
phrase accent, 8, 9, 11–22, 24, 28, 29, 32–36, 40
pitch accent, 8–12, 14–24, 28–31, 40
    !H* accent, 25, 28
    bitonal, 10, 24
    double accents, 20, 29, 30
    H+!H* accent, 23, 24, 26, 31
    low accent(L*), 8, 10, 12–16, 18–22, 24, 28, 29
    peak accent(H*), 10, 11, 14–25, 28, 31, 37
    Pierrehumbert's H*+L accent, 24, 25
    Pierrehumbert's H+L* accent, 24
    rising peak accent(L+H*), 10, 11, 14–17, 19, 21–25, 31
    scooped accent(L*+H), 22, 24, 26, 31
pitch range, 9, 15, 17, 18, 20, 23–25, 27, 28, 31, 33, 37
    downstep, 14, 23–28, 35
    HiF0, 9, 31
    upstep, 17, 18, 23, 24, 31, 33
pitch-tracking
    algorithms, 13, 14
    errors, 14
        pitch-doubling, 14, 17
        pitch-halving, 14
plateau-like regions, 17–19, 28
prenuclear accent, 11, 16, 17, 20
principles, 7, 9
prominence, 9–11
prosodic structure/grouping, 8, 9, 11, 31, 32
    mismatch with subjective boundary strengths, *see* break indices, break index 2, 32, 35
    phrases, *see* phrase, intermediate and phrase, intonation

scripts, 3–6
    examples, 4–6, 8
    exercises, 4–6, 8
    transcriber, 4, 5
segmental effects, 13
    flaps, *see* flapping
    glottal stops, *see* glottal stop
    voiced segments, 12, 13
    voiceless segments, 12, 13, 19, 31
stress, 8–16, 19–21, 23, 29
    levels of, 9, 10
    nuclear, 10, 11
    secondary, 20, 21
    stress shift, 30
stylized high-rise, *see* contour types, stylized high-rise
surprise-redundancy, *see* contour types, surprise-redundancy

tags
    tag question, 19
    vocative tag, 19, 20
tier
    break index, 8, 11, 18, 20, 31–34, 36, 38, 39
    miscellaneous, 8, 9, 38, 39
    orthographic, 8, 9, 31, 32, 38, 40
    tone, 8, 10–12, 14, 18, 22, 31–36, 38, 39
timing of accents, 15–17, 22, 23, 36
ToBI transcription, 2, 5, 8, 12, 34, 40
tools, 3–5, 7, 9, 12
    checker, 5, 12

unaccented word/syllable, 11, 20, 40
uncertainty in labelling, 24, 27–30, 37, 38
    *?, 12, 15, 24, 28
    '-' diacritic, *see* break indices, 37
    'p' diacritic, *see* disfluencies, 'p' diacritic
    boundary strength, *see* break indices, break index 2
    X*?, 15, 25, 28
upstep, *see* pitch range, upstep

voice quality, 13

breathy, 13, 14, 19
creaky, 13, 14

yes-no question, *see* contour types, yes-no question