Introduction to intonational phonetics

Online appendix to accompany *Intonational Phonology* (second edition)

Because it is now so easy to carry out acoustic analysis of recorded speech, many studies of intonational phonology are now based on instrumental data of one sort or another. In order to make sense of such data, and in order to design or select appropriate materials for instrumental study, it is important to have a basic understanding of the phonetics of fundamental frequency (F0) and speech timing. Providing an elementary introduction to these phenomena is the sole goal of this appendix, which should not be taken as a thorough review of the literature. For more detail on many of the topics treated here the reader is referred to the papers in Sudhoff *et al.* (2006).

Boldface numbers in square brackets (e.g. **[5.2.1]**) refer to sections of the published book.

I. Phonetics of F0

Certain features of speech F0 can be seen in instrumental data regardless of whether pitch is used lexically or post-lexically **[1.1.1]**. They appear to result from facts about speech production rather than language structure. They need to be taken into account by anyone undertaking instrumental analysis of spoken data.

1. Declination

The realisation of any given F0 event – pitch accent, lexical tone, etc. – is affected by its position in the phrase. Specifically, ***other things being equal,*** there is a universal tendency for F0 to become lower throughout a phrase. This phenomenon is known as declination. It is presumably related to the tendency of singers to “go flat” during unaccompanied performance.

There is evidence (e.g. Pierrehumbert 1979, Terken 1991, Yuen 2007) that listeners expect declination and compensate for it in speech perception (i.e. for two F0 peaks to sound equally high the later one has to be slightly lower acoustically). However, there is no clear understanding of the mechanisms that cause declination, despite a great deal of interest and research in the 1960s and 1970s (e.g. Lieberman 1967, Maeda 1976, Cooper and Sorensen 1981). A likely contributor to the effect is the lowering of subglottal pressure as air is expended during speech, but there may be others.

The most important reason it is so difficult to arrive at an understanding of the mechanisms is the innocent-sounding phrase “other things being equal”. Especially in languages without lexical tone, it is difficult to say that two F0 peaks are linguistically equal and represent the same underlying target. Much of the research in the 1970s ignored the possibility that some of the overall lowering of F0 during phrases and utterances might actually be intentional and/or linguistically significant. It now seems clear that many languages have phonological or quasi-phonological processes (such as downstep **[2.4]**) that cause otherwise identical pitch events to be realised at a lower F0 when they occur later in a phrase or utterance. Once these processes are taken into account, the amount of true (i.e. unintentional, biomechanically determined) declination appears much less than was assumed in the research of the 1970s.

2. Microprosody

The idealised course of fundamental frequency – the manifestation of the speaker's intonational intentions (or, in a tone language, the manifestation of the sequence of tones) – is not only subject to declination, but is also locally distorted or perturbed by a variety of effects related to the segmental string. These "microprosodic" effects **[1.3.2]** give rise to difficulties in interpreting visual representations of F0, which are discussed further in section III.3 below.

There are three main types of microprosodic effects:

- **obstruent perturbations:** F0 is often unstable during and immediately before and after obstruents. It tends to ***fall*** during the transition ***into*** an obstruent; it tends to ***rise*** during the transition out of a ***voiced*** obstruent (and/or to be lower than it would be otherwise for some or all of the following vowel) and to ***fall*** during the transition out of a ***voiceless*** obstruent (and/or to be higher than it would be otherwise for some or all of the following vowel). Where vigorous voicing is maintained during an obstruent (which it often is not in English), the F0 is generally lowered relative to the surrounding sonorants.

- **voiceless gaps:** You can't have F0 without voicing, so voiceless segments result in interruptions to the F0 contour. In a sense these represent an extreme kind of obstruent perturbation, with voicing considered as a matter of degree: voicing often becomes irregular or disappears altogether during "voiced" obstruents in some languages, and conversely, some voicing may persist into the beginning of a "voiceless" obstruent. The F0 on either side of a voiceless gap often suggests that F0 follows the course that it would have taken if the voiceless segments had been voiced, though this observation needs to be qualified by the facts about obstruent perturbations in the strict sense just introduced above.

- **intrinsic F0:** Other things being equal, high vowels (e.g. [i u]) have higher F0 than low vowels (e.g. [a]). If you say *to Lima* and *a llama* using the same intonation pattern and being careful not to raise or lower your voice between the two, the F0 peak on *Lima* will be objectively higher than that on *llama* even though they sound exactly the same. As with declination, however, it is not always easy to determine whether other things really are equal.

There is reasonably good agreement on the phonetic details of microprosody, and (as with declination) there is clear perceptual evidence that speakers normalise or compensate for microprosodic effects (e.g. Silverman 1986). However, the causes of microprosodic effects are considerably less well understood, though they appear to be traceable in some way to universal acoustic and/or biomechanical factors (e.g. Whalen & Levitt 1995). Thus obstruent perturbations are caused in part by irregular phonation as the voicing is suspended for the duration of an obstruent, or (in the case of voiced obstruents) by changes in airflow and glottis position as the speaker maintains phonation during partial or complete supraglottal closure. Intrinsic F0 appears likely to be due partly to influences of tongue position and/or jaw angle on the position of the larynx. However, it is also increasingly clear that the magnitude of microprosodic effects may vary from language to language, possibly for functional reasons (e.g. intrinsic F0 effects are reduced in Mambila, a language with four level tones (Connell 2000); segmental perturbations are quite large in Korean, a language with a three-way distinction of stop voicing/tenseness (Silva 2006)).

3. Alignment of F0 and segmental features

It is useful to distinguish between **association** (an abstract relation between two roughly simultaneous phonological events) and **alignment** (the actual temporal coordination of measurable phonetic landmarks) **[5.1.1]**. For example, pitch accents and lexical tones tend to belong with a specific syllable: this abstract phonological belonging is association. Recent instrumental research has made clear that such association can be manifested phonetically in a variety of subtly different ways, and that the precise temporal alignment of the phonetic properties of two associated phonological events may be affected by many factors. This appears to be true of both lexical and intonational pitch features.

Three basic types of findings appear well established:

- **segmental anchoring:** F0 movements such as rises and falls generally do not have a fixed slope or duration. Rather, identifiable points in the F0 contour (such as the beginning and end of an F0 movement) are ***anchored*** to identifiable landmarks in the segmental string (such as the beginning and end of the stressed syllable) **[5.1.2]**. The slope and duration of the F0 movement are therefore determined by the temporal interval between the segmental landmarks (Arvaniti, Ladd & Mennen 1998).

- **phonological conditioning:** The segmental landmarks to which F0 points are anchored can be characterised in the first instance in phonetic terms (e.g. "on average 15 ms. before the end of the accented vowel"), but the conditioning factors involved in segmental anchoring often refer to phonological features and abstractions such as vowel length, syllable structure, mora, etc. (Ladd, Mennen & Schepman 2000; Ishihara 2003; Prieto & Torreira 2007). The details also differ from language to language (Atterer & Ladd 2004) or even between varieties of the same language (Arvaniti & Garding 2007).

- **time pressure:** Notwithstanding the preceding points, there is also a considerable effect of phonetic time pressure on the slope, duration, and extent of F0 movements **[5.1.3]**. For example, an accentual rise that would normally end at the beginning of the vowel following the accented syllable obviously cannot do so if the accented syllable is the last syllable in the utterance. The best-established causes of such effects are closely adjacent F0 movements (e.g. cases of stress clash) and closely adjacent word-, phrase-, and utterance-boundaries (Silverman & Pierrehumbert 1990; Prieto, van Santen & Hirschberg 1995). There are unclear or contradictory findings about the effects of speech rate and overall pitch range.

It appears that listeners are very sensitive to such fine alignment effects (Rietveld & Gussenhoven 1995). This is consistent with the fact that alignment can be used as the basis of linguistic contrasts, as in the Swedish lexical accent distinction (Bruce 1977) **[1.3.3]**, or the difference between questions and statements in Neapolitan Italian (D'Imperio 2001).

II. Phonetics of duration and rhythm

Details of timing in speech are still far from well understood. At a first approximation, timing may be thought of in terms of individual segmental durations. A more sophisticated model will need to take the ***temporal coordination*** of various aspects of the signal into account (e.g. the alignment of pitch peaks with segmental events, just discussed). Either way, it is certain that even subtle disruptions of normal duration patterns are often extremely conspicuous perceptually, and that durational differences of a few tens of milliseconds can make the difference between one percept and another.

1. Segment duration

The duration of phonetic segments is known to be affected by a large number of factors. These include:

- supposedly universal intrinsic effects, such as: low vowels are generally longer than high vowels; fricatives are generally longer than other obstruents; taps are necessarily short; vowels before voiced consonants are generally longer than before voiceless consonants.

- language-specific intrinsic effects, e.g. the vowel of *bat* is much longer in American English than in Southern British English; the vowel of *bean* is much shorter in Scottish English than in other varieties.

- language-specific phonological distinctions: e.g. long and short vowels in German, Hungarian, Finnish, and many others; long and short consonants in Italian, Finnish, Japanese, and many others.

- effects of prosodic structure: e.g. in many languages segments are longer before prosodic boundaries (ends of phrases) than elsewhere; in many languages stressed vowels are markedly longer than unstressed vowels (e.g. English or Italian but *not* Finnish or Hungarian); and see further on rhythm below.

- and, of course, speech rate or tempo: the faster you talk, the shorter the segments – but the effects are not at all uniform, because vowels can be stretched and squeezed more easily than consonants.

It is also known that these factors ***interact*** in complex ways. To take a simple example, the syllable durations in *permit* [noun] and *permit* [verb] are quite different: in the noun, the first syllable is stressed, but the second syllable is word-final, and the two syllables are roughly equal in duration; in the verb, the second syllable is both stressed and word-final, so it tends to be markedly longer than the first.

2. Rhythm

Rhythm is an overall temporal structure in a sequence of physical events. In the simplest case, it involves the occurrence of some repeated event at a ***regular interval***. Crucially, it also seems to involve an ***alternation*** of some sort (e.g. between the occurrence of the repeated event and the silence or non-event in between occurrences). Music and poetry are often characterised by regular rhythms in which alternation between stronger and weaker events is important.

Rhythm in speech has long been assumed to involve the occurrence of some element of phonological structure at regular intervals of time. According to a long-standing way of looking at speech rhythm, some languages are supposed to be "syllable-timed" (i.e. having the syllables occur at regular intervals) while other languages are supposed to be "stress-timed" (i.e. having the stresses occur at regular intervals). English is the paradigm example of a stress-timed language.

There certainly is some tendency for the major prominences of an English utterance to be evenly spaced, as you can see if you compare *Tom thought Sue took milk* with *Tommy thought that Susie took milk* or *Tom thought that Sue took the milk.* The syllables *Tom, thought, Sue (Su-), took,* and *milk* occur at roughly the same intervals in all three versions of the sentence, i.e. regardless of the number or location of the unstressed syllables. The question is whether this arises from some overall rhythmic plan, or just happens because of the combined effect of a number of independent but interacting duration effects (specifically, vowel reduction in unstressed syllables, plus a tendency to lengthen a stressed syllable immediately followed by another stressed syllable).

If there is an overall rhythmic plan, it would make sense to expect various compensation mechanisms to maintain a strict rhythm (e.g. in a syllable-timed language, if one syllable is particularly long for intrinsic reasons, an adjacent syllable should be shortened to maintain the rhythm; in a stress-timed language a stressed syllable followed by two unstressed syllables should be shorter than one followed by only one; etc.). However, virtually all experimental attempts to demonstrate the existence of such compensation mechanisms in English and a number of other languages have failed.

Nevertheless, the notion of language-specific rhythmic types lives on anyway, because in some sense it is not completely wrong. For example, even though there is no clear evidence for actual temporal compensation mechanisms, English speech rhythm does seem to involve alternations between stressed and unstressed (or more stressed and less stressed) that may be more abstract than actual temporal intervals but still have the effect of keeping stresses more evenly spaced than we might expect otherwise. For example, the Rhythm Rule (e.g. Liberman & Prince 1977) **[7.2.2]** tends to shift the stress on prenuclear words to a prominent syllable earlier in the word:

citation form: *thirTEEN, submaRINE, interNAtional*
in context: *THIRteen PEOPle, SUBmarine WARfare, INternational AIRport*

Also, as just noted, the fact that English normally reduces vowels in lexically unstressed syllables is going to help make the stresses appear to be more regular in time than they actually are. So "stress-timing" and "syllable-timing" are better thought of as involving clusters of durational properties (e.g. if a language normally reduces vowels in unstressed syllables it is more likely to sound "stress-timed") (cf. Dauer 1983). This point of view seems to underlie recent attempts at quantifying the rhythmic characteristics of different languages (e.g. Ramus *et al.* 1999; Low *et al.* 2000; White & Mattys 2007), in the sense that they take many interacting durational parameters into account. In practice, though, these works have tended to select just two parameters as the basis of their metrics, so that rhythmic typology can be plotted in a two-dimensional space. This makes it very easy for people to continue to think in terms of stress-timing and syllable-timing as rhythmic types, the only modification of the traditional idea being that these notions become the endpoints of a continuum rather than fundamentally distinct classes.

In a few well-studied languages there is slightly stronger evidence that a unit such as the syllable or the mora plays a role in controlling actual durations. The duration of test utterances in Japanese is a straightforward linear function of the number of moras they contain (Port, Dalby, & O'Dell 1987). But even here the evidence for temporal compensation (e.g. squeezing one mora to make up for an adjacent long one) is pretty limited (Beckman 1982, Warner & Arai 2001, Ota *et al*. 2003). This question will definitely continue to motivate research for some time to come.

III. Methodological implications

1. Interpreting visual displays of F0

Obstruent perturbations and voiceless gaps: The existence of microprosodic effects on F0 means that visual displays of extracted F0 require some experience to interpret appropriately. This can be seen from a detailed comparison of Figures 1-3, which show F0 traces and waveforms of three ordinary short English yes-no questions spoken with what is intended to be the same intonation pattern. (To see how similar they are, listen to the sound files by clicking on the captions.)

The most obvious difference is that in Fig. 1 the contour is continuous, whereas in 2 and 3 there are many interruptions, due to voiceless gaps and obstruent perturbations. As listeners we are scarcely aware of these interruptions, but on the screen they are very conspicuous, especially because the F0 in the immediate vicinity of the interruptions tends to jump around a lot. Such effects can be seen clearly across the /s/ at the beginning of the third syllable of *Atkinson's* in Fig. 3: the extracted F0 before the interruption for the /s/ is much lower than that after the interruption, even though perceptually and linguistically there is only a smooth fall from the peak on the first syllable to the low turning point at the beginning of the third. The dip in F0 accompanying the /zð/ sequence in *is that* in Fig. 2, and the apparent discontinuity in F0 around the release of the initial consonant in *Jessica's* in Fig. 2, are similar. Even an alveolar tap (as in *Betty* in Fig. 3) often causes a brief local dip in F0; a glottal stop (at the end of *that* in Fig. 2) often causes a much greater local dip, although in some languages glottal stops may be associated with *higher* F0.

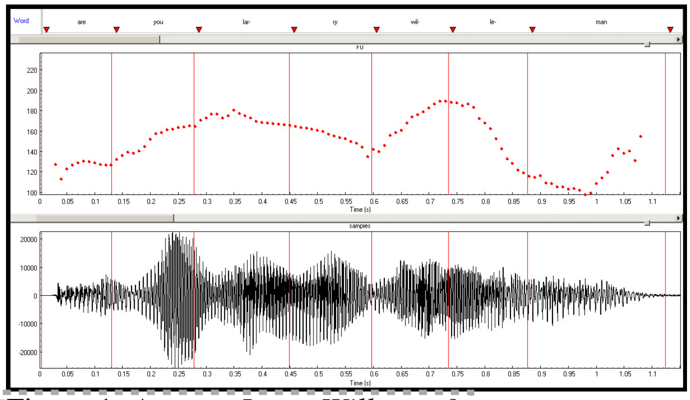

Figure 1: *Are you Larry Willeman?*

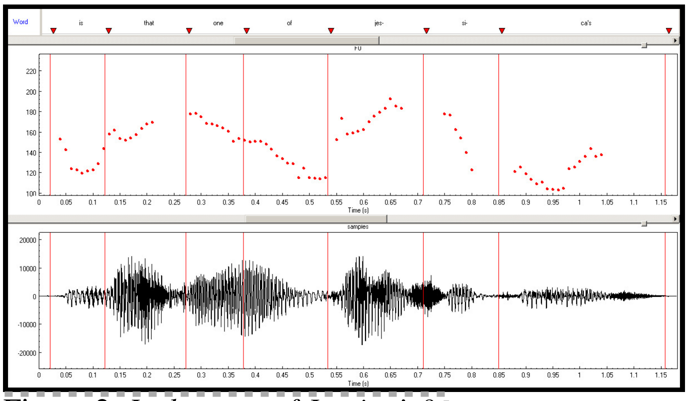

Figure 2: *Is that one of Jessica's?*

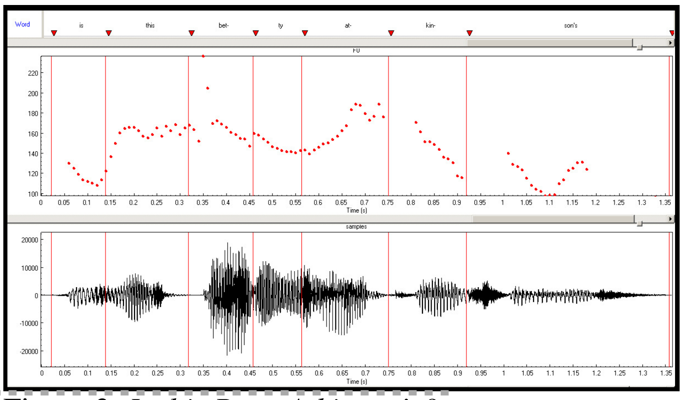

Figure 3: *Is this Betty Atkinson's?*

The visual consequence of these effects is often that the F0 contour on a vowel flanked by obstruents (like the second syllable of *Jessica's* in Fig. 2) looks like an abrupt fall on the visual display. The fact that the F0 falls abruptly on such a syllable does not mean that it has a pitch accent! Beginners tend to over-interpret what they see on the screen. In case of a conflict between what you see on the screen and what you hear, it is better to trust your ears.

F0 extraction errors: A second important fact to keep in mind about visual displays of F0 is that F0 extraction is based on mathematical algorithms applied to the digitised acoustic signal, not on human pattern recognition. These algorithms can occasionally be fooled and give spurious or meaningless F0 values. This often happens when the energy in the signal is very low, as at the beginning and end of phonation (which can be seen clearly at both ends of the contour in Fig. 1), or during weakly voiced obstruents (which can be seen during the [b] of *Betty* in Fig. 3). In general, it is safe to ignore outlying F0 values where voicing is weak; if it is essential to record a value for such stretches of speech, it is best to calculate F0 directly by measuring the duration of periods in the waveform.

The most important type of F0 extraction problem is that of octave errors, in which the extracted F0 value is exactly twice or exactly half what it should be (i.e. a musical octave above or below its true value). When they occur, octave errors often span many analysis frames, so that the F0 value plotted by the program is an octave too high or too low for as much as half a second or more. A good example of an octave error is seen in Fig. 4: by doubling the low extracted F0 values at the very end of the syllable -*mit* we arrive at values that are continuous with the end of the steep rising pitch contour. Fig. 5 shows exactly the same utterance analysed by a different program with a different F0 extraction algorithm, which did not make the octave error and shows continuously rising F0 at the end of the contour. Any abrupt change in extracted F0 such as the one at time 0.44 in Fig. 4 should be scrutinised carefully. If you can obtain values that are continuous with the preceding and/or following context by simply doubling or halving the extracted values on either side of the abrupt change, you should normally assume that an octave error has occurred. As with obstruent perturbations, it is important to trust your ears: if you don't hear a pitch jump of an octave, it almost certainly isn't there.

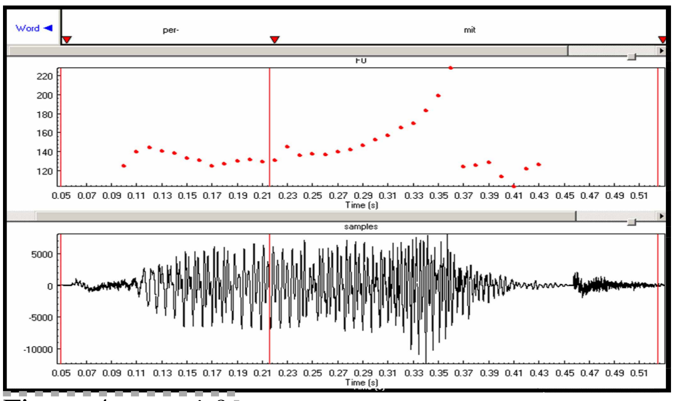

Figure 4: *permit?*

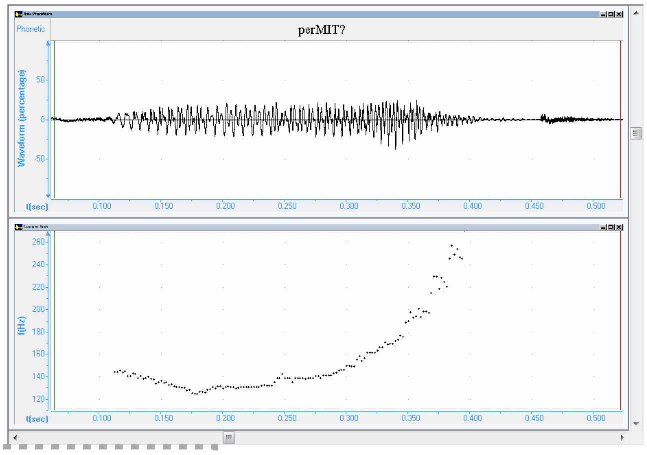

Figure 5: *permit?*

Octave errors can sometimes happen for no apparent reason, but they are often associated with slightly irregular phonation. In the case shown in Fig. 4, for example, there are two likely reasons for the irregularity of the phonation near the end of the vowel of *-mit*: first, syllable-final voiceless stops in English are often accompanied or preceded by some sort of glottal constriction, and second, at the end of any utterance the airflow is likely to decrease rapidly. Either of these could have induced the octave jump in the F0 extraction algorithm. In fact, it may not always be appropriate to consider octave jumps in extracted F0 as "errors", because they may be triggered by stretches of irregular phonation that are perceived as changes of voice quality (e.g. creaky voice) rather than as changes of pitch. This relation between pitch and voice quality represents a potentially fruitful area for research, but is well beyond the scope of this introduction.

Vertical and horizontal scale: Finally, because examining visual displays of F0 contours involves subjective human pattern recognition skills, it's useful to employ tricks to make it easier to use those skills. Most importantly, be sure there's a good balance between the horizontal (time) and vertical (F0 level) scales in the display. The contour from a male speaker's voice may look featurelessly monotonous if the vertical scale on the display goes from 0 Hz to 600 Hz, but will exhibit clear local accent peaks if the vertical scale is expanded so that it only shows values between 50 and 250 Hz. Similarly, pitch movements that look like clear local accent peaks in a display of, say, two seconds of speech may be hard to see if you zoom in on 200 ms. of a single accented syllable. Optimise the display for what you are trying to see.

In doing so, however, always keep in mind that there is a trivial explanation for some octave errors: some signal processing packages force all extracted F0 values to lie within the range specified by the user in analysis or display parameters. For example, in Praat, if you have set parameters for a male speaker and then analyse female speech without changing them, F0 values may be reported at half their true value if the true value is too high for the display window. For example, this was the first thing I considered as the explanation for the octave error in Fig. 4 – though it turns out not to be the explanation in this particular case.

2. Controlling for microprosodic and durational effects

The existence of multiple interacting effects on segment duration and of phonetic factors that disrupt the idealised path of F0 means that considerable care and even ingenuity must be exercised in designing any corpus of materials to be used as a source of instrumental data.

Any study involving duration measurements must either use a very large sample in which the multiple interacting effects will cancel each other out, or must carefully control potential sources of variation in segment duration. To investigate the effect of stress on syllable duration, for example, it is essential to compare segmentally similar stressed and unstressed syllables in prosodically similar contexts. Similarly, the methodological significance of intrinsic F0 is that if you want to study F0 scaling instrumentally, you need to control vowel quality. Don't try to compare measurements of mid tones and high tones if all the mid tones occur on [i] and all the high tones occur on [o]. It's essential to compare like with like.

The existence of obstruent perturbations and voiceless gaps means that the best samples of speech for making instrumental measurements of pitch are stretches containing only sonorants, where voicing is uninterrupted and F0 extraction is likely to be reliable. Obviously, however, there are countervailing issues of naturalness: speakers may tire of repeating sentences about Anna and Manny or about Warren, Ryan, and Allen, or your focus of interest may be on a class of words where obstruents are inevitable (for example, in Greek the "wh-words" – *who, why, where,* etc. – all contain voiceless stops). In any experiment on phonetic detail there is an inherent conflict between naturalness and experimental control; different investigators have different preferences, and different methods may suit different situations, different languages, and different research questions. However, it is worth noting in this connection that a control study by Lickley *et al.* (2005) showed that fine differences of detail in F0/segment alignment in Dutch were conditioned by stress in the same way in short sentences read aloud and in quasi-spontaneous (non-scripted) task-oriented dialogue, suggesting that read speech based on controlled experimental materials is not intrinsically invalid or misleading about phonetic detail.

3. Quantifying F0/segment alignment

The fact that F0/segment alignment exhibits clear phonetic regularities (cf. section I.3 above) raises the question of how best to describe alignment quantitatively. It seems fair to say that there is no clear agreement on this point. There are at least two distinct issues: how to define significant points in the F0 contour, and how to characterise their coordination with the segmental string.

Much published work on alignment deals with the alignment of F0 peaks – local maxima in the F0 contour. Ignoring the problem of spurious F0 values (discussed in the next section), F0 peaks are generally easy to identify objectively from a visual display. The only important issue here involves high accents or tones that are realised as relatively stable high-level stretches (sometimes called "plateaux"), where it is not clear whether to take the beginning, the middle, or the end of the plateau as the location of the peak; work by Knight & Nolan (2006) suggests that the end may most faithfully reflect our perception, but this is an area where further research is needed. When we are interested in F0 in locations other than peaks, however, the criteria for choosing a single F0 point as the representative value are much less clear. In particular, there is considerable discussion in the literature on how best to locate "elbows" – points in the contour where level F0 begins to rise or fall, or where rising or falling F0 begins to level off. A recent study by Del Giudice *et al*. (2007) suggests that various algorithmic methods (line-fitting, modelling, etc.) give results that are comparable to locating the elbow subjectively by eye. It seems likely that further research will eventually lead to the development of standard criteria.

As for the coordination of F0 points with the segmental string, there are two basic approaches. One characterises the location of the F0 point relative to a single segmentally identifiable ***point***, such as the offset of a stressed vowel (e.g. Arvaniti, Ladd & Mennen 1998); the other locates the F0 point relative to a segmentally identifiable ***span,*** such as the rhyme of a stressed syllable (e.g. Silverman & Pierrehumbert 1990). In the first case the alignment is expressed as a difference (e.g. 20 ms before the vowel offset, 15 ms after the vowel offset. etc.); in the second it is expressed as a proportion (e.g. the interval between the beginning of the rhyme and the F0 point is 0.65 of the duration of the rhyme as a whole – or alternatively, the F0 point is aligned 65% of the way through the rhyme). Both these methods have drawbacks. The point-based method is strictly empirical and is not based on any theoretical insight about what controls alignment other than the basic notion of segmental anchoring. Moreover, if we wish to compare different potential influences on alignment, it will not do to choose a segmental reference point that is too close to the F0 point, because random variation may mask systematic effects if the alignment values are very small (this problem is discussed at greater length by Atterer & Ladd 2004). The span-based method may have the opposite problem of locking us in prematurely to a theoretical assumption about the relevant segmentally-definable span (for further discussion see Schepman *et al*. 2006). Once again, it seems likely that the best way of quantitatively describing F0/segment alignment will emerge from further research.

Acknowledgements Thanks to Amalia Arvaniti, Nikolaus Himmelmann, and Bert Remijsen for comments, to Claudia Leto for help with the illustrations, and to Ziggy Campbell for help with the sound files.

References:

Arvaniti, Amalia & Garding, Gina (2007). Dialectal variation in the rising accents of American English. In J. Hualde & J. Cole (Eds.), *Papers in Laboratory Phonology 9* (Berlin: Mouton de Gruyter), pp. 547-76.

Arvaniti, Amalia; Ladd, D. Robert; & Mennen, Ineke (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3-25.

Atterer, Michaela & Ladd, D. Robert (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32: 177-97.

Beckman, Mary (1982). Segment duration and the 'mora' in Japanese. *Phonetica* 39: 113-35.

Bruce, Gösta (1977). *Swedish word accents in sentence perspective.* Lund: Gleerup.

Connell, Bruce (2000). The perception of lexical tone in Mambila. *Language & Speech* 43: 163-82.

Cooper, William & Sorensen, John (1981). *Fundamental frequency in sentence production*. Heidelberg: Springer.

Dauer, Rebecca (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11: 51–62.

Del Giudice, Alex; Shosted, Ryan; Davidson, Kathryn; Salihie, Mohammad; & Arvaniti, Amalia (2007). Comparing methods for locations pitch "elbows". *Proceedings of 16th ICPhS*, Saarbrücken, pp. 1117-20.

D'Imperio, Mariapaola (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33: 339-56.

Ishihara, Takeshi (2003). A phonological effect on tonal alignment in Tokyo Japanese. *Proceedings of 15th ICPhS*, Barcelona, vol. 1, pp. 615-18.

Knight, Rachael-Anne & Nolan, Francis (2006). The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association* 36 : 21-38.

Ladd, D. Robert; Mennen, Ineke; & Schepman, Astrid (2000). Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685-96.

Liberman, Mark & Prince, Alan (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249-336.

Lieberman, Philip (1967). *Intonation, Perception, and Language.* Cambridge, MA: MIT Press.

Low, Ee-Ling; Grabe, Esther; & Nolan, Francis (2000). Quantitative characterisations of speech rhythm: syllable timing in Singapore English. *Language and Speech* 43: 377-401.

Maeda, Shinji. 1976. *A characterization of American English intonation*. PhD thesis, MIT.

Ota, Mitsuhiko; Ladd, D. Robert; & Tsuchiya, Madoka (2003). Effects of foot structure on mora duration in Japanese. *Proceedings of 15th ICPhS*, Barcelona, vol. 1, pp. 59-63.

Pierrehumbert, Janet. 1979. The perception of fundamental frequency declination. *Journal of the Acoustical Society of America* 66: 363-69.

Port, Robert; Dalby, Jonathan; & O'Dell, Michael (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81: 1574-85.

Prieto, Pilar & Torreira, Francisco (2007). The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35: 473-500.

Prieto, Pilar; van Santen, Jan; & Hirschberg, Julia (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics,* 23: 429-51.

Ramus, Franck; Nespor, Marina; & Mehler, Jacques (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265-292.

Rietveld, Toni, & Gussenhoven, Carlos (1995). Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23: 375-385.

Schepman, Astrid; Lickley, Robin; & Ladd, D. Robert (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34: 1-28.

Silva, David J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23: 287-308.

Silverman, Kim (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43: 76-91.

Silverman, Kim, & Pierrehumbert, Janet (1990). The timing of prenuclear high accents in English. In J. Kingston and M. Beckman (Eds*.) Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press), pp. 71-106.

Sudhoff, Stefan; Lenertová, Denisa; Meyer, Roland; Pappert, Sandra; Augursky, Petra; Mleinek, Ina; Richter, Nicole; & Schließer, Johannes, eds. (2006). Methods in Empirical Prosody Research. Berlin: Walter de Gruyter.

Terken, Jacques (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89: 1768-76.

Warner, Natasha & Arai, Takayuki (2001) The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109: 1144-56.

Whalen, Douglas & Levitt, Andrea (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23: 349-366.

White, Laurence & Mattys, Sven (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35: 501-22.

Yuen, Ivan (2007). Declination and tone perception in Cantonese. In C. Gussenhoven and T. Riad (eds.), *Tones and Tunes, vol. 2: Experimental Studies in Word and Sentence Prosody* (Berlin: Mouton de Gruyter), pp. 63-77.