Notes on suprasegmentals and IPA-style transcription.

Following are some thoughts on the twin questions, when does the IPA not work well, and why are suprasegmentals such a problem?

Data reduction. The first set of reasons are related to the notion of data-reduction. The IPA is a basically a directed data-reducing technique, whereby we take in continuous events and classify them into a small number of discrete categories. In the process of classification, we eliminate aspects of the original event while preserving others.

How do we know whether we've thrown the right data away? The answer to this question, obviously, depends on what the question is. If, however, we are interested in a model of what people do, a very likely answer from my reading of the literature is that people don't throw any data away. This is not to say, however, that some aspects of speech information are not linguistically more important than others, or that any aspect of a linguistic event is the same as any other aspect. Different physical effects have different functions within a linguistic system, and hence will vary differently across phonetic contexts and between speakers. The IPA gives a rough classification as to many of the gross physical attributes which are likely to be central to lexical signification in the world's languages. However, it is a massive mistake to think that the IPA exhaustively classifies such information.

Discreteness. While physical events are strictly speaking never actually discrete, transcription is. This is not to say physical events are not organized in a discrete fashion. For example, the toner on this page is organized into letters which are chosen from a small set of 26 possible categories (plus punctuation).

The actual physical attributes of the toner have little or nothing to do with the letter categories, what matters is some aspect of shape and orientation with respect to other shapes on the page. The situation with transcription is a bit more difficult, since the symbols are associated with physical categories of speech behavior. Do such categories really exist? I believe the answer to this is 'yes'. Evidence for this comes from the extremely complicated relationship between speech actions and meaning. This complicated and conventional relationship apparently is supported by a systematic categorization of speech actions and sounds at a very fine time scale.

One difference between speech and orthography lies in the fact that the unit boundaries in speech are less well defined in both the paradigmatic and syntagmatic domains. Speech segments do not have edges. A second difference between speech and orthography which must also not be forgotten is that speech unfolds in time, while orthography is for all practical purposes static for the communicators.

One-dimensional segmentation. IPA transcription conflates speech onto classifications bundled together on one dimension. This approach naturally will run afoul of the syntagmatic organization of utterances, since syntagmatic organization is far more complicated than a commutable series of symbols organized on a single line would suggest. The basic problem is that the paradigmatic and syntagmatic domains are not orthogonal to one another, but rather interact in some fairly complex ways. This interaction gives rise to (at least) two types of problems, exemplified below:

a) the underspecification problem.

Examples:

i) unstressed vowels in English are often difficult to transcribe. In the paradigmatic domain, it's difficult to tell what the vowels are, or even whether there are any category differences of linguistic importance. In the syntagmatic domain, we often run into the problem of deciding whether there is really a vowel there at all, or just a noisy consonant release.

ii) consonant and vowel transcriptions divide up the speech space in different ways. Thus, for example, we note that an [h] is essentially a voiceless vowel. However, if we transcribe it as an [h], we have no categorical boundaries within the category, while if we transcribe it as a voiceless vowel, we have to decide which vowel it is. So, essentially there are 20 some different kinds of [h]'s. However, we don't mark the differences between them, if they pattern as consonants. The same situation arises with marking approximants and high vowels. In general, a close vowel with a non-syllabic marker under it is an alternative to using a consonantal approximant symbol. However, there are more close vowels than there are central approximants. For instance, we can mark the difference between an [i̯] and an [ɹ̩], but no such categorization difference exists between [j]'s. What is happening in these cases is that the location of the contrast in the syllable is impinging upon the categorization that the transcription system provides for.

iii) prosodic categories are expressed in terms of segmental qualities. Both stress and quantity distinctions exist as modulations of activity associated with coexisting segmental categories. Hence, these categories do not have any intrinsic physical quality. For example, in order to talk about quantity, we have to ask 'quantity of what?' The same applies to stress; stress is a quantitative aspect of some qualitative activity.

Even worse for a categorization scheme, these modulations themselves are modulated. For instance, lengthening is a modulation of a speech action or group of speech actions. However, lengthening is not all or nothing, rather it occurs in a continuously varying fashion. Even worse, lengthening effects are rarely, if ever, localized to a particular segment or action, but rather they affect different parts of an utterance, depending on what kind of lengthening we are talking about. Stress lengthening stretches vowel centers and consonant occlusions (at least in onsets), while lengthening due to consonant voicing typically stretches the motion into the closure for the consonants, and that more typically the latter half of the action.

iv) many categories are defined relative to some overall aspect of the speaker. The clearest example of this is tone, whereby what counts as a high tone is relative to what speaker is producing the utterance. Even worse, with respect to relativity, is that what counts as a high tone depends on the overall pitch range the speaker is using at a particular time (due to various pragmatic, proxemic, and socio-linguistic factors), as well as where in the unfolding phrase the tone is placed. For example, the occurrence of down-step systematically shifts tonal categories downward. A similar situation seems to obtain for vowel quality. Like tone, what counts as a particular vowel in terms of acoustics depends on the speaker. Unlike tone, however, what counts as a vowel has not been shown to vary significantly depending on location in an utterance.

b) the countability problem.

i) it is often difficult to determine the number of segments in a linguistic structure, much less in actual speech. The most obvious example of this problem is that of affricates, which sometimes should be treated as a single segment, e.g. for the purpose of explaining how they line up with other syntagmatically similar lexical items, and sometimes should be treated as two segments. One is tempted to suggest that this duplicitous behavior of affricates arises from the multi-faceted nature of the speech process. Perhaps, affricate articulation is analogous to the articulation of a single stop, involving a closure and release gesture. On the other hand, the acoustic results of this gesture involve two very different attributes, silence and noise, which both extend over a significant portion of time, such that either or both of these attributes can lead one to cross-classify affricates with combinations of stops and fricatives. Regardless of whether this speculation has any merit, the fact still remains that some speech events are ambiguous with respect to number of units.

Similar problems, of course, also occur with any of the complex segments discussed in the early feature geometry literature (such as Sagey's thesis): pre-nasalized stops, secondarily articulated stops, and diphthongs are the most salient examples.

ii) hierarchical organization. The issue of countability is a central driving force in prosodic organization. Prosodic organization groups speech actions into higher-level, countable units, such as syllables and intonationally marked phrases. One very clear and likely attribute of these prosodic structures is that they are hierarchical in nature; that is to say, units at one level are often grouped with other units at that level into a single higher level unit. Thus, the number of units an utterance is divided into will vary depending on what level one is examining. Thus, ambiguity in the number of units is inevitable, and any time we are dealing with attributes of these different unit levels, we will run into problems trying to conflate all of the attributes onto one stream of transcription units.

The result of this is that the IPA transcription techniques for juncture markers in particular are very poor, and often not used. Similar problems arise in using IPA transcriptions of phrase tones. Phrase tones cannot be transcribed with any clarity in the IPA system, simply because the tones are phonologically marking structure at anything like the time-scale of the segment. There have been some attempts at importing the

notion of edge-marking into the IPA, e.g. with 'minor group' and 'major group' markers. To the extent that these transcription symbols can be assigned to well-defined phonological levels in a particular language, the addition of these marks is a step in the right direction. However, a much more explicit characterization of the multiple levels of structure which are occurring will be more useful. The ToBI system has just such definitions. In addition, ToBI solves the segmental problem by placing tones on a different tier from the rest, and develops notational conventions for aligning the tones with the rest of the segmental material.

Note that this technique of developing complex conventions governing the relationship between transcribed items shows up in other places. For example, the marking of stress involves not only the conflation of stress differences to a countable number of levels, but also involves a convention as to where to put the stress mark. According to the convention, placing the stress mark, say, before the vowel, actually is to be interpreted as stress having affected all of the components of the syllable of which the vowel is the head. Assuming that we can define the domain of stress as the syllable (see, however, problem a,iii above) we can interpret the convention in a non-segment fashion to have an accurate rendition of the original. However, to the extend that stress domain is not universally predictable, the transcription system will simply break down.