'We speak prosodies and we listen to them.' (J. R. Firth, 1948)

1. Firth and Prosodic Analysis.

The title of my talk is taken from a seminal paper by J R Firth in 1948, *Sounds and Prosodies*. What I want to do today is to establish 'prosody' as a phonological concept, along Firthian lines, and to show how a prosodic approach to talk makes a more holistic analysis of language possible than conventional linguistic approaches.

From what I say, it will become clear that I don't understand 'prosody' in its narrow sense of 'intonation'. Nor do I understand it in an only slightly wider sense of 'suprasegmental', that is, things like timing, rhythm, intonation, voice quality, and other things which our alphabetic writing system doesn't easily represent. Instead, I would like to argue for a more formal, a more phonological, understanding of 'prosody', and one which I think fits well with the observation so often made among interactional linguists that we need to work holistically, and not concentrate on just one aspect of the language material we analyse. As soon as we work with real language data, we need to work holistically, and I want to argue that this is what the Firthian understanding of language lets us do.

In phonology the word 'prosody' is understood to mean a wide range of things, but in particular things referred to as 'suprasegmentals' on for example the IPA chart. This includes things like speed (or tempo), loudness, rhythmicality, voice quality and so on. A rather ungenerous description that isn't too far from the truth would be to say that prosody refers to any significant aspect of speech which there isn't a letter of the alphabet for. Instead, prosodic aspects of speech are usually trancribed in our writing system with punctuation.

The 'non-alphabetic' understanding of prosody is also one with a long and rather noble history in modern linguistics. In 1934, two Danish linguists, Louis Hjemslev and Hans-Jørgen Uldall, presented a paper at the International Phonetics Congress in London. In this paper they looked at things which demarcate particular stretches of speech, or are special to particular structures. The example they gave was the Greek diacritic for 'rough' and 'smooth' breathings, that is the presence of absence of /h/ at the start of vowel-initial words. Since this is a contrast only possible at the start of a word, they argued that it can be treated as a property of the word as a whole: when you hear /h/, you know you are at the start of a word. They called this a prosodic feature, after the Greek name for the diacritics written over the letters.

This conference was attended by J R Firth, who became the first professor of linguistics in Britain just after World War II; in his paper *Sounds and Prosodies*, he used exactly the same example, and made the same point. Firth also treated glottal stop in German as a prosody, because it marks off vowel-initial morphemes. For Firth and his co-workers, this approach became a way of working that set them apart from phonemic phonology. They started not

12/6/02

from minimal pairs, but from grammatical patterns, taking a more holistic approach to meaning in language; and for them, unlike the phonemicists, meaning was deeply embedded in their analysis, not separate from it. Their approach to conducting analysis came to be known as Prosodic Analysis. Their work was very influential in Britain even in the late 1960's. At York we have a special interest in this work, and it has influenced everyone who has worked on phonetics and phonology at York over the past two decades.

Let me say something now about the Firthian understanding of 'prosody'. I've written quite extensively about what prosodies are and are not, and if you're interested, I would refer you to some of the references on the bibliography.

Like Saussure and Hjelmslev, Firth saw paradigmatic and syntagmatic relations as fundamental in the analysis of language. The Firthians gave a privileged status to syntagmatic relations. Their phonological work shows their concern to find out for example how words are delimited from one another, and how they are held together. They were interested in how grammatical relations are expressed in the phonology. Prosodies for them are anything which relate to syntagmatic function. Of course, this includes, for example, intonation. But *a priori* there is no parameter which cannot be prosodic, because prosodies are determined not on a phonetic basis, but on a phonological basis, which for the Firthians included reference to how something functions.

Robins (1957) distinguishes two major functions of prosodies. I don't quite agree with him, but his distinction is helpful to newcomers to Prosodic Analysis.

According to Robins, prosodies can be 'extensional' or 'demarcative'. Extensional prosodies are a kind of glue which hold stretches of talk together. A canonical example is vowel harmony, whose domain is typically over a word. Vowel harmony involves having vowels of a particular kind (for example, just back vowels) within a given stretch. Vowel harmony, then, is one thing that holds the word together. But note that the thing that makes it prosodic is its function and its domain.

Demarcative prosodies offset or delimit chunks of talk. They are like Trubetzkoy's *Grenzsignale*. They tell you when one piece of structure is over and another one beginning. An example of this would be glottal stops at the start of V-initial morphemes in German.

Both kinds of prosody have a clear syntagmatic function: they create and delimit chunks of local coherence.

I'll give you examples of two structures of talk which have prosodic properties to illustrate these ideas. But before I do that, it's worth saying a little more about the Firthian approach to doing linguistics. I want to say this because on reading Firthian material from as far back as the 1930's up to the 1960's I am struck by the modernity of their ideas. It seems to me that modern linguistics frequently reinvents the wheel, and we should avoid doing this.

Among our collection of Firthian material at York, we have some lecture notes taken by

Elizabeth Anderson, who was to become Elizabeth (Betsy) Ulldal, wife of Hans-Jørgen. She attended lectures given by Firth in 1938-39. Among her notes are some extraordinarily modern insights—presumably Firth's—which chime in closely with many of our interests and methodology here at this symposium. In a section on Pareto, phonetics is described as "the study of social actions", which is full of abstract categories. Phonetics is described as containing a complex of information, the result of "interacting forces" which are "mutually influential" and can't be separated out. In here we see what was to become a principle of Firthian linguistics: the notion of 'context of situation': in other words the idea that language is deeply embedded in the occasions of its use. The social and interactional context in which it is used cannot, according to Firth, be separated from its structure and form. So we need to work out what is meant by 'context of situation'. Interactional linguistics, with its anthropological and sociological basis, provides us with a way into this.

In another comment in the notes, we read "you've got to find the values of the elements you analyse out. To be interpreted from the point of view of the person whose values they are and who is performing the action". Does this sound familiar to you? This is after all how conversation analysts go about establishing categories: by establishing the relevance of the categories to the participants themselves, rather than relying on the analyst's (or speaker's) intuitions, as in much of linguistics.

One of Firth's main points was that the contexts in which language is used are so deeply embedded within the shape of language in use, that you can't separate them out. So we need to have a properly worked out notion of 'context of situation'. As Firth also noted, in 1935, conversation was poorly understood. The organisation of conversation is now better understood, but much of linguistics has made rather little progress since 1935. Another of Firth's essential ideas was that although it was essential to work out analyses of language at different levels, it is also essential to put them together again, to produce an analysis at what he called 'congruent levels'. This means for example that syntax has to be relatable to phonology, and the way that the context of situation is embedded in language in use also needs working out. This is part of what he called 'renewal of connection'. Renewal of connection allows linguistic statements to be grounded again in the real world.

So this is what I take from Firthian linguistics:
(Firstly) A concern for the establishment of phonological systems with formal categories, taking into account both what is paradigmatic and contrastive, and what delimits chunks of material like words, phrases and turns. Prosodies are a phonological resource for producing coherence in talk.
(Secondly) A concern for linguistic statements made at a number of levels, but all mutually compatible. At some point we have to reintegrate our analyses at various levels. What works at one level needs to be relatable to other levels. Modern linguistics is rooted in approaches which make statements with respect to only one level at a time: for example phonology without syntax.
(Thirdly) An insistence that the shape of language is determined by the occasions of its use and the social actions it promotes; and the environment of language, its context of situation is not separable from it, but embedded within it.

12/6/02

What I'll do in the rest of this talk is to present to you some examples of prosodies in conversation. Some of them are 'prosodic' in the narrower, 'suprasegmental' sense, becuase they involve voice quality, tempo, amplitude, pitch and so on. But what I want to show is that we can understand 'prosody' as the 'glue' of speech, and this glue is not *just* suprasegmental things.

One way into this understanding of prosodic is to listen in a different way. People often listen to speech as if the task was to identify which segment they hear from an almost infinite set of segments. A better way to listen is parametrically. Instead of listening for, for example, an "m", you can listen for each of its ingredients separately, like labial closure, nasality, the way the closure is released, what the tongue body is doing, and so on. Doing this makes it easier to hear things which extend over longer stretches of speech than just a 'segment', and it's fundamental to the kind of approach we take at York. It also makes it difficult to define 'prosody' phonetically, because when you listen like this, you soon realise that 'segments' are an epiphenomenon: they're just a by-product of the way the articulators move in relation to each other.

2. Prosodies in Lists.

I'm going to start with lists. My thanks here to Betty Couper-Kuhlen and Margret Selting. I attended a workshop of theirs in Helsinki in 1999, and it was there that I first saw that there are sensible ways of working with natural language. They gave a wonderful workshop on lists, and one result from that workshop for me was to get some of my students collecting lists and working on them.

Lists have lots of nice properties. Some of these are what Gail Jefferson has called 'poetic'. In producing a list, speakers have to put together a set of things which will form a coherent whole. They do this in several ways:

Firstly, the items in the list are typically syntactically coherent because each item has the same kind of syntactic constituent.

Secondly, list items have to be pragmatically coherent. They aren't random collections of words, but of course things which have some pragmatic connection with each other. It might be things in the fridge to make dinner with, things you hope for for the future, what someone remembers about you, anything.

Thirdly, lists are often intonationally coherent. The list items are produced with characteristic 'list intonation', which usually includes a final item which is intonationally different from the others. There are issues here connected with the length of the list, the projection of its possible completion or continuation, and so on. But nonetheless "list intonation" is something that most people recognise.

Fourthly, lists are often temporally coherent. Of the sixty-odd lists that my students

collected, about half of these were produced in such a way that the list items were produced with onsets equally spaced in time. This is one way to project the collaborative production of a list.

Fifthly, lists often display a degree of alliteration. In phonetic terms, we can interpret this as the repetition of particular articulatory configurations. Commonly, two out of three list items begin with the same consonant.

Fragments 1-3 on the handout are lists taken from a radio phone-in programme on BBC Radio 4 in January 2000. Notice that in each list, two out of three items begin with voiceless stops with the same place of articulation. The third items start with voiceless stops too, but they have a different place of articulation. In Fragment 2, the speaker has misremembered the caller's names, and substituted a new name which alliterates with the others: there was no Pamela who called!

The last three things are clearly phonetic effects. It's easy to see intonation and timing as prosodic techniques for producing the effect of a list. Repetition of material produces coherence: the phonetic repetition is also iconic of the pragmatic and syntactic repetition. While the words in each list item are different, the point of a list is to group things together. Repeating the intonation and timing is one resource for doing this. So intonation and timing work as the 'syntagmatic glue' in lists—in Firthian terms, they have a prosodic function, because they join up the items in the list, as well as setting one item off from the others.

But then, why not include repetition of other sounds as well? If alliteration is used as a resource in list-building, then in function it is no different from intonation or timing. Alliteration, in other words, also contributes to the effect of coherence and as such, it can be seen as prosodic.

Fragment 4 on the handout is a list as produced by a Finnish speaker. She and her friend are talking about things they'd like for the future. Sanelma here produces a list in lines 4-7 which is as delicately a choreographed piece of phonetics as you can find. Here's how it goes. She produces a list of things that she'd like for the future. She chooses a career, a flat and a family. The structure of each list item is the same: Noun + and. The list items are produced with their onsets two seconds apart: this is precisely timed. It's quite remarkable that it's timed this way, since two seconds is quite a long time. If you use a logarithmic measure of time, you can expect more leeway between beats when the beats are further apart, without producing a percept of irregularity. But here the timing is very precise.

Notice how the first two list items project more to come. We can see several phonetic clues to this. One is that in lines five and six, at the end of the first and second list items, the speaker makes a closure which is held. The closure projects to the same place of articulation as the next list item. So even when she reaches the end of one list item, we have evidence that she knows at least how the next list item will start. The third list item on the other hand is breathy then creaky: as I've shown elsewhere, creak is one way for Finnish speakers to mark turn-ending. It's this which Mirja orients to, along with Sanelma's highly isochronous

production of the list, and Mirja now comes in with her own contribution to this list. This comes in exactly on beat. Like two of the words in Sanelma's list, it's vowel initial; but she produces it without voicing, and it gets repaired and offered again in line 10. So the collaboration in the production of this list is also a phonetic collaboration: Sanelma orients to Mirja's timing, the syntactic shape of Mirja's list items, and to their phonological shape: Vowel initial.

Notice also the intonation: it's very similar for all the list items.

So, here we have a very good example of a 'typical' list. It's held together syntactically, pragmatically, and phonetically. Notice that the phonetic resources are not just intonational. Considering all the things that Sanelma could have chosen for the future, why has she picked these three?

Whatever the answer, it is a fact that her list, like many other lists with an apparently open set of choices, contains two items which start in the same way. These vowel-initial words are also three syllables long. This makes the list entries four syllables long. Does this perhaps facilitate the production of the list items isochronously? You can't produce segments without producing duration, and you can't produce duration without producing segments. There's an intimate dependency here. Anecdotally, we noticed quite a number of lists in our collection had items which had the same length in terms of number of syllables.

So here, in this list, and in many other lists, we see prosodies of various kinds at work. They serve to join things together and to keep them apart. They mark structure, in this case the boundaries of list items; and the similarities between list items. Intonation, timing, and also metrical structure and so-called segmental detail work together, along with syntax and pragmatics, to produce the whole. To describe a list in terms of just one of these things is a bit like saying that a painting by van Gogh has vibrant colours. Yes, it does. But there's more to the painting than that. The colours make sense only in relation to other aspects of the painting, like the brushtrokes, the things represented in the painting, the shapes made on the canvas, and so on. So it is with speech, I want to argue: no point in looking at just the intonation, or just the voice quality, or just the segments. It is produced and heard as a whole.

3. Prosodies in Increments.

What makes increments prosodic is not the way they're done, so much as their relation to their host.

Now I'll move on to talk about increments. Increments in talk in interaction have been described and examined by Schegloff (REF) and Fox and Thompson (REF). There isn't time to give a full account, but the basic shape goes like this:

(a)   A produces a turn which is hearable as complete: I'll call this the HOST
(b)   Silence, and/or a response from B, which may be followed by further silence
(c)   A turn from A whose beginning is A's prior turn, which is hearable as reaching

completion; I'll call this the INCREMENT

The syntactic shape of an increment is such that it is parasitic on its host. It is not syntactically complete, but is built on the syntax of the host.

Interactionally, Ford Fox and Tompson argue that increments commonly seem to be used to display an orientation to some failure for the first turn to be taken up or understood.

Phonetically, increments pose an interesting problem: how to make something sound like a continuation of the prior turn, which was hearable as complete? In other words: how do you produce an increment so that it sounds coherent with its host? what phonetic resources do you use, alongside syntactic ones, to make the increment hearable as a continuation of the host?

One of my students, Gareth Walker, and I, have spent the last year working through a corpus of increments we collected from material recorded by our students. Four kinds of increment were identified on interactional grounds, each with their own phonetic characteristics. This isn't an exhaustive classification of increments, but the four categories cover a bit more than half our data. The increment types we identified not entirely different from one another phonetically, but they aren't all alike either.

Gareth established four kinds of increment: assessing increments, stance-modifying increments, post-response informational augments and relevance delimiting increments. I'll focus on one of these, but there is a summary of the phonetic properties identified for all four kinds of increment on your handout, and Gareth's MA dissertation is available on the web.

Stance modifying increments have the following structure:

> (a) A produces a complete turn which makes uptake from B relevant
> (b) B either does not produce any uptake, or produces a minimal response
> (c) A produces an increment which weakens the stance, belief or position put forward in the host.

These are labelled with a, b and c respectively in the examples on the handout. In fragment 5 on the Handout, the turn marked <a> at line 12 expresses an opinion, which the other speaker might be expected to align to. It is followed by a pause which the speaker, H, treats as an incipient misalignment by the other speaker, G. H produces an increment which weakens the stance of the turn produced in <a>, thus reducing the likelihood of a dispreferred response from G. She carries on with this action of qualifying her statement in line 17 onwards. In Fragment 6, C produces a turn which is also receipted with a gap. The increment *they said* at <c> recasts the news that was just given as being something second hand. Thus if there is any disagreement on its way, it will not be disagreement with with her directly. The turn also aligns C more closely to D, who has already shown herself to be uncertain about the previous night's events. I'll say more about Fragment 7 in a moment.

12/6/02

Some of the phonetic properties of these increments are:

      the increment has lower maximum pitch than the host
      the increment has the same pitch contour as the host
      the increment is as loud as, or louder than, the host
      the increment is the same speed as, or slower than, the host
      the voice quality of the increment matches that of the host
      the articulatory setting of the increment matches that of the host

Let's look in detail at Fragment 7.

This extract contains a stance-modifying increment in line 24 XX.

The preceding talk is about a friend of D, Sue, who is not known to S. D has been giving information about Sue and line 21 is a complex second pair part to S's first pair part. The answer to the question contains an expansion, *secretarial thing I think*. This expansion qualifies *I don't know*, offered as the first answer to the question, and recasts this *I don't know* as not completely accurate. There follows another expansion, the host + increment *I don't know... really*, which returns to the stanceof the first TCU in this turn. But the increment here modifies the absoluteness of *I don't know*. The phonetics and sequential placement of this second *I don't know* are such that a literal reading of *I don't know* is unlikely.

Stance modifying increments have the sae pitch contour as their hosts. This is shown in FIGURE 1 on the Handout with respect to FRAGMENT 7. Notice that the absolute values are not the same: these increments have a lower overall pitch than their host. So there are two constraints: (1) The pitch contour of the increment matches that of the host. (2) The overall pitch of the increment is lower than that of the host.

This kind of increment is as loud as, or quieter than, the host. In this example, the mean intensity of the host i s 64.7dB, while the increment is as at 57.1dB.

The tempo of these increments is about the same as that of the hosts or slowe. In this example, it is about the same, with the host at 5.2 and the increment at 5.3 syll/sec.

In all the data we have of this kind, the voice quality of ths host is matched by that of the increment. In this example, there is very obvious breathy voice.

There's a very noticeable shift in articulatory setting between *secretarial things* and *I think, I don't know*. There are at least two ways you can hear this. First compare the production of the fricative in *things* and *think*. In *things* there is endolabial friction which is voiceless. In *think*, there is voicig, breathines and probably nasality. Second, compare the two *I don't know*s. The first has diphthongs in *don't know*. It has a period of nasality with oral occlusion, with a glottal closure made half way through. The second one has a monophthong in *don't* and then a diphthong in *know*, with no occlusion during the nasal portion. In *really*, the articulation is more open or 'lax' than we might expect, with a minimal tongue-tip gesture, and

12/6/02

only a vestigial tongue-back gesture for /r/. So we can talk of *really* as being articulated in the same kind of way as the host, and of the expansino being articulated *as a whole* differently from the preceding talk.

What is is about increments then that makes them prosodic?

It's certainly true that many so-called suprasegmental properties, which regarded as prosodic, are of linguistic interest in these kinds of increments. We've observed that the pitch, voice quality, tempo and loudness are all resources used to match the increment to the host. But so are other kinds of articulatory features. The entire articulatory setting of the increment matches that of the host. To treat this as something separate is to miss an important generalisation: just as the syntax of the increment is fitted to the host, so is the phonetics. But we have to qualify this: each kind of increment we have established on interactional grounds has its own phonetic exponents: there is no one way for an increment to be fitted to its host. The prosodies that join hosts and increments are different for each type of increment, so we have a phonological system of host + increment joins, which maps on to the kind of action the host+increment perform.

If prosodies are the syntagmatic phonological 'glue' that produce locally coherent talk, then there's no value in treating the 'segmental' setting separately from the 'suprasegmental' one. They are all of a piece. What we have is phonetic parameters which extend over particular stretches of talk; and these phonetic parameters are the phonetic exponents of phonological categories which are relatable to other levels of linguistic statement, which includes syntax and the current interactional tasks

4. Conclusions.
What I've done is to show you some examples of phenomena from naturally-occurring talk which show a particular kind of orderliness. They show an orderliness which is about how things are put together.

Segments are an illusion which arise from the temporal co-ordination of articulatory gestures. In defining segments, we have to decide to ignore some gestures and attend to others. Another way to make observations is to listen parametrically and make observations about what we might think of as the phonetic ingredients of speech. Once we do this, as John Kelly and John Local showed in their book *Doing Phonology*, then a distinction between 'prosodic' or 'suprasegmental' features on one hand and 'segmental' ones on the other, vanishes. So defining 'prosody' phonetically doesn't work. There's no robust phonetic definition of prosody.

On the other hand, defining 'prosody' phonologically means that it has to be defined with reference to function. I've taken a very old-fashioned view, which is at the same time a very radical one. Prosodies are things that produce chunks of coherence in talk. They relate to the syntagmatic structure of talk.

I've shown you two examples of things I think of as prosodic. Both of them are hearable as

whole chunks. The problem that both kinds of structure pose for participants in conversation is how to produce things which somehow fit together to produce 'one' thing out of several.

Lists are hearable as lists. They are oriented to as lists by both speakers and hearers. Lists are coherent in all kinds of ways, and this includes what might be called 'segmental' detail: the coherence is generated not just through intonation and rhythmicality but also through repetition of other phonological material.

Increments provide similar evidence. The problem for a speaker in producing an increment is that of matching the increment to the host, so that it is hearable as a continuation of the host. Again, intonation, tempo, voice quality and so on are some of the resources available. But so is the more general articulatory setting.

In the data I played you, I've shown that there's more to prosody than prosody. As Firth pointed out, the task for us is to take language apart, but then put it back together again. This means that our phonetic observations need to be holistic. In short, we need to have a holistic linguistic theory which will account for the patterns and practices of everyday talk. Our phonological generalisations need to be relatable to generalisations we make about syntax and interaction, and all the other ingredients that make language what it is. Our phonetic observations need to be accountable to phonological generalisations. Like Firth, I think we should opt for prosody as a countable noun, not as an abstract noun. I hope I've shown you that we do indeed speak prosodies.

12/6/02