Discussion

Anne Cutler As the lone psychologist on this panel, I shall emphasise the "cognitive reality" part of our title by citing some psycholinguistic evidence that prosodic *structure* is psychologically real. In this limited discussion I shall confine my remarks to the temporal structure of English. English is said to exhibit a tendency towards isochrony, in that speakers adjust the duration of unstressed syllables so that stressed syllables occur at roughly equal intervals. There is very little evidence that English is in fact *physically* isochronous; however, the case for the *psychological* reality of isochrony is much stronger.

Firstly, English speakers certainly *perceive* their language as isochronous. In a recent study Donovan and Darwin (1979) presented listeners with sentences in which all stressed syllables began with the same sound, e.g. /t/, and asked them to adjust a sequence of noise bursts to coincide temporally with the /t/ sounds in the sentence. They could hear both sentence and burst sequence as often as they liked, but not together. Donovan and Darwin found that the noise bursts were always adjusted so that the intervals between them were more nearly equal than the intervals between the stressed syllables in the actual sentence — i.e., the listeners heard the sentences as more isochronous than they really were.

Secondly, there is the role of rhythm in syntactic disambiguation. Lehiste (1977) argues that speakers trade on listener expectations by breaking the rhythm of utterances to signify the presence of a syntactic boundary. Durational cues certainly seem to be the most effective at resolving syntactic ambiguities(see; e.g., Streeter, 1978); and recent work by Scott (forthcoming) has demonstrated that boundaries are indicated not merely by a pause or by phrase-final syllabic lengthening, but crucially by the rhythm — the fact that the foot (inter-stress interval) containing the boundary is lengthened with respect to the other feet in the utterance. Moreover, in a further study of syntactically ambiguous sentences (Cutler & Isard, in press), it was found that speakers tended to lengthen the foot containing the boundary to an integral multiple of the length of the other feet, i.e. "skip a beat" and thus *maintain* the rhythm.

Finally, there is relevant speech error evidence (Cutler, in press): when an error alters the rhythm of an utterance (a syllable is dropped or added, or stress shifts to a different syllable), it is almost always the case that the error has a more regular rhythm than the intended utterance would have had. In the following examples (syllable omission and stress error), each foot (marked by /) begins with a stressed syllable:

(1) /opering /out of a /front room in /Walthamstow
     (Target: /operating /out of a /front room in /Walthamstow)
(2) We /do think in /specific /terms
     (Target: We /do think in spe/cific /terms)

The number of unstressed syllables between the stressed syllables is more equal in the errors than in the target utterances. The consistent pattern of such errors supports the notion that isochrony in English is psychologically real: the speakers have adjusted the rhythm of their utterances to what they feel it ought to be.

John J. Ohala Selkirk reports on a very interesting attempt to unify prosodic (sentence-level) and syllabic phonology using rather abstract, almost mathematical

constructs. But if this is to be taken as something real, cognitively or otherwise, I would ask what recommends it over the dozens of other schemes offered to account for the same or similar data? Have the other schemes been disproved? In fact, none of them, including Selkirk's, have been the subject of empirical verification. As far as I can determine, none of the data cited, including that mentioned by Cutler in her remarks, unambiguously supports this model over any other.

I grant that the attempt to unify some of the phonological processes at the word and the phrase level is ingenious, but I wonder if all this machinery at the word level is necessary? A lexicon containing the pronunciation (including stress) of all existing words, both stems and derivations, can account for the data as well. The pronunciation of new words can be done by analogy (Ohala, 1974).

One of the examples mentioned in Selkirk's (pre-circulated) paper deserves careful examination. She notes that the word *rhythm* [rɪðm̩] with a syllabic [m̩] yields the derived forms *rhythmy* [rɪðmi] with the syllabic [m̩] retained, and *rhythmic* [rɪðmɪk] with the syllabicity of the [m] lost. The explanation offered for the different treatment of [m] is that the suffix -*y* [i] is a morphologically neutral suffix, and thus does not interact phonologically with the stem, whereas -*ic* [ɪk], being morphologically active, can affect the syllable structure of the stem. This explanation could be tested. We could examine speakers' pronunciation of new derivations of e.g., *prism* [prɪzm]: *prismy* and *prismic*. I can't say what the results will be but I feel I would render both with syllabic [m̩], i.e., [prɪzmi] and [prɪzm̩ɪk]. The pronunciation that would be predicted according to Selkirk, [prɪzmɪk], I would reject because I would doubt that my listeners could recover the stem *prism* from it. Only after extensive usage and long familiarity with such a form would I risk the pronunciation [prɪzmɪk] and then by the process of analogy, the models being *rhythmic, orgasmic, cataclysmic, logarithmic,* etc. There being no existing models (that I know of) for a pronunciation such as *[prɪzmi], i.e., with non-syllabic [m]. I would expect that *prismy* would always remain [prɪzm̩i] in spite of familiarity and long usage.

I find Ladd's paper well reasoned and, on the whole, quite convincing. Further evidence supporting his view that intonational contours tend to be language-specific and conventional comes from recent work by Larry Hyman and Jean-Marie Hombart on Cameroonian languages where, in some cases, they have found downdrift to be eliminated or constrained due to certain language-specific tonal traits (Hyman and Hombert, personal communication).

Nevertheless, it is still tempting to think that there is some kind of universal substrate on which language-specific uses of fundamental frequency are superimposed. Although it may not be an absolute universal, it seems very common that uncertainty and lack of self-assurance is signalled with a generally high Fo whereas certainty, self-assurance, even agression, is signalled with low Fo. It is interesting to note (as I did in Ohala, 1970) that much the same use of Fo is found in the animal kingdom as well, e.g., among dogs, raccoons, etc. That is, in an encounter between individuals, a low-pitched growl signals self-assurance and aggression whereas a high-pitched squeal is used to signal appeasement, surrender, lack of assuredness. Being an amateur ethologist, I would speculate that in emitting a high-pitched sound, the animal is trying to imitate the necessarily high-pitched sound of the young of the species in order, perhaps, to elicit some kind of maternal or paternal response from his antagonist. I say 'necessarily high-pitched' because younger individuals, being physically smaller, will have less massive vocal cords (or syringeal flaps in the case of birds) and consequently higher Fo. If we are carrying this type of innate programming around inside of us, it would not be surprising if some of it manifested itself in our linguistic use of Fo. Unfortunately, I can't think of any simple way to test these speculations.