Journal of Phonetics (1988) **16**, 381–383

Prosodic patterns in spoken English: studies in the correlation between prosody and grammar for text-to-speech conversion

By Bengt Altenberg

Lund University Press, 1987

Alex Monaghan & D. Robert Ladd

University of Edinburgh

The work presented in this book is part of the "*Text Segmentation for Speech*" project at Lund University's Survey of Spoken English. Its stated aims (p. 11) are "to explore the grammatical predictability of certain prosodic features in English for application in automatic text-to-speech conversions", with the secondary aim of augmenting the scant amount of empirical data available on English prosody. It is generally clear, well-written and well-produced, though there is some deterioration in clarity and coherence—and in the standard of typography—towards the end of the book.

The book consists of six studies in statistical patternings of various prosodic phenomena in spoken English, plus a general introduction and a summary of the conclusions. Of the six studies, the first (Chapter 2) is based on 10 5000-word texts from the London-Lund Corpus of Spoken English. The other five (Chapters 3–7) are based on one particular text chosen from the initial 10. We cite two examples to illustrate the kind of questions Altenberg asks of the data.

One study (Chapter 2) deals with the length of tone units (TUs). Altenberg reports that TUs are, on average, 1.9 seconds or 4.5 words long; a histogram of tone unit lengths in the corpus (Figure 2:1) shows a skewed distribution with a mode of three words. Another study (Chapter 5), which cannot be so readily summarized in a sentence or two, deals with the relative stressability of different word classes. In several pages of discussion Altenberg gives tables, graphs, and a "probabilistic stress rule", describing for each word class (determiner, non-lexical verb, adverb, noun, etc.) what percentage of the tokens in the corpus text have which degree of prominence (categorized as zero, stressed, booster, and nucleus).

Altenberg's statistical approach to prosodic features is of great potential interest for text-to-speech applications, but the direct usefulness of his work remains unclear. The tentative rules and categories which he presents are compared throughout with an informal model by Crystal (1975), but the workings of this model are fairly vague and there is no attempt to make them any more explicit. Thus, the reader has no objective grounds upon which to base any comparison between Crystal's model and Altenberg's rules and must therefore rely on the author's subjective evaluations. Indeed, Altenberg's own rules are often quite inexplicit and even the explicit diagram of his onset-assignment rule (p. 153) does not correspond to his description of the rule's performance.

Since no mention is made of any computer implementation of any of these rules, we assume that the statistics given for their applicability and success rates correspond to analyses by hand of the data. This is not a reliable method of simulating the performance

analyses by hand of the data. This is not a reliable method of simulating the performance of an automatic system, since the assumptions made by the analyst are never explicitly stated. These and other assumptions which Altenberg makes (such as the type and accuracy of information produced by his hypothetical parser, or the interaction of his rules in a practical system) would have to be made explicit and carefully examined before his results could usefully be applied to any automatic text-to-speech system.

Probably the most obvious specific shortcoming of Altenberg's work is the concentration, in Chapters 3–7, on a single 5 000-word corpus. Altenberg is quick to acknowledge this problem (p. 14): "restriction of the material to a single text . . . reduces both the reliability and validity of the results". The chosen text is an informal lecture given by a rural craftsman to his fellow villagers. No transcript, annotated or otherwise, of this text is given in the book but several idiosyncratic features of its prosody are mentioned: of the 10 texts from which it was chosen it has the slowest speech-rate in terms of words per second, with "surprisingly short" TU's; it is neither fully spontaneous nor formally scripted speech, and various specific prosodic and intonational phenomena are ascribed to "the speaker's tendency to linger" (p. 40), or to his 'slow and hesitant speech" (p. 63).

Even if we accept the limits imposed by the size of the corpus, and by the lack of any clear criteria for comparison between Altenberg's work and other models or theories, the book still feels unsatisfyingly incomplete. Altenberg describes it as "a starting point for further research and experimentation" but it seems unnecessary to present six exploratory studies before embarking on more detailed research. Despite his claims that a smaller corpus facilitates a much closer study, the results he presents do not appear to take account of several factors that such a small-scale study might be expected to examine in some detail. For example, it would have enhanced the book's usefulness to have included some discussion (as opposed to mere reporting of statistics) of the influence of context, word order and putative distinctions like "normal"/"contrastive" on the specific texts that form the basis of the statistical reports—say, if an unusually low rate of nuclear stress on nouns in a give study might have been attributed to a large number of nouns being repeated and contextually deaccented.

The overall character of Altenberg's work is nicely illustrated by the paper on the location of TU boundaries (Chapter 4, by far the book's longest). The conclusions of this paper are presented fluently and coherently, yet their validity and their applicability in a text-to-speech system are questioned both implicitly and explicitly. For example, having assumed the availability of complex syntactic and semantic knowledge for the operation of his predictive rules, Altenberg states in his conclusion (p. 120) that it is "uncertain to what extent an automatic parser can accomplish this". A few lines further on he writes that "the order of the rules is essential for obtaining maximal descriptive economy", but no such ordering was described in relation to the statistical results presented earlier, and, as we noted above, the whole question of explicit rule implementation and testing is left unaddressed throughout. Or again, in the chapter's final paragraph (pp. 122–3), Altenberg makes the suggestion that "slow speech can be said to reveal, more clearly than faster speech, the potential breaking points in utterances produced under certain communicative circumstances." Unfortunately, this claim is not pursued any further and the rest of the paragraph reads: "However, the exact conditions (in terms of speed and speech style) under which a TU boundary may be suspended at these potential breaking points must be the subject of further study." We were repeatedly frustrated by this tendency to stop short of a close examination of his more unexpected or surprising results and to settle for rather superficial conclusions.

To summarize, this book wavers between its primary goal, that of facilitating the automatic assignment of prosody in text-to-speech systems, and its secondary goal of providing empirical analyses of spoken English. Consequently, although much of interest and value is presented, neither goal is fully satisfied. Altenberg asks all the right questions, and redefines and refines many of them in useful ways, but his answers are few and they are outweighed by the points which he raises only to drop again.

References

Crystal, D. (1975). *The English Tone of Voice*. London: Arnold.