

Intonation: Analysis, Modelling and Technology

Antonis Botinis (editor)

(University of Skövde and University of Athens)

Dordrecht: Kluwer Academic Publishers (Text, speech and language technology, edited by Nancy Ide and Jean Véronis, volume 15), 2000, ix+396 pp; hardbound, ISBN 0-7923-6605-0, \$156.00, £90.00, €145.00; paperbound, ISBN 0-7923-6723-5, \$54.00, £34.00, €50.00

Reviewed by

Martine Grice

Saarland University

The study of intonation is an expanding field, extending beyond core linguistic disciplines such as syntax, semantics, and pragmatics into areas as wide-ranging as psycholinguistics, neurolinguistics, discourse analysis, and emotion research. Intonation is also currently the prime focus of attention in speech synthesis research and is rapidly gaining ground in speech recognition. This expansion has been reflected in a number of workshops and conferences devoted solely to intonation and its interfaces, one of which was the ESCA Workshop on Intonation in Athens in September 1997. The volume reviewed here is one of two collections of papers based on contributions to this workshop. The other collection has appeared as a special issue of *Speech Communication* (33(4), March 2001).

The book consists of four main sections—Prominence and Focus, Boundaries and Discourse, Intonation Modelling, and Intonation Technology—along with an introduction (Antonis Botinis) and a historical overview (Mario Rossi). In this review, I shall concentrate on papers that relate intonation to semantic, pragmatic, or discourse functions and leave the papers dealing solely with speech or phonetics for a review in a journal specializing in those areas.

In the Prominence and Focus section, Julia Hirschberg and Cinzia Avesani's paper, "Prosodic Disambiguation in English and Italian," investigates to what degree speakers of English and Italian use intonational means to disambiguate semantically and syntactically ambiguous sentences. The authors found that, with the exception of quantifier scope, semantic ambiguities were generally more clearly disambiguated than syntactic ones. This was true for both languages. Regarding the semantic ambiguities, the two languages had similar strategies: the scope of negation was disambiguated by phrasing, and the differences in scope of focus-sensitive operators were distinguished by means of pitch accent placement. The authors note that although speakers were aware of the distinctions, they often produced a neutral rendition that would be felicitous for either interpretation, presumably because they read the tokens within a context that already resolved the ambiguity. We learn from this that although context is necessary for eliciting the correct reading, it may at the same time dispense with the reason to disambiguate.

In the Boundaries and Discourse section, all three papers have something to offer for readers of *Computational Linguistics*. Vincent van Heuven and Judith Haan's "Phonetic Correlates of Statement versus Question Intonation in Dutch" is based on both

production and perception experiments. The authors show that although questions as a category have a number of acoustic properties that clearly distinguish them from statements, each question type has a distinct profile of its own in terms of F_0 (fundamental frequency, the perceptual correlate of which is pitch). These question types are *wh*-questions (lexical and syntactic marking), yes-no questions (which in Dutch are marked syntactically), and declarative questions (syntactically indistinguishable from statements). One major cue for the perception of questions is a sentence-final rise in pitch that was never found in statements. The smaller the number of lexicosyntactic indicators as to interrogativity, the higher the rise in pitch and the greater the incidence of such a rise. Other cues include the pitch range, height, and overall shape (e.g., downward or upward trends) across the whole sentence.

Monique van Donzel and Florian Koopmans-van Beinum's "Pitch Movements and Information Structure in Spontaneous Dutch Discourse" confirms previous findings that new information is more often accented than inferrable information. The authors establish the following hierarchy of accentability: new information > inferrable information > verbs > modifiers > discourse markers and evoked information. Discourse boundaries, assigned on the basis of a discourse model developed by the authors in earlier work, are realized with rising pitch (labeled as nonfinal by naive listeners) more often than previously reported. The authors also show that pitch height depends on neither newness nor the type of discourse boundary. Speakers varied a great deal with regard to how often they marked discourse boundaries with pitch variation as opposed to, say, pausing. However, despite these realizational differences, naive listeners perceived prominences and boundaries to a comparable extent across speakers, indicating that they are flexible enough to adapt their perceptual criteria to the current speaker.

Anne Wichman, Jill House, and Toni Rietveld's "Discourse Constraints on F_0 Peak Timing in English" is a double study of Southern Standard British English, using natural uncontrolled data and a carefully designed corpus of read paragraphs. The timing of F_0 peaks is shown to be dependent on where the accent falls within a discourse unit: peaks were later in paragraph-initial position (equivalent in this study to discourse-topic-initial position) than in paragraph-internal position, and sentence-initial accents were in turn later than sentence-final ones. The authors conclude that topic-initiality exerts a strong rightward push on F_0 peaks, even causing them to occur outside the accented syllable. Predictably, discourse structure is also found to affect F_0 peak height, topic-initiality leading to higher peaks than topic-mediality.

In the Intonation and Technology section, Gösta Bruce, Marcus Filipson, Johan Frid, Björn Granström, Kjell Gustafson, Merle Horne, and David House's "Modelling of Swedish Text and Discourse Intonation in a Speech Synthesis Framework" provides an overview of an intonation model that was originally based on single-utterance laboratory speech. The model has now been extended to cover dialogues and multispeaker conversations, incorporating information on lexical semantics and discourse and textual structure. An important step in the research program is model-based resynthesis, whereby a synthetic F_0 contour is superimposed on the original utterance. The F_0 values for resynthesis are calculated on the basis of the symbolic utterance-level representation (pitch accents and boundary tones only) of the original. The differences between calculated and original F_0 values, such as overall trends and shifts up and down in F_0 range and height, are related to the analysis of the text in order to extract parameter values that can be fed into the text-to-speech implementation.

It is clear that this volume is aimed at readers who already have a basic knowledge of intonation and know what they are looking for. Since most of the chapters deal with highly specialized topics, each one is likely to be read in isolation. However,

an incentive to read more could have been provided, had each of the four sections been accompanied by a synopsis of its main themes and common threads.

Martine Grice is an assistant professor at Saarland University, Germany. Her main research interest is intonation theory, in particular, the structure of tonal representations. She has also developed schemes for the database annotation of tonal and junctural phenomena, both for Standard German (GToBI) and for a number of varieties of Italian (IToBI). Grice's address is Institute of Phonetics, FR.4.7, Saarland University, P.O. Box 151150, D-66041 Saarbrücken, Germany; e-mail: mgrice@coli.uni-sb.de.