The PSI/PHI architecture for prosodic parsing

Dafydd GIBBON and Gunter BRAUN

Faculty of Linguistics and Literary Studies
University of Bielefeld
Postfach 8640
D – 4800 Bielefeld 1

Abstract

In this paper an architecture and an implementation for a linguistically based prosodic analyser is presented. The implementation is designed to handle typical prosodic input in the form of parallel input channels, and processes each input channel independently in a data-directed, phonologically motivated configuration of partly parallel, partly cascaded feature modules and module clusters, each implemented as finite transducers, producing intonationally relevant categories as output. The design criteria included maximal restriction of computational power (the system could be compiled into one massive finite transducer); relevance to computational linguistic formalisms with a view to developing an integrated model mapping prosodic structures on to textual structures; relatability to speech recognition algorithms, and to phonological theories. It was implemented in an object oriented environment with parallel processing simulation (CheOPS), and a linguistically interesting surface language (BATLAN).

1. Aims and design criteria

In this paper a new architecture for the parallel processing of feature systems, particularly in phonology, is presented and applied to data-directed prosodic parsing in English. It uses independent feature processing modules in configurations which allow Parallell, Sequential, Incremental (**PSI**) or Parallel, Hierarchical, Incremental (**PHI**) processing of phonetic data by the modules, with linguistically relevant output (in this case, prosodic categories such as *pitch accent*).

The domain of prosodic features (especially *intonation*, *stress* or *accent*, *tone*) has not yet received significant attention from computational linguists. It poses problems which are rather different from the phoneme or letter concatenation models typically used in computational models of written language. In particular, the problems concern the parallel processing of segmental and suprasegmental features in functionally and structurally partly autonomous tiers; an architecture for this purpose has to be able to cope with a variety of different synchronisation protocols for feature modules and module clusters. In addition, a prosodic parser has to translate from a detailed phonetic feature representation into a more abstract prosodic structure suitable for mapping on to lexical items in syntactic, semantic and pragmatic contexts.

Several styles of computational architecture could be used for this purpose, from relatively *ad hoc* blackboard architectures to the kind of virtual, abstract parallelism for feature processing in unification grammars. The selection criteria used here included:

1 Maximal restriction of computational power in terms of time and space bounds.
2 Conceptual compatibility with computational linguistic formalisms used at other linguistic levels, with a view to developing an integrated model.
3 As far as possible, relatability to speech recognition algorithms.
4 Suitability as a simulator for phonological theories of the autosegmental type.

It would appear at first sight to be a nearly impossible task to fulfil all these criteria at once. However, study of each of these areas revealed that a concept using finite automata (in particular, finite transducers) to simulate the feature modules, and partly parallel, partly cascaded configurations of these to simulate feature clusters and autosegmental tiers comes close to satisfying them, including the speech recognition requirement (cf. the *hidden Markov models*, Levinson 1986).

A secondary aim was to develop a versatile workbench for such descriptions. A linguistically interesting and useful surface language for transition networks (BATLAN, Eikmeyer/Gibbon 1983) was selected for formulating the feature modules, and provided with control mechanisms for parallel and (unidirectionally) cascaded modules, clusters and tiers with different modes of synchronisation and interaction. The implementation was developed in an object oriented programming environment with facilities for the simulation of parallel processing (CheOPS, Eikmeyer 1986).

2. The PSI/PHI architecture and prosodic parsing

The PSI/PHI concept is suitable for application to other linguistic domains which can be modelled with parallel feature processing. A pure PSI system (using finite automata in feature modules) is weakly equivalent to massive single finite automaton, though obviously with greater expressive power, while a pure PHI system (using push-down automata in feature modules) could be used to process languages up to and including context sensitive indexed languages. The prosodic parsing application is a pure PSI system. The PHI facility is

not used at present; neither is a set of augmentations designed to provide ATN-like facilities if required. The feature module automata are formulated as finite transition networks.

In the prosodic parsing application, the PSI system has a two-level cascade structure, each consisting of parallel tiers of features and feature clusters. The configuration used at present in a "stress parser" for German is shown in Figure 1.

Input to the parser consists of parallel channels of digitised signal parameters such as intensity or fundamental frequency (other spectral parameters could also be used). The initial feature detector (FD) level plays a specific functional role; it has the task of simulating the classical five tasks of a feature detector: parameter identification, time window specification, smoothing function, segmentation algorithm, and classification (value assignment) algorithm. Fairly simple feature models for acoustic edge-detection (zero crossing, slope maxima) and contour detection (peak etc.) are formulated as transition network transducers. Since the input signal is a continuous stream of indefinite length, the transducers are not assigned special finite states, but can be stopped anywhere.

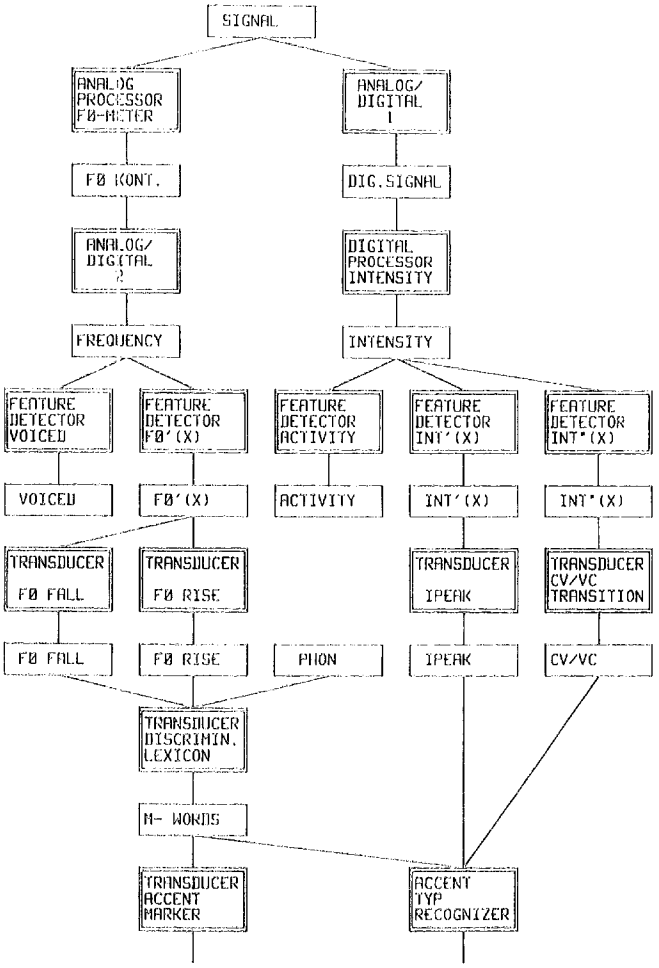
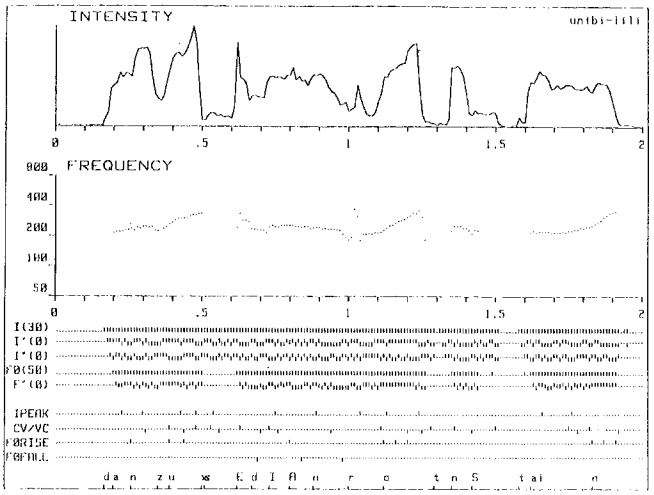


Figure 1: "stress parser"



d/a n/z/u x s E\d/I\A n/r//o t n S t/ai//n

dann suchst du dir einen roten stein

Figure 2: "feature module representation"

In lieu of segmental feature detectors, a phonetic transcription is assigned manually using a resynthesis of the digital data for fine labelling purposes; the phonetic transcription is effectively regarded as an abbreviation for the relevant tiers of segmental feature modules and module clusters.

The second level takes the primitively segmented output of the FD level and filters out the relevant prosodic categories and category sequences.

The two remaining parallel levels of accent sequences (in the case of the present parser) and segmental phonetic transcription (without syllable or word boundaries) are both fed into the textual mapping component. The main component so far implemented is the lexicon, formulated as a classical discrimination net transducer. The output from the lexicon is at present an orthographic representation with underlining of accented words.

The feature module representations defined by these levels are shown in Figure 2. At the top is a representation of two digitised signal parameters. The first group of "feature tapes" shows the output of the FD level, with upward, downward and central spikes representing binary or tertiary valued features, and dots representing null output. The second group shows the output of the FC level, with considerably more sparse representation of data-driven abstraction hypotheses about possible occurrences of prosodic categories. The other two levels show outputs within the textual mapping component.

Close attention has been paid to the empirical basis of this application. Results of experimental phonetic studies were used in formulating the FDs, and accent perception tests were conducted in order to verify the output of the PSI system against native listeners (Braun & Jin 1987). The tests yielded a satisfyingly high rate of approx. 85%. Within a homogeneous dialect group the parser is speaker-independent; the data are "raw" instructions for constructing blocks worlds, and include hesitations and repairs.

The application is being developed within a project group financed by the DFG (Deutsche Forschungsgemeinschaft) to include further prosodic categories and a suitable syntax with strategies for coping with special speech processes such as slips and repairs. The pragmatic-semantic level of theme-rheme and focus structures has already been defined for restricted blocks worlds dialogues (Pignataro 1987) and will be incorporated into an automatic focus assignment system.

Other, structurally different, or less expressive, or more heterogeneous systems using finite devices (particularly in phonology and phonetics), are being studied with a view to extending areas of application of the PSI/PHI architecture (cf. t'Hart & Collier 1975, Pierrehumbert 1980 in intonation; Church 1980, 1983 in syntax and segmental phonology; Kay & Kaplan 1981, Kay 1987 in phonology and morphology; Bolc & Maksymienko 1981, Chomyszyn 1986 in a Polish text-to-speech interface with rules by Steffen-Batogowa; Koskenniemi 1983 in morphology; Gibbon 1981, 1987 in intonation and tonology; Berwick & Pilato 1987 in syntax acquisition).

References

Berwick, R.C. S. Pilato, 1987. "Learning syntax by automata induction." *Machine Learning* 2, 9–35.

Bolc, L. & M. Maksymienko, 1981. *Komputerowy system przetwarzania tekstow fonematycznych.* U Warsaw Press.

Braun, G. & Jin, F., 1987. Akzentwahrnehmung und Akzenterkennung. "Prosodische Kohäsion" Project Report U Bielefeld.

Chomyszyn, J., 1986. "A phonemic transcription program for Polish." *Int. J. Man-Machine Studies* 25, 271–293.

Church, K.W., 1980. Memory limitations in natural language processing. Master's thesis, M.I.T.

Church, K.W., 1983. Phrase Structure Parsing. A method for taking advantage of allophonic constraints. Ph.D. thesis, M.I.T.

Eikmeyer, H.J., 1986. "CheOPS: an object-oriented system in PROLOG." User Manual. Bielefeld.

Eikmeyer, H.J. & Gibbon, D., 1983. "BATNET: ein ATN-System in einer Nicht-LISP-Umgebung." *Sprache und Datenverarbeitung* 7, 26–35.

Gibbon, D., 1981. "A new look at intonation syntax and semantics". In: A. James, P. Westney, eds., *New Linguistic Impulses in Foreign Language Teaching.* Tübingen: Narr.

Gibbon, D., 1987. "Finite state processing of tone systems." In: *Proc. 3rd Conf. European Chapter of ACL,* Copenhagen, 1–3 April 1987, 291–298.

't Hart, J. & Collier, R., 1975. "Integrating different levels of intonation analysis." *J. Phonetics* 3, 235–255.

Kay, M., 1987. "Nonconcatenative Finite-State Morphology." *Proc. 3rd Conf. European Chapter of ACL,* Copenhagen, 1–3 April 1987, 2–10.

Kay, M. & Kaplan, R., 1981. "Phonological rules and finite-state transducers." Paper at Annual Meeting of ACL, 28.2.1981, NYC. (Cited by Koskenniemi).

Levinson, S.E., 1986. "Continuously variable duration hidden Markov models for automatic speech recognition." *Computer Speech and Language* 1, 29–45.

Pierrehumbert, J., 1980. The Phonology and Phonetics of English Intonation. Ph.D. thesis, M.I.T.

Pignataro, V., 1987. Ein Sprachgenerierungsmodell mit Topik und Fokus. "Prosodische Kohäsion" Project Report, U Bielefeld.