THE SYMBOLIC CODING OF SEGMENTAL DURATION AND TONAL ALIGNMENT: AN EXTENSION TO THE INTSINT SYSTEM.

Daniel Hirst
CNRS & Université de Provence. email: daniel.hirst@lpl.univ-aix.fr
http://www.lpl.univ-aix.fr/~hirst

ABSTRACT

This paper presents work based on an analysis-by-synthesis approach which aims to develop a reversible coding system for prosody, capable of deriving a 'linguistic-like' surface phonological representation directly from acoustic data that is sufficient to reproduce a synthetic version of the original utterance without significant loss of linguistic information. With such a coding system, capable of representing any significant prosodic distinctions, the task of predicting such representations would be greatly simplified, becoming one of mapping between sets of symbolic representations. This approach has already been applied to the stylisation and symbolic coding of fundamental frequency curves by means of the INTSINT transcription system. An automatic version has also been proposed. This paper presents a preliminary proposal for an extension to the INTSINT system to cover segmental duration and the relative alignment of phonematic and tonal segments.

1. INTRODUCTION: INTSINT TRANSCRIPTIONS

INTSINT (an **IN**ternational **T**ranscription **S**ystem for **INT**onation) was developed during the preparation of a study of the intonation of twenty languages [11] and was used to transcribe examples of intonation patterns for nine of these: British English, Spanish, European Portuguese, Brazilian Portuguese, French, Romanian, Russian, Moroccan Arabic and Japanese. It aims to capture the surface distinctions used in different languages for building distinctive intonation patterns. Unlike many other transcription systems, including in particular ToBI [20], [19], INTSINT is entirely concerned with the representation of prosodic form rather than of prosodic function. In this sense it can be thought of as a prosodic equivalent of a narrow IPA transcription system for segmental transcriptions.

INTSINT represents an intonation pattern as a sequence of 'tones', coding the relative height of the significant "target points" of the pattern. Three of these tones: *Top*, *Mid* and *Bottom* are assumed to be defined globally with respect to the speaker's pitch range. Three other tones: *Higher*, *Same*, *Lower* are defined locally with respect to the preceding tone. Two further tones *Upstepped* and *Downstepped* are similar to *Higher* and *Lower* but imply a smaller interval with respect to the preceding tone. Typically, *Upstepped* and *Downstepped* are used in iterative sequences whereas *Higher* and *Lower* will generally correspond to peaks and valleys. Table 1 (from [14]) shows the orthographic and iconic symbols used in INTSINT.

Table 1: Orthographic and iconic symbols for the INTSINT coding system.

ABSOLUTE	T □	M □	B □
RELATIVE Non-Iterative	H ↑	S □	L □
Iterative	U <	•	D >

The choice of tonal symbols implies a quantification of the frequency domain. The alignment of the two sets of symbols, however, has not yet been the object of similar quantification. Instead, typically, the symbols are aligned graphically and analogically as in the following, a transcription of the French utterance "Il faut que je sois à Grenoble Samedi vers quinze heures." (I have to be in Grenoble by Saturday 3 p.m.):

(1) [ilfok@Z@swazagR@nObl][samdivERk~Ez9R]
[□ □ ↑ □ □][□ □ ↑ □]

The fact that INTSINT codes prosodic form rather than prosodic function means that it can be used as a data-driven tool for automatically extracting 'linguistic-like' information from acoustic data. In conjunction with MOMEL, which provides an automatic stylisation of F0 curves [12], [13] INTSINT has been used as a reversible coding system for F0 curves for a number of languages [1], [15], [21]. and the representation system developed has now been implemented in two text-to-speech systems for French [4], [22],

In the rest of this paper, rather than the iconic symbols used in (1), I use the orthographic symbols (T, M, B etc.) which (with the exception of D and L) have been integrated into the SAM phonetic alphabet SAMPA[23]. This alphabet is particularly suitable for computer-coding since it makes use only of symbols in the ASCII/ANSI range 32 to 126 and can consequently be used without problem for transferring transcription files between computers using different systems, where higher ASCII number characters are usually not compatible. For a recent proposal to extend SAMPA (X-SAMPA) to cover all current IPA symbols, see [24].

2. SURFACE REPRESENTATIONS OF SEGMENTAL DURATION AND TONAL ALIGNMENT.

INTSINT does allow for some degree of tonal alignment by the use of square brackets as in (1), to indicate points of synchronisation. Here I explore some possible extensions to the INTSINT system which would provide for a completely symbolic transcription of an utterance as a linear sequence of symbols including features of duration (lengthening etc) and tonal alignment. Such a notational system should ideally be as independent as possible of any particular prosodic theory. Indeed one of the aims of such a system would be to provide a means to compare alternative hypotheses in a common notational form. Implemented in conjunction with a text-to-speech system it would then be possible to evaluate fairly directly the relative merits of different theories.

The INTSINT system can be extended to provide symbolic coding for segmental duration by coding durations of segments with the symbols [-, +] for short and long respectively Segments with no explicit duration symbol are assumed to be of average length. Extra degrees of lengthening or shortening can be indicated by repeating the duration symbol (a--, a++, a---,a+++ etc.) [1].

[1] This is a slight divergence from SAMPA which proposes to use '-' as a separator. The symbol '.' (not used in X-SAMPA) could however be used for this purpose. It seems preferable to keep paired symbols such as '+' and '-" for paired interpretations such as that proposed here. Note that the colon is used in IPA to represent distinctive phonological length, not the actual physical duration of the segment. Thus /i:-/ would represent a shorter than average /i:/ segment.