# A PRAAT PLUGIN FOR MOMEL AND INTSINT WITH IMPROVED ALGORITHMS FOR MODELLING AND CODING INTONATION.

*Daniel Hirst*

Laboratoire Parole et Langage (LPL), CNRS UMR 6057
Université de Provence, Aix-en-Provence
daniel.hirst@lpl.univ-aix.fr

## ABSTRACT

This paper presents a revised version of an implementation of the Momel and INTSINT algorithms for the automatic modelling and symbolic coding of intonation patterns. The algorithms are implemented as external functions which are seamlessly integrated into the Praat speech manipulation software by means of the recently proposed *plugin* facility for Praat. Pitch detection is carried out using a subroutine to calculate optimal values of maximum and minimum F0 automatically. The implementation of the Momel algorithm incorporates an improved treatment of the modelling of pitch contours in the vicinity of onsets and offsets of voicing. The version of the INTSINT algorithm implemented is the two parameter robust version described in recent publications.

**Keywords:** automatic, model, coding, intonation, algorithm

## 1. INTRODUCTION

Different versions of the Momel and Intsint algorithms have been developed in the LPL in Aix en Provence over the last twenty years [14][16][17][15] and have been used for the phonetic modelling and symbolic coding of the intonation patterns of a number of languages (including English [1], French [24][2][25], Italian [10], Catalan [8], Brazilian Portuguese [9], Venezuelan Spanish [21], Russian [23], Arabic [22] and isiZulu [18]). Up until recently, the most accessible implementation of these algorithms was in a Unix environment as a C program (Momel) and a Perl script (Intsint) using MES, the Motif-based speech editor developed at the LPL [7]. This has been an obstacle for the wider use of the algorithms by phoneticians and linguists, who very often do not have access to this environment. This was partly overcome by the recent implementation [1] of a Praat script allowing users to run the C program and Perl script directly from this very widely used speech manipulation software.

The recent introduction of a plugin mechanism for Praat has made it possible to develop a more user-friendly environment for the algorithms. In the following sections, I describe the general structure of the Momel-Intsint plugin and then give details on the improvements integrated into the algorithms.

## 2. GENERAL STRUCTURE

The recently introduced *plugin* mechanism makes it possible to add new functions to Praat without requiring the user to manipulate scripts directly. Instead, after the plugin has been installed in the Praat preference file, the scripts are called directly from the menus of Praat, just like the other functions which are directly implemented in the program.

The modelling and coding algorithms have been implemented as a set of Praat scripts, each corresponding to a specific step in the process. Each step applies to all the files in a specified directory making it possible to apply manual correction and evaluation at each step.

### 2.1. Organisation of data

The data to be analysed is placed in a subdirectory of the user's working directory. Each recording file is contained in a separate directory with the name of the file and each derived file is placed in the same directory and has the same name with a specific extension.

### 2.2. Analyses

- *Maximum and minimum F0*. The user can specify manually the maximum and minimum values of F0 to be used in the detection and analysis. Optimised values may also be calculated by means of a subroutine described below.

- *Momel target calculation*. The stylisation of the measured F0 curve by means of a quadratic spline function is carried out for the whole subdirectory.

- *Momel target correction*. The user can visualise the signal of a given recording together with the estimated pitch targets, which can be displayed with either linear or quadratic interpolation. The target points can then be manipulated manually and the resulting pitch curve can be compared to the original via Psola re-synthesis. The automatically detected target points and the manually corrected targets are stored as seperate files.

- *INTSINT coding*. The automatic or manually corrected target points of the Momel modelling are coded using the version of the *INTSINT* algorithm described below.

- *Momel and INTSINT manipulation*. This option allows the user to make a threeway comparison between the original recording, that modelled with the Momel target points and that coded by the Intsint algorithm after being reconverted to Momel targets.

## 3. IMPROVED ALGORITHMS

### 3.1. F0 detection

The quality of the F0 modelling crucially depends on the quality of the F0 detected. In particular it is essential to use appropriate values for the maximum and minimum F0 values when performing pitch detection. Empirical experiment [6] has shown that, at least for read speech, a reasonably robust estimate of the optimal maximum and minimum can be obtained from the 1st and 3rd quartiles respectively of the f0 values estimated using the default maximum and minimum (ie 75 and 600). The formulae implemented are

f0 max = 1.5 * q3
f0 min = 0.75 * q1

where q3 and q1 represent the 3rd and 1st quartiles respectively. This gives satisfactory values for many cases but the user has the possibility to provide manual values if necessary.

### 3.2. Momel target detection

The quadratic spline function used to model the macro-melodic component is defined by a sequence of target points, (couples <s, Hz>) each pair of which is linked by two monotonic parabolic curves with the spline knot occurring (by default) at the midway point between the two targets. The first derivative of the curve thus defined is zero at each target point and the two parabolas have the same value and same derivative at the spline knot. This, in fact, defines the most simple mathematical function for which the curves are both continuous and smooth.

The Momel algorithm derives what I refer to as a phonetic representation of an intonation pattern, which is neutral with respect to speech production and speech perception since, while not explicitly derived from a model of either production or perception, it contains sufficient information to allow it to be used as input to models of either process. The relatively theory-neutral nature of the algorithm has allowed it to be used as a first step in deriving representations such as those of the Fujisaki model [20], ToBI [19][26] or INTSINT.

A recent evaluation of the algorithm [4] was carried out on recordings of the continuous passages of the Eurom1 corpus [5] for five languages (English, German, Spanish, French, Italian), in all, a total of 5 hours of speech. The evaluation estimated a global efficiency coefficient (as calculated by the F-measure) of 95.5% by comparison with manually corrected target point estimation.

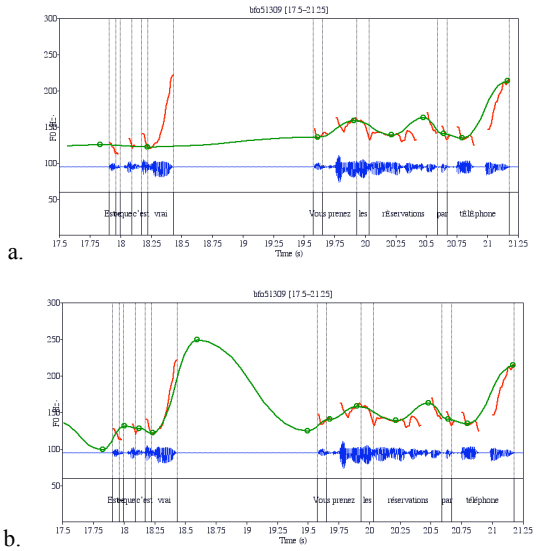
The *F*-measure is calculated as the harmonic mean (i.e. the product divided by the arithmetic mean) of the measure of *recall* (percent detected of total correct) and that of *precision* (percent correct of total detected). The *F*-measure is commonly used in the field of information retrieval as a global estimate of efficiency.

Compared to the 46982 target points provided by the automatic analysis, 3179 were added manually by the correctors and 1107 removed. The algorithm gave only slightly less efficiency (93.4%) when applied to a corpus of spontaneous spoken French.

The majority of the corrections involved systematic errors, in particular before pauses (especially preceded by a concave rising movement), which a recent improvement of the

algorithm usually manages to eliminate. Figure 1a gives an example of a passage (taken from the French version of the Eurom1 corpus) where a rising pitch before a pause is completely missed by the algorithm whereas it is correctly detected by the revised algorithm as shown in Figure 1b.

**Figure 1:** Raw (red) and modelled (green) fundamental frequency for the extract "Est-ce que c'est vrai? vous prenez les réservations par téléphone?" (Is it true? you take bookings by phone?). a. Old version b. New version




The improvement to the algorithm had in fact already been implemented in the Unix-based version for final and initial pitch movements. In Figure 1a the pitch rise on "vrai" consists almost entirely of the concave part of the rise. This contrasts with the rise on "telephone" where there is both a concave and a convex part and where the target point is correctly detected. The corrected algorithm extrapolates the final concave rise and estimates the closest target point that will produce such a rise. In the present implementation this feature is extended to all occurrences of pitch movements occurring before a silent pause, the minimum duration of which by default is 250ms.

### 3.3. Intsint coding

In an earlier version of the INTSINT algorithm, (that described in [17] and implemented in [7]), the estimation was based on a statistical analysis of the distribution of target points based on their local configuration. This required the optimisation of 10 different parameters followed by a recoding of the targets when this improved the fit.

More recently it was shown [15] that a phonetic interpretation of the INTSINT tonal segments can be carried out using two speaker dependent (or even utterance dependent) parameters of the pitch domain.

**key**: like a musical key, this establishes an absolute point of reference defined by a fundamental frequency value (in Hertz).

**range**: this determines the interval (in octaves) between the highest and lowest pitches of the utterance.

The targets **T**, **M** and **B** are defined 'absolutely' without regard to the preceding targets

$\mathbf{T} = key*\sqrt{(2\text{^}range)}$
$\mathbf{M} = key$
$\mathbf{B} = key/\sqrt{(2\text{^}range)}$

Other targets are defined with respect to the preceding target:

$\mathbf{H} = \sqrt{(P_{i-1}*T)}$
$\mathbf{U} = \sqrt{(P_{i-1}*\sqrt{(P_{i-1}*T)})}$
$\mathbf{S} = P_{i-1}$
$\mathbf{D} = \sqrt{(P_{i-1}*\sqrt{(P_{i-1}*T)})}$
$\mathbf{L} = \sqrt{(P_{i-1}*B)}$

A sequence of tonal targets such as:

[M T L H L H D B]

assuming values for a female speaker of *key* as 240 Hz and *range* as 1 octave, would be converted to the following F0 targets:

[240 340 240 286 220 273 242 170]

with appropriate time values derived from the representation of the alignment (omitted here). This sequence of target points can then be used to generate a quadratic spline function modelling the macroprosodic curve of the utterance.

The particular values used for calculating the value of **D** and **U** are chosen so that in a sequence [**T D**] for example, the **D** tone is lowered by about the same amount with respect to the **T** as the **H** tone in the sequence [**T L H**]. In many phonological accounts, Downstepped tones are analysed as a High tone which is lowered by the presence of a "floating" low tone, so that the surface tone [**D**] can be considered as underlyingly [**L H**].

The algorithm implemented optimises both the sequence of tonal segments and the key and range for the recording within the parameter space mean ±20 Hz for key and [0.5…2.5 octaves] for range.

An evaluation of the two versions of the algorithm on two hours of read speech (Korean) [12] has shown that the new algorithm performs systematically better than the earlier version.

## 4. CONCLUSIONS

The integration of the Momel and Intsint algorithms as a Praat plugin will, it is hoped, provide a useful tool for linguists and phoneticians working on prosodic analysis.

A mailing list for users of Momel and Intsint where it is possible to post questions and comments about the use of these algorithms. may be found at the following address:

http://tech.groups.yahoo.com/group/momel-intsint

The list will also provide updated information about latest versions and where to find them.

## 5. REFERENCES


[1] Auran, C. 2004. *Prosodie et anaphore dans le discours en anglais et en français : cohésion et attribution référentielle.* Doctoral thesis, Université de Provence.

[2] Bertrand, Roxane 1999. *De l'hétérogénéité de la parole : analyse énonciative de phénomènes prosodiques et kinésiques dans l'interaction interindividuelle.* Doctoral thesis, Université de Provence

[3] Boersma, P. & Weenink, D. 2006. Praat: doing phonetics by computer (Version 4.4.23) [Computer program]. Downloadable from http://www.praat.org/

[4] Campione, E. 2001. *Etiquetage prosodique semi-automatique de corpus oraux : algorithmes et méthodologie.* Doctoral thesis.. Aix-en-Provence: Université de Provence.

[5] Chan, D., Fourcin, A., Gibbon, D., Granström, B., Huckvale, M., Kokkinas, G., Kvale, L., Lamel, L., Lindberg, L., Moreno, A., Mouropoulos, J., Senia, F., Trancoso, I., Veld, C., & Zeiliger, J. 1995. EUROM: a spoken language resource for the EU. *Proceedings of the 4th European Conference on Speech Communication and Speech Tecnology, Eurospeech '95*, (Madrid) 1, 867-880.

[6] De Looze, C. in progress. I*nfluence de l'empan temporel sur les variations prosodiques en anglais contemporain*. Doctoral thesis, Université de Provence

[7] Espesser, Robert 1996. *Mes Signaix package Speech Signal processing tools* (v 1.0) [http://aune.lpl.univ-aix.fr/ext/projects/mes_signaix.htm/]

[8] Estruch, Monica 2000. Évaluation de l'algorithme de stylisation mélodique MOMEL et du système de codage symbolique INTSINT avec un corpus de passages en Catalan. in *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence* , vol. 19, pp. 45-61

[9] Fernandez-Cruz, Regina. 2000. L'analyse phonologique et acoustique du portugais parlé par des communautés noires de l'Amazonie. Doctoral thesis, Université de Provence.

[10] Giordano, Rosa 2005. Analisi prosodica e transrizione intonativa in INTSINT. in Leoni & Giordano (eds) 2005. *Italiano parlato : analisi di un dialogo*. (Liguori editore, Naples). 231-256. [Italian]

[11] Hirst, D.J. & Auran, C. 2005. Analysis by synthesis of speech prosody. The ProZed environment. in *Proceedings of Eurospeech/Interspeech* Lisbon, 2005.

[12] Hirst, D.J., Cho, H., Kim, S. & Yu, H. 2007. Evaluating two versions of the Momel pitch modelling **algorithm** on a corpus of read speech in Korean. *Proceedings of Eurospeech/Interspeech,* Antwerp 2007.

[13] Hirst, D.J. & Di Cristo, A. (eds) 1998. *Intonation Systems. A survey of Twenty Languages*. (Cambridge, Cambridge University Press). [ISBN 0 521 39513 S (Hardback); 0 521 39550 X (Paperback)].

[14] Hirst, D.J. 1987. *La description linguistique des systèmes prosodiques. Une approche cognitive.* Thèse de Doctorat d'Etat, Université de Provence

[15] Hirst, D.J. 2005. Form and function in the representation of speech prosody. in K.Hirose, D.J.Hirst & Y.Sagisaka (eds) *Quantitative prosody modeling for natural speech description and generation* (=*Speech Communication* 46 (3-4)), 334-347

[16] Hirst, Daniel & Robert Espesser 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* 15, 71-85.

[17] Hirst, Daniel, Albert Di Cristo & Robert Espesser 2000. Levels of representation and levels of analysis for intonation. in M. Horne (ed) *Prosody : Theory and Experiment*. Kluwer Academic Publishers, Dordrecht. 51-87

[18] Louw, J.A. & Barnard, E. 2004. Automatic modeling with INTSINT. in. *Proceedings of the 15th Annual Symposium of the Pattern Recognition Association of South Africa*, Grabouw, November 2004, pp. 107-111.

[19] Maghbouleh, A., 1998. ToBI accent type recognition. In: *Proceedings ICSLP 98.*

[20] Mixdorff, H., 1999. A novel approach to the fully automatic extraction of Fujisaki model parameters. In *Proceedings ICASSP 1999*.

[21] Mora Gallardo, E. 1996. *Caractérisation prosodique de la variation dialectale de l'espagnol parlé au Venezuela.* Doctoral thesis, Université de Provence.

[22] Najim, Z. 1995. *Prosodie de l'arabe standard parlé au Maroc : analyse historique, sociolinguistique et expérimentale.* Doctoral thesis, Université de Provence.

[23] Nesterenko, Irina, 2006. *Analyse formelle et implémentation phonétique de l'intonation du parler russe spontané en vue d'une application à la synthèse vocale.* Doctoral thesis, Université de Provence.

[24] Nicolas, P. 1995. *Contribution de la prosodie à l'amélioration de la parole de synthèse : cas du texte lu en français.* Doctoral thesis, Université de Provence.

[25] Portes, Cristel. 2004. *Prosodie et économie du discours : spécificité phonétique, écologie discursive et portée pragmatique du patron d'implication.* Doctoral thesis, Université de Provence.

[26] Wightman, C. & Campbell, N., 1995. Improved labeling of prosodic structure. *IEEE Trans. on Speech and Audio Processing*.